

Hamburg Studies on Multilingualism 7



# Language Contact and Contact Languages

EDITED BY Peter Siemund  
and Noemi Kintana

JOHN BENJAMINS PUBLISHING COMPANY

## Language Contact and Contact Languages

## *Hamburg Studies on Multilingualism (HSM)*

Hamburg Studies on Multilingualism (HSM) publishes research from colloquia on linguistic aspects of multilingualism organized by the Research Center on Multilingualism at the University of Hamburg.

### **Editors**

Monika Rothweiler

Juliane House

Peter Siemund

University of Hamburg

Research Center on Multilingualism

### **Volume 7**

Language Contact and Contact Languages

Edited by Peter Siemund and Noemi Kintana

# Language Contact and Contact Languages

*Edited by*

Peter Siemund

University of Hamburg

Noemi Kintana

University of Hamburg

John Benjamins Publishing Company

Amsterdam / Philadelphia



<sup>TM</sup> The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

#### Library of Congress Cataloging-in-Publication Data

Language contact and contact languages / edited by Peter Siemund and Noemi Kintana.

p. cm. (Hamburg Studies on Multilingualism, ISSN 1571-4934 ; v. 7)

Includes bibliographical references and index.

1. Languages in contact. 2. Linguistic change. I. Siemund, Peter, 1965- II. Kintana, Noemi.

P130.5.L333 2008

306.44--dc22

2008005166

ISBN 978 90 272 1927 5 (Hb; alk. paper)

© 2008 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands

John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

The production of this series has been made possible  
through financial support to the Research Center on Multilingualism  
(Sonderforschungsbereich 538 "Mehrsprachigkeit")  
by the Deutsche Forschungsgemeinschaft (DFG).



# Table of contents

List of contributors	IX
----------------------	----

## Part I. Introduction

Language contact: Constraints and common paths of contact induced language change <i>Peter Siemund</i>	3
---	---

## Part II. Typology

Inflectional morphology and language contact, with special reference to mixed languages <i>Bernard Comrie</i>	15
Contact-induced word order change without word order change <i>Bernd Heine</i>	33
Remodeling grammar: Copying, conventionalization, grammaticalization <i>Lars Johanson</i>	61
Contact-induced change: The case of the Tamangic languages <i>Michael Noonan</i>	81
Total reduplication vs. echo-word formation in language contact situations <i>Thomas Stolz</i>	107

## Part III. Diachrony

Variability within the French interrogative system: A diachronic perspective <i>Martin Elsig</i>	135
Verb-late word order in Old Swedish subordinate clauses: Loan, <i>Ausbau</i> phenomenon, or both? <i>Steffen Höder and Ludger Zeevaert</i>	163



Contact-induced phonological changes in the Catalan spoken in Barcelona <i>Conxita Lleó, Susana Cortés and Ariadna Benet</i>	185
---	-----

Prepositional aspect constructions in Hiberno-English <i>Lukas Pietsch</i>	213
---	-----

#### Part IV. Acquisition

Acquisition of Basque in successive bilingualism: Data from oral storytelling <i>Margareta Almgren, Leire Beloki, Itziar Idiazabal and Ibon Manterola</i>	239
Interrogative inversion in non-standard varieties of English <i>Michaela Hilbert</i>	261

#### Part V. Translation

Linguistic variation through language contact in translation <i>Nicole Baumgarten and Demet Özçetin</i>	293
Empirical studies of translations as a mode of language contact – “explicitness” of lexicogrammatical encoding as a relevant dimension <i>Erich Steiner</i>	317
Author index	347
Language index	351
Subject index	355

## List of contributors

Dr. Margareta Almgren  
University of the Basque-Country  
Av. Universidades, 5, 01006  
Vitoria-Gasteiz, Spain  
xamaoik@terra.es

Prof. Dr. Nicole Baumgarten  
University of Southern Denmark  
Alsion 2, 6400  
Søndeborg, Denmark  
nicole.baumgarten@sitkom.sdu.dk

Leire Beloki  
University of the Basque-Country  
Av. Tolosa, 70, 20018  
Donostia-San Sebastián, Spain  
leire.beloki@ehu.es

Ariadna Benet  
University of Hamburg  
Max-Brauer-Allee 60, 22765  
Hamburg, Germany  
ariadna.benet@uni-hamburg.de

Prof. Dr. Bernard Comrie  
Max Planck Institute  
for Evolutionary Anthropology  
Deutscher-Platz 6, 04103  
Leipzig, Germany  
comrie@eva.mpg.de

Dr. Martin Elsig  
University of Hamburg  
Max-Brauer-Allee 60, 22765  
Hamburg, Germany  
martin\_elsig@msn.com

Dr. Susana Cortés  
University of Hamburg  
Max-Brauer-Allee 60, 22765  
Hamburg, Germany  
susana.cortes@uni-hamburg.de

Prof. Dr. Bernd Heine  
University of Cologne  
Meister-Ekkehart-Straße 7, 50923  
Cologne, Germany  
bernd.heine@uni-koeln.de

Michaela Hilbert  
University of Hamburg  
Von-Melle-Park 6, 20146  
Hamburg, Germany  
Michaela.Hilbert@uni-hamburg.de

Steffen Höder  
University of Hamburg  
Max-Brauer-Allee 60, 22765  
Hamburg, Germany  
steffen.hoeder@uni-hamburg.de

Prof. Dr. Itziar Idiazabal  
University of the Basque-Country  
Av. Universidades, 5, 01006  
Vitoria-Gasteiz, Spain  
itziar.idiazabal@ehu.es

Prof. Dr. Lars Johanson  
University of Mainz  
Jakob-Welder-Weg 20, 55128  
Mainz, Germany  
johanson@uni-mainz.de

Prof. Dr. Conxita Lleó  
University of Hamburg  
Max-Brauer-Allee 60, 20146  
Von-Melle-Park 6, 22765  
Hamburg, Germany  
lleo@rrz.uni-hamburg.de

Ibon Manterola  
University of the Basque-Country  
Av. Universidades, 5, 01006  
Vitoria-Gasteiz, Spain  
ibon.manterola@ehu.es

Prof. Dr. Michael Noonan  
University of Wisconsin-Milwaukee  
Curtin Hall 439, PO Box 413, 53201  
Milwaukee, USA  
noonan@csd.uwm.edu

Demet Onuktav  
University of Hamburg  
Max-Brauer-Allee 60, 22765  
Hamburg, Germany  
demet.onuktav@uni-hamburg.de

Dr. Lukas Pietsch  
University of Hamburg  
Max-Brauer-Allee 60, 22765  
Hamburg, Germany  
lukas.pietsch@uni-hamburg.de

Prof. Dr. Peter Siemund  
University of Hamburg  
Max-Brauer-Allee 60, 22765  
Von-Melle-Park 6, 20146  
Hamburg, Germany  
peter.siemund@uni-hamburg.de

Prof. Dr. Erich Steiner  
University of Saarbrücken  
Bau A2 2 Postfach 15 11 50, 66041  
Saarbrücken, Germany  
E.Steiner@mx.uni-saarland.de

Prof. Dr. Thomas Stolz  
University of Bremen  
Bibliothekstraße Postfach 33440, 28334  
Bremen, Germany  
stolz@uni-bremen.de

Dr. Ludger Zeevaert  
University of Hamburg  
Max-Brauer-Allee 60, 22765  
Hamburg, Germany  
zeevaert@uni-hamburg.de

PART I

**Introduction**



# Language contact

## Constraints and common paths of contact-induced language change

Peter Siemund

University of Hamburg, Research Center on Multilingualism/  
Department of English and American Studies

It is a perhaps trivial truth that language contact over time will change the languages involved, but despite this seemingly obvious fact, robust models accounting for contact-induced change are still at a premium. We know that languages can influence one another in a situation of contact, but predicting the outcome of a language contact situation remains an immensely challenging task. And even though various substantial contributions to the field (Weinreich 1953; Thomason & Kaufman 1988; Aikhenvald & Dixon 2001; Curnow 2001; Ross 2001; Thomason 2001; Johanson 2002; Winford 2003; Heine & Kuteva 2005; Aikhenvald 2006) have increased our knowledge – both empirical and theoretical – considerably, the principles and mechanisms involved in language contact as well as their interaction remain difficult to capture.

Current interest in language contact research has clearly shifted from the analysis of individual situations of language contact to a systematic comparison of and across language contact situations, perhaps seriously starting in 1953 with Weinreich's seminal monograph, which is still widely quoted. Generally speaking, this comparative approach aligns itself with cross-linguistic typological work where interest does not focus on the grammatical systems of individual languages, but rather on the properties such systems have in common across languages. In the same way as typological work is interested in identifying constraints on the general architecture of languages and thus in restricting the scope of logically conceivable languages, the systematic comparison of language contact situations aims to identify constraints on the quality and quantity of mutual influence between languages. In short, it tries to restrict the scope of possible influence through contact that languages can exert on one another and to identify the constraints involved as well as the paths such contact can take. Another issue concerns matters of likelihood

and relative sequence with which linguistic items and structures are affected in a language as the consequence of contact with another, which are generally held to be principled and not random. Moreover, language contact research inspired by typological work is typically interested in providing explanations and motivations why some paths of contact-induced change are common while others are absent or rare, and which assumptions concerning the speakers – the eventual locus of language contact – are necessary in order to explain the phenomena observed.

From a cross-linguistic perspective, language contact appears to be influenced by – if not constrained by – various social parameters of the contact situation, the modules or levels of language involved (phonology, morphology, etc.) as well as the overall architecture of the languages in contact. It has also been hypothesized that issues of markedness and complexity play a role. Since the locus of language contact is the speaker, results from language acquisition research are highly relevant for language contact and need to be included too. Previous research into language contact has unearthed a multitude of formal and functional constraints determining and restricting the outcome of language contact and has also worked towards the identification of the relevant mechanisms. Most of these constraints and motivations have the status of hypotheses rather than proven facts and a lot more work needs to be done to show that they are really warranted. A crucial problem for the analyst is that the factors introduced above interact with one another in intricate ways. Many of these issues are taken up by the contributors to the current volume in one way or another. In what follows, I will make an attempt to address these problems briefly while, at the same time, weave in the perspectives, findings and generalizations laid out by the authors in their contributions.

As far as the social parameters of language contact situations are concerned, it has been observed that the number of speakers in the respective linguistic groups, the relative social status of the groups involved as well as the relative prestige of the languages to a great extent determine the linguistic outcome of language contact. In addition, it matters a lot how long two communities with different languages stay in contact and, above all, how intense the social and linguistic contact between the groups is. One of the best predictors of contact-induced language change probably is the degree of bilingualism found across the communities in contact and, to give an extreme case that may be more widespread than often assumed, whether one community is gradually shifting to the language of the other community, thus instantiating a case of language shift. Such social factors are considered the primary determinants of contact-induced change by Thomason & Kaufman (1988). Clearly, anybody interested in the linguistic principles determining contact-induced language change is required to filter out the noise produced by these social factors, but this appears to be an almost insurmountable task because language contact situations usually differ in terms of their social

constellations and are thus strictly speaking not comparable. It would be desirable, for analytic purposes at least, to have a set of contact situations that are comparable with respect to the social factors involved, but such a wish is obviously difficult to grant.

The relevance of the social factors summarized in the preceding paragraph immediately manifests itself in the modules of language preferably affected by language contact. For example, it has been found that in a situation of contact between two stable linguistic communities, the replication of material in the one language from the other begins with lexical units and only much later begins to affect grammatical units. Conversely, if a community for whatever reason decides to shift to the language of another group, the modules most likely to be affected are those of phonology and syntax. Thomason (2001) distinguishes these two scenarios by reserving the term 'borrowing' for the first case and 'shift-induced interference' for the second.

Most of the contributions assembled in the present volume illustrate the relevance of these social factors and even explicitly address them. For example, the linguistic results of language contact in Nepal are impossible to understand without a discussion of the social factors involved. I felt quite puzzled by the complexities of the social constellations described in Michael Noonan's paper. The relevance of social factors is also explicitly addressed in the contributions by Michaela Hilbert (Indian English, Singapore English), Conxita Lleó, Susana Cortés and Ariadna Benet (Catalan, Spanish) and Lukas Pietsch (Irish, English).

It has also been hypothesized that typologically closely related languages are highly susceptible to contact effects while more distantly related languages are assumed to be less likely to influence one another (Thomason 2001). In keeping with fundamental assumptions underlying research into linguistic universals, this hypothesis is based on the idea that languages are not arbitrary collections of structural properties, but obey certain laws of language architecture. As a corollary of such a principled nature of language, it follows that in situations of language contact properties that do not fit into the receiving languages cannot be transmitted or need to be modified so as to fit the receiving language. Provided these assumptions about the structure of language are correct, contact-induced language change should be restricted by linguistic universals (e.g. implicational universals) in a principled way. In accordance with results from typological studies one could hypothesize that grammatical hierarchies (animacy hierarchy, accessibility hierarchy) are more strongly complied with in situations of language contact than harmonic universals (Gass & Ard 1984; Croft 2004). Interestingly enough, some phenomena appear to spread from language to language no matter how closely or distantly related the relevant contact languages are, as Thomas Stolz shows in his study on reduplication.



The set of language contact situations discussed in the present volume is quite diverse and it may be left to the reader to decide which of them classify as contact between typologically close or distant languages. I will here restrict myself to a listing of the language pairs analyzed, even though this listing is not fully complete since too great a number of contact situations have been addressed (a surprisingly high number of language pairs involves English): Arabic/Domari, Basque/French, Basque/Spanish, Catalan/Spanish, English/Breton, English/Chinese, English/Dutch, English/German, English/Greek, English/Hindi, English/Inuktitut, English/Irish, English/Latin, English/Malay, English/Montana Salish, Finnish/Russian, German/French, German/Latin, Kadiwéu/Portuguese, Low German/Danish, Molisean/Italian, Nahuatl/Spanish, Polish/Latin, Romani/Greek, Standard French/Colloquial French, Swedish/Latin, Takia/Waskia, Tamangic/Nepali, Turkish/Arabic, Turkish/Azerbaijani, Turkish/Georgian, Turkish/Greek, Turkish/Kurdish. Mixed languages like Michif and Copper Island Aleut are also discussed.

Another issue concerns the notions ‘markedness’ and ‘complexity’ which are widely held to influence or to be influenced by language contact in spite of the fact that these notions are difficult to define and different definitions can produce highly diverging results. Following somewhat simplifying assumptions that markedness concerns the formal expression of semantic and functional oppositions and that degree of complexity roughly correlates with the number of formal distinctions available, it has often been suggested that an expectable result of contact-induced language change is a reduction in terms of markedness and complexity. In other words, in contact situations languages are assumed to lose formal oppositions, grammatical categories, and the like. These assumptions are used as the basis for the analysis of the contact situation between Catalan and Spanish by Conxita Lleó, Susana Cortés and Ariadna Benet, who argue that precisely the marked vowels of Catalan are being lost due to influence from Spanish. Even though these assumptions have a certain intuitive appeal, there is enough documented evidence that marked features can also be transmitted through language contact and that languages can acquire new categories and means of formal expression thus adding to their complexity. In his discussion of language contact between Tamangic languages and Nepali, Michael Noonan provides ample illustration of this statement albeit mainly in the areas of morphosyntax and rhetorical strategies.

Language contact by necessity has to occur in the heads of the speakers involved and hence a summary of the major factors determining contact-induced language change cannot be complete without at least briefly addressing the complex topic of language acquisition. Abstracting away from important differences in the processes of acquisition of a first or a second language, the simultaneous

or successive acquisition of two languages, the balanced or unbalanced acquisition of two languages as well as some others, a highly relevant finding of language acquisition research for contact linguistics appears to be the fact that the mutual influence of two simultaneously acquired languages is much lower than usually expected (Meisel 2007). This is a surprising finding from the perspective of research into language contact because it is usually taken for granted that bilingualism leads to contact-induced language change. Moreover, cases of contact-induced change have amply been attested in previous research activities. In the context of the present volume this general paradox is exemplified in the contribution by Margareta Almgren, Leire Beloki, Itziar Idiazabal and Ibon Manterola, who show that L2 speakers of Basque (children) who acquired Basque through immersion programmes yield nearly native-like production data and show virtually no influence from their Spanish L1. Martin Elsig describes how in French, largely due to internal developments, two varieties (Standard French versus Colloquial French) have emerged which are quite distinct in terms of their interrogative systems and, in spite of extremely close contact, show no mutual influence. Apparently, learners of French have no difficulty in keeping these two varieties of French apart. In sharp contrast to this, however, Michaela Hilbert in her study of interrogative inversion in Indian English and Singapore English documents how apparently universal strategies of L2 acquisition have led to similar structural patterns in the two varieties of English, thus documenting contact-induced change through language acquisition albeit precisely not through mutual influence of the languages involved.

Before turning to a discussion of more specific formal and functional restrictions on contact-induced change, I would briefly like to address the mechanisms underlying the replication (borrowing or transfer) of material from one language in another. A very naïve view of language contact would probably hold that speakers take bundles of formal and functional properties, semiotic signs so to speak, from the relevant contact language and insert them into their own language. To be sure, this view is much too simplistic and not seriously maintained any longer. A probably more realistic view held in language contact research is that whatever kind of material is transferred in a situation of language contact, this material necessarily experiences some sort of modification through contact. In his contribution to this volume Lars Johanson explicitly states that “Copies are never ‘imported’ or ‘transferred’ foreign elements and never true replicas of their models. There are always dissimilarities in substance, meaning, contextual applicability and frequency between models and copies.”

It is also clear that the copying of linguistic signs just by themselves is not the only process of copying or replication, and on the whole probably the least interesting one. Semantic and functional properties of some structural unit of the model language may be copied onto a structural unit of the receiving language

that can in some aspect be related to the structure of the model language. In addition, it has been observed that even the combinatorial properties of some unit as well as its frequential properties may be copied or replicated. The relevance of frequential copying is nicely illustrated in the contribution by Nicole Baumgarten and Demet Özçetin, who show that the frequency of first person plural pronouns in German business communication has increased, apparently under influence from English where the use of such pronouns in this text type is more common. A similar point is made by Erich Steiner who in his contribution argues that parameters like 'the proportion of explicit to implicit referents', 'the proportion of phoric to fully lexical (auto-semantic) phrases', 'the number of newly introduced discourse referents per discourse segment', etc., which give rise to typical text types, are open to contact-induced language change due to different frequencies in the contact language. Frequential copying may be accompanied by structural reanalysis and it appears also feasible that this kind of copying can delay or accelerate processes of internal language change, often referred to as 'reinforcement'. There are also claims to the effect that the frequency of occurrence of some item positively correlates with the likelihood of its replication.

In a recent contribution by Heine & Kuteva (2005) an even more subtle mechanism of contact-induced language change is proposed whereby speakers are assumed not only to replicate material, functional, combinatorial or frequential properties of some unit in the model language, but also processes of grammaticalization. Heine & Kuteva (2005) suggest that speakers of what they call the 'replica language' can identify grammaticalization processes that have taken place in the 'model language' and use lexical material available in the replica language to grammaticalize some grammatical element found in the model language. This model of contact-induced grammaticalization is taken up for discussion in the contribution by Steffen Höder and Ludger Zeevaert, who on the basis of subordinating clauses in Old Swedish argue that the innovation of this clause type can be interpreted as a grammaticalization process triggered by functionally similar, but structurally not equivalent constructions in Latin.

In the wide body of literature on language contact (Weinreich 1953; Thomason & Kaufman 1988; Aikhenvald & Dixon 2001; Curnow 2001; Ross 2001; Thomason 2001; Johanson 2002; Winford 2003; Heine & Kuteva 2005; Aikhenvald 2006), we also find a number of more specific formal and functional constraints, which shall be briefly reviewed in what follows. There appears to be increasing agreement these days that some domains of grammar are more sensitive to contact-induced language change than others, even though there is no consensus which grammatical domains exactly are sensitive or insensitive in this sense. A tentative hypothesis could be that structural properties which

very much belong to the core of the grammatical system are less susceptible to influence through language contact than more peripheral grammatical domains, where core grammar could refer to parameters like basic word order, head-dependent order, inflectional categories and the like, i.e. altogether structural elements of language whose function is not particularly transparent to the linguistically naïve speaker. Non-core grammar, in contrast, could refer to topicalization structures, periphrastic constructions, sentence connectors, discourse markers and perhaps also some derivational morphology (cf. Ross 2001). It needs to be made clear once again that we are here dealing with hypotheses, not facts, and that these hypotheses are not even very precise.

Nevertheless, some of the contributions to this volume address this problem explicitly and appear to lend support to the idea of grammatical domains that are more or less sensitive to contact-induced language change. For example, even though Bernd Heine neatly documents various cases of language contact (Arabic/Domari, Finnish/Russian, Kadiwéu/Portuguese, English/Inuktitut, etc.) where the basic word order of one of the contact languages is changed, the relevant changes all appear to have happened indirectly, namely through the exploitation and generalization of an already existing word order pattern. Bernd Heine states quite succinctly that “In the survey data discussed in this paper we have not come across a single case where speakers really produced a word order entirely alien to the language concerned – hence, there is not really a change from one word order to a new order.” In a similar vein Bernard Comrie argues that inflectional morphology – clearly a core area of grammar – can in principle be borrowed, but is typically reinterpreted as agglutinative morphology, i.e. borrowed in one specific meaning only. He illustrates this claim on the basis of English, German, Polish, Romani, Michif, Copper Island Aleut, among others.

Another formal prerequisite for contact-induced language change frequently voiced in the relevant literature concerns the availability of formal correspondences between the languages in contact. In other words, it is assumed that for contact-induced change to happen speakers must be able to identify formally equivalent elements in the languages involved. Obviously, this does not concern the transfer of complete linguistic signs from one language to another, as is the case with lexical borrowing, but refers to those cases where the meaning or function of some structural unit of a language, or its frequency of use for that matter, is transferred to some unit of the relevant contact language. The crucial question here is how close or how distant such formal correspondences can be or have to be. In the chapter contributed by Lukas Pietsch, based on the contact situation between Irish and English, it is argued that such processes of identification can also occur on the level of constructions leading to the transfer of complete

constructional schemata that may surface in various different realizations. Such identification and mutual adoption of constructional schemata can lead to convergence, i.e. a high number of matching structures.

As far as functional constraints on replication or transfer in situations of language contact are concerned, it has already been hinted at in various places throughout the preceding discussion that linguistic units with a transparent function are more readily replicated than those whose function is opaque. This observation ties in with what was said above about sensitivity to contact-induced language change. Moreover, this notion of transparency of function may be extended so as to embrace the wider concept of cognitive saliency, so that speakers could be hypothesized to replicate preferably those structures whose degree of saliency is relatively high. To be sure, saliency is an extremely broad concept and with respect to language contact it may comprise notions as different as functional transparency, prosodic prominence and maybe even social prestige. Functional transparency and saliency are topics that are addressed in various contributions to the present volume, but perhaps most explicitly in the study on reduplication by Thomas Stolz. There it is argued that reduplicated structures spread so easily in language contact precisely because they are highly transparent functionally and prosodically salient, since the word-initial position is usually exploited. These issues are also taken up for discussion in the studies by Bernard Comrie, Bernd Heine, Lars Johanson, Michael Noonan and Lukas Pietsch.

A final problem that needs to be touched upon briefly is the proper motivation of contact-induced language change, i.e. why do speakers replicate items from one language in another in the situation of language contact? A frequently heard answer to this question is that speakers perceive something in the contact language that they feel is missing in their own language. In short, language A lacks what language B has and speakers replicate items to compensate for the perceived gap. Michael Noonan and Erich Steiner – rather cautiously – address this problem within the realms of the current volume. This view has a lot of plausibility for lexical borrowing, but appears to lose in plausibility the less transparent the function of the replicated items becomes. As a matter of fact, it is outright contradictory to an important precondition on replication introduced above, namely that speakers must be able to establish formal and functional correspondences between structural units in the contact languages. Taking this condition seriously we would in effect predict that languages replicate structures and categories they already possess, and – curiously enough – this is what we often find. Some illustration is provided in the contributions by Bernard Comrie, Michael Noonan and Lukas Pietsch.

Most of the contributions to the present volume originate in papers read at an international colloquium on Language Contact and Contact Languages, which took place from 6 – 8 July 2006 at the University of Hamburg, and was organized

by the Collaborative Research Centre on Multilingualism. The Collaborative Research Centre on Multilingualism draws most of its funding from the *Deutsche Forschungsgemeinschaft* (German Research Foundation), whose generous support is gratefully acknowledged.

The editorial process leading to the current volume could not have been successful without the assistance of no fewer than 26 reviewers (anonymous to the chapter authors), whose comments and advice helped tremendously to increase the quality and the coherence of the volume. Both Noemi Kintana and myself would like to express our gratitude for their support.

## References

- Aikhenvald, Alexandra Y. 2006. Grammars in contact: A cross-linguistic perspective. In *Grammars in Contact: A Cross-Linguistic Typology*, A. Y. Aikhenvald & R. M. W. Dixon (eds), 1–66. Oxford: OUP.
- Aikhenvald, Alexandra Y. & Dixon, Robert M. W. (eds). 2001. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. Oxford: OUP.
- Curnow, Timothy Jowan. 2001. What language features can be ‘borrowed’? In *Areal Diffusion and Genetic Inheritance: Case Studies in Language Change*, R. M. W. Dixon & A. Y. Aikhenvald (eds), 412–436. Oxford: OUP.
- Croft, William. 2004. *Typology and Universals*. Cambridge: CUP.
- Gass, Susan & Ard, Josh. 1984. Second language acquisition and the ontology of language universals. In *Language Universals and Second Language Acquisition*, W. E. Rutherford (ed.), 33–68. Amsterdam: John Benjamins.
- Heine, Bernd & Kuteva, Tania. 2005. *Language Contact and Grammatical Change*. Cambridge: CUP.
- Johanson, Lars. 2002. *Structural Factors in Turkic Language Contacts*. London: Curzon.
- Meisel, Jürgen M. 2007. Exploring the Limits of the LAD. *Arbeiten zur Mehrsprachigkeit Folge B/ Working Papers in Multilingualism Series B*, Nr. 80. University of Hamburg: Sonderforschungsbereich Mehrsprachigkeit.
- Ross, Malcolm. 2001. Contact-induced change in Oceanic Languages in north-west Melanesia. In *Areal Diffusion and Genetic Inheritance: Case Studies in Language Change*, R. M. W. Dixon & A. Y. Aikhenvald (eds), 134–166. Oxford: OUP.
- Thomason, Sarah G. 2001. *Language Contact: An Introduction*. Edinburgh: EUP.
- Thomason, Sarah G. & Kaufman, Terrence. 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley CA: University of California Press.
- Weinreich, Uriel. 1953. *Languages in Contact. Findings and Problems* [Publications of the Linguistic Circle of New York 1]. New York NY: Linguistic Circle of New York.
- Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Oxford: Blackwell.



## PART II

# Typology





# Inflectional morphology and language contact, with special reference to mixed languages

Bernard Comrie

University of California Santa Barbara,

Max Planck Institute for Evolutionary Anthropology

What happens to inflectional morphology in cases of language contact, including both instances of borrowing and instances of mixed languages? Examination of data from various languages, including in particular Romani, Michif, and Copper Island Aleut, suggests that it is preferred to integrate agglutinative morphology, including the reinterpretation of fusional morphology as agglutinative. However, integration of fusional morphology is possible, especially in a segregated component of a mixed language, or in patterns of fusional morphology that match those already present in the recipient language.

**Keywords:** borrowability, Copper Island Aleut, inflectional class, inflectional morphology, irregularity, Michif, mixed language, morphological typology, productivity, Romani

## 1. Introduction<sup>1</sup>

Inflectional morphology is generally, and I believe correctly, held to be one of the least borrowable parts of a language's structure. And within inflectional morphology, there is generally held to be a hierarchy of borrowability, with agglutinative morphology more borrowable than fusional morphology. Field (2002: 40–48), for instance, argues that a language can only borrow inflectional morphology if it

---

1. An earlier version of this paper was presented to a meeting of the *Sonderforschungsbereich Mehrsprachigkeit* (Collaborative Research Center on Multilingualism) at the University of Hamburg in October 2005; the Copper Island Aleut material was discussed at a meeting of the Association for Linguistic Typology in Cagliari in September 2003. I am grateful to all those who participated in the ensuing discussions and to two anonymous referees.

already has inflectional morphology, and that it can only borrow fusional inflectional morphology if it already has fusional inflectional morphology.

In this article I wish to examine various aspects of this phenomenon. In Section 2, I look at borrowed morphology in some major European languages, i.e. at cases that clearly involve borrowing and not, for instance, language shift or language mixing. In Sections 4 and 5 I examine the result of language mixing involving inflectional morphology in two languages that have figured prominently in the discussion of mixed languages, namely Michif and Copper Island Aleut. Section 3 examines an intermediate case, that of Romani, a language that is not a mixed language (or at least, is not generally considered as such), but that involves a much higher incidence of borrowed inflectional morphology than is the case with the major European languages discussed in Section 2.

While the details are often intricate and idiosyncratic, a general pattern emerges whereby inflectional morphology certainly can be borrowed and certainly can participate in language mixing; however, there is nonetheless a tendency for such inflectional morphology to be agglutinative rather than fusional, including instances of fusional morphology reinterpreted as agglutinative, with clear cases of fusional inflectional morphology being either marginal or in cases where the recipient language (in the case of borrowing) or both component languages (in the case of language mixing) have fusional morphology.

## 2. Major European languages

Major European languages typically have large numbers of lexical items borrowed from Latin and Greek, some of which form part of the general vocabulary of educated speakers. For present purposes, the discussion is restricted to nouns, since this is the word class where borrowed inflectional morphology turns out to be most relevant. In some instances, these borrowed nouns, either obligatorily or optionally, have some inflectional morphology that has also been borrowed from Latin or Greek.

### 2.1 English

In English, general educated usage requires the Latin or Greek plural for some lexical items. In my own usage, the Greek- or Latin-origin plural is required for *criterion*, plural *criteria*; *crisis*, plural *crises*; *curriculum*, plural *curricula*; and optional for *syllabus*, plural *syllabuses* or *syllabi*; *formula*, plural *formulas* or *formulae*. In some cases different plural forms belong to different meanings, e.g. *index*

has the Latinate plural *indices* in the technical sense of ‘indicator’ (e.g. economic indices), but indigenous *indexes* in other senses (e.g. the indexes of a book).

Almost without exception, these borrowed plural markers are agglutinative suffixes in English, i.e. they are added to a stem that is formed by removing the corresponding singular suffix, and can be further extended by the genitive marker (e.g. PL GEN *criteria*’s). This is in contrast to Latin or Greek, where a plural suffix like *-a*, for instance, also includes case (nominative–vocative–accusative, contrasting with genitive and dative[–ablative]). The suffixes in English are agglutinative in the sense that they can readily be segmented from the stem, although the choice of plural suffix is dependent on inflectional class – there are pairings like singular *-um*, plural *-a*; singular *-a*, plural *-ae* – which thus introduces a departure from prototypical agglutination, but one that does not involve fusion of the plural suffix with other grammatical categories of the language. Of the items cited, the one exception is *index*, plural *indices*, where the form of the stem changes between singular and plural, but stem change between singular and plural is of course possible in the indigenous lexicon, even with the regular plural suffix, e.g. *wife*, plural *wives*.

In educated usage, some Greek and Latin plural patterns are fully assimilated, in the sense that regular plurals would violate the norm (e.g. forming a plural \**criteria*s or \**crises*s). But there is a substantial penumbra of variability and even uncertainty. Thus: I may cringe when my students treat *criteria* as singular, with a regular plural *criteria*s. I may smirk inwardly when a colleague tries to form the plural *stati* from *status*. But I grew up with *data* as a singular mass noun, the corresponding singulative and count noun being *piece of data*; and although at some point in my scientific career I opted for *data* as a plural count noun in my conscious usage, I have no strong negative feelings about my earlier usage (and may well still use it in less monitored moments), and would still consider the Latinate singular *datum* (rather than *piece of data*) pedantic. And anyone who tried to replace the plural of English *museum*, namely *museums*, by *musea* would strike me as at best pedantic, and more likely simply wrong. So the details can indeed be intricate and idiosyncratic, but the general pattern remains that English is tolerant of borrowed plural markers, provided they are agglutinative, in the sense of readily segmentable from the stem, and even tolerates a certain amount of stem allomorphy, as in its indigenous lexicon.

## 2.2 Polish

Polish presents a more complex set of data, since Polish has a rich productive category of case as well as number, with the exponent of case and number being

fused, e.g. singular genitive of most masculine inanimate nouns has the suffix *-u*, singular dative *-owi*, while the plural genitive has *-ów*, plural dative *-om*. It turns out that for the translation equivalent of *museum*, Polish does use a Latinate singular–plural pairing, i.e. singular *muzeum*, plural *muzea*, these being the forms found in the nominative–vocative–accusative. But the adaptation to Polish case morphology is more complex. In the singular, nouns of this type are treated as indeclinable; inanimate indeclinable nouns are neuter in Polish, and this is true for nouns of this type (even though a noun ending in a consonant would normally be masculine and declinable); this also matches the Latin gender assignment. The plural suffix *-a* fits in with this gender assignment, since it is the plural suffix for neuter nouns even in the indigenous lexicon. In the plural, however, nouns of this type are not indeclinable, but in the genitive, dative, locative, and instrumental cases take suffixes from the general run of declinable nouns, as set out in (1).<sup>2</sup>

(1)	SG	PL
NVA	muzeum	muzea
GEN	muzeum	muzeów
DAT	muzeum	muzeom
LOC	muzeum	muzeach
INS	muzeum	muzeami

Polish has thus created a unique inflectional class for nouns borrowed from Latin neuter nouns in *-um*. However, the inflectional properties of this new inflectional class are either agglutinative, or take over fusional suffixes that are already available in Polish. The stem is invariably *muze-*. The singular suffix *-um* is an addition to the inventory of Polish neuter singular suffixes, but it is clearly agglutinative, encoding only number and not case. In the plural, the suffixes fuse the expression of number and case, but only in ways that are already found in indigenous inflectional morphology, indeed using only exponents that are found there.

### 2.3 A note on German, and some preliminary conclusions

German will be mentioned only in passing here, as an indication of how the influence of the donor language can be rather well hidden in certain cases. German *Professor* ‘professor’ is a loan from Latin, but forms its plural using an inflectional

2. The dative, locative, and instrumental suffixes are the same for regular nouns of all declension classes in Polish. The genitive plural suffix *-ów* is found primarily with masculine nouns, but may be selected as being the most distinctive of the various genitive plural markers (which also include *-i* and zero).

suffix that is widespread in the indigenous lexicon and is not of Latin origin, namely *-en*, to give *Professoren*. However, there is a further difference between singular and plural that is not shown in the orthography, namely a stress shift from *Proféssor* to *Professóren* (and also lengthening of the stressed *o* vowel in the plural). This finds no parallel in indigenous morphology, but rather replicates the stress and vowel length shift found in Latin *proféssor*, plural *professórēs*. The influence of the Latin model is thus subtle, but nonetheless quite far-reaching: Although the plural suffix *-en* of the German form is agglutinative, a stem change is involved that is not matched in the indigenous lexicon. (German does, however, have other stem changes in plural formation in the indigenous lexicon, in particular umlaut, e.g. *Maus* ‘mouse’, plural *Mäuse*.) Incidentally, stress-shift patterns similar to German *Professor*, plural *Professoren*, are also found in other continental Germanic languages.

The examples from major European languages make clear that borrowing of inflectional morphology is possible, even to the extent of creating new inflectional classes for loanwords. However, the suffixes themselves are basically agglutinative (encoding only one grammatical category, though perhaps varying for different inflectional classes), or (as in the case of the Polish fused number and case suffixes) use fused suffixes that already exist in the recipient language. Although new patterns of stem allomorphy may be borrowed, these simply add to patterns of stem allomorphy already found in the language.

### 3. Romani

Although Romani is not usually considered a mixed language, given the solid Indo-Aryan foundation to all major components of the language, it has undergone borrowing to a much greater extent than the major European languages discussed in Section 2. In particular, the period spent in close contact with Greek before the dispersal of Romani speakers across much of Europe led to massive borrowing from Greek. The account given below is based on Matras (2002: 80–85, in turn based on Elšík 2000), and follows Elšík & Matras (2006: 72) in using the terms ‘oikoclititic’ for inflectional morphological paradigms inherited from Indo-Aryan and ‘xenoclititic’ for those borrowed from Greek (and other languages in the later history of individual Romani varieties).

In (2) below I have given Matras’s reconstruction of the four basic inflectional forms of the Romani noun (representing the intersection of the singular/plural number opposition and the nominative/oblique case opposition) in Early

Romani, i.e. during the period of contact with Greek referred to above.<sup>3</sup> Although there are some similarities between oikoclitic and xenoclitic classes – perhaps most noticeably the constant plural oblique *-en* (though with a variant *-jen* restricted to the oikoclitic classes), an inflection that is clearly indigenous – there are also striking differences. Several of these clearly reflect direct Greek influence, for instance the prevalence of plural nominatives in *-i*, and the striking generalization in masculine nouns to carry across the vowel of the singular nominative into the singular oblique.

(2)	NOM		OBL		
	SG	PL	SG	PL	
oikoclitic					
M	kher	kher-a	kher-es-	kher-en-	‘house’
	čačipen	čačipen-a	čačipen-as-	čačipen-en-	‘truth’
	vast	vast	vast-es-	vast-en-	‘hand’
	šer-o	šer-e	šer-es-	šer-en-	‘head’
F	pan-i	pan-ja	pan-jes-	pan-jen-	‘water’
	ǰuv	ǰuv-a	ǰuv-a	ǰuv-en-	‘louse’
	suv	suv-ja	suv-ja	suv-jen-	‘needle’
	pir-i	pir-ja	pir-ja	pir-jen-	‘pot’
xenoclitic					
M	for-o(s)	for-i	for-os-	for-en-	‘town’
	pap-u(s)	pap-i	pap-us-	pap-en-	‘grandfather’
	sapun-i(s)	sapun-ja	sapun-is-	sapun-en-	‘soap’
F	cip-a	?	cip-a-	cip-en-	‘skin’

The inflectional suffixes shown in xenoclitic nouns in (2), including suffixes borrowed from Greek, clearly involve fusion of number and case, but this is equally true of the oikoclitic classes, i.e. of the indigenous lexicon. The fused exponents of number and case are in turn, however, rigidly agglutinative with respect to the lexical stems, something that is true of both the oikoclitic and xenoclitic classes. In other words, while contact with Greek led to massive importation of new suffixes, these new suffixes simply behave exactly like corresponding indigenous suffixes, fitting into the same morphological positions in inflectional forms of nouns.

This possibility of replacing an existing suffix with one from a contact language remains productive as different Romani varieties have come into contact with other languages. Thus, the plural nominative of the noun *planet* 'planet' uses

3. Given the extent of variation across different varieties of Romani in the formation of the feminine plural nominative in the xenoclitic classes, Matras does not offer a single reconstruction. Forms found in different varieties include *-e*, *-i*, *-ǰ*, *-es*, as well as extension of oikoclitic *-a*.

a Rumanian-origin suffix in Vlach *planeturi*, a Serbian-origin suffix in Serbian Gurbet *planete*, and a Polish-origin suffix in Polska Roma *planeti*; in eastern Balkan varieties of Romani, one finds a newer layer of Greek loan morphology in the plural suffix illustrated by *sepeči* ‘basket weaver’, plural *sepečides* (Matras 2002:204). But the crucial point remains the same: Despite the high level of borrowing of inflectional morphology, it is always the case that a borrowed inflectional suffix replaces an equivalent existing inflectional suffix, i.e. the morphological typology remains constant. One might compare with this the claim by Heine (this volume) that “borrowed” word order must match an existing, even if minority pattern in the recipient language.

#### 4. Mixed languages

A mixed language is one which reflects, historically, either the combination of different major components from different historical sources (e.g., in the case of Caribbean and Indian Ocean creoles, the lexicon of a European language and a syntax that is, depending on different current hypotheses, either from a substrate language or languages or based on universal principles), or the combination of different major parts of major components (e.g. Cree verb lexicon, morphology and clause structure with French noun phrase lexicon, morphology, and syntactic structure in Michif, to be discussed in more detail in Section 4.1). It should be noted, however, that it is not a major aim of this article to contribute to the characterization of mixed languages. In the languages considered in Sections 2 and 3, it is reasonably clear that we are dealing with the continuous history of one language that has borrowed from one or more other languages. In the languages considered in Section 4, there is no single source that can be considered the basic source of the mixed language in question. Rather, the aim is to examine how inflectional morphology behaves under conditions of far more radical language mixing than was the case in the languages examined in Sections 2 and 3.

##### 4.1 Michif

The account of Michif that follows is essentially an interpretation of material presented in Bakker (1997, especially the grammatical sketch in Chapter 4), which should also be consulted for the social background to the formation of this language. Michif is spoken in communities in North Dakota and nearby parts of Canada. It arose from a mixture of Cree (with some contribution from other Algonquian languages) and French. In general, Michif clause structure



follows Cree patterns (Bakker 1997: 87), and verbs are of Cree origin and take Cree inflectional morphology (Bakker 1997: 97–102); in the terms of Matras (2003), Cree is its INFL-language. Noun phrase structure follows French patterns, with nouns and adjectives being of French origin (Bakker 1997: 102–110). A typical example of the result of this mixture is given in (3) (Bakker 1997: 87), where the noun phrases *li žwal* ‘the horse’ and *su liku* ‘its halter’ derive from French, while the verb form *ki:-wanIst-a:w* derives from Cree. Historically, Michif can thus be regarded essentially as Cree clause structure into which French noun phrases are plugged.

- (3) *ki:-wanIst-a:w*                      *li*                      *žwal*                      *su*                      *liku*.  
       PST-lose.it-3AN>3INAN    the.M.SG    horse    his.M.SG    halter  
       ‘the horse lost its halter.’

In addition, personal pronouns and demonstratives are basically of Cree origin (Bakker 1997: 104, 108–109), while articles, quantifiers, and adjectives are basically of French origin (Bakker 1997: 103–104, 106–107, 109–110), the latter reflecting the French-origin nature of noun phrase structure.

There are, however, some exceptions to this neat bifurcation. Thus, kinship nouns are often of Cree origin (Bakker 1997: 104–105), and even nouns of French origin sometimes take Cree obviative inflectional morphology (see below) (Bakker 1997: 89). Michif has some French-origin verbs, or perhaps rather verb forms, since their morphology is often “incorrect” from the viewpoint of French (Bakker 1997: 100–102). Michif has both Cree adpositions (pre- and postpositions) and French prepositions, though their semantics usually matches the distinctions made in Cree (Bakker 1997: 110–112). Discourse particles are mainly Cree, though sometimes English (Bakker 1997: 113). Coordinating conjunctions are taken from both Cree and French (Bakker 1997: 113). Interestingly, the phonologies of items from the two different sources (Cree, French) are different (Bakker 1997: 80–86).

It should be noted that the sources are only recognizable as such from a historical viewpoint. Most present-day speakers of Michif speak neither French nor Cree or any other Algonquian language (Bakker 1997: 130–133), but have English as their second language (and indeed use of Michif is now receding before English). However, the generation that created Michif must have had a high degree of fluency in both French and one or more Algonquian languages, as many fine details of both French and Algonquian grammar have carried over into Michif, giving rise to some surprising complexities in that language.

It is important to note that both Cree- and French-origin morphology is productive in Michif (with the possible exception of French-origin verb forms, which may be fossilized). This can be seen, for instance, in the gender system. Internally

to the noun phrase, French patterns are followed (Bakker 1997: 102–108), so that there is a masculine/feminine gender opposition, with nouns denoting male humans being masculine, those denoting female humans feminine, and non-human nouns distributed across the two genders. Gender is shown, for instance, in agreement of the definite and indefinite article and of prenominal adjectives. Contrast masculine *li bwa* ‘the wood’, *ǎ spjũ* ‘a spy’, *ǎ gru sarpā* ‘a big snake’ with feminine *la mǎzũ* ‘the house’, *en fam* ‘a woman’, *la grus tāt* ‘the big tent’ (Bakker 1997: 103, 106–107). It is thus not the case that the article is a frozen part of the Michif noun, since the same noun will appear now preceded by the appropriate-gender definite article, now the indefinite article, now separated from its article by an appropriate-gender prenominal adjective. Reinforcing this is the observation that for some nouns the gender differs from that found in French, e.g. Michif *li šev(r)* ‘the goat’ (French *la chèvre*, feminine) (Bakker 1997: 103), while others are nouns that do not exist in French, e.g. *la dam* ‘the dam’, a loan from English (Bakker 1997: 105). French-origin gender is thus a productive category within the Michif noun phrase. Michif does lack one property of French, since in Michif postnominal adjectives do not show gender agreement (Bakker 1997: 106). However, this involves a reduction of the scope of gender agreement, not loss of productivity. One might compare number agreement within the English noun phrase, essentially reduced to the demonstrative adjectives (singular *this*, *that*; plural *these*, *those*); while this is a massive reduction compared to Old English, the residual number agreement is nonetheless productive.

In terms of the external syntax of noun phrases, however, it is the Cree gender pattern that holds, with its distinction between animate and inanimate (whereby a number of nouns denoting inanimate objects, e.g. means of transport and some trees, are nonetheless treated grammatically as animate) (Bakker 1997: 99). In example (3), for instance, the verb form indicates that an animate subject is acting on an inanimate object, i.e. for the purposes of verb morphology *žwal* and *liku* are of different genders, even though noun-phrase internally they are both of the French-origin masculine gender. Indeed, the morphology of the verb form in (3) encodes yet another grammatical category of Cree noun phrase external syntax, namely the distinction between proximate and obviative (Bakker 1997: 88–89), whereby “proximate” indicates roughly the most salient participant in a discourse segment, although there are also grammatical constraints on the occurrence of proximate and obviative, e.g. it is not possible for subject and object of a transitive verb to be both proximate. In Algonquian languages at least animate nouns have a distinct suffixally marked obviative form, but while marking of obviation on nouns is rare in Michif, a verb form as in (3) does mark this category, in the case of the verb form in (3) indicating that a proximate subject is acting on an obviative object.

The Cree morphology found in Michif is partly agglutinative (e.g. the prefix *ki-* of the verb form in (3)), partly fusional (e.g. the suffix on the verb in (3)). Rather few French-origin items have distinct gender forms, given the restriction to articles and pronominal adjectives, but those that do betray no obvious formal pattern (except that, as in French, one gets a fair amount of leeway out of assuming that some masculine forms involve deletion of the final consonant of the corresponding feminine form), so that one is dealing with a system at least some distance from canonical agglutination, and probably in large measure fusional. But the pieces of fusional morphology are strictly segregated between Cree-origin and French-origin structures. In other words, the historical mixture that gave rise to present-day Michif includes substantial pieces of fusional morphology from each of the major donors.

## 4.2 Copper Island Aleut

Copper Island Aleut (Mednyj Aleut) is a mixed language arising historically from the contact of Aleut<sup>4</sup> and Russian. It is the language of Copper Island (Russian: Mednyj ostrov), one of the Russian Commander Islands (Komandorskie ostrova), although it is now heavily endangered and receding before Russian. There are several recent accounts of the language, including of the social factors that led to its creation, such as Golovko (1994, 1996), and if in the present article I base myself largely on Sekerina (1994), this is primarily because this source includes some relevant statistical information based on a corpus.

A typical example of a Copper Island Aleut sentence is given in (4) (Sekerina 1994: 22). The noun phrase *taana-â* is pure Aleut, both in its lexical stem and in the suffixal inflection. The verb complex *ni-buud-iš ukuu-tʃ* consists of an Aleut lexical stem (*ukuu-* ‘see’), but everything else is Russian, including not only the future auxiliary verb (stem *buud-*, negative prefix *ni-*, and person–number suffix *-iš*), but also the infinitive suffix *-tʃ*; in the terms of Matras (2003), Russian is its INFL-language.

- (4) *taana-â    ni-buud-iš    ukuu-tʃ*  
       land-SG   NEG-FUT-2SG   see-INF  
       ‘you will not see the land.’

---

4. As a basic source of information on (non-mixed) Aleut, I have used Bergsland (1997). For ease of comparison, I have to some extent homogenized the transcriptions for Russian, traditional Aleut, and Copper Island Aleut; note in particular that *j* has its IPA value.

As in the case of Michif, the generation that created Copper Island Aleut must have had good fluency in both source languages, here Aleut and Russian. However, even this simple example shows that the details of the mixture are rather different from those seen in Michif in Section 4.1; in particular, the verb combines an Aleut stem with Russian inflectional morphology. Nonetheless, as with Michif it will be useful to consider nouns and verbs, and noun and verb morphology, separately. Incidentally, at least in traditional Copper Island Aleut usage, some unusual syntactic properties of Aleut, quite distinct from Russian, are present, including the possibility of having verb agreement with the person–number of a possessor interpreted as topic rather than with the possessum, the head noun of the noun phrase (Golovko 1996: 68–70).

Lexically, nouns can be either Aleut- or Russian-origin, with 61.5% being Aleut-origin in Sekerina’s corpus (Sekerina 1994: 29). Derivational morphology of nouns is Aleut-origin, as is their inflectional morphology (Sekerina 1994: 22–23, 26–27). This inflectional morphology, set out in (5), involves categories that are absent from Russian, such as a distinct dual number (in the expression of the number of the possessum), and possessive forms expressing the person–number<sup>5</sup> of the possessor. (The first line of forms in (5) is used when no possessor is expressed.)

(5)	SG	DU	PL
	- $\hat{x}$	-x	-n, -s
1SG	-ŋ	-ki-ŋ	-ni-ŋ
2SG	-n	-ki-n	-t <sup>6</sup>
1PL	-mis	-ki	-mis
2PL	-ci	-ki	-ci

The morphology of Aleut (i.e. the “unmixed” language Aleut, in contrast to Copper Island Aleut) is usually characterized as agglutinative, and indeed most of the forms in (5) can be so analyzed, though with constraints on possible combinations.

5. Cysouw (2003: 66–98) argues at length that the traditional treatment of person and number (specifically, singular versus non-singular) as orthogonal dimensions in personal pronouns and pronominal markers on verbs is flawed. From this one can conclude that “fusion” of person and number in pronouns and in pronominal markers on verbs is not a true instance of fusional morphology, but rather that such combinations as “first person plural” should be treated as a single category.

6. Sekerina (1994: 23) actually has *-m* for the second person plural, but comparison with other sources suggests that this is a typographical error for *-t*; perhaps the italic form of Cyrillic *t*, which is identical to Latin *m*, was carried over. The discussion of this form below, incidentally, remains the same whether it is *-t* or *-m*.

The nonpossessed plural suffixes *-n*, *-s* seem to be in free variation, and reflect an original dialect distinction within Aleut (Bergsland 1997:47). The singular suffix is *-x̂*, but this disappears in combination with a possessive suffix. The dual suffix is *-x* when nonpossessed, but *-ki* when possessed, and does not allow overt combination with a plural possessive suffix (the number suffix, in its possessed form, appearing alone), although it does combine agglutinatively with the singular possessive suffixes. The plural suffix is *-n* or *-s*, but *-ni* before a possessive suffix, although the only combination is with the first person singular possessive suffix, in which case the combination is regularly agglutinative. If the plural suffix of the possessum would be expected to combine with a suffix indicating a plural possessor, only the possessive suffix indicating the possessor shows up overtly. The combination of plural possessum and second person singular possessor, *-t*, is clearly fusional. The possessive suffixes can be identified as first person singular *-ŋ*, second person singular *-n* (but fusing with number of the possessum in the plural, as noted above), first person plural *-mis*, and second person plural *-ci*, the last two lacking the possibility of overt expression in the presence a dual possessum suffix. Basically, each suffix encodes a unique and invariable grammatical category, as would be expected in a pure agglutinative system. However, there is a clear instance of fusion in the suffix *-t* (plural possessum, second person singular possessor), and presumably also in the distinction between the two dual suffixes, *-x* when there is no possessor, *-ki* when there is one – while this could be regarded as purely allomorphic in the combinations with overt singular possessive suffixes, the bare suffix *-ki* contrasts with *-x* in that the former allows recovery of a plural possessor, while the latter does not.

Lexically, verbs can again be either Aleut- or Russian-origin, but the skewing in favor of Aleut in Sekerina's corpus is much greater here, with 94% of verbs being Aleut-origin (Sekerina 1994:29). Derivational morphology is again Aleut (Sekerina 1994:26–27). Nonfinite inflectional morphology is also Aleut (except for the infinitive, to be discussed below) (Golovko 1996:68), although in more recent varieties of the language Aleut nonfinite verb morphology appears to be increasingly lost, not, however, in favor of Russian nonfinite verb morphology, but rather in favor of constructions using finite verb morphology (Sekerina 1994:29). But despite the above heavy bias towards Aleut in verbs and verb structure, the finite inflectional morphology of Copper Island Aleut is exclusively Russian, as also is the infinitive. Example (6) below compares verb morphology in Bering Island Aleut, a non-mixed variety, with that of Copper Island Aleut, using the verb *awa-/aba-* 'work' as illustration; the Bering Island forms are taken from Menovščikov (1968:396–398), though there are no significant differences from those cited by Bergsland (1997). The root in both languages is essentially the same and of Aleut

origin, but the inflectional suffixes are radically different, Aleut in the case of Bering Island Aleut, Russian in the case of Copper Island Aleut.

(6)		Bering Island Aleut	Copper Island Aleut
PRS	1SG	awa-ku-q	aba-ju
	2SG	awa-ku- $\hat{x}t$	aba-iš
	3SG	awa-ku- $\hat{x}$	aba-it
	1PL	awa-ku-s	aba-im
	2PL	awa-ku- $\hat{x}t$ -xičix	aba-iti
	3PL	awa-ku-s	aba-jut
PST	3SG	awa-na- $\hat{x}$	
	M.SG		aba-l
FUT	3SG	awa-ŋan ana- $\hat{x}$	buud-it aba-tj
IMP	2SG	awa- $\hat{d}a$	aba-j

As noted above, Aleut is generally characterized as agglutinative, whereas Russian is generally characterized as fusional, and indeed elements of agglutinative structure are clear in the Bering Island Aleut forms in (6), e.g. the tense suffixes present *-ku* and past *-na*, and person–number suffixes such as third person singular *- $\hat{x}$* . However, in characterizing the Russian-origin verb morphology of Copper Island Aleut, it is important to consider not so much the morphology of Russian, but rather the morphology of the Russian-origin forms as filtered through Copper Island Aleut. As will emerge from the discussion below, Copper Island Aleut is, or at least can be analyzed as being, considerably more agglutinative than its Russian origin.

While Russian verb morphology is rather complex, in particular in the number of allomorphic variants, including some outright irregularities, much of this complexity is concerned with the behavior of consonant-final stems and with the distinction between the two major conjugations when the inflectional suffixes are stressed. Copper Island Aleut essentially avoids this problem by having only vowel-final stems – this seems to be categorical for Aleut-origin verb lexemes – and by having stem-stress to the exclusion of desinence stress. These two points will be taken up in order.

In Russian, the regular past tense suffix is *-l*, which shows up with no further suffix in the masculine singular, as in *čita-l* ‘read-PST’. However, after a consonant, either the *l* drops (e.g. *nes* ‘carry-PST’), or the final consonant drops (e.g. *ve-l* ‘lead-PST’, stem *ved-*), depending on the final consonant; in the feminine and neuter singular and in the plural, the *l* marking past tense reappears in the former type, but the stem-final consonant does not in the second, e.g. *nes-l-a* ‘carry-PST-F’, *ve-l-a* ‘lead-PST-F’. Since Copper Island Aleut has basically only vowel-final stems, these complications are avoided.

Russian has two “conjugations” (inflection classes), differentiated by the inflectional person–number suffixes in the present tense (except in the first person singular, where the form *-(j)u* is used in both conjugations). The difference is in the vowel of the ending. In the third person plural, this is first conjugation *-(j)ut* versus second conjugation *-(j)at*, irrespective of stress.<sup>7</sup> In other person–number combinations, the difference is clear under stress, where the first conjugation has the vowel /o/, as in third person singular *-(j)ot*, while the second conjugation has /i/, as in third person singular *-it*. When atonic, however, the difference between /o/ and /i/ is either neutralized or, in careful speech, almost neutralized ([ɪ] versus [i]). Copper Island Aleut has only unstressed person–number suffixes, and only forms with the vowel /i/, either reflecting the second conjugation or, more probably, neutralization of the two conjugations. Copper Island Aleut thus ends up with present tense inflection as in (7) (Sekerina 1994: 25), with the Russian etyma given below (in phonetic, rather than orthographic form).

(7)	1SG	2SG	3SG	1PL	2PL	3PL
CIA	-ju	-iš	-it	-im	-iti	-jut (?-jat)
Russian	-ju	-iš	-it	-im	-iti	-jut/-jat

There are, however, some complications that need to be noted. First, while *-jut* seems to be the usual third person plural suffix in Copper Island Aleut, some sources suggest that *-jat* also occurs, but without indicating the principle(s) governing their distribution. Second, Copper Island Aleut does have at least one consonant-final stem, namely the future auxiliary *bu(u)d-*, which takes present-tense person–number inflections (Sekerina 1994: 25). In the first person singular and in the third person plural it takes *-u*, *-ut*, rather than *-ju*, *-jut* (Sekerina 1994: 28), i.e. Copper Island Aleut may either have this as an irregular verb, or have a generalization that different allomorphs are used after consonant-final stems.<sup>8</sup>

Turning to the past tense, Russian has a radically different set of morphological oppositions, reflecting the fact that the present-day Russian past tense finite forms were originally nonfinite participles. Russian has no person(–number) contrast in the past, only gender and number, with a four-way opposition: masculine singular, feminine singular, neuter singular, and plural. The past tense suffix is *-l*, to which is then added feminine *-a*, neuter *-o*, plural *-i*; absence of any of these

7. The distribution of *-(j)ut* versus *-(j)at* is somewhat different in different dialects of Russian, and even the distribution in standard pronunciation shifted in the course of the twentieth century. There is therefore no guarantee that the distribution was the same in the variety of Russian that contributed to Copper Island Aleut and in the current standard.

8. The distribution in Russian is more complicated, since some consonant-final stems take *-ju*, *-jut*, *-jat*.

additions indicates masculine gender. In Russian, the opposition between unstressed *o* and *a* is neutralized, so that Copper Island Aleut would in any case be expected to lose the distinction between feminine and neuter. However, although gender is not a category of Aleut, the Russian-origin distinction among masculine singular, feminine singular, and plural is present in Copper Island Aleut, which thus ends up with past tense inflections as in (8) (Sekerina 1994: 25, 28), with their Russian etyma below.

(8)	M	F/N	PL
CIA	-l	-l-a	-l-i
Russian	-l	-l-a/-l-o	-l-i

But Copper Island goes one stage beyond this. In (non-mixed) Aleut, it is possible to cliticize non-third person pronominal forms to predicates that would otherwise lack person–number marking (Bergsland 1997: 57). This possibility is carried over into Copper Island Aleut, but using Russian forms, so that the forms in (9) result (masculine forms only given in the singular), with the proviso that cliticization of the subject pronouns is optional (Sekerina 1994: 25).

(9)	1SG	2SG	3SG	1PL	2PL	3PL
	-l-ja	-l-ti	-l	-l-i-mi	-l-i-vi	-l-i

The morphemes that occur in the past tense inflection of Aleut verbs can thus be given morphological glosses as in (10).

(10)	-l	past
	-a	feminine
	-i	plural
	-ja	first person singular
	-ti	second person singular
	-mi	first person plural
	-vi	second person plural

As in an agglutinative system, each formative encodes a single grammatical category. Note that there is no exponent of masculine: Absence of both feminine and plural implies masculine.

Finally under verb morphology, the imperative and infinitive forms may be noted. The Copper Island Aleut imperative takes the suffix *-j*, the infinitive the suffix *-tʃ*, the same suffixes as are used in Russian (in particular, with stem-stressed vowel-final stems) (Sekerina 1994: 25).

As a conclusion to the discussion of Copper Island Aleut verb forms, we may observe that the Russian system, usually characterized as fusional, has been by



and large reanalyzed as agglutinative. Even the tense distinction between present and past can be analyzed in this way: Presence of *-l* indicates past tense, its absence is interpreted as present. The person–number suffixes would, however, have different “allomorphs” in present and past (e.g. second person singular *-iš* versus *-ti*), these suffixes being obligatory in the present, optional in the past.

Finally, it may be noted that Copper Island Aleut has both Russian-origin and Aleut-origin personal pronouns, with the forms in (11) being cited by Sekerina (1994: 23, 24). For a given person–number combination both Russian-origin and Aleut-origin forms are usually available. Some gaps are presumably accidental, but the absence of Aleut-origin nominative pronouns is systematic, since such forms are effectively lacking in (non-mixed) Aleut (Bergsland 1997: 56–59). Although the Aleut-origin forms show some agglutination, in particular in the accusative forms, there is also fusion, indeed suppletion, in the dative form.<sup>9</sup> The Russian forms show considerable irregularities in morphology, including suppletion (the relation between nominative and oblique stems, at least in the first and third persons).

(11)	NOM		ACC		DAT
	SG	PL	SG	PL	SG
	Russian-origin:				
1	ja	mi	min'a	nas	
2	ti	vi	tib'a		tibe
3	on	ani	ivo	ix	
	Aleut-origin:				
1			tiŋ	timis	ŋun
2			tin		
3			tin		

As in many languages, pronoun morphology involves substantial irregularities, although in this Copper Island Aleut does not differ significantly from either Russian or Aleut, and has taken forms directly from both, including instances of fusion and suppletion.

To sum up the discussion of Copper Island Aleut: Where both source languages have fusional morphology, then Copper Island Aleut also allows fusional morphology, as in the case of personal pronouns. Otherwise, there is a strong tendency, perhaps exceptionless in verb morphology, to reanalyze fusional morphology found in only one of the input languages as agglutinative. (In the

9. For an account of the forms in (non-mixed) Aleut, including the reconstructed historical basis of the irregularities, see Bergsland (1997: 67–69).

case of the small amount of fusion found in Aleut-origin possessive suffixes on nouns, this may reflect the fact that Russian simply has no equivalent, rather than having an equivalent that is agglutinative).

## 5. Conclusions

In cases of borrowing of inflectional morphology, tested in this article only against recipient languages that already have inflectional morphology, there is a strong preference for borrowing agglutinative morphology, which includes the reinterpretation of fusional morphology of the donor language as agglutinative (as when Latin plural nominative–vocative–accusative formatives are reinterpreted in English as plural formatives). New stem alternations can be borrowed from the donor language, although in the examples considered here these always involve category pairs (in fact, singular/plural in English and German) where stem alternation is already a possibility in the recipient language. Fusional morphology can be borrowed, as illustrated by Greek-origin formatives combining number and case in Romani, although the new fusional morphology is interpreted in such a way that new fused formatives fill exactly the same positions and roles as are already filled by existing fused formatives.

In the case of mixed languages, possibilities for combining morphological complexities are richer. In Michif, where Cree-origin and French-origin components are rather strictly segregated (clause structure versus noun phrase structure, respectively), so that Cree-origin morphemes rarely have to be attached to French-origin lexical items or vice versa, the complexities of the donor languages are essentially transmitted unchanged. This is also true in the case of Copper Island Aleut personal pronouns, where both Aleut-origin and Russian-origin forms have been transmitted essentially unchanged. But Copper Island Aleut verbs present a rather different case, given that they normally involve combination of an Aleut-origin stem and a Russian-origin inflectional suffix, and here considerable simplification of the idiosyncrasies of Russian verb morphology has taken place, giving rise to a system that is overall agglutinative, thus departing from the largely fusional nature of Russian verb morphology.

In sum, when morphologies collide, there is a tendency to prefer agglutination over fusion, but fusion can survive, in particular in a segregated component of a mixed language (as with Michif) or in patterns that match patterns of fusion already present in the recipient language (as with Romani).

## Abbreviations

AN – animate; CIA – Copper Island Aleut; DAT – dative; DU – dual; F – feminine; FUT – future; GEN – genitive; IMP – imperative; INAN – inanimate; INF – infinitive; INS – instrumental; LOC – locative; M – masculine; N – neuter; NEG – negative; NOM – nominative; NVA – nominative–vocative–accusative; OBL – oblique; PL – plural; PST – past; SG – singular.

## References

- Bakker, Peter. 1997. *A Language of Our Own: The Genesis of Michif, the Mixed Cree-French Language of the Canadian Métis*. Oxford: OUP.
- Bergsland, Knut. 1997. *Aleut Grammar*. Fairbanks AK: Alaska Native Language Center.
- Cysouw, Michael. 2003. *The Paradigmatic Structure of Person Marking*. Oxford: OUP.
- Elšík, Viktor. 2000. Romani nominal paradigms: Their structure, diversity, and development. In *Grammatical Relations in Romani: The Noun Phrase*, V. Elšík & Y. Matras (eds), 9–30. Amsterdam: John Benjamins.
- Elšík, Viktor & Matras, Yaron. 2006. *Markedness and Language Change: The Romani Sample*. Berlin: Mouton de Gruyter.
- Field, Frederic W. 2002. *Linguistic Borrowing in Bilingual Contexts*. Amsterdam: John Benjamins.
- Golovko, Evgenij. 1994. Mednyj Aleut or Copper Island Aleut: An Aleut-Russian mixed language. In *Mixed Languages: Studies in Language and Language Use*, P. Bakker & M. Mous (eds), 113–121. Amsterdam: IFOTT.
- Golovko, Evgenij. 1996. A case of nongenetic development in the Arctic area: The contribution of Aleut and Russian to the formation of Copper Island Aleut. In *Language Contact in the Arctic: Northern Pidgins and Contact Languages*, E. H. Jahr & I. Broch (eds), 63–77. Berlin: Mouton de Gruyter.
- Matras, Yaron. 2002. *Romani: A Linguistic Introduction*. Cambridge: CUP.
- Matras, Yaron. 2003. Mixed languages: Re-examining the structural prototype. In *The Mixed Language Debate: Theoretical and Empirical Advances*, Y. Matras & P. Bakker (eds), 151–177. Berlin: Mouton de Gruyter.
- Menovščikov, Georgij A. 1968. Aleutskij jazyk. In *Jazyki Narodov SSSR V: Mongol'skie, Tungusomančžurskie i Paleaoaziatskie Jazyki*, P. Ja Skorik et al. (eds), 386–406.
- Sekerina, Irina A. 1994. Copper Island (Mednyj) Aleut (CIA): A mixed language. *Languages of the World* 8: 14–31.

# Contact-induced word order change without word order change<sup>\*</sup>

Bernd Heine  
University of Cologne

In studies on language contact, word order has been a prominent topic. The linear arrangement of phrases and words has been shown to be vulnerable when people speaking different languages interact regularly, and there is a wide range of works suggesting that such interactions lead people to adopt arrangements from another language with which they are in contact.

In the present chapter it is argued on the basis of a crosslinguistic survey that it is hard to identify cases where language contact resulted in people creating really new word orders; rather, what appears to happen commonly is that contact induces people to choose among the discourse options that are available in one of the languages in contact one that most readily corresponds to the structures they find in the other language.

**Keywords:** decategorialization, desemanticization, focus construction, grammaticalization, narrowing, pragmatic unmarking, replica language, replication, restructuring, model language, word order

---

\* Work on the present paper was carried out during my stays at various research institutions. I am indebted in particular to the Institute for Advanced Study, LaTrobe University and the Research Centre for Linguistic Typology, Melbourne and its directors Bob Dixon and Sasha Aikhenvald, the Netherlands Institute for Advanced Study (NIAS) and its Rector Wim Blockmans, and to Peter Siemund and Lukas Pietsch and the *Sonderforschungsbereich Mehrsprachigkeit* (SFB 538) of the University of Hamburg. Earlier versions of this paper were presented at the universities of Groningen, Amsterdam, and Hamburg, and at the Max Planck Institute for Evolutionary Anthropology, Leipzig, and I am grateful for the many valuable comments that I received from the audiences there. Finally, I wish to also express my gratitude to Tania Kuteva and Christa König for exchanges on issues leading to this paper.

## 1. Word order in language contact

That word order is among the linguistic phenomena that are most likely to be affected by language contact has been pointed out independently by a number of authors (e.g., Thomason & Kaufman 1988: 88; Thomason 2001c; Winford 2003). Thomason (2001b: 69–71) observes that ignoring vocabulary borrowing, word order is among “the next easiest things to borrow”, and Dryer (1992: 83) sees the effects of linguistic diffusion to be particularly pervasive in the area of word order (see also Nettle 1999: 138; Zeevaert 2006: 2–3). The following is a selection of the many cases that have been named as examples for changes in sentence word order resulting from language contact; for a wealth of additional cases, see Johanson (1992: 254–259):<sup>1</sup>

- Contact with Germanic and Slavic languages, having SVO (= subject – verb – object) order, is said to have been a strong contributing factor in the shift of the Western Finnic and Hungarian languages from SOV to SVO word order (Kahr 1976: 142; Thomason & Kaufman 1988: 55; Thomason 2001b: 88).
- Akkadian, a Semitic language that inherited VSO order from Proto-Semitic, acquired SOV word order under Sumerian influence (Thomason & Kaufman 1988: 55; Thomason 2000, 2001b: 88).
- When the Indo-Aryan language Romani (Romanes) came into contact with languages of the Balkans, it is said to have replaced the verb-final (SOV) order inherited from its Indo-Aryan past by SVO (and VSO), which is characteristic of the Balkan languages (Matras 1996: 64).
- Indic Indo-European languages are claimed to have turned rigidly SOV and rigidly postpositional as a result of Dravidian influence (Kahr 1976: 143).
- The Western Oceanic language Takia is said to have changed from SVO to SOV order under the influence of the Papuan language Waskia (Ross 2001) and, more generally, Austronesian languages are claimed to have changed from SVO to SOV in New Guinea (Thomason & Kaufman 1988: 55).
- The Wutun language of the Chinese family borrowed from Tibetan a rigid verb-final word order and postpositional ordering (Li 1983; Thomason & Kaufman 1988: 92).
- Imitation of Chinese word order is said to have introduced “significant changes into the word order of Japanese” (Miller 1967: 45).

---

1. In the remainder of the paper we will use shorthand phrasings like “language X has changed from Y to Z”. It goes without saying that languages cannot do such things; rather, that it is speakers that are responsible for any changes that may happen.

- Contact with Cushitic languages in northeastern Africa is blamed for a shift from a hypothetical SVO, or the VSO of Proto-Semitic, to SOV in Amharic and other Ethio-Semitic languages (Leslau 1945, 1952).<sup>2</sup>
- The Ma'a language of northeastern Tanzania is claimed to have shifted from SOV to SVO under Bantu influence (Thomason & Kaufman 1988:55).

But contact-induced word order change is by no means confined to sentence structure; it can be found in the same way in noun phrase and other structures, as has been argued for in cases such as the following:

- On the model of Indo-European Balkanic languages such as Macedonian and Albanian, speakers of Turkish dialects on the Balkans have reversed the genitive and its head in possessive constructions; e.g., *babası Alinin* 'the father of Ali' instead of Standard Turkish *Ali'nin babası* (Friedman 2003:61). More examples can be found in dialects of West Rumelian Turkish spoken in Macedonia (Friedman 2003:50ff.).
- Western Oceanic languages commonly have prepositions but Takia has lost the prepositions, having created postpositions on the model of the postpositions of the Papuan language Waskia (Ross 2001).
- Bombay Hindi has switched its question particle from sentence-initial to sentence-final position under Marathi influence (Thomason & Kaufman 1988:98).

Not all of these claims are backed by appropriate empirical evidence. Take the African cases mentioned, for example: We know virtually nothing about the word order of Ma'a prior to its Bantu contacts; accordingly, a claim that Ma'a experienced a word order change from pre-contact SOV to post-contact SVO must remain of doubtful value. And similar observations can be made in other cases; it is in fact widely believed that the Ethio-Semitic languages of northeastern Africa acquired their SOV order from Cushitic (or Omotic) languages; but the evidence is not all that clear, especially in light of the alternative hypothesis that the Ethio-Semitic SOV syntax may be a retention rather than a contact-induced innovation (Grover Hudson, p.c.). In a number of other cases there is no really convincing evidence to establish whether, or how, a hypothesized instance of word order change came about, and I will have little to say about such cases. Still, overwhelmingly, there

---

2. "Examples are the changes in Ethiopic Semitic from VSO, Aux-Verb, Noun-Adjective (probably), and Head Noun-Relative Clause word orders with prepositions to SOV, Verb-Aux, Adjective Noun, and Relative Clause-Head Noun word orders with postpositions to match the patterns of the substrate Cushitic languages" (Thomason & Kaufman 1988:131).

can be little doubt that word order syntax is fairly vulnerable in situations of language contact.

The present paper is meant to show that the linear arrangement of words is in fact a phenomenon that is likely to be affected by language contact. It would seem, however, that what looks superficially like word order change is likely to be a process that does not really lead to a new word order in the language concerned.

## 2. Grammatical replication

Language contact may have a wide range of implications for the languages involved, and it may affect virtually any component of language structure (Thomason & Kaufman 1988). It manifests itself in the transfer of linguistic material from one language to another, typically involving the following kinds of transfer:

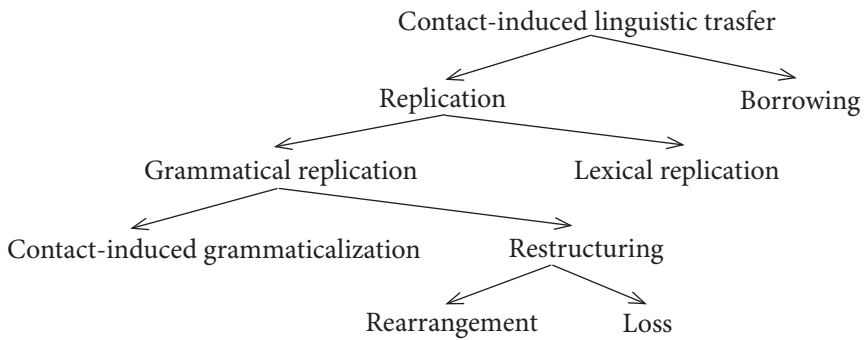
- (1) Kinds of linguistic transfer
  - a. Form, that is, sounds or combinations of sounds,
  - b. Meanings (including grammatical meanings) or combinations of meanings,
  - c. Form-meaning units or combinations of form-meaning units,
  - d. Syntactic relations, that is, the order of meaningful elements,
  - e. Any combination of (a) through (d).

My interest in this paper is with (1d), although we will see that the phenomena looked at cannot be reduced to (1d). Following Weinreich ([1953] 1964: 30–31; see also Heine & Kuteva 2003, 2005, 2006), the terms model language (M) and replica language (R) are used for the languages being, respectively, the source and the target (or the donor and the recipient) of transfer, and his term replication stands for kinds of transfer that do not involve phonetic substance of any kind, that is, for (1b) and (1d) – for what traditionally is referred to with terms such as “structural borrowing” or “(grammatical) calquing”. Thus, by grammatical replication I mean a process whereby speakers create a new grammatical meaning or structure in language R on the model of language M by using the linguistic resources available in R.

The term borrowing is reserved for transfers involving phonetic material, either on its own (1a) or in combination with meaning (1c).<sup>3</sup> Furthermore, I will

---

3. There are many alternative terminologies; for example, Thomason & Kaufman (1988) or Thomason (2001b: 93) use borrowing, source language and receiving language for both kinds of transfer.



**Figure 1.** Main types of contact-induced linguistic transfer.

distinguish between replication restricted to the lexicon (= lexical replication) and replication that concerns grammatical meanings or structures (= grammatical replication). As has been shown in Heine & Kuteva (2003, 2005, 2006), grammatical replication is essentially in accordance with principles of grammaticalization; however, there are a few cases that are not, and the term restructuring has been proposed for the latter. The model of contact-induced transfer used here can summarily be represented as in Figure 1.

Studies on language contact in the past have focused in particular on linguistic areas or sprachbunds. While it is true that these are paradigm products of language contact, for an analysis of grammatical replication they are as a rule of limited value, for the following reason: Sprachbunds, irrespective of whether they concern the Balkans, Meso-America, Ethiopia, or South Asia, are the result of a long and complex history, involving a range of different languages and of historical processes that took place at different periods in the development of the sprachbund (see e.g. Tosco 2000 for the Ethiopian sprachbund), and it remains in many cases unclear which of the factors, historical processes and/or languages exactly contributed what to some particular change. Another important source of information has been seen in creoles, which, like sprachbunds, owe their existence to language contact. But like that of sprachbunds, the history of creoles is the result of an interaction of a variety of different factors, such as the various “substrate” languages, the “superstrate” language(s), possible “adstrate” languages, as well as of a sequence of historical events, and to determine reliably what each of these factors contributed to produce a given grammatical change is more often than not near to impossible. Accordingly, I will have little to say on sprachbunds and creoles and rather concentrate on cases of language contact that took place



more recently and where it is fairly uncontroversial which the model and which the replica language was.

### 3. Strategies

The main purpose of this paper is to study documented cases of what have been argued to be changes of word order induced or influenced by language contact. On the basis of evidence from such cases it would seem that there is a limited range of strategies that jointly can be held responsible for what appear to be instances of “word order change”, namely the following: (a) narrowing (of options), (b) shift from one construction type to another, (c) pragmatic unmarking, and (d) extension and frequency. I will now deal with each of these in turn.

#### 3.1 Narrowing

One way of replicating a word order arrangement found in another language is by narrowing down the range of discourse options available by choosing among the use patterns that are available in the replica language the one that most readily corresponds to the one in the model language and making it the regular one – using it more frequently and in a wider range of contexts.

The following examples may illustrate the central role played by narrowing in the rise of new contact-induced use patterns. Kadiwéu, a Waikurúan language of Brazil, has quite free word order, attested orders being OVS, VOS, SOV, OSV, VSO, and SVO. But Kadiwéu-Portuguese bilinguals translate Portuguese sentences into Kadiwéu with SVO word order, and Thomason interprets this as “an adjustment to the basic SVO word order of Portuguese”<sup>4</sup> (Thomason 2001a: 1642, 2001b: 89). What these bilingual speakers appear to be doing is to turn one of the minor use patterns (SVO) into a major one – one that is used more frequently at the expense of the alternative orders, cf. Figure 2.

That language contact may lead to a narrowing of syntactic options in language contact is also suggested by observations on diaspora speakers of Slavic languages. Sussex (1993: 1020) found that in émigré Slavic languages there is a decline in variety of word order (which he attributes to some extent to a decline

---

4. In Thomason (2003: 700) there is a slightly different rendering of this situation: She observes that bilingual Kadiwéu speakers usually have SVO order when translating sentences from Portuguese. Thomason (2001b: 143) observes that this is not an actual ongoing but “potential change”.



**Figure 2.** Word order in Kadiwéu-Portuguese language contact (Thomason 2001a: 1642).

of case inflections), where SVO order becomes the major use pattern. Already in first-generation Slavic émigré adults this order is more common than in the homeland, but it is overwhelmingly dominant in the speech and written language of children who either emigrated before adulthood or who were born outside their Slavic homeland.

Another example of narrowing is the following: When working on Montana Salish in 1999, Thomason (2001b: 82) asked elders to translate English sentences into Salish, and she was given more translations with English-like SV(O) word order than with the more typical Salish VS(O) order.<sup>5</sup> In naturally occurring Salish texts, SV(O) order is one of the options available to Salish speakers, and it is one that is not especially rare.

In colloquial Turkish, there is a volitive construction with the predicate in the optative mood which can be either left- or right-branching (2). But speakers of the Turkic language Azerbaijani of Iran follow the Persian model by generally using right-branching, cf. (3) (Kıral 2005: 287).

(2) Colloquial Turkish (Kıral 2005: 287)

*buraya            gelsin            istemiyorum.*  
here.DAT    come.OPT.3.SG    want.NEG.PRES.1.SG

OR

*istemiyorum            buraya    gelsin.*  
want.NEG.PRES.1.SG    here.DAT    come.OPT.3.SG  
'I don't want him to come here.'

(3) Azerbaijani (Kıral 2005: 287)

*mən    istemirəm            jëlä            bura.*  
I            want.NEG.PRES.1.SG    come.OPT.3.SG    here  
'I don't want him to come here.'

5. Note, however, that this is an instance of *spontaneous* rather than regular replication: The sentences came from just one elder while another objected, and a third commented, "Well, you could say it this way (i.e. SVO) too" (Thomason 2001b: 82).

Such examples appear to be fairly common in situations of language contact: Speakers of the replica language select among the structural options that are available in their language the one that corresponds most closely to a structure that they find in their model language. What “selection” means is that that option is used more frequently and acquires a wider range of contexts. In the end – that is, in extreme cases – this may turn into the only structure used, eliminating all the other options that used to be available. Narrowing in language contact is by no means restricted to word order phenomena, but it is here where it is perhaps easiest to identify.

### 3.2 Shift from one construction type to another

Another kind of presumed word order change can be portrayed as an epiphenomenal product of a change in construction type in situations of language contact. The strategy that appears to be used is to recruit a construction that matches best the word order arrangement of the model language, even if it serves a function in the replica language that is not exactly that of the model construction.

In Germanic languages, noun-noun compounding is a highly productive mechanism, while French and other Romance languages lack such a mechanism. When speakers of Germanic languages are exposed to intense contact with French they tend to decrease the amount of compounding and to increase the use of attribute patterns on the model of French. The result is that in situations where speakers of languages such as German or Flemish are regularly exposed to French as a dominant language,<sup>6</sup> they tend to shift to some extent from compounding to attributive use patterns. Examples of this change have been found on the one hand in the contact situation between French and Germanic languages in Belgium. German speakers in eastern Belgium may use the German compound *Herbstzeit* (lit.: ‘autumn time’) ‘autumn’, but they also use an attributive/possessive pattern instead, saying *Zeit des Herbstes* (‘time of the autumn’) on the model of French *le temps d’automne* (‘the time of autumn’). On the other hand, there are also examples from South Tyrol, northern Italy, where the main official language is Italian, and where German speakers tend to develop their possessive pattern into a major use pattern where in Standard German compounding would be preferred. Accordingly, in an attempt to replicate the possessive construction of Italian, German speakers say e.g. *das Bündel von Trauben* ‘the bunch of grapes’ instead of *das Traubenbündel* (‘the grape bunch’) on the model of Italian *il grappolo d’uva* (Riehl 2001).

---

6. Concerning the term “dominant language”, see Johanson (1992).

In no case has this bilingual behavior led to category shift, but it appears to have given rise to the preference of one kind of word order over another: Compounding in German and Flemish has modifier-head order whereas the prevailing pattern of attributive possession is head-modifier. While there are no quantitative data at my disposal, there is reason to assume that among these German and Dutch speakers attributive possession, and hence head-modifier order, is more frequent than compounding and the modifier-head syntax associated with it.

Another kind of shift from one construction type to another is provided by Matras & Sakel (2007) on Domari, an Indo-Aryan language spoken in the Middle Eastern Dom area. The principal contact language of Domari speakers is Arabic. In Arabic, adjectives follow the noun they modify, cf. (4), while in Domari they precede, cf. (5a). Now, Domari has a predicative construction, illustrated in (5b), where the adjective follows the noun and is in turn followed by an enclitic predicative marker (PRED). This predicative construction is often preferred when introducing nominal entities in Domari discourse, as it matches the word order of the Arabic noun-adjective construction, cf. (5c).

- (4) Arabic (Afroasiatic; Matras & Sakel 2007)

*l- walad l- kbīr*  
 DEF- boy DEF- big.M  
 'the big boy'

- (5) Domari (Indo-Aryan; Matras & Sakel 2007)

- a. *till- a zara*  
 big- M boy  
 'the big boy'
- b. *zara till- ēk.*  
 boy big- PRED.M  
 'the boy is big.'
- c. *er- a zara till- ēk.*  
 came- M boy big- PRED.M  
 'The big boy arrived.' (Lit.: 'The boy, being big, arrived.')

Obviously, the only reasonable motivation underlying this behavior of Domari speakers must have been to find an equivalent to the linear arrangement of Arabic. The evidence that this behavior was contact-induced is of the following kind. First, Arabic is an important L2 for Domari speakers of this community, second, there are no postposed nominal modifiers in Domari, and third, Domari structure has been influenced by Arabic in a number of other ways as well.

Like the preceding example, this case does not represent a change from one form of categorization to another; rather, it also involves a change in preference of one construction over another. But unlike the preceding case, this one shows that

a construction can be used for quite a different purpose, in that language contact appears to be responsible for the extension of a predicative construction to also serve as a nominal modifier.

The Aztec language Nahuatl provides an example of a gradual shift from one construction to another, apparently in an attempt to match the word order alignment of the model of the dominant language Spanish: There is a process that, at least on the surface, may be interpreted as leading from pre-verbal to post-verbal marking of object nouns. In Classical Nahuatl, incorporating nominal objects placed pre-verbally was highly productive, as illustrated in (6). In modern Hispanized varieties of Nahuatl, incorporation is disfavored, the preferred construction being one that also was available, where the object follows the verb, as in (7) (Flores Farfán 2004). There is little doubt that this change was influenced by contact with Spanish, although language attrition is also held responsible for the process.

- (6) Classical Nahuatl (Flores Farfán 2004: 86)

*ni-        xo chi-    te moa- Ø-        Ø.*  
1.SG-    flower- seek-    PRES-    SG  
'I seek flowers.'

- (7) Hispanized Xalitla Nahuatl (Flores Farfán 2004: 86)

*ni-        teemoa-    Ø-        Ø    xoochi-    meh.*  
1.SG-    seek-        PRES-    SG    flower-    PL  
'I seek flowers.'

Once again there is no real word order change; rather, what happened is that a productive pattern loses in productivity in favor of another existing pattern that matches the word order of the Spanish model. As Flores Farfán (2004: 86) argues, with the activation of post-verbal object placement, the old construction of object incorporation has acquired a new significance in monolingual varieties of Nahuatl, serving "topicalization" (presumably more appropriately: focus marking), in that in such varieties (7) would be unmarked whereas (6) could be translated as 'It is flowers what I'm looking for'.

### 3.3 Pragmatic unmarking

As we saw in Section 3.2, language contact may induce people to "copy" arrangements of meaningful elements from one language into another, in accordance with (1d) (Johanson 1992, 2002). But the "new word order" does not arise as a result of (1d) but rather of (1b) without there actually being any word order change. This is the case, for example, when in the process of grammaticalization some structure is reinterpreted as some other structure, with the result that a seemingly

new word order arises. We had cases of grammaticalization above, like when we were dealing with a construction of predication that is pressed into service to mark nominal modifiers in Domari. But perhaps the main driving force for adjusting one's word order to that of another language is to select a pragmatically marked use pattern that exhibits an ordering corresponding to that of the model language and to grammaticalize that pattern into an unmarked syntactic pattern; note that a development from pragmatically marked to syntactic constituent is a fairly common grammaticalization process (Givón 1979a, 1979b, 1979c, 1984, 1995).

The present section will illustrate the role of pragmatics in grammatical replication by looking at a number of cases of language contact. I will first deal with phrase structure and subsequently with sentence word order.

### 3.3.1 *Phrase structure*

**Noun modifier.** An incipient shift in the order of attributive possession can be observed among certain groups of Russian L1 speakers in Finland, as studied by Leisiö (2000). In Finnish (and Finland Swedish) the genitive modifier precedes its head in attribute possession, while in Russian, the genitive modifier follows the head but – and this is relevant to our discussion – in colloquial speech, the reverse order is more common when the possessor is a specific person and/or is in focus. Leisiö distinguishes two groups of Russian speakers in Finland: On the one hand there are what she calls “dialect speakers” (i.e. the so called Kyyrölä Russians), whose speech has Northern Russian dialect features; on the other hand there are speakers of “nondialect Russian”. The former group differs from the latter in particular in the fact that it has a long history of close contact with Finnish. A quantitative analysis of text corpora of the two groups shows a significantly higher frequency of Finnish-type modifier-head uses for “dialect speakers”. As can be seen in Table 1, modifier-head order is predominant among “dialect speakers” (88%), while among “nondialect speakers” the Standard Russian order head-modifier prevails (59%).

What the analysis by Leisiö (2000) suggests is, first, that intensity of language contact with Finnish has resulted in a significantly higher rate of word order of Finnish-type genitive constructions, and second, that the change involved was not

**Table 1.** Relative frequency of genitive and possessive-adjective constructions in two groups of L1 speakers of Russian in Finland (based on Leisiö 2000: 309).

	Head-modifier	Modifier-head	Total (Absolute figures)
“Nondialect speakers” (40)	59%	41%	100% (141)
“Dialect (Kyyrölä) speakers” (28)	12%	88%	100% (103)

one from one word order to another but rather from minor (= less frequent) to major (= more frequent) use pattern, and, third, that the change has a pragmatic base: The preposed modifier position in colloquial Russian is said to be associated with animacy, prominence, and/or focal information. In other words, what used to be a pragmatically marked placement appears to have turned into the normal, that is, unmarked one in “dialect speakers” of Kyyrölä Russian.

It would seem that this is not an isolated case. Like Russian, Polish has head-modifier order in genitive constructions but in spoken Polish, the modifier is sometimes preposed, especially when the noun refers to a person, e.g., *naszego kolegi siostra* ‘our friend’s sister’. This modifier-head pattern appears to gain in frequency among émigré Poles in close contact with English; thus, Sussex observes:

One also finds copies of English word order in structures like preposed possessives in émigré Polish: *mojej siostry tata* ‘my sister’s father’ (standard Polish: *tata mojej siostry*). (Sussex 1993: 1020)

Molisean, the variety of a Croatian minority in Molise, southern Italy, provides a similar example of structural change within the noun phrase. Rather than preceding the nominal head, nominal attributes in Molisean tend to be postposed, thereby matching the structure of the Italian model language. That we are dealing with a contact-induced change is suggested e.g. by the fact that Italian may use word order to express a functional distinction between a differentiating (postposing) and descriptive (preposing) use of attributes. According to Breu (1996), exactly this distinction has been replicated by Molisean speakers, cf. (8).

(8) Molise Croatian (Breu 1996: 31)

- a. *jena mala hiža*  
one big house  
‘a big house’
- b. *jena hiža mala*  
one house big  
‘a big house (not a small one)’
- c. *una casa grande*  
‘a big house’ (Italian)

Evidence in support of a contact hypothesis is provided by the following observations: First, speakers of this Slavic variety have had a 500-years history of intense contact with the host language Italian; second, Molise Croatian has replicated a wide range of Italian structures as a result of this contact and, third, as Breu (1996)

argues convincingly, an account of this change of preference of head-modifier order in terms of an inclusively internal development is implausible.

The following is a more complex example of grammaticalization of a pragmatic structure into a new syntactic structure. Ross (1996, 2001) describes a situation where two genetically unrelated languages spoken on Karkar Island of the north coast of Papua New Guinea have become semantically and syntactically largely intertranslatable while each of the two has retained its own lexical material – a situation he proposes to call *metatypy*. The model language is Waskia, a Papuan language of the Trans-New Guinea type, and the replica language Takia, a Western Oceanic language of the Bel family of the North New Guinea cluster. In the process of contact, “Takia speakers have increasingly come to construe the world around them in the same way as the Waskia” (Ross 2001: 144). In an attempt to assimilate their language to the Papuan language Waskia, speakers of the Western Oceanic language Takia largely adopted the syntax of the Papuan language Waskia. For example, while in Proto-Western Oceanic the determiner (article) preceded the head noun, cf. (9a), it follows the head noun in both Takia (9b) and Waskia (9c).

(9) Determiner – noun order (Ross 2001: 142)

- |    |                         |              |                           |
|----|-------------------------|--------------|---------------------------|
| a. | Proto-Western Oceanic   | <i>*a</i>    | <i>tam<sup>w</sup>ata</i> |
|    |                         | DET          | man                       |
|    |                         |              | ‘the man’                 |
| b. | Takia (Western Oceanic) | <i>tamol</i> | <i>an</i>                 |
|    |                         | man          | DET                       |
|    |                         |              | ‘the man’                 |
| c. | Waskia (Papuan)         | <i>kadi</i>  | <i>mu</i>                 |
|    |                         | man          | DET                       |
|    |                         |              | ‘the man’                 |

The way Takia changed from preposed to postposed determiner is the following: Proto-Western Oceanic had a set of three deictic morphemes, one of them being *\*a* (‘that’, near speaker). When one of these was used attributively, it followed the Proto-Western Oceanic adjective syntax, taking a pronominal suffix agreeing in person and number with the head noun, the result being a structure as in (10a). This structure underwent a canonical grammaticalization process from demonstrative to definite determiner, resulting in the Takia structure (10b) (= (9b)). In this process, the construction was subject to the usual mechanisms of grammaticalization, that is, loss of deictic force (desemanticization), of its status as an inflectable constituent (decategorialization), and loss of the preposed article *\*a* and reduction of the postposed determiner *\*a-ña* > *an* (erosion).



## (10) Grammaticalization (based on Ross 2001: 142)

- |                          |              |                           |           |           |
|--------------------------|--------------|---------------------------|-----------|-----------|
| a. Proto-Western Oceanic | <i>*a</i>    | <i>tam<sup>w</sup>ata</i> | <i>a-</i> | <i>ña</i> |
|                          | DET          | man                       | that-     | 3.SG      |
|                          |              | 'that man'                |           |           |
| b. Takia                 | <i>tamol</i> | <i>an</i>                 |           |           |
|                          | man          | DET                       |           |           |
|                          | 'the man'    |                           |           |           |

As the description by Ross suggests, the change in Takia from preposed to postposed determiner did not involve any change in the order of constituents; rather, in order to adapt to the postposed determiner order of the model language Waskia, Takia speakers drew on a construction that was available to them, namely what appears to have been a pragmatically marked postposed deictic determiner, and they grammaticalized it into their new determiner.

This example illustrates a process that we observed already earlier: Following a pattern in the model language, speakers draw on some existing use pattern in the replica language that corresponds most closely to the model, frequently one that until then was more peripheral and of low frequency, and they activate it – with the effect that that peripheral pattern turns into the regular equivalent of the model, acquires a high frequency of use, and eventually may emerge as a fully grammaticalized category, equivalent to the model category.

**From prepositions to postpositions.** Prepositions may change diachronically into postpositions and vice versa. But do they really do so in language contact? Another case of “word order change” discussed by Ross (1996, 2001) concerns a set of postpositions exhibiting a similar semantic patterning in the two languages. Western Oceanic languages commonly have prepositions but Takia speakers have lost the prepositions, or use them no longer productively. In an attempt to replicate the postpositions of Waskia, Takia speakers developed postpositions by grammaticalizing inalienably possessed relational nouns. In this way, a Proto-Western Oceanic construction illustrated in (11) turned into a postpositional construction (12) in Takia (note that Takia has given up the possessee-possessor order of Western Oceanic and adopted the possessor-possessee order of Waskia). In accordance with mechanisms associated with grammaticalization, the prepositional phrase *\*i lalo-ña* lost its nominal structure and turned into an adposition (decategorialization) and was phonetically reduced to *lo* (erosion). Thus, in order to establish equivalence with the model language Waskia, speakers of the replica language Takia had recourse to a grammaticalization process. One crosslinguistically fairly widespread conceptual schema leading to the rise of new possessive constructions is referred to by Heine (1997: 144) as the Topic Schema, taking the form [(As for) X, X's Y] (e.g., something like *I saw John, (actually) his car*), where

the possessee follows the possessor and agrees with the latter in the form of a possessive attribute, and turn it into a possessive construction (*John's car*). In using this schema as a pragmatically marked option, Takia developed a relational noun for 'inside' into locative (inessive) adposition.

- (11) Proto-Western Oceanic (Ross 1996: 189, 2001: 143)

*\*i lalo- ña a Rumaq*  
 PREP inside- its ART house  
 'inside the house'

- (12) Takia (Western Oceanic; Ross 1996: 190, 2001: 143)

*ab lo*  
 house in  
 'in the house'

Predictably, in languages using this schema the possessee follows, rather than precedes, the possessor – irrespective of any word order constraints that may characterize the language concerned. It would seem that Takia speakers drew on this schema, with the effect that the possessee follows the possessor. Now, when the possessee phrase *\*i lalo-ña* was grammaticalized, the expected result was a postposition on the erstwhile possessor rather than a preposition. This suggests that Takia's history from prepositions to postpositions did not involve any word order change; rather, the prepositions were lost and the postpositions arose via the creation of a new possessive construction. While this was seemingly a language-internal process, we concur with Ross that it was at the same time contact-induced, in that the choice of the Topic Schema provided Takia speakers (or their ancestors) with a strategy to match the postpositional structure of the model language Waskia.

**Genitive word order alignment.** That the Topic Schema provides an important means in the grammatical replication of word order characteristics can also be demonstrated with the following examples on attributive possession. Unlike fellow Western Oceanic languages, such as Arop-Lokep in (13), Takia (14) follows the Papuan language Waskia in placing the possessor before the possessee (15), and once more it seems that it was the Topic Schema that allowed them to replicate the possessor-possessee syntax of Waskia: This schema appears to have provided the most convenient tool for preposing the possessor, where (14) can be paraphrased diachronically roughly as '(as for) Kai, his house'.

- (13) Arop-Lokep (Western Oceanic; Ross 2006)

*rumu ke tool in*  
 house ABL man that  
 'that man's house'

- (14) Takia (Western Oceanic; Ross 2006)

*Kai sa- n ab*  
 Kai CLASS- his house  
 'Kai's house'

- (15) Waskia (Papuan; Ross 2003: 184, 2006)

*Kai ko kawam*  
 Kai ABL house  
 'Kai's house'

A strikingly similar situation can be found in the Qashqa-Darya dialect of Arabic spoken in Uzbekistan. Having been separated from the Arabic-speaking world for many centuries, this dialect (together with Bukhara, another Arabic dialect; see below) has been in contact with the Indo-European languages Tajik and Dari and the Turkic languages Uzbek and Turkmen; Qashqa-Darya speakers are particularly fluent in Uzbek. Chikovani (2005) notes a number of contact-induced changes that Qashqa-Darya Arabic has undergone under the influence of Uzbek. In attributive possession, the head-modifier order of Arabic is still widely used (16a). But under the influence of Uzbek, a new order possessor-possessee has evolved via the grammaticalization of the Topic Schema, where the possessor precedes and is cross-referenced on the possessee by means of a possessive suffix (16b).

- (16) Qashqa-Darya Arabic (Chikovani 2005: 131)

- a. *şōḥb il-bāgir*  
 owner the-cow  
 'the owner of the cow'
- b. *äfāndi morta*  
 effendi wife.his  
 'the effendi's wife'

### 3.3.2 Sentence order

**From SOV to SVO: West Rumelian Turkish.** One effect that the change from minor to major use pattern may have is that it leads from what is usually described as a marked structure to an unmarked structure. Turkish is commonly portrayed as a verb-final language, but there are pragmatically marked structures where the verb precedes its complement. For example, Friedman (2003: 66) notes that a sentence such as (17a), where the verb occurs in non-final position, would have a pragmatically marked meaning like 'It is Erol who is the good student' in Standard Turkish. But Turkish varieties spoken on the Balkans have been strongly influenced by Balkan languages, and in West Rumelian Turkish dialects spoken in Macedonia, the above sentence would be an unmarked sentence equivalent of

English ‘Erol is a good student’ – corresponding to the unmarked Macedonian sentence (17c). This suggests that under the influence of Macedonian and perhaps other Balkanic verb-medial (= SVO) languages, speakers of West Rumelian Turkish dialects have developed one of their pragmatically marked minor use patterns into an unmarked pattern – thereby establishing syntactic equivalence with the language or languages of their Balkanic neighbors. Consequently, in these Turkish dialects of Macedonia the verb occurs far more frequently in a non-final position than it does in Standard Turkish.<sup>7</sup>

- (17) Turkish and Macedonian (Friedman 2003: 66; no glosses provided)
- |    |                                       |                       |
|----|---------------------------------------|-----------------------|
| a. | <i>Erol’ dur iyi öğrenci.</i>         | Standard Turkish      |
|    | ‘It is Erol who is the good student.’ |                       |
| b. | <i>Erol’ dur iyi öğrenci.</i>         | West Rumelian Turkish |
|    | ‘Erol is a good student.’             |                       |
| c. | <i>Erol e dobar učenik.</i>           | Macedonian            |
|    | ‘Erol is a good student.’             |                       |

That language contact was a contributing factor in this development is suggested, first, by the fact that West Rumelian Turkish has been influenced massively by Macedonian, second, that an internal development from SOV to SVO is unlikely in this case and, third, that using a pragmatically marked option to establish an equivalence relation with the word order of the model language appears to be a crosslinguistically common strategy.

**From SOV towards pragmatically unmarked order in Eskimo.** Rather than “word order change”, certain varieties of Eskimo are experiencing a new orientation in presenting sentence participants as a result of language contact. Fortescue (1993) describes “standard Eskimo narrative discourse”, with special reference to West Greenlandic, in terms of the word order arrangements listed in (18), where X stands for adverbial material modifying the verb, and the optional constituents are: P1 = an NP with the function of a given topic (anaphoric), Pø = foregrounded/highlighted material, which may include e.g. a new topic NP or heavy material modifying a head NP, or other material expressing high “newsworthiness”, and Tail = afterthought or clarifying/elaborating material. He observes that deviations

---

7. In more general terms, Johanson (1992:255) observes: “In türkischen Sprechsprachen des balkanischen Areals hat das ständige Kopieren die Wortstellung so umgestaltet, dass z.B. Komplemente und freie Angaben gewöhnlich dem Prädikatskern folgen. Besonders das Gausische ahmt die Rechtsverzweigkeit der umgebenden slawischen Sprachen und des Rumänischen nach [...]” [In Turkic languages of the Balkan area, constant copying has transformed word order to the extent that, for instance, complements and adjuncts usually follow the predicate nucleus. Gausic in particular imitates the right branching of the surrounding Slavic languages and of Rumanian].

from the order in (18a) are statistically not very frequent, while deviations from (18b) and (18c) are not tolerated, at least not in West Greenlandic.

(18) Neutral word order in Eskimo (Fortescue 1993)

- a. (P1) S-O-X-V (Pø) (Tail)
- b. possessor – possessee
- c. head – modifier

Fortescue observes that there are also what he calls “non-neutral” orderings deviating from the neutral orders in (18), which he finds to be particularly frequent in Eskimo varieties most strongly exposed to contact with English, characterized by a high degree of bilingualism, that is, in North Alaska Inupiaq, Central Alaskan Yup’ik, and Canadian Inuktitut, and they are less pronounced in older than in younger texts. In the former two varieties, the orders VO, possessee – possessor, and modifier – head, all deviating from the orders in (18) but corresponding to the order in English, appear to be common and to lack “contextual markedness” (Fortescue 1993:282). That English contact influence is a contributing factor is suggested also by the observation that it is in recent texts of Canadian Inuktitut “from areas where English has made the strongest inroads that the proportion of ‘non-neutral’ clausal orderings increases dramatically” (p. 283). Fortescue concludes:

We can return now to the paradoxical suggestion that the influence of rigidly SVO English, in those areas where its impact has been great, should have resulted in ‘freer’ word order. What seems to have happened there is that the original fairly labile but by no means ‘free’ word order principles of Eskimo have been nudged (through the mechanism of bilingualism) in the direction of favouring those patterns, once contextually ‘marked’, that correspond to common English patterns, to the detriment of those that are alien to English. The effect of English in these areas has thus been to loosen up the word order, obscuring the pragmatic factors that originally determined the relative positioning of clausal constituents when ‘neutral’ conditions did not apply, but without imposing its own rigid ordering patterns on the receiving language. (Fortescue 1993:285)

The main factor in these changes was apparently the weakening of pragmatic significance attached to word order. It seems that there is a new use pattern VO in Alaskan Eskimo where post-verbal objects are no longer given special emphasis, by being foregrounded/highlighted, that is, where this ordering is less pragmatically than syntactically motivated. These changes are clearly more marked in the spoken than in the written forms of language use, and they are more marked in the Eskimo-English contact situations of western Canada and Alaska than in Eskimo-Danish contacts in West Greenland, where the colonizing language at no point has “been forced upon the native population as has been the case at various

times and places in North America” (Fortescue 1993:287). The overall effect of these changes is not that Eskimo has acquired a new word order but rather that the pragmatics of information structure has lost its significance and that linear arrangement now is looser and more readily corresponds to arrangements in the model language English.

**From VO to OV via topicalization in Bukhara Arabic of Uzbekistan.** That speakers react to language contact by molding pragmatically marked use patterns in the direction of the model language can also be shown with the following example. Ratcliffe (2005:143–145) argues that in the Bukhara dialect of Arabic of Uzbekistan there is a common syntactic OV pattern where transitive verbs have an encliticized pronoun referring back to a nominal object, cf. (19), and he notes: “This type of construction is by no means alien to other forms of Arabic, where a word can be topicalized by being moved to the first position of a sentence, with its syntactic role indicated by a resumptive pronoun”, as in the Egyptian example of (20).

- (19) Bukhara Arabic (Ratcliffe 2005:144)

*akina xadā- ha.*  
knife (he) took- it  
‘He took a knife.’

- (20) Egyptian Arabic (Ratcliffe 2005:145)

*il- fustān gibt- u.*  
the dress I.got- it  
‘I got the dress.’

What appears to have happened in Bukhara Arabic is that a topicalization strategy within VO syntax was grammaticalized to a pragmatically unmarked OV syntax, with “the reanalysis of a resumptive pronoun as a verbal inflection agreeing with the object” (Ratcliffe 2005:145). The result is a pattern that matches the OV order of the model languages Tajik and Uzbek, and Ratcliffe (2005:145) comments that the “... shift from VO to OV is not immediate, but is mediated through a variant word-order pattern available to the language in the stage where VO is the unmarked order”.

**From VSO to SVO in Breton.** There is an abundance of information on how in situations of contact one language may adopt syntactic properties of another language. But it may as well happen that the replication of a grammaticalization process has some effects on the syntactic structures concerned, leading e.g. to the rise of new syntactic relations and new arrangements of meaningful elements.

Breton differs from fellow Celtic languages in a number of respects. Geographically, it is the only modern Celtic language spoken in mainland Europe.

Linguistically, it has some properties that also set it apart from other insular Celtic languages. One of these properties is the presence of a fully grammaticalized possessive perfect (or ‘have’-perfect) akin to what is commonly found in mainland western Europe. Another property concerns word order syntax: While the Celtic languages Irish, Scottish Gaelic, and Welsh are characterized by a verb-initial (VSO) syntax, Breton is not. For Breton, a number of different descriptive taxonomies have been proposed, ranging from one in terms of an underlying VSO-syntax with surface SVO-structures, to one in terms of a basic SVO-syntax with relics of VSO-structures (see Ternes 1999). Irrespective of how one wishes to categorize the overall structure of Breton clauses, it seems fairly uncontroversial to say, first, that Breton has salient SVO-structures that are absent in its closest relative Welsh in particular and in other Celtic languages in general. The following examples from Welsh (21a) and Breton (21b) illustrate this difference.

(21) Celtic languages (Ternes 1999: 238)

a. Welsh

*Mae 'r tywydd yn braf.*  
 is the weather AP nice  
 ‘The weather is nice.’

b. Breton

*An amzer a zo brav.*  
 the weather that is nice<sup>8</sup>  
 ‘The weather is nice.’

Second, it also appears uncontroversial to argue that this difference is the result of more recent developments whereby Breton became typologically dissimilar from other insular Celtic languages, increasingly replacing its earlier VSO-syntax by SVO-structures. While this is seemingly an instance of syntactic change from one kind of word order to another, thereby making Breton syntactically similar to the majority language French, as a matter of fact it is not; rather, it is the result of a pragmatic process whereby a focus construction, having the structure of a (bi-clausal) cleft construction, is grammaticalized to a new (mono-clausal) syntactic pattern where the (focalized) subject is placed sentence-initially (see Harris & Campbell 1995: 155–157; Ternes 1999 for details). Thus, a sentence of the form (21b) is diachronically derived from a cleft construction of the type (22).

---

8. Ternes (1999: 238) describes the particle *a* as a “verbal particle” (VP). I follow Harris & Campbell (1995: 155–156) in glossing it as a relative clause marker (see below).

## (22) Focus construction

[It is X [that Y]]

where [It is X] = copular matrix clause, and [that Y] = relative clause.

That this reconstruction is correct is suggested in particular by the fact that the verb phrase in sentences such as (21b) appears to be historically a relative clause: Throughout the history of Breton, the particle *a* is used in relative clauses in which the subject or direct object is relativized, and *zo* (or *so*) is a special third person singular form of the verb 'be', which originated in relative clauses (Harris & Campbell 1995:155). Thus, colloquial Breton has replaced its earlier verb-initial clause structure by a cleft construction where the subject noun phrase is focalized ([It is X]) and the rest of the sentence takes the form of a relative clause ([that Y]). While this grammaticalization process is still in its intermediate stages in written Breton, colloquial (dialectal) varieties of Breton are described as presenting the final stages of the process (Ternes 1999:248).

But this process does not appear to have happened in a vacuum; rather, it also took place in the general area of Romance languages where Breton is spoken, specifically in colloquial varieties of French and in Gascon. In Gascon it has reached a stage of grammaticalization that is not unlike that found in Breton (Haase 1997; Ternes 1999:248–249). Thus, the grammaticalized Breton structure has its equivalents in these Romance languages (Wehr 1984:86f.; Haase 1997:218), as the following examples (23) and (24) illustrate, where the subject and the verb phrase are connected by what is diachronically a relative clause marker (The asterisk in the glosses indicates that the forms *que* and *qui* are hypothesized to be historically relative clause markers):

## (23) Colloquial French (Wehr 1984:79)

*Ton nez qui coule.*

your nose \*REL run

'Your nose is running.'

## (24) Gascon (Haase 1997:218f.; cited from Ternes 1999:249)

*Lo monde que van tribalhar.*

the world \*REL go work

'The people go to work.'

Given the fact that Breton has been strongly influenced by its Romance-speaking neighbors, there is reason to adopt the hypothesis suggested by Ternes (1999) according to which Breton speakers replicated a grammaticalization process that



they observed in colloquial French and perhaps other Romance varieties.<sup>9</sup> This hypothesis is based on the following facts:

- a. This process did not take place in languages genetically closely related to Breton, that is, in other insular Celtic languages.
- b. Breton shares a grammaticalization process with its immediate Romance neighbors.
- c. Since Breton is known to have had an extended period of language contact with its Romance neighbors, resulting in massive linguistic transfers, it appears plausible that we are dealing with yet another instance of contact-induced transfer.
- d. The grammaticalization of bi-clausal focus (cleft) constructions to monoclausal constructions is cross-linguistically not entirely uncommon (see e.g. Heine & Reh 1984; Harris & Campbell 1995: 152–162; Harris 2003); still, if it is found in neighboring languages then language contact offers the most plausible hypothesis to account for this fact.

To conclude, Breton is experiencing a change that makes it typologically more similar to French, the major second language of the region where Breton is spoken, bringing it in line with the SVO word order of French. But this change in word order appears to have been an epiphenomenal product of something else, namely a shared grammaticalization process that appears to have been responsible for this typological alignment from pragmatic to syntactic marking.

### 3.4 Extension and frequency

Of all the factors discussed in this paper, the most pervasive effects on grammatical replication can be seen in the extension of existing structures to new contexts and in an increased frequency of use. These may be concomitant results of other factors such as pragmatic forces (Section 3.3), but they may as well be the only effects of language contact, and in the present section I am concerned with the latter.

---

9. There is an alternative hypothesis according to which the grammaticalization that Breton underwent is not the result of transfer from Romance languages but can be traced back to an earlier Celtic structure (Jost Gippert; p.c.). While such a possibility cannot entirely be ruled out, the evidence available suggests that the Breton structure cannot be traced back to earlier Celtic; rather, it is the result of processes that happened more recently, that is, within the last millennium after the split of Breton from its insular Celtic relatives. Nevertheless, even if one were to assume that that latter hypothesis is correct, this would not change the basic fact that there was a transfer of a grammaticalization process from one language to another – be it from Romance to Celtic or the other way round.

What frequently happens is that speakers draw on a minor use pattern – one that has a more marginal status, being used rarely and/or only in specific contexts only to build a new major use pattern by increasing the frequency of use and extending the range of contexts in which it may occur.

**From SVO to free word order and SOV preference.** In the following example, a Turkic language, Uzbek, was not the replica but the model language. As we noted above, the Qashqa-Darya dialect of Arabic, spoken in Uzbekistan, has been separated from the Arabic-speaking world for many centuries. Together with Bukhara, another Arabic dialect, it has been in contact with the Indo-European languages Tajik and Dari and the Turkic languages Uzbek and Turkmen, but especially with Uzbek; note that Qashqa-Darya speakers are particularly fluent in Uzbek. The description by Chikovani (2005) suggests the following changes, which he attributes to the influence of the verb-final Turkic language Uzbek: The Qashqa-Darya speakers, first, acquired a free word order, where the main constituent orders are SVO, OVS, and SOV, and, second, the verb-final order SOV, illustrated in (25), has become the most frequent word order type in this Arabic dialect.

(25) Qashqa-Darya Arabic (Chikovani 2005: 131)

*boy i bint hušrūya gāl-ki.*  
 bey girl beautiful said  
 ‘The bey said to the beautiful girl ...’

This case can be considered to be another instance of narrowing (Section 3.1), in that out of a range of different word orders available, one that immediately corresponds to that of the model language is selected as the primary order. The reason for treating it here is that on the basis of the data available it would seem that frequency of use was the primary factor for shaping the status of word order alignment.

**Extension of preposed adjective order.** In concluding, one may mention an example of a gradual shift from postposing to preposing adjective placement. In the contact situation between English and the French/Norman dialect Guernésiais on the Channel Island of Guernsey, speakers of Guernésiais appear to be replicating the preposed-adjective order of English by expanding the existing use pattern of adjective – noun ordering. Guernésiais speakers tend to prepose adjectives distinctly more frequently than would be expected historically: Jones (2002: 154) found 70% of all adjectives to be preposed. In this case, no grammaticalization from one meaning or pragmatic function to another appears to have been involved; rather, what happened is that an existing use pattern was extended. That contact played a role in this change is suggested e.g. by the fact that not only

simple adjectives but also compound adjectives, which are likely to follow the noun in Mainland Norman, are being preposed, e.g.,

- (26) Guernésiais (Jones 2002: 156)  
*ses anti-rouoyalistes principes*  
'his anti-royalist principles'

#### 4. Conclusions

In the survey data discussed in this paper we have not come across a single case where speakers really produced a word order entirely alien to the language concerned – hence, there is not really a change from one word order to a new order. While I do not wish to claim that the latter is impossible, this does not seem to be what speakers normally do. What they do, rather, is to recruit existing structures, redefine them, and create new structures that mirror the word order characterizing the model language.<sup>10</sup> The main strategies for creating new word order arrangements matching those of the model language are the following:

- a. Select among the word order alternatives that exist one that matches the order to be found in the model language.
- b. Use an existing construction and assign it a new function.
- c. Use a pragmatically marked construction and develop it into a pragmatically unmarked construction.
- d. Extend an existing use pattern to new contexts.
- e. Use an existing use pattern more frequently.

This classification, meant primarily for descriptive purposes, is far from satisfactory, especially since the strategies are by no means mutually exclusive; on the contrary, there is almost invariably more than one strategy involved in a given case of “word order change”. But more importantly, there appears to be an entailment relationship in that – as far as the data that we were able to survey suggest – (a), (b), and (c) entail (d) and (e), in that the latter two are involved in some way or other in all other strategies. Conceivably, these strategies are different manifestations of a more general process; more research is required on this issue.

---

10. Note the following observation made by Johanson (1992:255): “Bevor Schlüsse auf Fremdeinfluß im Sinne kopierter Wortstellungsmuster gezogen werden, sind jedoch die bereits vorhandenen sprachinternen Variationsmöglichkeiten unbedingt zu beachten”. [Before conclusions can be drawn on copied word order patterns, it is necessary to take the already existing language-internal possibilities of variation into account].

While the speakers' goal underlying the strategies described in this paper is apparently to develop some form of correspondence or equivalence relation between the replica and the model language, the outcome of the process can be at variance with this goal. For example, we observed that some communities of Eskimo speakers in North America have changed their word order behavior in the direction of the model language English; but rather than moving towards the fairly rigid SVO order of English, the result was 'freer' word order in Eskimo.

In more general terms, the observations made in this paper suggest that speakers recruit material available in R (the replica language) to create new structures on the model of M (the model language) and that, rather than being entirely new, the structures created in R are built on existing use patterns and constructions that are already available in R. Accordingly, even if there is a clearly syntactic goal, such as copying a word order characteristic of another language, the strategy employed to achieve this goal is not really syntactic but rather semantic or pragmatic in nature.

Fourth, the observations made in this paper show why grammatical replication is – and presumably will remain – a notoriously controversial field of study: The grammatical changes described in this paper concern processes of grammaticalization that could in principle have happened as well *internally*, that is, without language contact – in other words, it is not possible to “prove” that contact was a contributing factor. Nevertheless, I hope to have demonstrated that in the cases examined contact must have been involved, either as a triggering or an accelerating factor, or both.

And finally, the discussion may have shown that there appears to be one fundamental difference between the two main types of contact-induced transfer, namely borrowing and replication (see Figure 1). Whereas the former involves a transfer of substance from one language to the other, where “substance” may take the form of loanwords, borrowed phonetic units or properties, etc., there is not really any transfer in the case of replication; rather, what speakers do is that they employ the grammatical means that are available in the replica language in order to create novel structures that correspond, or are believed to correspond, to “appropriate” structures in the model language.

## Abbreviations

ABL = ablative; ACC = accusative; AP = adverbial particle; ART = article; CLASS = classifier; DAT = dative; DEF = definite marker; DET = determiner; F = feminine; M = masculine; NEG = negative marker; OPT = optative marker; PRED = predicative

marker; PREP = preposition; PRES = present tense; REL = relative clause marker; SG = singular; 1, 2, 3 = first, second, third person.

## References

- Breu, Walter. 1996. Überlegungen zu einer Klassifizierung des grammatischen Wandels im Sprachkontakt (am Beispiel slavischer Kontaktfälle). *Sprachtypologie und Universalienforschung* 49(1): 21–38.
- Chikovani, Guram. 2005. Linguistic contacts in Central Asia. In *Linguistic Convergence and Areal Diffusion: Case Studies from Iranian, Semitic and Turkic*, E. Csátó, A. Bo Isaksson & C. Jahani (eds), 127–132. London: Routledge/Curzon.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68: 81–138.
- Flores Farfán, José Antonio. 2004. Notes on Nahuatl typological change. *Sprachtypologie und Universalienforschung* 57: 85–97.
- Fortescue, Michael. 1993. Eskimo word order variation and its contact-induced perturbation. *Journal of Linguistics* 29: 267–289.
- Friedman, Victor A. 2003. *Turkish in Macedonia and Beyond: Studies in Contact, Typology and other Phenomena in the Balkans and the Caucasus* [Turcologica 52]. Wiesbaden: Harrassowitz.
- Givón, Talmy (ed.). 1979a. *Syntax and Semantics, 12: Discourse and Syntax*. New York NY: Academic Press.
- Givón, Talmy. 1979b. From discourse to syntax: Grammar as a processing strategy. In *Syntax and Semantics, 12: Discourse and Syntax*, Talmy Givón (ed.), 81–111. New York NY: Academic Press.
- Givón, Talmy. 1979c. *On Understanding Grammar*. New York NY: Academic Press.
- Givón, Talmy. 1984. *Syntax: A Functional-Typological Introduction*. Volume I. Amsterdam, Philadelphia: John Benjamins.
- Givón, Talmy. 1995. *Functionalism and Grammar*. Amsterdam: John Benjamins.
- Haase, Martin. 1997. Gascon et basque: Bilinguisme et substrat. *Sprachtypologie und Universalienforschung* 50 (3): 189–228.
- Harris, Alice. 2003. Crosslinguistic perspectives on syntactic change. In *The Handbook of Historical Linguistics*, B. D. Joseph & R. D. Janda (eds), 529–551. Oxford: Blackwell.
- Harris, Alice C. & Campbell, Lyle. 1995. *Historical Syntax in Crosslinguistic Perspective*. Cambridge: CUP.
- Heine, Bernd. 1997. *Possession: Sources, Forces, and Grammaticalization*. Cambridge: CUP.
- Heine, Bernd & Kuteva, Tania. 2003. Contact-induced grammaticalization. *Studies in Language* 27: 529–572.
- Heine, Bernd & Kuteva, Tania. 2005. *Language Contact and Grammatical Change*. Cambridge: CUP.
- Heine, Bernd & Kuteva, Tania. 2006. *The Changing Languages of Europe*. Oxford: OUP.
- Heine, Bernd & Reh, Mechthild. 1984. *Grammaticalization and Reanalysis in African Languages*. Hamburg: Buske.
- Johanson, Lars. 1992. *Strukturelle Faktoren in türkischen Sprachkontakten* [Sitzungsberichte der Wissenschaftlichen Gesellschaft an der Johann Wolfgang Goethe-Universität Frankfurt am Main Vol. 29(5): 169–299]. Stuttgart: Franz Steiner.

- Johanson, Lars. 2002. *Structural Factors in Turkic Language Contacts*. London: Curzon.
- Jones, Mari C. 2002. Mette a haout dauve la grippe des Angllais: Convergence on the Island of Guernsey. In *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors* [Contributions to the Sociology of Language 86], M. C. Jones & E. Esch (eds), 143–68. Berlin: Mouton de Gruyter.
- Kahr, Joan Casper. 1976. The renewal of case morphology: Sources and constraints. *Working Papers on Language Universals* (Stanford) 20: 107–151.
- Kiral, Filiz. 2005. Modal constructions in Turkic of Iran. In *Linguistic Convergence and Areal Diffusion: Case Studies from Iranian, Semitic and Turkic*, E. A. Csató, B. Isaksson & C. Jahani (eds.), 285–293. London: Routledge/Curzon.
- Leisiö, Larisa. 2000. The word order in genitive constructions in a diaspora of Russian. *International Journal of Bilingualism* 4: 301–325.
- Leslau, Wolf. 1945. The influence of Cushitic on the Semitic languages of Ethiopia. A problem of substratum. *Word* 1(1): 59–82.
- Li, Charles N. 1983. Language contact in western China. *Papers in East Asian Languages* 1: 31–51.
- Matras, Yaron. 1996. Prozedurale Fusion: Grammatische Interferenzschichten im Romanes. *Sprachtypologie und Universalienforschung* 49(1): 60–78.
- Matras, Yaron & Sakel, Jeanette. 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31(4): 829–865.
- Mehlig, Hans Robert (ed.). 1994. *Slavistische Linguistik 1993*. Munich: Otto Sagner.
- Miller, Roy Andrew. 1967. *The Japanese Language*. Chicago IL: University of Chicago Press.
- Nettle, Daniel. 1999. *Language Diversity*. Oxford: OUP.
- Ratcliffe, Robert R. 2005. Bukhara Arabic: A metatypized dialect of Arabic in Central Asia. In *Linguistic Convergence and Areal Diffusion: Case Studies from Iranian, Semitic and Turkic*, E. A. Csató, B. Isaksson & C. Jahani (eds), 141–159. London: Routledge/Curzon.
- Riehl, Claudia Maria. 2001. *Schreiben, Text und Mehrsprachigkeit: Zur Textproduktion in mehrsprachigen Gesellschaften am Beispiel der deutschsprachigen Minderheiten in Südtirol und Ostbelgien* [Tertiärsprachen. Drei- und Mehrsprachigkeit 4]. Tübingen: Stauffenberg.
- Ross, Malcolm D. 1996. Contact-induced change and the comparative method: Cases from Papua New Guinea. In *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, M. Durie & M. D. Ross (eds), 180–217. Oxford: OUP.
- Ross, Malcolm D. 2001. Contact-induced change in Oceanic languages in North-West Melanesia. In *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, A.Y. Aikhenvald & R. M. W. Dixon (eds), 134–66. Oxford: OUP.
- Ross, Malcolm D. 2003. Diagnostic prehistoric language contact. In *Motives for Language Change*, R. Hickey (ed.), 174–198. Cambridge: CUP.
- Ross, Malcolm D. 2006. Calquing and metatypy. *Journal of Language Contact* 1.
- Sussex, Ronald. 1993. Slavonic languages in emigration. In *The Slavonic Languages*, B. Comrie & G.G. Corbett (eds), 999–1035. London: Routledge.
- Ternes, Elmar. 1999. Ist Bretonisch SVO oder VSO? Typologische Überlegungen zu einer umstrittenen Frage. In *Akten des Zweiten Deutschen Keltologen-Symposiums*, S. Zimmer, R. Ködderitzsch & A. Wigger (eds), 236–253. Tübingen: Niemeyer.
- Thomason, Sarah Grey (ed.). 2000. *Contact Languages: An Introduction*. Washington DC: Georgetown University Press.
- Thomason, Sarah Grey. 2001a. Contact-induced typological change. In *Language Typology and Language Universals: An International Handbook*. Vol. 2. [Handbücher zur Sprach-

- und Kommunikationswissenschaft 20(2), M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds), 1640–1648. Berlin: Walter de Gruyter.
- Thomason, Sarah Grey. 2001b. *Language Contact*. Edinburgh: EUP.
- Thomason, Sarah Grey. 2001c. *Language Contact: An Introduction*. Washington DC: Georgetown University Press.
- Thomason, Sarah Grey. 2003. Contact as a source of language change. In *The Handbook of Historical Linguistics*, B.D. Joseph & R. D. Janda (eds.), 686–712. Oxford: Blackwell.
- Thomason, Sarah Grey & Kaufman, Terrence. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley CA: University of California Press.
- Tosco, Mauro. 2000. Is there an Ethiopian language area? *Anthropological Linguistics* 42(3): 329–365.
- Wehr, Barbara. 1984. *Diskurs-Strategien im Romanischen: ein Beitrag zur romanischen Syntax*. Tübingen: Narr.
- Weinreich, Uriel. [1953] 1964. *Languages in Contact*. The Hague: Mouton.
- Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Oxford: Blackwell.
- Zeevaert, Ludger. 2006. *Variation und kontaktinduzierter Wandel im Altschwedischen* [Arbeiten zur Mehrsprachigkeit Folge B/ Working Papers in Multilingualism Series B 74]. Universität Hamburg: Sonderforschungsbereich 538 Mehrsprachigkeit.

# Remodeling grammar

## Copying, conventionalization, grammaticalization

Lars Johanson

University of Mainz

The paper discusses processes involved in copying of grammatical elements. The Code Interaction framework is applied in the description and analysis of contact induced phenomena. This framework distinguishes between different types of copying. Extensive Selective Grammatical Copying creates new morpho-syntactic elements which may be conventionalized and result in grammatical remodeling. It is, however, claimed that diachronic processes are not copiable and that Selective Grammatical Copying is not a process of unidirectional contact-induced grammaticalization. The importance of one specific type of copying, namely Frequential Copying, is stressed.

**Keywords:** code copying, combinational copying, conventionalization, frame-changing, frequential copying, global copying, habitualization, selective grammatical copying, target of copying, target of grammaticalization

### 1. Introduction

The topic of this paper is how grammar may be restructured through conventionalized results of Code Copying, and how these results relate to grammaticalization processes.

The goal is to show how Selective Grammatical Copying creates contact-induced morphosyntactic elements, which may be conventionalized and thus lead to linguistic change in the sense of grammatical remodeling. The stages of grammaticalization of the elements involved in the relevant contact processes will be discussed. It will be claimed that diachronic processes are not copiable and that Selective Grammatical Copying is not a process of unidirectional contact-induced grammaticalization. These discussions will be followed by concrete examples of analyses according to the theoretical concepts. The often ignored importance of



Frequential Copying will be stressed. Finally, the overall results of grammatical re-modeling through extensive Selective Grammatical Copying will be summarized.

The issues will be dealt with in a specific Code Interaction framework, an integrated model for the coherent description and analysis of various kinds of contact-induced phenomena and their interrelations. This framework is intended to make the complexity, dynamics and creativity of code interaction visible with a minimum of technical machinery.

While a detailed account of this framework, which has been presented in a number of publications since 1992, e.g. in Johanson (1992, 1998, 1999a, 1999b, 2002a, 2002b, 2005a, 2005b), is not possible within the limits of the present paper, its main tenets will be summarized in the first sections.

The structure of the paper is as follows: Sections 2–4 present hypotheses and methods following from the Code Copying framework. Sections 5 and 6 present the specific concept of Selective Grammatical Copying. Section 7 provides examples of analyses. Section 8 deals with the effects of Frequential Copying. Section 9 summarizes the results of extensive Selective Grammatical Copying.

## 2. Insertional code copying

The following is a very brief summary of the principles of the Code Copying framework.

One important component of the Code Interaction framework is *code copying*, a common creative resource available to ease the tension between interacting linguistic systems, whether related or unrelated.

The term *code* refers to languages and language varieties (dialects, sociolects, idiolects, registers). It is used in the same sense as in *code switching*, which means alternation between two or more languages or language varieties. The difference is that copying does not imply alternation of codes, but insertion of elements of one code into another code. All contact-induced processes to be discussed here are subsumed under this kind of code interaction.

Contact-induced copying means that users of a Basic Code copy elements of a Model Code. The Model Code is the 'source', 'donor' or 'diffusing' code, while the Basic Code is the 'recipient' or 'replica' code.

The term *copying*, which covers what is otherwise called 'borrowing', 'diffusion', 'transfer', 'interference', 'replication', etc., is chosen to stress that copies are per definition *not identical* to their models. Copies are never 'imported' or 'transferred' foreign elements and never true replicas of their models. There are always

dissimilarities in substance, meaning, contextual applicability and frequency between models and copies. It is true that the Basic Code becomes more similar to the Model Code than it was before the act of copying, but identity between models and copies is in principle excluded, all similarities being partial.

The differences between the model and its copy are essentially due to *adaptation*, modification in the direction of the Basic Code because of various kinds of incongruence between the two codes. Properties of the models may thus be replaced by properties typical of the Basic Code. The modifications in question include adaptation to the phonological system, to the lexical semantic system (causing denotative and connotative differences), to the morphosyntactic system, etc.

*Insertional copying* means that copies are inserted into Basic Code clauses. On the basis of structural and conceptual similarities, equivalence relations are established, consciously or intuitively, between elements in the Model Code and elements in the Basic Code.

Basic Code clauses provide a morphosyntactic frame of native and nativized grammatical markers and patterns. Inserted copies are immediately part of the Basic Code and subject to its internal processes. According to our theoretical concept, all languages are ‘contact languages’ in this sense.

Insertional copying is an innovative act, not necessarily the result of imperfect learning. The differences should not be viewed as failure to meet the norms of the Model Code. They are rather due to creative restructuring according to cognitive principles and communicative needs. Copying often serves as a creative technique of enriching codes and enhancing their functionality; cf. Csato’s remarks (1998) on “the creativity of code-copying” as observed in the Europeanized Turkic language Karaim. Copies do not necessarily fill gaps, but can be inserted even in cases where the Basic Code possesses close equivalents.

The motivation for copying is mostly social, i.e. the wish to communicate in a prestigious way or in the same manner as the social environment. The structural, cognitive and psychological factors involved need close attention.

Insertional copying may proceed in two directions:

*Take-over insertion* means that code users take over copies from a secondary code in their primary code, which is then the Basic Code. For example, speakers may copy elements from English when using their primary code Swedish.

*Carry-over insertion* means that code users carry over copies from their primary code into their variety of a secondary code, which is then the Basic Code. For example, speakers may copy elements from their primary code Swedish when using their more or less idiosyncratic variety of English.

Secondary codes shifted to by code users may retain carried-over copies as substrate phenomena. Complex situations including both take-over and carry-over insertion are often observed in intensive language contacts.

### 3. Global and Selective Copying

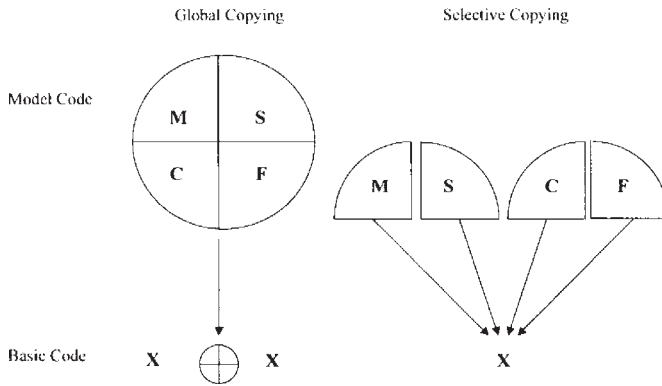
*Global Copying*, a process often called ‘transfer’, means that units, i.e. morphemes and morpheme sequences, of the Model Code are copied *globally*, as a whole, including their material shape (substance) and properties of meaning, combinability and frequency. The copy is inserted into a position felt to be equivalent to that filled by the model in the Model Code.

Global Grammatical Copying produces grammatical function units such as English *they*, *their*, *them*, global copies from Scandinavian, the Karaim superlative prefix *nay-*, a global copy from Slavic, Danish *hvis* ‘if’ < *hwes*, a global copy from Middle Low German, German *in puncto* ‘in respect of’, a global copy from Latin. Persian *ham* ‘and, also’, Arabic *amma* ‘but’, Russian *i* ‘and’, *no* ‘but’ are examples of grammatical function units that have been copied globally into numerous contact codes, e.g. in various Turkic languages.

*Selective Copying* means that only individual selected properties – material, semantic, combinational and frequential properties – of elements of a Model Code are copied onto units of a Basic Code. This term should not be confused with selection processes in the sense of adaptation, which are common to all kinds of copying and which result in differences between models and copies (see Section 2). An equivalence relation is established, consciously or intuitively, between an element in a Model Code and a suitable Target of Copying in the Basic Code, a native element onto which the relevant properties can be copied. The target is reanalyzed with respect to these properties, its functions becoming more similar to those of the element in the Model Code. Various kinds of Selective Grammatical Copying will be discussed below.

The Code Copying framework allows the study of Global Copying and Selective Copying within one and the same framework. These are simply two well-defined types of copying rather than different types of code interaction. The relation between them may be illustrated in Figure 1. On the left side, we find a whole ‘globe’ of properties, i.e. of material, semantic, combinational and frequential properties. On the right side, we only find loose parts of a corresponding ‘globe’, i.e. selected material, semantic, combinational and frequential properties.

Selective Copying may thus comprise Material Copying, Semantic Copying, Combinational Copying and Frequential Copying.



M = Material properties (substance); S = Semantic properties (also including pragmatic meanings); C = Combinational properties (in syntax and word structure); F = Frequency properties; X = elements of a Basic Code.

Figure 1.<sup>1</sup> Synoptic representation of Global and Selective Copying.

#### 4. Momentary, habitualized and conventionalized copies

Insertional copying acts may be momentary, ephemeral phenomena, corresponding to Weinreich's 'nonce-borrowing' (1953: 11). They begin in individuals, more precisely in the mind of individuals, if we do not take this to exclude intuitive, unconscious copying. Copying grammatical function units and patterns is certainly less conscious than copying content units, which speakers tend to be more aware of. On the other hand, individual speakers who perform an originary act of copying normally have some degree of knowledge of the Model Code. Selective grammatical copying presupposes some degree of familiarity with Model Code structures and some ability to analyze them.

Code copies may habitualize, i.e. become used habitually, and thus have more lasting effects, with various degrees of recurrence in the individual. Code copies may also become more or less conventionalized, i.e. have effects on the linguistic behavior of speech communities of varying sizes. Code copying starting as momentary performance phenomena may thus have diachronic effects. The path leading to conventionalizing is a continuum of changes in the sociolinguistic status, of degrees of recurrence in the speech community.

So-called 'integration', which, according to some scholars, leads from structurally less 'integrated' to more 'integrated' copies, whatever this may mean, is in

1. Figure 1 was first presented in my talk "Code copying in historical linguistics: The case of Selective Grammatical Copying", at the 17th International Congress of Linguists in Prague, 2003.

principle a different development, and not an unequivocal or necessary criterion for distinguishing degrees of conventionalization.

As soon as a copy is conventionalized, the degree of the individual speaker's knowledge of the Model Code is irrelevant. Deviations originally perceived as 'interference' may establish themselves as new sets of norms, even replacing their Basic Code equivalents. The processes lead to linguistic change. New systems are created in which conventionalized copies form integral parts. They are now facts of linguistic history and thus objects of historical linguistics. The development may include further diffusion, the code-copying variety becoming the unmarked way of speaking in ever-expanding groups. Copies may, however, also be unsuccessful or only reach a very limited diffusion. They can of course also be subject to deconventionalization and loss.

## 5. Selective Grammatical Copying

Selective Grammatical Copying means that copies of grammatical markers and patterns of a Model Code are inserted into the morphosyntactic frame of a Basic Code. Semantic, combinational and frequential properties are copied, without accompanying material properties, onto Basic Code elements. Copying of semantic and combinational properties is traditionally referred to as 'grammatical calquing', 'loan semantics', 'loan translation', 'loan syntax', 'structural borrowing', 'grammatical replication', 'loanshift', 'indirect morphosyntactic diffusion', etc.

Selective Grammatical Copying creates new or modified grammatical elements. Semantic copying yields grammatical meanings, combinational copying affects the combinability, e.g. causing extended or narrowed applicability to contexts, and frequential copying leads to increased or decreased occurrence of a given element.

Copies may include grammatical markers such as voice, aspect, mood, tense, case and number markers, articles, pronouns, junctors for word- and clause-combining, or ordering patterns relevant for word-formation, phrase-, clause- and sentence formation, clause-combining, text construction, text subdivision, other morphosyntactic or morphosemantic structures, and patterns of pragmatic organization. Some aspects of Selective Grammatical Copying are dealt with in Johanson (2003, 2005b), etc.

Thus the following situation obtains: Global Grammatical Copying involves copying of material properties (substance). Neither these nor other properties of the model are, however, simply 'transferred' from one code to the other. In the case of Selective Grammatical Copying, the relevant properties are in a similar way just copied and not 'transferred'. Basic Code users employ means that are

available in their own code to create elements that are felt to be equivalent to elements in the Model Code.

The results of Selective Grammatical Copying may vary:

1. a new element may be *added*;
2. the Basic Code equivalent may be *replaced* by the copy;
3. the Basic Code equivalent may be *retained*, while assuming modified functions.

In all these cases, copying leads to morphosyntactic convergence of the interacting codes, but never to identity of models and copies. It must be emphasized here again that all cases involve adaptation to the grammatical system of the Basic Code. It is the task of the linguist to describe similarities and differences as accurately as possible.

What operational procedures make Selective Grammatical Copying possible? Users of a Basic Code become aware of a certain grammatical element in a Model Code. On the basis of some conceptual similarity, an equivalence relation is established, consciously or intuitively, between the given Model Code element and a suitable Target of Copying in the Basic Code. The Target of Copying is a native lexical or grammatical element that seems to match the grammatical element of the Model Code and onto which the relevant properties of that element can be copied most naturally. The Target of Copying is reanalyzed and remodeled, e.g. assigned the relevant properties.

Properties of a dual marker in a Model Code may for example be copied onto a numeral *two* as a Target of Copying to create a dual marker in a Basic Code:

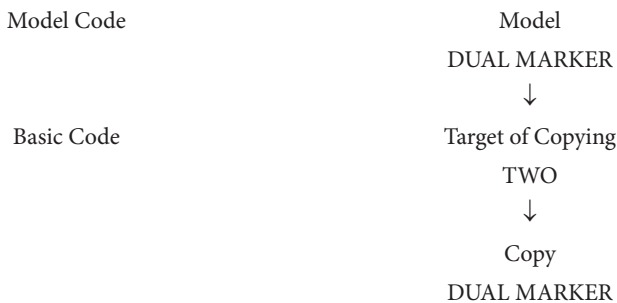


Figure 2. Properties of a dual marker copied onto a numeral 'two'.

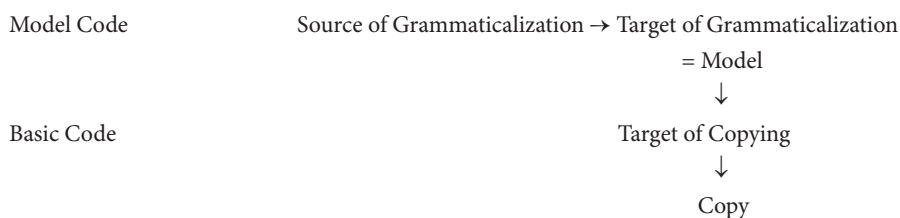
This example is not imaginary, but empirically attested. For example, speakers of Tayo, a French-based creole in New Caledonia, have copied properties of dual

markers from the two Melanesian languages Drubéa and Cèmuhi onto the Basic Code unit *-de* 'two' (< French *deux*) (Corne 1995).

## 6. Grammatical copying and stages of grammaticalization

The relations between Selective Grammatical Copying and stages of grammaticalization raise intricate problems and merit special attention. They are discussed in highly stimulating ways in Aikhenvald's book on language contact in Amazonia (2002), in Aikhenvald (2003), in Aikhenvald & Dixon (2001), and in Heine & Kuteva (2003, 2005).

I refer to the following schematic picture in order to illustrate my own understanding of the relationship between Selective Grammatical Copying and stages of grammaticalization:



**Figure 3.** Relationship between Selective Grammatical Copying and degrees of grammaticalization.

An element of the Model Code, the model, is copied at a specific stage of its code-internal development along a specific grammaticalization path. It goes back to a lexical Source of Grammaticalization and is thus a Target of Grammaticalization within the Model Code.

Similarly, each Target of Copying in the Basic Code has its own history and occupies a certain position in its own code-internal development. Each instance of copying is a product of this coincidence. The elements involved represent certain stages of grammaticalization, the results of previous processes reached at the moment of coincidence.

The interpretation of this situation is the only essential point in which I cannot follow Heine & Kuteva (2003, 2005) in their attempts to relate grammaticalization theory to contact-induced language change. These authors (Heine & Kuteva 2005:92) take contact-induced changes to follow the same principles of grammaticalization as changes that do not involve language contact. They contend that speakers of “language R”, the replica language or Basic Code,

replicate a grammaticalization process which they assume to have taken place in “language M”, the Model Code. Basic Code users would thus assume a certain grammaticalization process to have occurred in the Model Code and actually copy this assumed process.

I maintain that diachronic processes are not copiable and that Selective Grammatical Copying is not a process of unidirectional contact-induced grammaticalization.

### 6.1 Diachronic processes are not copiable

There are certainly cases in which the chosen Target of Copying is similar to the Source of Grammaticalization of the model. For example, the dual marker of the Model Code may have developed from the numeral *two*, and Basic Code users may have chosen their corresponding numeral as the Target of Copying. This establishes an analogous relation, but it does not mean that the Target of Copying has undergone or undergoes the grammaticalization process that has taken place in the Model Code. The act of copying does not imply a repetition of this gradual process from less grammaticalized to more grammaticalized items. It is even irrelevant whether the choice of the Target of Copying is motivated or not by knowledge of the diachronic Source of Grammaticalization of the model, i.e. whether the choice is guided or not by the etymological relation between the model and its Source of Grammaticalization.

It follows from our theoretical concept that diachronic processes are not copiable, even if they happen to be recoverable. The process from the Source of Grammaticalization to the Target of Grammaticalization cannot be copied, and no corresponding process takes place in the Basic Code as a result of the act of copying. This act immediately turns the Target of Copying into a grammatical element. What is copied is just the *result* of a grammaticalization process as assumed for the Model Code element. The claim that diachronic processes may be copied in certain contact situations cannot be empirically verified and must mean something else, i.e. that analogous processes may take place. Needless to say, observations to this effect are highly interesting in their own right and merit our full attention.

### 6.2 Fresh copies are less advanced

Fresh copies often represent less advanced stages of grammaticalization than their models with respect to semantic, combinational and frequential properties.



Properties of less grammaticalized elements of the Model Code are not likely to be copied onto more grammaticalized Targets of Copying. This would be an odd instance of reversed directionality.

The semantic functions of copies have often not reached the stage of grammaticalization of their models. They have not undergone the stages of grammaticalization that their models have undergone. Their use is often pragmatically determined. They often have lower text frequencies than their originals. Their use is often optional rather than obligatory. They often have a lower degree of combinability, being less generally applicable to contexts. It might sometimes even be difficult to decide whether fresh copies already represent fully grammaticalized categories or not. For example, the use of newly copied aspectual-temporal elements is often contextually restricted and optional. This is also true of Sorbian, Czech and Slovenian definite articles copied from Germanic, or the Basque indefinite article copied from Romance. Heine & Kuteva (2005: 101) note: “wherever there is sufficient evidence, it turns out that the replica construction is less grammaticalized than the corresponding model construction”.

If we would take the act of copying itself to imply a grammaticalization process, these phenomena would have to be viewed as backward movements, cases of reversed directionality, violations of the unidirectionality principle that is generally assumed for grammaticalization paths.

### 6.3 Code-internal development

Copying is a punctual interaction across code borders. The further development of copies is code-internal and gradual. Once a copy is conventionalized, it is subject to the internal processes of the Basic Code and may be dealt with in the same way as non-copied innovations. Selective grammatical copies may be further grammaticalized, acquiring properties typical of advanced stages, developing more general grammatical meanings, being used in wider ranges of contexts, increasing their use and degree of obligatoriness, being decategorialized into clitics and affixes, eroded in the sense of loss of phonetic substance, etc. All this is solely a matter of code-internal development. Copies representing advanced stages of grammaticalization have few chances to develop further, except for taking on highly general functions and being subject to material erosion.

In their further developments, copies often become indistinguishable from native elements. They may survive and develop further, even after their models have vanished. A copy may also become more and more similar to its model, which makes it increasingly difficult to tell which of two similar grammatical

elements in chronologically remote contact situations was the model and which one was the copy.

## 7. Examples of analyses

In what follows, I will give a few examples of analyses of copying of selective grammatical properties onto various Targets of Copying. To make comparison with other frameworks possible, I deliberately choose well-known instances dealt with in previous literature, particularly in Heine & Kuteva (2005).

### 7.1 Nongrammatical elements as Targets of Copying

Targets of selective grammatical copying may be nongrammatical words, e.g. nouns, verbs and adverbs, onto which grammatical properties are copied.

Selective properties are often copied onto units that are materially similar to the model, e.g. Dutch *in termen van* on the model of English *in terms of*. These types represent relatively early stages of grammaticalization. Complexes in which the lexical element is still clearly identifiable may later, in the copying language, develop into more advanced grammatical function units, e.g. adpositions of the type ‘in the house of’ > ‘at, beside’: French *chez*; cf. Latin *casa* ‘house’, Swedish *hos*; cf. *hus* ‘house’.

A Model Code item may have both a nongrammatical function and a grammatical function. A verb *to have* may have a nongrammatical function and a grammatical function as an auxiliary. A noun meaning *body* may have a nongrammatical function and a grammatical function as a reflexive marker *self*. A numeral *one* may have a nongrammatical function and a grammatical function as an indefinite article. A relational noun such as *front* may have a nongrammatical function and a grammatical function as an element of an adposition *in front of*.

Basic Code users may copy the grammatical function onto a nongrammatical item that corresponds to the Model Code lexeme. The result will exhibit the same kind of polysemy as found in the Model Code. The ambiguity between the nongrammatical function and the grammatical function of the model is mirrored by the copy. For example, speakers of Pipil, an Aztec language of El Salvador, have copied combinational and semantic properties of Spanish prepositions onto relational nouns to use them as adpositions (Heine, Claudi & Hünemeyer 1991). Speakers of Basque in southwestern France have copied properties of Gascon and

French *un* onto their numeral *bat* 'one', to use it as an indefinite article as well (Haase 1992: 59–61, 71).

Models and targets do not necessarily belong to the same word class. Properties of one word class in the Model Code may be copied onto a target of another word class, e.g. verbal properties onto adverbs. In the English-based creole Sranan, spoken in Suriname, *doro*, a global copy of Dutch *door* 'through', can be used as a verb, 'to put through', 'to go through' (Heine & Kuteva 2005: 242, quoting Adrienne Bruyn).

## 7.2 Grammatical elements as Targets of Copying

The Target of Copying may also be a more or less grammaticalized element. Its functions are restructured, which may entail modifications, e.g. an extended or a narrowed functional range.

A certain function, e.g. that of a relative clause marker, may be copied onto a native Target of Copying. Let us assume that the model is polysemic, functioning both as a question marker and as a relative clause marker. Basic Code users may choose a Target of Copying that corresponds to the question marker function. The copy thus becomes polysemic in the same way as the model is. This has occurred in many languages. For example, Gagauz, a Turkic language spoken mainly in the southern region of the Republic of Moldova, has copied properties of Slavic interrogatives onto the question marker *ani* 'where' in order to use it as a relative clause marker 'which, who' (Menz 2001). Basque varieties have copied properties of Spanish relative clause markers onto *zein* 'which, who?' in a similar way (Trask 1998: 320). Users of the North Arawak language Tariana have copied properties of Portuguese interrogative pronouns to create relative clause markers (Aikhenvald 2002).

The Rhaeto-Romance language Sursilvan possesses a permissive auxiliary (*la*)*schar* 'to allow to do'. The properties of a corresponding German model *lassen* 'to let', which also has the causative meaning 'to cause to do', are copied onto the auxiliary, which can then also be used in a causative function (Stimm 1984). Swedish immigrants to the USA have copied properties of English *have* onto the Swedish verb *ha* 'have' to use it as a causative auxiliary as well (Myers-Scotton 2002: 104).

A Model Code marker with comitative and instrumental function may be copied onto a corresponding instrumental marker in a Basic Code, so that the latter can be used in both functions. Karaim *-Ba*, the enclitic form of a postposition meaning 'with', has extended its use as a grammatical case marker due to

copied semantic and combinational properties of the Russian instrumental case (Csató 1998).

On the model of the German polite second person pronoun *Sie* 'you', which is homophonous with (and developed from) the third person pronoun *sie* 'they', the third person plural pronoun of the local Polish dialect of Silesia has extended its use to polite address (Weinreich 1953:40).

### 7.3 Grammatical patterns as Targets of Copying

Basic Code users may be able to analyze a grammatical pattern found in a Model Code and to copy its combinational and semantic properties onto a corresponding Target of Copying, an existing construction of the Basic Code.

Speakers of several languages have copied properties of Turkish reduplication patterns onto native elements in order to use them as intensives. They have, for example, analyzed *dümdüz* 'entirely even' as a reduplicated form of the adjective *düz* 'even' and copied the relevant properties onto a corresponding adjectival Target of Copying, e.g. Serbian and Croatian *ravravno* from *ravno* 'even' (Weinreich 1953:42). (The consonant *m* in *dümdüz* is one of four consonants that may occur in Turkish partial reduplications of this kind.)

With regard to pronominal referencing by means of anaphora, language contact frequently results in over- and undermarking (Johanson 1992:181–182, 2002a:16–17). A Model Code may represent a 'pro-drop' type of clause patterns, which makes limited use of anaphora, e.g. subject and object pronouns, for referents recoverable from the cotext and the situational context. Combinational and frequential properties of Model Code patterns that apply more explicit pronominal referencing may be copied onto corresponding elements of a Basic Code. Conversely, properties of Model Code clause patterns of the 'pro-drop' type may be copied onto a Basic Code of a more explicit type. Both kinds of copying lead to the violation of anaphoric conventions.

Certain Anatolian Turkish dialects exhibit overmarking that most likely results from Greek influence. Speakers of Greek in the East Black Sea area seem to have carried over copies of combinational properties of anaphoric clause patterns to their own brand of Turkish (Brendemoen 1993). Turkish children and adolescents growing up in northwestern Europe are prone to an increased use of anaphora, evidently under the influence of the surrounding majority languages.

Copies of this kind may originally be less grammaticalized, and their use in terms of zero anaphora conventions may be generalized through code-internal

development, entailing a change from a 'pro-drop' structure to a 'non-pro-drop' structure, or vice versa.

Combinational copying may also affect constituent ordering patterns. Users of the Basic Code may copy an ordering pattern found in the Model Code onto a selected Target of Copying, an existing structure which is thus redefined. For example, a Model Code pattern of the SVO type *I love you* may be copied onto an equivalent Basic Code pattern of the SOV type *I you love*. This may give rise to variation. The variation may vanish in favor of a new fixed word order, i.e. the old Basic Code pattern may be replaced by SVO, which is then the only pattern. Another result may be that a new pattern is added, while the old pattern is retained, but assigned a modified function, e.g. becoming a pragmatically marked pattern.

## 8. Frequential copying

Frequential patterns of Model Code elements can be copied onto Basic Code elements, leading to increased or decreased use of the latter. Frequential copying may, in its initial, weak forms, violate pragmatic norms, while not leading to manifest linguistic change.

As noted above, language contact frequently results in over- or undermarking. Frequential copying may thus increase or decrease the use of anaphora (Johanson 1992: 254, 2002a: 110–111). The use of one linguistic option may increase at the expense of another option. This may mean increased use of plural markers in codes with restricted plural marking patterns, or of anaphoric pronouns in 'pro-drop' codes (see 7.3). The use of an element may also decrease due to frequential copying. A two-way system of definite articles may, for example, lose ground in favor of a one-way system. Speakers of Dutch in Australia tend to copy combinational properties of the English definite article *the* onto the Dutch article *de*, gradually giving up the article *het* (Clyne 2003: 22, 31).

Frequential copying is a major factor in supporting constituent ordering patterns. Under the influence of an equivalent ordering pattern in the Model Code, frequential properties are copied onto existing patterns available in the Basic Code. An existing ordering pattern may thus be extended to new contexts and used more frequently. A pre-existing pattern may be overgeneralized, i.e. reinforced by frequential copying, while the frequency of use of an unsupported alternative diminishes accordingly. This reinforcement by analogy gives preference to properties which the contact languages seem to have in common (Johanson 1992: 182–183, 215, 252, etc., 2002a: 18, 61, 108, etc.).

Thus the resulting constituent ordering patterns are not really new; only their frequency and their extended or narrowed contextual occurrence are new. This analysis corresponds to Heine's scenario (2006, this volume) of "word order change without word order change". After his overview and discussion of numerous data in various languages, Heine (2006: 19) notes that he has "not come across a single case where speakers produced an entirely new word order". This conclusion is perfectly in line with my understanding of frequential copying.

A simple example: The Basic Code may have the two alternative constituent order patterns, SOV *I you love* and SVO *I love you*. SVO order may be acceptable, but more marked. The frequency of a Model Code SVO pattern *I love you* is copied onto the equivalent Basic Code SVO pattern *I love you*, i.e. onto the alternative that matches the Model Code order. Due to this frequential copying, the SVO pattern is extended to new contexts and used more frequently. It may develop from a pragmatically marked pattern into a pragmatically unmarked pattern. The alternative SOV pattern *I you love* may eventually be replaced by SVO *I love you*, which is then the only pattern available. Alternatively, it may be retained, while assigned modified functions. Gagauz provides a good example of frequential copying affecting the use of an existing SVO order in the sense of increased frequency relative to a normal SOV order (Menz 1999: 40–41, 2006: 141).

Heine (2006: 20) notes that the word order changes dealt with by him concern processes of grammaticalization "that could in principle have happened as well *internally*, that is, without language contact – in other words, it is not possible to 'prove' that contact was a contributing factor". I have claimed that a 'natural' or 'universal' tendency (e.g. towards analytic constructions as against synthetic constructions) can be more or less reinforced by frequential copying. A certain natural structure, which is already latently present in the Basic Code, can be strengthened through contact with a Model Code that possesses an equivalent structure. The introduction of the structure in the Basic Code is facilitated or accelerated through frequential copying. As I have suggested, we must thus supplement the simple question '... copying or independent tendency?' by the further alternative '... or a natural tendency reinforced by the contact language?' (Johanson 1992: 278–279, 2002a: 142).

## 9. Remodeling the Basic Code frame

With increasingly more conventionalized grammatical copies, the Basic Code frame may change considerably.

Changes of the Basic Code frame, into which copies are inserted, are produced by global copies of grammatical markers, selective copies of grammatical markers, and selective copies of combinational patterns. Successive copying processes of different kinds promote each other, and may lead to 'snowball effects' with respect to the remodeling of the Basic Code frame. Copying of constructional patterns goes hand in hand with progressive selective copying of grammatical markers. Older copies prepare the ground for new copies. At each stage of development, the Basic Code frame must be described anew. Every given new norm is deviated from by new 'marked' copying. All new systems offer new frames for further insertion and conventionalization. On frame-changing developments, see in particular Johanson (1999a).

Combinational Copying creates new equivalence positions for further copying of grammatical markers. On the other hand, copied markers trigger copying of combinational properties, e.g. of new constructional patterns in which the markers are contained. The use of a copied grammatical marker may be tied to a certain position in a pattern. For example, Tariana of northwestern Brazil has copied Portuguese subordination markers, which has triggered the copying of a Portuguese combinational pattern, according to which a complement clause follows the main clause predicate instead of preceding it (Aikhenvald 2002: 182).

In high-copying codes, complex and combined processes may essentially reshape the morphosyntactic frame, adding new typological properties to it. The codes can acquire seemingly disparate properties. It is even "perfectly possible for a language to copy structures that might appear to be 'typologically inconsistent' with the rest of its structure" (Comrie 2002: xi). A precondition for this is that the new properties are not imported or transferred foreign elements, but just copies adapted to Basic Code structures.

Gradual processes extending over centuries may involve multiple layers and complex combinations of copying, conventionalization and subsequent code-internal grammaticalization. Established copies which are considered part of the 'inherited' inventory may be the final result of long series of changes, and they cannot easily be traced back to the copying act.

Long-lasting intense contacts between one specific Basic Code and one specific Model Code – with high, perhaps increasing, degrees of bilingualism among the Basic Code users – lead to increased isomorphism between the codes, convergent developments irrespective of their genealogical affiliations. Structural adaptation of the copies to the Basic Code, on the other hand, reduces the degree of isomorphism. Convergence means that the codes become more similar to each other. If the convergence is due to bilateral influence, the codes move toward each other, acquiring more and more common characteristics. Convergence is, however, mostly due to unilateral influence, a one-sided inclination, with one code

approaching the other and becoming more similar to it (cf. Clyne 2003:79; Johanson 2005a:4).

The opposite development is divergence. Two codes may bilaterally draw apart from a common norm, acquiring dissimilar properties. Codes descended from a common ancestor may evolve into different forms when used under different conditions. The divergence may be caused unilaterally, with one of the codes deviating and branching off from a common norm.

Selective Grammatical Copying may create highly isomorphic structures that make codes more compatible and intertranslatable. To reduce the planning effort in forming sentences in the two codes, groups of advanced bilinguals often develop strategies that indicate common mental procedures for arranging the information (Matras 2006). Gradual isomorphic processes are typical of intense communication areas, in which codes develop shared constructional patterns and morphosyntactic markers regardless of genealogical boundaries. Even codes that differ strongly from each other in their vocabulary may become increasingly isomorphic.

Many types of Selective Grammatical Copying serve isomorphism by creating convenient translation equivalents in the interacting codes. This may mean combinational and/or semantic reorganization of grammatical devices in certain subsystems, e.g. in those of case markers and adpositions. Other types of Selective Grammatical Copying entail a more radical restructuring of syntactic constructions on Model Code lines. These larger structures, e.g. constituent ordering patterns, tend to be copied in varieties spoken by advanced bilinguals, i.e. in cases where the codes are already morphosyntactically similar and include copies of the less radical type. The two kinds of copies, which correspond to Ross's (2005) dichotomy "grammatical calquing" vs. "metatypy", differ in degree, but they do not seem to be distinct contact phenomena with precise demarcation lines. While "grammatical calquing" was earlier included in "metatypy" (Ross 2001), Ross (2005) now conceives of the two types as distinct kinds of contact phenomena.

Extensive copying processes may lead codes to converge strongly with others, including unrelated codes, and also to diverge strongly from their genealogical heritage. On the other hand, codes do not completely abandon their heritage and do not fuse with unrelated contact codes. No high-copying code seems to have turned into the Model Code it has copied extensively from (cf. Johanson 2002c).

## Acknowledgement

Work on this paper, a contribution to the Collaborative Research Center (SFB) 295 of the Deutsche Forschungsgemeinschaft at the University of Mainz, was



carried out while the author was Fellow-in-Residence at the Swedish Collegium for Advanced Study (2005–2006). I want to thank two anonymous reviewers for valuable advice.

## References

- Aikhenvald, Alexandra Y. 2002. *Language Contact in Amazonia*. Oxford: OUP.
- Aikhenvald, Alexandra Y. 2003. Mechanisms of change in areal diffusion: New morphology and language contact. *Journal of Linguistics* 39: 1–29.
- Aikhenvald, Alexandra Y. & Dixon, R. M. W. 2001. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. Oxford: OUP.
- Brendemoen, Bernt. 1993. Pronominalsyntax in den türkischen Schwarzmeerdialekten – syntaktische Innovation oder Archaismus? In *Sprach- und Kulturkontakte der türkischen Völker* [Veröffentlichungen der Societas Uralo-Altaica 37], J. P. Laut & K. Röhrborn (eds), 51–73. Wiesbaden: Harrassowitz.
- Clyne, Michael. 2003. *Dynamics of Language Contact: English and Immigrant Languages* [Cambridge Approaches to Language Contact]. Cambridge: CUP.
- Comrie, Bernard. 2002. Introduction. In *Structural Factors in Turkic Language Contacts*, L. Johanson, vii–xii. London: Curzon.
- Corne, Chris. 1995. A contact-induced and vernacularized language: How Melanesian is Tayo? In *From Contact to Creole and Beyond*, P. Baker (ed.), 121–148. London: University of Westminster Press.
- Csató, Éva Á. 1998. Should Karaim be ‘purer’ than other European languages? *Studia Turcologica Cracoviensia* 5: 81–89.
- Haase, Martin. 1992. *Sprachkontakt und Sprachwandel im Baskenland*. Hamburg: Buske.
- Heine, Bernd. 2006. *Contact-induced Word Order Change without Word Order Change* [Working Papers in Multilingualism, Series B, 76]. University of Hamburg: Sonderforschungsbereich 538: Mehrsprachigkeit.
- Heine, Bernd, Claudi, Ulrike & Hünemeyer, Friederike. 1991. *Grammaticalization: A Conceptual Framework*. Chicago IL: University of Chicago Press.
- Heine, Bernd & Kuteva, Tania. 2003. On contact-induced grammaticalization. *Studies in Language* 27: 529–572.
- Heine, Bernd & Kuteva, Tania. 2005. *Language Contact and Grammatical Change*. Cambridge: CUP.
- Johanson, Lars. 1992. *Strukturelle Faktoren in türkischen Sprachkontakten* [Sitzungsberichte der Wissenschaftlichen Gesellschaft an der J. W. Goethe-Universität Frankfurt am Main, 29:5]. Stuttgart: Steiner.
- Johanson, Lars. 1998. Code-copying in Irano-Turkic. *Language Sciences* 20: 325–337.
- Johanson, Lars. 1999a. Frame-changing code-copying in immigrant varieties. In *Bilingualism and Migration* [Studies on Language Acquisition 14], G. Extra and L. Verhoeven (eds), 247–260. Berlin: Mouton de Gruyter.
- Johanson, Lars. 1999b. The dynamics of code-copying in language encounters. In *Language Encounters across Time and Space*. B. Brendemoen, E. Lanza & E. Ryen (eds), 37–62. Oslo: Novus Press.

- Johanson, Lars. 2002a. *Structural Factors in Turkic Language Contacts*. London: Curzon.
- Johanson, Lars. 2002b. Contact-induced linguistic change in a code-copying framework. In *Language Change: The Interplay of Internal, External and Extra-linguistic Factors* [Contributions to the Sociology of Language 86], M. C. Jones & E. Esch (eds), 285–313. Berlin: Mouton de Gruyter.
- Johanson, Lars. 2002c. Do languages die of 'structuritis'? The role of code-copying in language endangerment. *Italian Journal of Linguistics* 14: 249–270.
- Johanson, Lars. 2003. Code Copying in historical linguistics: The case of Selective Grammatical Copying. Unpublished talk given at the 17th International Congress of Linguists, Prague, July 24–29, 2003.
- Johanson, Lars. 2005a. Converging codes in Iranian, Semitic and Turkic. In *Linguistic Convergence and Areal Diffusion. Case Studies from Iranian, Semitic and Turkic*, É. Á. Csató, B. Isaksson & C. Jahani (eds), 3–31. London: Routledge/Curzon.
- Johanson, Lars. 2005b. On copying grammatical meaning. *Sprachtypologie und Universalienforschung* 58: 75–83.
- Matras, Yaron. 2006. Layers of convergent syntax in Macedonian Turkish. In *Turkic Languages in Contact* [Turcologica 61], H. Boeschoten & L. Johanson (eds), 46–62. Wiesbaden: Harrassowitz.
- Menz, Astrid. 1999. *Gagausische Syntax: Eine Studie zum kontaktinduzierten Sprachwandel* [Turcologica 41]. Wiesbaden: Harrassowitz.
- Menz, Astrid. 2001. Gagauz right-branching propositions introduced by the element *ani*. *Turkic Languages* 5: 234–244.
- Menz, Astrid. 2006. On complex sentences in Gagauz. In *Turkic Languages in Contact* [Turcologica 61], H. Boeschoten & L. Johanson (eds), 139–151. Wiesbaden: Harrassowitz.
- Myers-Scotton, Carol. 2002. *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: OUP.
- Ross, Malcolm. 2001. Contact-induced change in Oceanic languages in north-west Melanesia. In *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, A. Y. Aikhenvald & R. M. W. Dixon (eds), 134–166. Oxford: OUP.
- Ross, Malcolm. 2005. Metatypy. In *Encyclopedia of Languages and Linguistics*. 2nd edn, K. Brown (ed.), 95–99. Oxford: Elsevier.
- Stimm, Helmut. 1984. Eigenheiten der Kausativkonstruktionen im Surselvischen. In *Das Romanische in den Ostalpen* [Veröffentlichungen der Kommission für Linguistik und Kommunikationsforschung 15], D. Messner (ed.), 329–350. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Trask, Robert L. 1998. The typological position of Basque. *Language Sciences* 20: 313–324.
- Weinreich, Uriel. 1953. *Languages in Contact. Findings and Problems* [Publications of the Linguistic Circle of New York 1]. New York NY: Linguistic Circle of New York.



# Contact-induced change

## The case of the Tamangic languages

Michael Noonan

University of Wisconsin-Milwaukee

The Tamangic languages are a small group of languages in the Tibeto-Burman family spoken in Nepal. They have a long history of language contact with languages outside their group, but members of the family have been in contact with different languages. In this paper, we will describe the sorts of contact-induced changes the Tamangic languages are presumed to have undergone in recent times and to examine the sociolinguistic situations prevailing between speakers of the recipient and donor languages which existed at the time of the change. The data from the Tamangic languages will be examined in light of some hypotheses concerning both the relationship between kinds of contact situations and kinds of contact-induced change, and the grammatical effects of contact-induced change.

**Keywords:** borrowing, complexity, language contact, language shift, metatypy

### 1. Introduction

The Tamangic languages are a small and uncontroversial genetic grouping spoken entirely within Nepal.<sup>1</sup> Since the arrival of this Tibeto-Burman subgroup in Nepal from Tibet, perhaps 1500 years ago, speakers of these languages have been in contact with speakers of languages belonging to a number of different phyla, some of which are also Tibeto-Burman (e.g. the Kham-Magar family, Ghale, and the Tibetan Complex) and others not (e.g. Indo-European Nepali and the isolate Kusunda). This contact has resulted in a number of changes in these languages, though the source, type, and degree of change varies among the languages of the group. For example, Chantyal has been greatly affected by contact with Magar and

---

1. The work reported on in this paper has been supported by the following grants from the National Science Foundation: DBC-9121114, SBR-9600717, and SBR-9728369.

especially Nepali (Noonan 1996, 2003a), while others – e.g. Nar-Phu, Manange, and Seke – have been much less affected by contact-induced change, and what influence they have undergone has come from other sources (for these languages, the Tibetan Complex).

This paper has two goals. The first is to describe the sorts of contact-induced changes the Tamangic languages are presumed to have undergone in recent times and to examine the sociolinguistic situations prevailing between speakers of the recipient and donor languages which existed at the time of the change. The second goal is to examine the data from the Tamangic languages in light of some hypotheses concerning both the relationship between kinds of contact situations and kinds of contact-induced change, and the grammatical effects of contact-induced change. As Haspelmath (2004) has pointed out, diffusional linguistics is still in the ‘hunting and gathering stage’, with relatively little attention yet paid to the evaluation of models or competing hypotheses. Such hypotheses exist, however, and it is important to begin the process of examining them against linguistic data.

## **2. Background: Language, history, and demographics**

### **2.1 Nepal**

Nepal is a very complex area ethnically and linguistically. There are at least 140 languages spoken in the country with the indigenous languages divided between various branches of the Tibeto-Burman family and the Indic branch of Indo-European; in addition, there is one language isolate, Kusunda. About half of the population are mono-lingual speakers of Indo-European Nepali, the national language. The rest are all fluent in varying degrees in Nepali, and Nepali continues to encroach steadily on the other languages of the country, both in terms of the continued rise in the percentage of effectively monolingual speakers and in terms of contexts of use among those who continue to speak the other languages. The Nepali government has, in a variety of ways, encouraged this development (Noonan 2006a).

### **2.2 The Tamangic Languages**

The Tamangic languages differ considerably among themselves as to the degree and manner of contact with other languages, in particular with Nepali. In the chart below, I’ve tried to summarize a few salient facts. In §2.3 I’ll discuss the histories of two of the communities, the Chantyal and the Nar-Phu, in more detail

**Table 1.** Contact with Non-Tamangic Languages.

<b>Language</b>	<b>Degree and manner of contact with speakers of Nepali</b>	<b>Speakers of other languages absorbed into the community</b>	<b>Possible sources of linguistic influence besides Nepali</b>
<i>Chantyal</i>	Intense, for at least two centuries	Nepali, Magar, Thakali, Newari, and probably others	Most likely some influence from Magar at an early period
<i>Gurung</i>	Steady influence for at least two centuries	Probably speakers of Ghale	In the early period, Tibetan and possibly Zhangzhung; <sup>2</sup> Ghale
<i>Tamang</i>	Sporadic influence until recently	Quite possibly speakers of Central Himalayish languages	Central Himalayish?
<i>Thakali</i>	Sporadic influence until recently		Tibetan
<i>Manange</i>	Little, until recently	Ghale? Tibetan?	Tibetan
<i>Seke</i>	Little, until recently		Tibetan
<i>Nar-Phu</i>	Little, until quite recently	Ghale? Tibetan?	Tibetan and possibly Ghale; Zhangzhung?

as these represent the extremes of outside contact – in particular, contact with Nepali – among these languages.

In Table 1 above, languages are ordered informally by degree of Nepali influence, with Chantyal placed first. The influence of languages besides Nepali and Tibetan cannot easily be assessed and in most cases must be noted as simply probable. It's also worth noting that Chantyal is the only language of the group known to have absorbed speakers of Nepali.

### 2.3 Histories of the Chantyal and Nar-Phu communities

In discussing contact situations, it's important not simply to present information about the linguistic effects of contact, but also to present information about the sorts of social situations that obtained during the period of contact that produced the linguistic effects. Since the social situations are rather complex, and since this paper has to be kept to a reasonable length, I will focus here on the Chantyal and Nar-Phu languages, representing as they do the extremes of contact-induced influence within the family. The Chantyal situation is, in any case, rather unusual, and deserves special attention.

---

2. Zhangzhung is the language of the Bon scriptures.

### 2.3.1 *The Chantyal people*

The Chantyal people are a relatively small ethnic group, numbering about 10,000 people, historically centered in the Baglung and Myagdi districts within the Dhaulagiri Zone of west-central Nepal. The Baglung Chantyal ceased to speak the Chantyal language sometime in the 19th century; the majority of the Myagdi Chantyal continue to speak the Chantyal language. The Chantyal who do not speak Chantyal speak the national language, Nepali.

The Chantyal people originated as a separate group in the region of the upper Kali-Gandaki Valley and at some point they became specialists in mining. The first trace of them as a separate group finds them living to the west of their present location, working in mines and quarries and moving gradually eastward. They seem to have arrived in their present homeland in the Myagdi and Baglung districts in the late 18th century or early in the 19th century, having a patent from the King of Nepal to mine the copper found there.

The Chantyls who arrived in the Baglung and Myagdi districts were a very different group from the people currently calling themselves Chantyal. There is strong evidence indicating that a single clan served as the nucleus for the entire group. The twelve other clans derive from non-Chantyls who were inducted into the Chantyal ethnic group. Chantyal oral tradition preserves memory of the origin of most of these clans. This memory is reinforced by the fact that some of the clan names are identical to clan names of other nearby ethnic groups.

The groups inducted into the Chantyal ethnic group spoke a variety of languages, Nepali among them, and only one of these languages, Thakali, was closely related to Chantyal. The language of interethnic communication in this region of Nepal during this period was Nepali, which was already well established locally in the 18th century. Clearly, then, when new groups, including those who ultimately were inducted into Chantyal ethnicity, encountered the core Chantyal, the language they communicated in would have been Nepali. One thing that is clear, however, is that it must have been very attractive at that time to be a Chantyal; the profits from mining must have been considerable and the role of the Chantyls as miners was so valuable to the Nepalese state that Chantyls were exempt from the military draft during the period of wars with British India early in the 19th century. Further, the patent assigned to the Chantyls specified that only they could do the mining, so the Chantyls themselves must have been in need of additional labor to deal with the extensive copper deposits in the considerable area over which they had been granted a patent.

The social and linguistic effects of this influx of non-Chantyls into the community must have been considerable. The Chantyls had for long been a small group living among Nepali speakers – the west of Nepal is primarily Nepali speaking. But after moving into the Dhaulagiri Zone, native speakers of Chantyal

became a minority within their own communities. The massive influx of non-Chantyal speakers, many or most of whom must have had Nepali as their native language, resulted in the large-scale use of Nepali within Chantyal villages. This overwhelmed the Chantyal language in the Baglung District. In the Myagdi District, the Chantyal language survived, but the presence of so many non-native speakers in the community had profound effects on the lexicon, the phonology, and the syntax.<sup>3</sup>

The Chantyal people are a minority in both the Baglung and Myagdi districts. Nepali-speaking Brahmins and Chetris, Nepali-speaking Magars, and a few Thakalis and Newars make up the bulk of the population. Within the Chantyal-speaking villages, alongside the Chantyal, live Nepali-speaking Kamis, an untouchable caste of blacksmiths, who are always addressed in Nepali. When Chantyls and other peoples live on the same slope, the Chantyal farm the highest land: on the slopes below them live Magars and members of the Hindu castes, all of whom are addressed in Nepali.

In sum, what is being claimed here is that the Chantyal language was a fairly typical Tamangic language at the time of its appearance in the Dhaulagiri Zone, after which the induction of large numbers of non-Chantyal speaking peoples into the ethnic group in a very short period had a large-scale impact on the community language, resulting in its replacement by Nepali in most villages and its retention in greatly changed form in the remainder. In §3, we will examine the nature of the changes to Chantyal in the communities where it survived.

### 2.3.2 *The Nar-Phu people*

The Nar-Phu people also seem to have originated in the Kali-Gandaki Valley, and moved from there to their present location in the valley of the Nar River in the Manang District of Nepal, north of the Annapurna massif next to the Tibetan border. The region is very high, with altitudes beginning around 3500m.

The Nar-Phu language is spoken in two villages, called in Nepali *Nar* and *Phu*. The small valley in which Nar is located contains three Buddhist monasteries adhering to the Nyingmapa sect of Tibetan Buddhism; one large Nyingmapa monastery is located right in the middle of Phu. The monks and nuns at these monasteries may be local people, or may be from other parts of Nepal or Tibet. In the past, important lamas would visit – and sometimes settle – there from the Kyirong region of Tibet.

---

3. The information in this section was drawn from Noonan (1996), which can be consulted for a further analysis of the contemporary sociolinguistic situation among the Chantyal. Noonan (2006a) provides a description of current attempts to revive and promote Tibeto-Burman languages, including Chantyal, in west-central Nepal.



Until quite recently, outside contacts were largely limited to speakers of Tibetan dialects and the closely related Tamangic languages Manange and Gurung. Since Gurung and especially Manange are largely mutually intelligible with Nar-Phu, the Nar-Phu people of the two villages have a 'secret language', the point of which is to confound Mananges and Gurungs who might otherwise understand their conversations. Contact with the government of Nepal and Nepali speakers generally was quite limited until the 1970s. As a result, the Nar-Phu lexicon had few Nepali borrowings, at least until the last decade or so. There was rather more influence from Tibetan, however, as there have been speakers of Tibetan dialects in the villages almost continually for a long period. Because the Tamangic group and Tibetan are related languages, it is not always easy to identify Tibetan borrowings, particularly older ones.

All the inhabitants of Nar and Phu speak the Nar-Phu language, though with decreasing fluency as the national language, Nepali, has come to be used more and more commonly among the Nar-Phu diaspora and even within the villages of Nar and Phu.<sup>4</sup> When Nar-Phu is spoken, it is now frequently mixed with a good deal of Nepali. In addition, almost all adults past thirty know Tibetan as well, though knowledge of Tibetan is decreasing as people spend less time in the villages and as the importance of Nepali has increased. Until quite recently, literacy in Classical Tibetan was more common than literacy in Nepali. Tibetan literacy is taught in the monasteries; Nepali literacy is picked up either outside the region or in the local, government-run school, which has been open only intermittently due to the difficulty of retaining qualified teachers in this area where living is hard, even by the standards of rural Nepal. Still the importance of Nepali literacy is obvious to all, and all younger people seem to be literate in Nepali.

The language that we will be concerned with here is the speech of older, more fluent speakers, who still speak a version of the language which is relatively uninfluenced by Nepali.

### 3. The linguistic effects of contact

In this section, we will consider briefly the linguistic effects of contact on the Tamangic languages, surveying the lexicon, the phonology, the morpho-syntax, and rhetorical strategies.<sup>5</sup>

---

4. This decline in fluency is marked, but is quite recent, and affects at this stage only younger people.

5. Sources of information on the Tamangic languages are: Chantyal (Noonan 2003a), Gurung (Glover 1974; Glover et al. 1977), Manange (Hoshi 1986; Hildebrandt 2003b), Nar-Phu

### 3.1 Lexicon

Chantyal has borrowed massively from Nepali: 71% of the morphemes recorded in Noonan et al. (1999) are of Nepali origin. All of the other Tamangic languages have borrowed from Nepali, but in no other Tamangic language has the native linguistic stock been so reduced as in Chantyal and in so many lexical domains. In Chantyal, even very basic lexical items – many names for body parts, many kinship terms, many common verbs, the entire number system – have been borrowed from Nepali. In many respects, it's difficult to establish a lexical boundary between Nepali and Chantyal since virtually any lexical (as opposed to grammatical) morpheme may be used in a Chantyal sentence, replacing even surviving (and still commonly occurring) Chantyal words. This is not code-switching [though that may happen occasionally as well], since the borrowed words are fully integrated grammatically into the Chantyal sentence.

Of the other languages, Gurung has been borrowing the most Nepali vocabulary over the longest period. Not only have Gurung speakers been in contact with speakers of Nepali in and around their home territory for a long period, but Gurung men have traditionally served in the Gurkha units of the British and Indian armies, in which Nepali was the working language.

The Tamangic languages in the Tibeto-sphere, i.e. the languages whose speakers were in contact with Tibetans and practice Tibetan Buddhism (Nar-Phu, Manange, Seke, and Thakali), have borrowed vocabulary from Tibetan in varying degrees. This appears to be most extensive in Nar-Phu, which has also borrowed some numerals, especially higher numerals and ordinals, from Tibetan.

### 3.2 Phonology

#### 3.2.1 *Tone, voice quality, and voicing*

The Tamangic languages are best described as having a basic four-tone system (Mazaudon 1973, 1978a, 1978b, 1993–1994; see also Noonan 2003b for a sketch of such a system in Nar-Phu). In the prototype pattern, two of the tones have modal voice (these are referred to as 'high register'), while two are often associated with phonetic murmur<sup>6</sup> (these are referred to as 'low register'). Consonant voicing is not contrastive:<sup>7</sup> initial stops and affricates in the high register tones may be either

---

(Noonan 2003b), Tamang (Mazaudon 1973, 1993–1994, 2003), Thakali (Georg 1996), Seke (Honda 2003).

6. Murmur is also known as 'breathy voice'.

7. Voicing is contrastive with liquids, however, in the non-murmured tones. See below.

voiceless or voiceless aspirated; voice quality and glottal timing is not contrastive in the low register tones, which, depending on the language, may be phonetically murmured or voiced. Intervocally, voiceless consonants are voiced. All native roots are C(R)V(C), and the final consonant is not contrastively voiced or aspirated. The system of initial stop and affricate consonants is summarized below:<sup>8</sup>

Table 2. Initial consonants and syllabic quality.

Tone	Initial Consonant	Syllabic Quality
1	C or C <sup>h</sup>	Modal
2	C or C <sup>h</sup>	Modal
3	C <sup>h</sup>	Murmured
4	C <sup>h</sup>	Murmured

(Where C = voiceless, C<sup>h</sup> = voiceless aspirated, and C<sup>h</sup> = murmured or voiced.)

In contrast, Nepali is toneless; stops and affricates have a four-way contrast of voiceless, voiceless aspirated, voiced, and murmured variants initially, medially, and finally. The two systems, then, have some areas of overlap, but many differences. The outlines of the two systems are contrasted below:

Table 3. Phonological systems of Nepali and Tamangic languages.

Feature	Nepali	Tamangic Prototype
<i>Tone</i>	No tones	Four-way tonal contrast
<i>Murmur</i>	Feature of consonants; found only with stops and affricates	Feature of tones; note that initial consonant in a low-register tone can have any consonantal manner of articulation
<i>Voiceless &amp; voiceless aspirate distinction</i>	Distinctive in all positions [initial, medial, final]	Distinctive in word-initial position in high register tones
<i>Voicing</i>	Distinctive in all positions	Redundantly specified for stops and fricatives in initial position in low register tones and with phonemically voiceless stops and affricates intervocally

The outcome in Chantyal, the language most affected by contact with Nepali, was a system different from either model. First, the tonal system as such was a casualty of the contact situation. Contour pitch distinctions, which characterized the

8. Mazaudon's analysis of the prototype system is slightly different, but the differences do not affect the essentials of the argument that follows.

earlier Tamangic system were completely lost. Even under less extreme contact situations, Tamangic tone systems are often casualties of contact with Nepali: Manange (Hildebrandt 2003a) and Nar-Phu have both witnessed reduction in the number of tonal distinctions made by some speakers as a result of increased use of Nepali, even in speech among co-ethnics. Further, most of the people inducted into Chantyal ethnicity were native speakers of non-tonal languages: of the native languages of the new Chantyls, only Thakali, another Tamangic language, is tonal. In any case, the massive influx of Nepali vocabulary would likely have overwhelmed the tone system.

Despite the loss of the tone system, the high-low register distinction survived, but assumed a different status in the grammar. Low-register tones were associated with murmured voice quality: murmured voice quality was retained, and, as in Nepali, came to be analyzed as a feature of consonants. In Nepali, murmur is contrastive only with stops and affricates; in Chantyal, any sort of consonant (stop, affricate, fricative, liquid, and glide) can be associated with distinctive murmur. Further, in a development whose origins are not completely understood, murmur came also to be associated with voiceless stops and affricates and with voiceless aspirated stops.<sup>9</sup>

Because of the large number of Nepali borrowings, voice, aspiration and murmur are now distinctive in all consonantal positions: initially, medially, and finally. The murmured consonants not found in Nepali are only realized as murmured word-initially.

In sum, Chantyal lost its tone system, but in its place it now has a new set of phonemic contrasts. No other Tamangic language, nor indeed any other language in the region, has this combination of a lack of a tone system and the presence of all of these phonemic contrasts involving murmur: indeed no other language in the region has the phonetic sequence of an aspirated stop followed by a phonetically murmured vowel.<sup>10</sup>

Of the other languages, Gurung had undergone the most change as a result of Nepali influence. Like Chantyal, Gurung came to allow voicing and murmur contrasts in positions other than initial position and permitted voicing contrasts in some tones. It should be noted that in the last two decades, most of the Tamangic languages have come under intense pressure from Nepali and the rate of lexical borrowing has increased considerably.

---

9. Actually, only /thf/ and /khf/ are attested; \*/phf/ is not. Phonetically, the murmur in these cases is realized on the following vowel, though is associated phonologically with the consonant.

10. In the Magar language, it is possible to have a murmured vowel preceded by a voiceless aspirated consonant, but this appears to be a recent development with quite different origins; it has, in any case, an entirely different basis phonologically (Grunow-Härsta in preparation).

### 3.2.2 Voicing opposition in liquids

Historically, the Tamangic languages contrasted voiced and voiceless liquids, as we see in the examples in (4) from Nar-Phu:

- |     |                          |         |                          |          |
|-----|--------------------------|---------|--------------------------|----------|
| (4) | hlæ [ʰ <sup>53</sup> la] | ‘month’ | hru [ʰ <sup>44</sup> ou] | ‘thread’ |
|     | læ [ʰ <sup>53</sup> la]  | ‘do’    | ru [ʰ <sup>44</sup> ru]  | ‘squint’ |

Nepali has no such contrast. Chantyal also lacks this contrast, and it is at least possible that this distinction was lost due to contact with Nepali. The other Tamangic languages still contrast voiced and voiceless liquids, at least in some dialects, though the etymological sources for the voiceless consonants may not line up in all the languages.

### 3.2.3 Retroflex consonants

In the South Asian region, there are phonetically two sorts of oppositions that are referred to as ‘retroflex’: dental *vs* true retroflex (apical post-alveolar) and dental *vs* alveolar, with the (apical) alveolar being affricated with a rhotacized off-glide such as [ɽ]. The first is characteristic of Indic languages like Nepali, the second characteristic of the Bodish group, such as the Tamangic languages (see Michailovsky 1988 for discussion).

Chantyal lacks a retroflex series of any sort, again making it unique among the Tamangic languages. Historically, the Magar languages also lacked a retroflex series, and this may account for the fact that the regional Nepali in the Dhaulagiri Zone also lacks retroflex consonants: Nepali had replaced Magar as the main regional language by the early 19th century. In any case, Chantyal has aligned with the regional (but not Standard) Nepali in lacking a retroflex series.

### 3.2.4 Phonemic opposition between an alveolar and an alveopalatal series of fricatives and affricates

The Tamangic group historically had a phonemic opposition between an alveolar and an alveopalatal series of fricatives and affricates. Spoken Nepali lacks such an opposition – though a few purists pronounce written श <ś> as an alveopalatal. In Nepali the opposition is allophonic, with alveopalatal variants pronounced before front vowels and /y/. Chantyal conforms to the Nepali pattern, as does Tamang. Gurung preserves the opposition marginally, and Thakali preserves it only in the affricates.

### 3.2.5 Word-initial /ŋ/

Chantyal apart, the Tamangic languages allow the velar nasal to appear in word initial position; Nepali does not. Chantyal has converged with Nepali in not

allowing /ŋ/ to appear word initially, though unlike Nepali, in Chantyal /ŋ/ contrasts with other nasals medially and finally.

### 3.2.6 *Λ ~ v allophony*

This refers to a characteristic of Nepali which has been passed on to a number of other Nepalese languages. In Nepali, the mid-central phoneme /ə/ has two allophones, a mid-central vowel and a low back rounded vowel in more-or-less free variation. This feature has been borrowed into Chantyal for its phoneme /ə/; it has also been borrowed into the Tamangic language Thakali (Georg 1996) and may well be found in other Tamangic languages, though the published descriptions of these languages are not clear on this point. (It is not found in Nar-Phu.)

### 3.2.7 *Variable word stress*

In Standard Nepali, stress is phonemic, though it is largely predictable from the orthography, which writes distinctions in vowel length that are no longer pronounced. In the Tamangic languages, stress was predictable and was generally fixed on the root. Where languages have borrowed large amounts of Nepali vocabulary, the borrowed items have accommodated to the Nepali stress pattern. Of the languages in our sample, this is most evidently true of Chantyal, though it is true to lesser degrees for most of the other languages.

## 3.3 Morpho-Syntax

### 3.3.1 *Word order*

Nepali and the Tamangic languages both allow AN word order, but the Tamangic languages also permit NA order. Chantyal alone stands out among the Tamangic languages in not allowing NA order, conforming to Nepali preferences.

Nepali allows only Num N order, and the Tamangic languages, again excepting Chantyal, allow only N Num order. Here too, Chantyal has conformed to the Nepali norm.

### 3.3.2 *Dative subjects*

The 'dative subject' construction is one in which the most animate argument is rendered in the case ordinarily assigned to indirect objects and, moreover, acquires many of the characteristics of subjects in the language. Semantically, dative subjects are typically non-volitional experiencers. See Masica (1991) for an extended discussion.

The dative subject construction is a prominent feature of Nepali syntax, and while examples of dative subjects can be found in all contemporary Tamangic

languages, it would appear that the construction has spread from Nepali into these languages as its extent and frequency can be correlated with the degree of contact with Nepali. Predictably, Chantyal exhibits the construction with the widest range of predicates (though still not as wide as Nepali), and Nar-Phu shows the least, with the construction there only sporadically attested. Some examples follow:

- (5) a. Nepali  
       *mə-lai bʰok lag-yo*  
       1s-DAT hunger attach-3s.PERF  
       ‘I got hungry.’  
   b. Chantyal  
       *na-ra khyana kha-i*  
       1s-DAT hunger come-PERF  
       ‘I got hungry.’  
   c. Nar-Phu  
       *ŋê čhê-čʰin*  
       1s hungry-PERF  
       ‘I got hungry.’

### 3.3.3 *Morphological valence increasing strategy*

We refer here to derivational processes which increase valence (applicative or causative); Nepali and the Tamangic languages have periphrastic causative constructions, and we are not considering these here.

Nepali has productive valence increasing derivational morphology, but the Tamangic languages do not.<sup>11</sup> Chantyal has borrowed a valence increasing derivational process from Nepali, along with (the very numerous) borrowed verbs, though interestingly the morphology and the semantic categories it encodes are not identical to that found in Nepali (Noonan 2003a); it is restricted to Nepali borrowings.

### 3.3.4 *Tense distinctions in non-finite verbals*

Nepali has a set of non-finite verbals which encode secondary [or relative] tense. The Tamangic languages lack such forms, except for Chantyal, which has innovated by utilizing the sequential converbal affix *-si* in combination with the nominalization affix *-wa*, resulting in a combination *-si-wa* which is assigned a secondary

---

11. There are a very few lexical traces of such a strategy in the Tamangic languages inherited from Proto-Sino-Tibetan, but no productive morphological processes.

past tense and is used in all of the many functions of nominalizations in Chantyal: see Noonan 1997 for a summary of these functions.<sup>12</sup>

The path by which *-si* came to be used in combination with the nominalizing suffix *-wa* is interesting and worth discussing in this context. Nepali has a sequential converb (referred to as a ‘conjunctive participle’ in traditional descriptions of Nepali) in *-erə*. As it happens, Nepali has a perfect participle in *-e* and a coordinative conjunction in *-rə*. Chantyls borrowed the coordinative conjunction from Nepali and at some point analyzed Nepali *-erə* as consisting of the perfect participle in *-e* and the coordinate conjunction *-rə*, and equated their sequential converb *-si* with Nepali *-e*. One result was that *-si*, which in other Tamangic languages is used alone as a sequential converb, is now found in combination with *-rə* in most of its uses. Once *-si* was reanalyzed as an anterior suffix in the new combination, it was available for combining with the nominalizer to create a new anterior nominalization *-si-wa*.

### 3.3.5 *Periphrastic tense-aspect with copular verbs*

Nepali has a number of non-finite verbals which combine with the two copular verbs to create a variety of tense-aspect distinctions, often including evidential and/or modal senses. The basic Tamangic pattern lacks this completely: Tense-aspect-mood distinctions are rendered by a small set of inflectional suffixes, though Nar-Phu has innovated an evidentiality system using copular verbs (Noonan 2003b), and a number of languages use nominalizations as finite verbs (Noonan 1997, *forthc.*).<sup>13</sup>

Chantyal has innovated in the direction of Nepali. It now has a considerable number of periphrastic verbal constructions composed of its two copular verbs with an expanded (relative to other Tamangic languages) set of non-finite verbals. The Chantyal system is as complex as the Nepali system, but does not map onto it perfectly, although there is considerable overlap. One reason that the systems do not match perfectly is that non-finite verbals in Chantyal have a substantially different syntax than those in Nepali: Chantyal relies much more on converbs than Nepali does (see Noonan 1999 for a list of uses of converbs in Chantyal) and nominalizations in Chantyal have a much wider array of functions than they do in Nepali (Noonan 1997).

12. Nar-Phu has also innovated a tense distinction in nominalizations, but from a completely different source: present tense *-pɛ* is distinguished from past tense *-pi*; *-pi* appears to consist of *-pɛ* together with the Tamangic past tense suffix *-i* [*<\*-ci*].

13. The Tamangic present/imperfect suffix *-m* derives from one of the copular verbs. It combines not with a non-finite verbal (as its counterpart does in Nepali), but rather with the bare verbal root, consistent with the usual pattern found in Bodish languages.



### 3.3.6 *Honorific verb stems*

Within Nepal, honorific noun and verb stems are characteristic of the Bodish group, including the Tamangic languages. They are well preserved only in those languages spoken by people who adhere to the Tibetan Buddhist faith – and can be considered a linguistic marker of that faith.<sup>14</sup> Chantyal are nominally Hindu, although their daily practice is better described as shamanistic. In any case, Chantyal has only traces of the honorific system left. Nepali has no honorific nouns or verbs.

Honorific nouns and verbs are best preserved in Nar-Phu, Manange, Thakali, and Seke. See Noonan (2003b) for exemplification of the phenomenon in Nar-Phu.

### 3.3.7 *Numeral classifiers*

Numeral classifiers are entirely absent from the Tamangic structural profile, but are present in Nepali. The Nepali classifier system is quite simple, consisting only of a human/non-human distinction.

Chantyal has borrowed the entire Nepali numeral system, and along with this system they have borrowed the Nepali system of classifiers. However, the human classifier is seldom used except in very formal speech, and this usage is consistent with the local colloquial Nepali. See Noonan (2003a) for discussion.

### 3.3.8 *Correlative constructions*

The correlative construction that concerns us here is a complex construction formed with a relative pronoun in the first clause and a demonstrative in the second: *who believes my argument, that person will be enlightened*. The Tamangic languages natively lacked this construction; it is, however, characteristic of Nepali. Chantyal has borrowed this construction from Nepali, as has Tamang; I have no evidence of this construction in any other Tamangic language.

### 3.3.9 *Additional grammatical borrowings*

The Tamangic languages have borrowed some additional grammatical morphemes and constructions.

*Relational Morphemes.* Chantyal, Gurung, and Tamang (and perhaps some others<sup>15</sup>) have borrowed relational morphemes from Nepali. All three borrowed a

---

14. It would appear that this feature characterizes Bon communities as well.

15. Grammar writers tend to underreport contact-induced phenomena when the contact situations are relatively recent. That is, they tend to report only structures which are perceived as

comparative morpheme and a benefactive ‘for’; Gurung and Tamang borrowed an allative; Chantyal borrowed a comitative; and Gurung borrowed a dative.

In the case of the comparative and benefactive there was no simple form to express these meanings, so far as I can tell: certainly there are no forms which one could reconstruct for Proto-Tamangic. In these cases, the forms could be viewed as filling in a gap. This sort of functional rationale is not available for the other borrowings: Chantyal, for example, has two other comitatives, so it is difficult to explain the borrowing in these terms. In the case of the Gurung dative, the borrowed Nepali form is *-lāi*; the Tamangic languages, however, have an inherited dative in *\*la/ra*, so it is likely that in Gurung the native form was remodeled on the Nepali one.

*Conjunctions.* All the languages except Nar-Phu, Manange, and Seke have borrowed conjunctions meaning ‘and’ and ‘or’ from Nepali. In addition, Tamang has borrowed the complementizer *ki* from Nepali (Poudel 2005).

*Periphrastic Evidential Marking.* Nar-Phu has borrowed a periphrastic evidential construction involving the copula from the Tibetan Complex. This construction is central to the organization of the verbal system and signals that the information so reported was obtained indirectly, through inference, hearsay, etc. (Noonan 2003b). All the other Tamangic languages and Nepali employ sentence-final particles to signal evidentiality; Nar-Phu uses sentence-final particles too, along with the obligatory expression of evidentiality in the verb complex.

### 3.3.10 *Features not borrowed into the Tamangic languages from Nepali*

For the sake of a more balanced presentation, it is worth considering features that were not borrowed into Chantyal from Nepali.

Nepali features that were not borrowed into Chantyal:

1. *Person-number features on verbs:* The Central Himalayish languages as well as Nepali have person-number marking on verbs, but the Tamangic languages have not developed this feature.
2. *Honorific grades of second and third person pronouns:* This feature has been borrowed into Chantyal and some of the other languages in a modest way: There is some tendency to use the 2PL pronoun for singular honorific reference. However, there is nothing comparable to the multiple-level honorific system of Nepali in the Tamangic languages.

---

‘native’ or ‘nativized’ if obviously borrowed. This means that much evidence for contact phenomena is systematically excluded from grammars and grammar sketches.

3. *Split-ergative syntax*: Except for Nar-Phu and Manange, the Tamangic languages demonstrate remarkably consistent ergative marking on transitive subjects. Nar-Phu and Manange have ergative marking that is based partly on discourse factors, for example topicality. This feature may have been influenced by Tibetan where a similar situation exists, but the matter requires more study.
4. *Valence decreasing strategy*: No Tamangic language has borrowed a valence decreasing strategy (e.g. passive) from Nepali or developed one independently.
5. *Finite subordinate clauses (except as complements of 'say')*: Nepali has finite subordinate clauses of various sorts, but the Tamangic languages as a group do not employ finite subordination except as complements of 'say' (the relative construction discussed above does not involve straightforward subordination). Poudel (2005) reports that Tamang has borrowed a finite subordinate construction from Nepali along with the complementizer *ki*.

Features not found in Nepali that were retained in Tamangic languages:

1. *Compound case*: The Tamangic languages (and other Bodish languages) utilize compound case (Noonan 2005, 2008). In fact, by the standards of the region, Chantyal makes extensive use of this feature, more than other Tamangic languages.
2. *Verbal with nominal and adjectival functions*: The Tamangic languages have retained this feature, and have apparently extended the use of this verbal in ways that mark the languages as increasingly different from Nepali (Noonan 1997).
3. *Double demonstratives*: Chantyal, and to a small extent some other languages in the group, employ 'double demonstratives' (Noonan 2001). Nepali does not have this feature.

### 3.4 Rhetorical strategies

Very little research has been carried out on rhetorical strategies in the languages of Nepal – or, indeed, from a typological perspective, on languages generally. Nonetheless, some observations can be made on the effects of contact on certain rhetorical strategies employed in the Tamangic languages.

#### 3.4.1 Referential density

It has long been recognized that languages differ in the percentage of possible arguments that, in continuous text, are given overt expression versus those that are

not. The percentage of overtly expressed core arguments to possible (i.e. notional) core arguments is referred to as 'referential density' (RD). We know, for example, that languages long resident in the western littoral of Eurasia have high RD values, whereas languages in East Asia, in particular languages in the Sino-Tibetan family, have low RD values. Two studies (Bickel 2003; Noonan *ms*) have investigated differences in referential density between Indo-European and Tibeto-Burman languages of Nepal, and both found that, in the terms of Noonan (*ms*), Indo-European Nepali scores in the moderate RD value range, whereas the Tibeto-Burman languages, including Chantyal, score in the low range.<sup>16</sup> In Noonan's (*ms*) crosslinguistic investigation, the lowest RD values of any language were recorded for the Tamangic language Manange: one discourse had an RD value of 34%.

In that same study, some Chantyal discourses, however, were in the moderate range, approaching percentages characteristic of Nepali, which averages 62%, while others had lower RD values. For the discourses with the moderate RD values, it is not clear whether any possible influence comes from Nepali or from English, as the speakers producing these discourses are fluent in English and use it frequently. Still, the mean RD values for all Chantyal discourses in the study, 50.3%, are somewhat higher than those of the Tamangic languages influenced by Nepali only recently, Manange and Nar-Phu, 38% and 40%, respectively.<sup>17</sup> So, while the evidence for Nepali influence is not striking, we can tentatively conclude that Nepali has had some slight influence on Chantyal with regard to RD.<sup>18</sup>

### 3.4.2 *Direct speech style*

The 'direct speech style' (Noonan 2006b) exploits the device of the direct quote in order to accomplish narrative objectives that in other languages/genres might be accomplished in other ways (or, indeed, be left unexpressed). In other words, the device of the direct quote is often best interpreted as signaling not someone's actual speech, but rather information about supposed motives, plans, decisions, etc. Given space limitations, the reader is referred to Noonan (2006b) for extensive discussion. An illustrative example follows:

---

16. The somewhat arbitrary ranking used in Noonan (*ms*) is: low is less than 50% of possible arguments overtly expressed, moderate is 50–70%, and high is above 70%.

17. The mean percentage figures given here are based on small samples and should be interpreted accordingly. Nonetheless, they accord with my general impressions of how discourses work in these languages.

18. Bickel (2003) and Noonan *ms* would predict on structural grounds that Chantyal would have lower RD values than Nepali, and in general these predictions are borne out by the analyzed data.

- (7) *phəlpʰul-ra ni məstəy thʰan-si-rə*  
 fruit-DAT little a+lot store-ANT-SEQ  
 ‘‘Having stored the fruit a little longer.’’
- pari-wa pəri-m” bʰi-si-rə*  
 make.happen-NOM happen-NPST say-ANT-SEQ  
 ‘[we] must make [raksi] happen”, having said.’
- nʰi-sə ənnə-bʰənda phəlpʰulce məstəy səmmə thʰana-m*  
 we-ERG grain-COMP fruit little a+lot until store-NPST  
 ‘we store the fruit a little longer than grain.’  
 ‘Because we need to ferment the fruit a little longer in order to make raksi,  
 we keep the fruit a little longer than grain.’

Chantyal makes extensive use of the direct speech style. Published data on the other Tamangic languages make it difficult in most cases to evaluate the extent to which they use it too, though it is clear that Nar-Phu and Manange do not make much use of it. Nepali makes extensive use of the direct speech style, especially in colloquial registers.

#### 4. Hypotheses on the linguistic effects of language contact

In this section, we will consider a number of hypotheses proposed in the last few years concerning the linguistic outcomes of language contact situations. We will compare the predictions made by these hypotheses against the data we have been examining from the Tamangic languages. The hypotheses will be divided informally into five sorts:

- (8) a. Factors relating to communities in contact and linguistic change;
- b. General constraints on contact-induced change;
- c. Specific predictions regarding borrowing of morphemes;
- d. General predictions regarding borrowing of grammatical constructions;
- e. Specific predictions regarding borrowing of grammatical constructions.

Many hypotheses have been proposed in the last few years, and we will only consider here those that bear on the kind of data we have been considering from the Tamangic languages.

a. Factors relating to communities in contact and linguistic change

1. Nichols (1992: 193)

It can be concluded that contact among languages fosters complexity, or, put differently, diversity among neighbouring languages fosters complexity in each of the languages.

The Chantyal data would seem to confirm this, at least in the areas of morphosyntax and rhetorical strategies, which became more complex overall. In Table 4 I list the morphosyntactic and rhetorical features discussed in §3 and assess informally the effects of contact as either increasing or decreasing complexity:

**Table 4.** Increase or decrease in complexity as a result of change.

Feature	Increase or decrease in complexity as a result of change
Allow NA order	<b>Decrease:</b> Chantyal now allows only AN order
<i>N Num order</i>	<b>Neutral:</b> Chantyal substituted one word order for another
<i>Dative subjects</i>	<b>Increase:</b> there is a new construction in Chantyal
<i>Morphological valence increasing strategy</i>	<b>Increase:</b> no comparable morphological strategy existed pre-contact
<i>Tense distinctions in non-finite verbals</i>	<b>Increase:</b> the contact-induced reanalysis of the old sequential converbal morpheme has resulted in a new set of positions
<i>Periphrastic tense-aspect with copular verbs</i>	<b>Increase:</b> a large number of TAM distinctions are now available with the rise of a set of new constructions
<i>Honorific verb stems</i>	<b>Decrease:</b> the contrast is no longer available
<i>Numeral classifiers</i>	<b>Increase:</b> there is a new construction in Chantyal
<i>Correlative constructions</i>	<b>Increase:</b> there is a new construction in Chantyal
<i>Referential density</i>	<b>Neutral:</b> changes in RD cannot straightforwardly be associated with either an increase or a decrease in overall complexity
<i>Direct speech style</i>	<b>Increase:</b> the adoption of the direct speech style allowed for an increase in expressiveness and, probably, the addition of new constructions

The phonology, however, would appear not to have become more complex as a result of contact. Chantyal lost its tone system as a result of contact with Nepali, and appears to have lost other phonemic contrasts as well. Other languages more recently, e.g. Manange and Nar-Phu, have also found their tone systems weakened as a result of contact with Nepali (Hildebrandt 2003a): if these languages survive for much longer, their tone systems will likely be lost.

Overall, then, Nichols' hypothesis is confirmed by the Tamangic data with regard to morphosyntax and rhetorical strategies, but not with regard to phonology.

2. Thomason (2001:78–79).

If the shifting language group is large relative to the target language group, then the shifting language group's linguistic features will become established in the target language.

This hypothesis seems to be referring to substratic influence. So far as we know, this hypothesis would apply among the Tamangic languages only to Chantyal, where a large percentage of the population at one point were not native speakers of the language. Given that the bulk of the shifting population spoke Nepali, then this hypothesis would be confirmed as regards the phonology, which definitely aligned itself with Nepali in many respects. This seems to have occurred in other aspects of grammar as well. We can say, therefore, that while this hypothesis is rather vague, it is consistent with the data.

3. Ross (2001), Curnow (2001), et al.

In cases of language shift, i.e. where one community adopts the language of another, the affected elements in the recipient language are phonology and morphosyntax, but not lexicon.

This hypothesis says, in effect, that substratic influence is not accompanied by lexical borrowing. I can think of many instances which would seem to confirm this (Irish English, Lango (whose speakers shifted from a Teso-like language to an Acholi-like one: Noonan 1991)), but it is not confirmed in the case of Chantyal, where massive lexical borrowing is found along with a large shifting population. It is, of course, possible that the massive lexical borrowing occurred after the period of shift, but this cannot be confirmed at this stage.

b. General constraints on contact-induced change

Thomason (2001)

- (i) Languages that are typologically very different are likely to follow the borrowing scale (70–71) closely, while languages that are very similar are likely not to do so in all respects.

This hypothesis suggests that borrowing is more constrained – and more limited to lexical borrowing – when languages are typologically different than when they are typologically similar. Nepali and the Tamangic languages have a number of typological differences, but overall are more similar than they are different: both are SOV languages, with similar sorts of case-marking systems; constituent word

order is more-or-less similar; both favor converbals for structuring narratives (Tamangic rather more than Nepali); both are mostly agglutinative (the Tamangic languages more so than Nepali); etc. So, borrowing, apart from lexical borrowing, would be predicted to occur in contact situations between Nepali and Tamangic, and that is what we find.

- (ii) Marked features in a language are less likely to be learned by the shifting group (because they are harder to learn). (Thomason 2001: 76)

This would seem to be confirmed by the data: tone, the voiced/voiceless opposition in liquids, and honorific nouns and verbs are marked features that were not learned by the shifting group in the case of Chantyal.

c. Specific predictions regarding borrowing of morphemes

Moravcsik (1978)

- (i) Non-lexical properties of a language cannot be borrowed unless lexical properties have been borrowed first.

This hypothesis contradicts a proposal by Ross below, which states that lexical and syntactic borrowing are independent. The data from our region are more consistent with Ross's hypothesis than Moravcsik's.

- (ii) No member of an unaccentable class (i.e. bound morphemes) can be borrowed unless a member of an accentable class which contains the unaccentable member (e.g. an inflected word) is borrowed first.

This is confirmed by our data: no language has borrowed bound morphology without borrowing nouns, verbs, etc. first.

- (iii) A noun must be borrowed before any non-nominal lexemes can be borrowed.

This aligns with Haugen's (1950) implicational scale and is generally confirmed by our data.

- (iv) Grammatical morphemes must be borrowed with their linear order with respect to their head.

This is confirmed by our data, but since the word orders of the Tamangic languages and Nepali are overall rather similar – in particular with regard to the placement of inflectional morphology – it is confirmed weakly.

d. General predictions regarding borrowing of grammatical constructions

Ross (2001: 139).

Lexical borrowing is independent of 'syntactic borrowing'.



Our Tamangic data confirm this hypothesis weakly, since we have cases of languages which have borrowed vocabulary from Nepali without borrowing any syntactic constructions (e.g. Nar-Phu), as well as languages which have borrowed both (Chantyal). We do not, however, have any instances of Tamangic languages which have borrowed only syntactic constructions without borrowing vocabulary.

What this means is that we have no instances of purely substratic influence in the development of the Tamangic languages so far as we can tell at this stage. However, within our region we can find cases which would confirm the other half of Ross's hypothesis, namely instances of syntactic borrowing in the absence of lexical borrowing. The regional Nepali in the Dhaulagiri Zone has almost no lexical borrowing from Tibeto-Burman languages, but in the most colloquial versions (i.e. those relatively uninfluenced by Standard Nepali) we can find a certain amount of syntactic influence, namely consistently ergative syntax (i.e. not the standard aspect-based split ergativity of Standard Nepali), avoidance of finite subordination, etc. This, and various phonological features are almost certainly evidence of substratic influence from the Tibeto-Burman languages which were formerly spoken in the region. So, regionally, Ross's hypothesis is confirmed.

e. Specific predictions regarding borrowing of grammatical constructions

1. Ross (2001: 149–50).

The sequence of changes associated with metatypy is: lexical semantic patterns > discourse structure > clause linkage > clause-internal structure > phrase structure > word-internal structure.

Ross was primarily concerned with instances of metatypy, but we will interpret his hypothesis here more generally.

Because Chantyal and Nepali were relatively close typologically, even at the point of first contact, it isn't obvious whether or not Chantyal has undergone metatypy, despite borrowings of all sorts. Nonetheless, we can find in Chantyal examples of most of the items in Ross's implicational list. The one interesting exception is clause linkage: Chantyal has resisted borrowing constructions involving clause linkage. That is, Chantyal has not borrowed constructions involving finite subordination from Nepali, and it has retained the Bodic conflation of nominalization and relativization (Noonan 1997, *forthc.*) rather than conforming to Nepali, which distinguishes the two.

Other Tamangic languages, however, seem to conform to Ross's hypothesis a bit better. Tamang, for example, borrowed constructions involving the first three, but has not (so far as I can tell) gone further.

2. Haig (2001:220).

- (i) Patterns of clause linkage and of basic constituent order are features which diffuse quickly across large areas, cross-cutting genetic groupings.

Nepali and Tamangic constituent order aligns fairly well, but as noted there are differences in the noun phrase: numeral/noun, adjective/noun. With the exception of Chantyal, the Tamangic languages retain their historic ordering possibilities, though Chantyal now aligns with Nepali. It is not clear, therefore, how to evaluate Haig's hypothesis with regard to constituent order.

With regard to clause linkage, we discussed the matter above in our discussion of Ross's implicational scale. Clause linkage patterns have not diffused quickly across the Tamangic languages from Nepali, since there are only a few cases where such patterns have diffused despite at least two hundred years of intense contact for half of the Tamangic languages.

- (ii) Linear alignment will proceed from larger to smaller units, starting perhaps with the narrative organization, means of expressing direct speech, topic introduction and tracking, and progressing down through clause coordination, subordination, and constituent order in the clause. I would, however, exclude grammatical subsystems such as the numerals from the domain of these generalizations. (Haig 2001:220)

As was noted with regard to Ross's implicational scale, which Haig's scale resembles, this sort of development is mostly confirmed – or at least not disconfirmed – by the data.

## 5. Conclusions

This paper has had two goals: To present data on the effects of language contact on the Tamangic languages and to determine the degree to which hypotheses concerning the outcomes of language contact are either supported or not supported by the Tamangic data.

Many of the hypotheses discussed in §4 were confirmed by the data, though some only weakly or in part. Clearly much work remains to be done, both in terms of formulating hypotheses that are testable and in terms of putting together a sufficiently diverse collection of data against which to test them. Fortunately, there are now a number of projects in which scholars are attempting to do just that.

## Abbreviations

1s = first person singular; 3s = third person singular; ANT = anterior; COMP = comparative; DAT = dative; ERG = ergative; NOM = nominative; NPST = non-past; PERF = perfective; SEQ = sequential.

## References

- Bickel, Balthasar. 2003. Referential density in discourse and syntactic typology. *Language* 79 (4): 708–736.
- Curnow, Timothy Jowan. 2001. What language features can be ‘borrowed’? In *Areal Diffusion and Genetic Inheritance: Case Studies in Language Change*, R. M. W. Dixon & A. Y. Aikhenvald (eds), 412–436. Oxford: OUP.
- Georg, Stefan. 1996. *Marphatan Thakali* [Lincom Studies in Asian Linguistics 02]. München: Lincom.
- Glover, Warren. 1974. *Sememic and Grammatical Structures in Gurung (Nepal)*. Norman OK: Summer Institute of Linguistics.
- Glover, Warren, Glover, Jessie R. & Gurung, Deu Bahadur. 1977. *Gurung-Nepali-English Dictionary, with English-Gurung and Nepali-Gurung Indexes* [Pacific Linguistics, Series C, Nr. 51]. Canberra: Research School of Pacific Studies, ANU.
- Grunow-Härsta, Karen. In preparation. A Descriptive Grammar of Two Magar Dialects. PhD dissertation. University of Wisconsin-Milwaukee.
- Haig, Geoffrey. 2001. Linguistic diffusion in present-day East Anatolia: From top to bottom. In *Areal Diffusion and Genetic Inheritance: Case Studies in Language Change*, R. M. W. Dixon & A. Y. Aikhenvald (eds), 195–224. Oxford: OUP.
- Haspelmath, Martin. 2004. How hopeless is genealogical linguistics, and how advanced is areal linguistics? Review of R. M. W. Dixon & A. Y. Aikhenvald (eds), *Areal Diffusion and Genetic Inheritance: Case Studies in Language Change* [Studies in Language 28:1], 209–223. Oxford: OUP.
- Haugen, Einar. 1950. The analysis of linguistic borrowing. *Language* 26: 210–31.
- Hildebrandt, Kristine. 2003a. Manange Tone: Scenarios of Retention and Loss in Two Communities. PhD dissertation, University of California.
- Hildebrandt, Kristine. 2003b. A grammar and dictionary of the Manange language. In *Tibeto-Burman Languages of Nepal: Manange and Sherpa*, C. Genetti (ed.), 2–189. Canberra: Pacific Linguistics.
- Honda, Isao. 2003. A sketch of Tangbe. In *Themes in Himalayan Languages and Linguistics*, T. R. Kansakar & M. Turin (eds). Kathmandu: South Asia Institute and Tribhuvan University.
- Hoshi, Michiyo. 1986. An outline of the Prakaa Grammar: A dialect of the Manang language. In *Anthropological and Linguistic Studies of the Kathmandu Valley and the Bandaki Area in Nepal* [Monumenta Serindica No. 15], H. Ishii et al. (eds), 187–317. Tokyo: Institute for the Study of the Languages and Cultures of Asia and Africa.
- Masica, Colin. 1991. *The Indo-Aryan Languages* [Cambridge Language Surveys]. Cambridge: CUP.

- Mazaudon, Martine. 1973. *Phonologie Tamang*. Paris: Société d'études linguistiques et anthropologiques de France.
- Mazaudon, Martine. 1978a. Consonantal mutation and tonal split in the Tamang sub-family of Tibeto-Burman. *Kailash: A Journal of Himalayan Studies* 6(3): 157–79.
- Mazaudon, Martine. 1978b. *Tibeto-Burman Tonogenetics. Linguistics of the Tibeto-Burman Area* 3.2 [Monograph issue].
- Mazaudon, Martine. 1993–1994. *Problèmes de comparatisme et de reconstruction dans quelques langues de la famille tibéto-birmane*. Thèse: Université de la Sorbonne nouvelle: Paris III.
- Mazaudon, Martine. 2003. Tamang. In *The Sino-Tibetan Languages*, R. LaPolla and G. Thurgood (eds), 291–313. London: Routledge.
- Moravcsik, Edith. 1978. Language contact. In *Universals of Human Language, Vol. 1*, J. Greenberg, C. Ferguson & E. Moravcsik (eds), 93–122. Stanford CA: Stanford University Press.
- Michailovsky, Boyd. 1988. Phonological typology of Nepal languages. *Linguistics of the Tibeto-Burman Area* 11(2): 25–50.
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago IL: University of Chicago Press.
- Noonan, Michael. 1991. *A Grammar of Lango*. Berlin: Mouton de Gruyter.
- Noonan, Michael. 1996. The fall and rise and fall of the Chantyal language. *Southwest Journal of Linguistics* 15(1–2): 121–136.
- Noonan, Michael. 1997. Versatile nominalizations. In *Essays on Language Function and Language Type. In Honor of T. Givón*, J. Bybee, J. Haiman & S. Thompson (eds), 373–394. Amsterdam: John Benjamins.
- Noonan, Michael. 1999. Converbial constructions in Chantyal. In *Topics in Nepalese Linguistics*, Y. P. Yadava (ed.), 401–420. Kathmandu: Royal Nepal Academy.
- Noonan, Michael. 2001. The 'double demonstratives' of Chantyal, *Linguistics of the Tibeto-Burman Area* 24(2): 173–188.
- Noonan, Michael. 2003a. Chantyal. In *The Sino-Tibetan Languages*, R. LaPolla & G. Thurgood (eds), 315–335. London: Routledge.
- Noonan, Michael. 2003b. Nar-Phu. In *The Sino-Tibetan Languages*, R. LaPolla & G. Thurgood (eds), 336–352. London: Routledge.
- Noonan, Michael. 2005. Spatial reference in Chantyal. In *Contemporary Issues in Nepalese Linguistics*, Y. Yadava (ed.), 151–188. Kathmandu: Linguistic Society of Nepal.
- Noonan, Michael. 2006a. The rise of ethnic consciousness and the politicization of language in West-Central Nepal. In *Globalization, Technological Advances, and Minority Languages in South Asia*, A. Saxena & L. Borin (eds), 161–174. Berlin: Mouton de Gruyter.
- Noonan, Michael. 2006b. Direct speech as a rhetorical style in Chantyal. In *Himalayan Linguistics, Himalayan Linguistics Journal* 6, 1–32.  
<http://www.uwm.edu/Dept/CIE/HimalayanLinguistics/index.html>
- Noonan, Michael. 2008. Case compounding in the Bodic languages. In *Case and Grammatical Relations*, G. Corbett & M. Noonan (eds). Amsterdam: John Benjamins.
- Noonan, Michael. Forthc. Nominalizations in Bodic languages, to appear in the proceedings of NRG3, edited by E. Seoane & M. J. Lopez-Couso. Amsterdam: John Benjamins.
- Noonan, Michael. *ms.* A crosslinguistic investigation of referential density. Downloadable at: <http://www.uwm.edu/~noonan/>
- Noonan, Michael, Bhulanja, Ram Prasad, Chhantyal, Jag Man & Pagliuca, William. 1999. *Chantyal Dictionary and Texts*. Berlin: Mouton de Gruyter.

- Poudel, Kedar Prasad. 2005. Clause combining in Dhankute Tamang. In *Contemporary Issues in Nepalese Linguistics*, Y. Yadava et al. (eds), 193–204. Kathmandu: Linguistic Society of Nepal.
- Ross, Malcolm. 2001. Contact-induced change in Oceanic Languages in north-west Melanesia. In *Areal Diffusion and Genetic Inheritance: Case Studies in Language Change*, R. M. W. Dixon & A. Y. Aikhenvald (eds), 134–166. Oxford: OUP.
- Thomason, Sarah. 2001. *Language Contact: An Introduction*. Washington: Georgetown University Press.

# Total reduplication vs. echo-word formation in language contact situations\*

Thomas Stolz  
University of Bremen

This contribution addresses the question how a potential universal behaves in language-contact situations. Instances of total reduplication are reviewed for a variety of mainly Indo-European languages of Europe. In these varieties, total reduplication is attested with an unexpectedly high token frequency – a property they share with their genetically unrelated neighbors. Thus, contact-induced increase of frequency cannot be ruled out as a possible explanation. However, the evidence is often controversial. This is different with echo-word formation – a marked sub-species of total reduplication. It is shown that echo-word formation has an areally biased geo-linguistic distribution which can be explained convincingly only via language contact. In a separate section, the properties enhancing borrowability of echo-word formation patterns are discussed.

**Keywords:** reduplication, echo-word formation, frequency of use, sprachbund, Turkish, Yiddish, Hindi, syllable structure, areal linguistics

## 1. Introduction

Given that total reduplication (= TR) is a potential universal (Moravcsik 1978: 328; Stolz 2006), how do we deal with the phenomenon in studies devoted to language contacts? I intend this contribution to show that the assumed universality of a feature does not necessarily mean that it remains unaffected by contact-induced

---

\* I am grateful to Andreas Ammann, Birsal Karakoç, Tamari Khizanishvili, Nataliya Levkovich, Conxita Lleó, Rachel Montague, Mercedes Pérez, Lukas Pietsch, Reyhan Şahin, Stavros Skopeteas, Christel Stolz, the editors, and two anonymous reviewers. This study is part of the project *Totale Reduplikation – die Arealität eines Universales* financed by the *Deutsche Forschungsgemeinschaft* (DFG). Disclaimer: I bear the sole responsibility for the contents and wording of this article.

developments. I concur with Koptjevskaja-Tamm & Wälchli (2001:627) “that the frequency at which a feature occurs throughout an area is an important and often under-estimated factor in areal linguistics.” The increase/decrease of type and token frequencies associated with a given feature may be the result of convergence triggered by language contact. This applies to TR (Stolz 2004) as I argue in Section 2 by way of demonstrating that the role of TR has been largely underestimated for the languages of Europe. However, the evidence of contact-induced quantitative and/or qualitative changes in the realm of TR is often inconclusive. There always remains the possibility that the languages under scrutiny have developed independently of each other. The shape they have given to the domain of the potential universal TR might look the same only incidentally, cf. Section 3. The dilemma of how to handle a universal like TR in contact linguistics without stating only the trivially obvious can be solved by investigating the areal distribution of variants of TR which deviate from the prototype of TR. In Section 4, I discuss the evidence from total-reduplication-*cum*-variation (= TRCV) in order to prove that not all kinds of TR pose the same problems to students of language contact. If marked variants of TR can be shown to diffuse via contact then nothing prevents us from assuming the same for proper TR. Why TRCV is particularly suited for transfer in language contact situations is the topic of Section 5 where I focus especially on phonological issues. In the conclusions, I explicate the lessons these findings teach areal linguists, specialists of language contact, typologists and universals researchers.

The two variants of TR have the following properties:<sup>1</sup>

- a. proper TR:<sup>2</sup> The prototypical pattern requires two phonologically identical and immediately adjacent strings of segments (called original and copy), each of which represents a syntactic word equipped with identical content in the object language. No syntactic boundary runs between the two strings of segments. They are either inflected identically or share a common slot for inflectional information. The combination of the two strings of segments has a constructional meaning which is different from the meaning associated with a single string of segments.

---

1. In the literature on reduplicative patterns, it is often unclear whether the authors distinguish repetition from reduplication systematically as they form a continuum (Gil 2005).

2. In what follows I adhere to a traditional view of total reduplication. Morphological doubling theory (Inkelas & Zoll 2005) is not particularly well suited for an explanation of the phenomena I discuss as it has a very wide scope and assumes that reduplication is basically semantic in nature.

- b. TRCV comes close to this definition but fails to meet one criterion, namely absolute phonological identity, because it consists of two immediately adjacent strings of segments which differ as to the quality of the filler of one of their slots, meaning: they share all segments but one. This partial phonological identity has another effect which jeopardizes the word-hood of the string of segments which constitutes the copy in the pair of strings. The string which displays one segment which is not copied from the original string often is a non-word and thus only exists in combination with the original string (i.e. it is bound to the co-occurrence of the original and appears only in this particular construction). The non-word itself has no autonomous semantics. The most commonly used term for reduplication of this kind is echo-word formation.

TR has already been the focus of various language-contact oriented studies (Abbi 1992; Stolz 2002, 2004; Kouwenberg 2003; Stolz & Sansò [forthc.]) which emphasize the susceptibility of TR to language contact and its areality.

## 2. Painting the European map anew...

Rubino (2005) surveys the synchronic distribution of reduplicative patterns on the world map. His findings largely corroborate the widely held belief that TR is not restricted by genetic and structural factors. The noticeable areal dispreference of his European sub-sample for reduplicative patterns is caused by a misinterpretation of the silence on the part of the consulted descriptive grammars with regard to TR (Bollée 1978). Discounting the dubious evidence of languages lacking reduplication altogether (= 56 languages or 15%), one observation Rubino (2005: 115) makes in his conclusion is particularly telling: "Western Europe is one area where reduplication does not play a critical role in the morphology. However, many Indo-European languages in the east, which are in contact with other language families, do have reduplicative morphemes." Indirectly he claims that reduplication has either been borrowed by certain easterly Indo-European languages from their differently affiliated neighbors or has been retained as an inheritance from the ancestor language exactly because of the parallels in the contiguous languages. This interpretation of the geo-linguistic distribution also implies that more westerly Indo-European languages have lost erstwhile productive reduplicative patterns without acquiring new ones. Many of those Indo-European languages which Rubino classifies as lacking reduplicative patterns turn out to employ TR.

On Rubino's world map, there are thirteen languages spoken in countries which are presently part of the European Union (ergo: "western" languages). With the exception of Hungarian, all of them lack productive reduplication according



to the criteria Rubino employs. This shared absence of reduplication looks like another areal feature of European languages as Indo-European languages of various phyla (Germanic, Romance, Slavic, Greek) share this feature with Finnish and Basque. Genetic affiliation does not preclude the existence of TR. On the other hand, areal factors seem to be more important (= allocation in Europe vs. Africa). Moreover, a glance at actual texts reveals that many of Rubino's examples of reduplication-less languages turn out to make frequent use of TR. (1)–(3) stem from Modern Greek, Italian and Basque – three of the 56 languages without reduplication in Rubino's sample.

- (1) Modern Greek [LPP Greek 109]

*Kápote-kápote*                      *tha ton*                      *katadhikazeis se thánato*  
 sometimes-sometimes    FUT    3SG.ACC    condemn.2SG in    death.ACC  
 'Every once in a while you will sentence it [= the rat] to death.'

- (2) Italian [LPP Italian 46]

*Se i*                      *camini*                      *sono*                      *ben puliti*                      *bruciano*  
 if    DET.M.PL    chimney.M.PL    be.3PL    well    clean.PART.M.PL    burn.3PL  
*piano piano*  
 slow    slow  
 'If the chimneys are kept clean, they burn smoothly.'

- (3) Basque [LPP Basque 52]

*begi-ra-begi-ra*                      *geratu*                      *zitzaion*                      *Printze txiki-a*  
 eye-ALL-eye-ALL    hold    3SG.ABS/3SG.DAT.PAST.be    Prince    small-DEF  
 'The little prince stared at him very intently.'

For many Mediterranean languages, TR is not exceptional. On the contrary, TR is a frequently employed means of expression in the Circum-Mediterranean languages (Wierzbicka 1986) whereas partial reduplication is a marginal phenomenon in Europe (Rainer 1998). In Stolz (2002, 2004) and Stolz & Sansò (forthc.), it is shown that languages with different genetic and typological background spoken around the Mediterranean and the immediate hinterland lean towards TR. This is reflected by the type and token frequencies of TR which I have determined via the statistical analysis of a parallel literary corpus based on the French original and 64 translations of *Le Petit Prince* in European languages/regional varieties (including those of Anatolia and the Trans-Caucasus). In Table 1, I present the type and token frequencies of TR in the sample text. I have included all of the languages with the same or higher type and token frequencies as the above three languages. An asterisk identifies those languages which are situated in the part of Europe for which Rubino postulates the general absence of reduplicative patterns.

**Table 1.** Type and token frequency of total reduplication (LPP corpus).

tokens			types		
rank	language	n	language	n	rank
1	Basque*	45	Basque*	28	1
2	Azeri	36	Azeri	26	2
3–4	Corsican*, Maltese*	23	Turkish	13	3
			Armenian, Corsican*, Sardinian*	9	4–6
5	Turkish	19			
6	Sardinian*	18			
7	Armenian	13	Hungarian*	8	7
8	Greek*	12	Greek*	7	8–9
9	Hungarian*	11	Maltese*		
10	Albanian*	10	Aragonese*	5	10
11	Aragonese*	7	Italian*	4	11
12–13	Georgian, Italian*	6			
Sum	13	229	11	125	Sum

All the languages listed in Table 1 belong to the Mediterranean and neighboring regions. Nine of the thirteen languages with high token frequency of TR and eight of the eleven languages with high type frequency of TR are situated in Europe, disproving Rubino's assumption that "western" Europe in its entirety behaves differently from the rest of the world.<sup>3</sup> Moreover, the evidence is such that a division of Europe into a relatively reduplication-unfriendly north and a reduplication-friendly south appears more plausible although even this division can only be accepted with a grain of salt.<sup>4</sup> Given that languages from the European north differ from their southern counterparts as to their readiness to employ TR, the Mediterranean evidence speaks in favor of an areal phenomenon because the predilection to use TR is shared by representatives of different (macro-)phyla and an isolate. Since close relatives of some of the languages included in Table 1 (e.g. French vs. Italian, Finnish vs. Hungarian) fail to reach the same frequencies, it is

3. For those languages which are spoken on the eastern fringes of Europe, Rubino's (2005) classification is more accurate – with the exception of Lezgian (Haspelmath 1993).

4. There are studies on reduplication phenomena in northerly languages like English (Thun 1963), German (Wiese 1990) and Swedish (Lindström 1999). These investigations reveal two things: (a) in these languages, lexicalized cases of TR are also attested but they do not form any productive patterns, and (b) even in sizeable corpora covering all kinds of genres, the frequency of TR is minimal.

very likely that the high frequency of TR has been diffused via language contact and not handed down from a common ancestor language of a given phylum.

TR was possibly already there before contact started to have its effects. Prior to contact, frequency values were relatively low, but they rose considerably when contact was established. We need qualitative evidence to verify/falsify this hypothesis.

### 3. Hard to prove

According to Rubino (2005), the languages of the British Isles are all of the same type, namely languages without reduplication. This is in stark contrast to observations made in Viereck, Viereck & Ramisch (2002: 119) who argue that in Contact or Island English, spoken on the Hebrides, intensification of adjectives is expressed by TR because in Scots Gaelic, the repetition of the adjective counts as intensification, thus *I have a good good house*. Likewise, it is assumed that the English dialect of Pembrokeshire (Southwest Wales) displays similar constructions copied from Welsh. The way in which the phenomena are presented suggests that the authors exclude TR from the inventory of possible constructions of “proper” English. Where they nevertheless occur they are attributed to a Celtic substratum. This, in turn, is tantamount to postulating TR as a relatively common device in Scots Gaelic and Welsh (both of which are marked as reduplication-less languages on Rubino’s map). In Calder (1980: 16), all the reader is told about reduplication is that “[t]he repetition of a word (or part thereof) has always been a feature of the Gaelic language.” However, tangible proof of Scots-Gaelic TR is hard to come by. For Welsh, the situation is less complicated as TR is indeed frequently attested in modern prose, cf. *bach bach* {small} {small} ‘very small’, cf. Thomas (1996: 216–217).

Since the dialect grammars of English varieties in Scotland and Wales mentioned in the *dtv-Atlas Englische Sprache* do not provide evidence for TR, these cases remain somewhat dubious. In addition, the one example given above from Island English contains a reduplication pattern which would pass the acceptability test for spoken English anyway, at least for some speakers. Thus, there is no compelling evidence for assuming a contact-induced increase of the frequency of TR in regional varieties of English.

Another difficult case is Cymbric, the Germanic variety spoken in a handful of North-Italian villages. Cymbric is still relatively close to the Bavarian dialects of German and thus we would not expect to find TR on a grand scale in Cymbric texts. In German, TR of a very small set of attributive adjectives (such as *ein alter alter Mann* ‘an old old man’) and a restricted number of adverbs (especially

intensifiers *sehr sehr schön* 'very very beautiful') is possible, although not necessarily common. German shares these options with spoken English and other members of the Germanic phylum. In Cymbrian, TR seems to be used slightly more often than expected from the Germanic vantage point. Popular texts contain instances of TR which would be considered acceptable but somewhat stylistically marked by standard speakers, e.g. *laise laise* {quietly} {quietly} 'very slowly' which is reminiscent of Italian *piano piano* with the same meaning. Cymbrian texts contain many examples of this kind, but none would be ruled out in colloquial German (or the fairy-tale genre). What distinguishes Cymbrian from spoken standard German is the frequency with which TR is used. Whether this can be directly attributed to the influence of the Italian patterns remains an open question. In light of the fact that Cymbrian bears evidence of heavy Italianisation, it is probable that speakers of Cymbrian employ TR much more often because of the high frequency of the phenomenon in their second language.

The same argument holds for other autochthonous minority languages in Italy although Greek and Albanian varieties in southern Italy do not only converge with their Romance neighbours, but also have the feature in common with the standard varieties of Modern Greek and Albanian in their countries of origin. Rohlfs (1967), who observed a constant increase of TR the further south one looks on the Italian dialect map, assumed that TR was introduced into Italo-Romance via Byzantine Greek. Sgroi (1986) saw another possible source in Siculo-Arabic, whose patterns survived in Sicilian (from where they spread northwards). Since TR is commonplace in Sardinian and Byzantine Greek and given that Arabic linguistic influence on Sardinian was most probably rather marginal (Wagner 1957), other scenarios must be invoked (e.g. independent parallel development, etc.). Thus, the diffusion of TR via language contact is a possibility, but there is as yet no way of telling donor and recipient languages apart.

The above riddle is difficult to solve as the functional domains of TR in the languages under scrutiny are largely identical – as is the text frequency of TR which is high for all of the languages. To demonstrate that TR in language A has been influenced by language B, it is helpful to find functions which are bona fide instances of loans or calques. One such case is discussed in Horcajada Diezma & Sánchez-Prieto Borja (1999). The authors look at distributive reduplication of numerals in medieval Spanish texts. The evidence they present is impressive: constructions such as *tres tres bestias e non más* 'three animals each and nothing more' abound. Not only are constructions of this kind unheard of in modern Spanish, but distributive numerals based on TR of this kind are absent from the western Mediterranean outside North Africa (Stolz 2004: 40). They occur in the eastern Mediterranean and along the North African coast. Horcajada Diezma & Sánchez-Prieto Borja (1999) conclude that the ephemeral existence of distributive

reduplication in Spanish depended on direct translation from Semitic sources. This practice failed to catch on because TR in general is marginal in Spanish and is always considered expressive-emotional – a connotation not usually associated with distributive constructions (Horcajada Diezma & Sánchez-Prieto Borja 1999:298–299). The authors show that the short-lived existence of distributive reduplication in medieval Spanish has an identifiable source, namely texts originally written in Arabic which had to be translated into Spanish. The pattern is an Arabism – the construction is richly attested in practically all Neo-Arabic varieties from Morocco to Syria (Mörth 1997).

Similarly, Burkhart (1985:135) hypothesizes that TR in the languages of the Balkans spread from Turkish during the long domination of southeast Europe by the Ottomans. She assumes that TR is a characteristic trait of Turkish which was copied by bilingual speakers of non-Turkic languages to become (at times only mildly) productive in (varieties of) Albanian, Bulgarian, Greek, Rumanian and Serbian. However, Burkhart (1985:135) also concedes that the Turkish patterns were not completely new to the speakers of recipient languages which already had their own, though restricted, inventory of reduplicative patterns. The evidence from medieval Spanish and the possibility of a Turkish-derived Balkanism suggest that we are dealing with two Orientalisms which in the course of time diffused into the European part of the Mediterranean.<sup>5</sup> Interestingly, the potential Turkish origin of TR in Armenian, Georgian and other languages of the Caucasian region has been rejected as Old Armenian (Godel 1945; Leroy 1986) and Old Georgian (Fähnrich 1994) already made use of this device several centuries before language contact with Turkish was established. Whether Turkish patterns had a reinforcing effect on the autochthonous ones cannot be determined with certainty.

A glimpse at Afrikaans concludes this section. As a member of the Germanic phylum we expect the language to resemble its closest relative Dutch in the pronounced aversion to TR. However, Afrikaans makes regular and frequent use of TR (Botha 1988). Scholars of Afrikaans concur as to the contact-induced origin of the phenomenon. TR of this kind and with this type/token frequency cannot be understood as an inheritance from 17th century Dutch (den Besten, Luycks & Roberge 2003:279), cf. *gaap-gaap* {yawn}-{yawn} '(while) yawning'. There are competing hypotheses about which language served as the donor of this construction. Den Besten, Luycks & Roberge (2003:285) conclude that Indo-Portuguese and Malay are the main sources of this phenomenon – with additional ingredients stemming from Khoekhoe. Thus, there is no single source, but a network of various adstratal (and substratal) sources. Independent of the exact determination of

---

5. Sephardic (Ladino) employs TR liberally and thus diverges from Iberian Spanish. Hetzer (2001) considers Sephardic reduplication a Balkanism.

the source language, one thing is clear: Afrikaans acquired this feature via contact with non-European speech-communities. Thus, genetic affiliation is overruled by language-contact factors. Afrikaans is isolated from the bulk of the Germanic phylum in Europe.

#### 4. Echo-word formation

In the subsections to follow I discuss three types of TRCV in all of which the first segment of the copy is not identical with the first segment of the original. They also share the pragmatic interpretation of deprecation and related connotations. Whether the segmental changes on the copy can be considered sub-morphemic regularities or even count as full-blown morphological devices is an issue I reserve for a separate publication.

##### 4.1 Yiddish-Shmiddish

In his classic monograph on languages in contact, Weinreich (1953:34) mentioned an instance of borrowing of TRCV: “A favourite Yiddish morphological device for the expression of disparagement, consisting in the repetition of a word with substitution of /šm-/ for the initial consonant [...], has been applied by Yiddish speakers to other languages”. Formations like Yiddish *libe-šmibe* (= love-X [X indicates that the copy is a non-word]) ‘love and the like’ contain a copy which obligatorily has an initial syllable head consisting of the binary segmental sequence /šm/. This sequence is added to vowel-initial stems and may also delete all surplus consonants of the syllable head of the original (Weinreich’s definition was not precise enough). According to Weinreich’s original wording, this feature has been transferred to the foreign languages forming part of the repertoire of Yiddish speakers. The wide-spread use of the pattern made especially in American English is indicative of a dissociation of this kind of echo-word formation from Yiddish bilinguals. Currently, English utterances as e.g. *fat-shmat*, *as long as she’s happy* are common in American English and may be produced actively by native speakers of American English without background in Yiddish. Much the same can be said for Russian where *blinčki-šminčki* (= pancake.PL-X.PL) ‘pancakes and such stuff’ is equally wide-spread among native speakers of Russian who do not know any Yiddish (Plähn 1987). In American English and (colloquial) Russian, *shm*-reduplication has become mildly productive.

In these cases it is relatively easy to determine the donor language and the way that the phenomenon was transferred to the recipient. Bilingual speakers of

Yiddish and English, Yiddish and Russian, respectively tended to employ the Yiddish pattern in their second language – presumably also when conversing with individuals who had no competence in Yiddish. The speakers of the adstrate languages considered the pattern attractive enough to copy it wholesale into English and Russian.

## 4.2 Turki-Murki

In his survey of Turkic language contacts, Johanson (2002: 81) mentions in passing that the so-called *reduplication with m* (“m’li ikilime” = “mühleme”) has been copied from Turkish by many languages, especially in the Balkans. The *mühleme* is an instance of TRCV whose basic pattern requires the copy to start with the bilabial nasal /m/. As the description in Müller (2004: 62–70) suggests there are no cases of m-initial words being subject to *mühleme*. Lewis (1967: 21) states that in this case a completely different construction is used in which the noun *falan/filân* ‘the lot’ fulfils the same function as the copy in *mühleme*-constructions. Example (4) illustrates the Turkish pattern.

- (4) Turkish (Lewis 1967: 237)  
*dergi mergi oku-mu-yor*  
 journal X read-NEG-HAB  
 ‘He does not read magazines or stuff like that.’

*Dergi* ‘journal’ is a normal Turkish lexeme whereas its echo-form *mergi* is not. *Mergi* has no separate existence. It has no meaning of its own. Only in combination with *dergi* is *mergi* an acceptable part of a Turkish utterance. The segmental chains of the original and copy are identical with one exception: their initial consonants differ and the leftmost segment of the copy is the bilabial nasal /m/. This pattern is fully productive in Turkish and can thus be applied to practically every Turkish noun unless the noun itself is m-initial.

A cursory look at languages in the neighborhood of Turkish (including former possessions of the Ottoman empire) reveals that patterns which closely resemble *mühleme* can be found to the south, east, north and west, in languages for which relatively close contacts with Turkish are almost self-evident or historically well documented. These languages belong to different phyla and macrophyla. Moreover, they are also typologically quite different from each other, cf. (5)–(8).

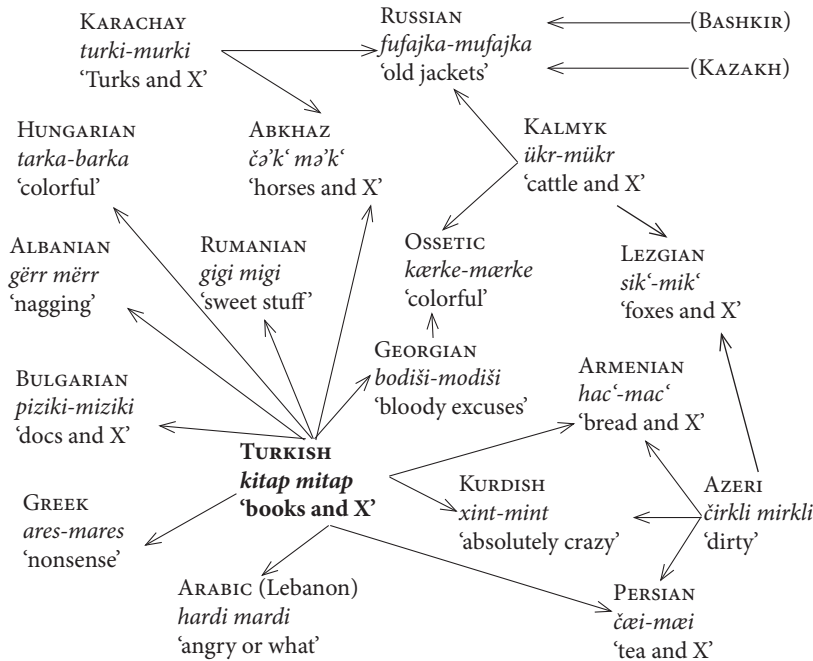
- (5) Lebanese Arabic (Jiha 1964: 163)  
*layš ‘āmil ardi mardi*  
 why make angry X  
 ‘Why do you pretend to be that angry?’

- (6) Kurdish [Zazaki] (Paul 1998:55)  
*ti biyē xint=mint*  
 you be.CONJ.2SG crazy=X  
 ‘Are you crazy or what?’
- (7) Georgian (Tamar Khizanishvili p.c.)  
*šeni bodiṣ-i=modiṣ-i ar m-č’irdeb-a*  
 your excuse-NOM=X-NOM NEG O.1SG-need-3SG  
 ‘I don’t need your empty excuses.’
- (8) Aromunian (Capidan 1932:524)  
*Astădzî mi duṣ s-acumpăr*  
 today O.1SG go.PART SUB-buy.1SG  
*zahăre-mahăre sare-mare carne-marne fărină-mărină culistră-mulistră*  
 sugar-X salt-X meat-X flour-X distaff-X  
 ‘Today I went to buy sugar and similar stuff, salt and similar stuff, meat and similar stuff, flour and similar stuff, distaffs and similar stuff.’

In addition to the East Armenian formations of the type *hac’-mac* ‘bread and similar stuff’, *avel-mavel* ‘brooms and similar stuff’ (Minassian 1980:63), there is also robust evidence for the existence of similar patterns in non-standard varieties of Bulgarian (Grannes 1978) and Modern Greek (Lukas Pietsch p.c.). The pattern is also attested in Persian, Ossetic, Abkhaz, Lezgian, and to a minor extent in Albanian, non-standard Rumanian and non-native Russian (Grannes 1973). Outside of this group of languages, the phenomenon is found in various Turkic languages (Azeri, Karachay, Bashkir, Kazakh) and Kalmyk. There are similar though not completely identical phenomena in Hungarian – and, surprisingly, Basque displays a pattern which comes close to *mühleme* in Turkish.

This geo-linguistic distribution – with the exception of Basque – did not escape the notice of earlier linguists. Müller (2004: 325–329) surveys the extant literature on this topic but is not able to give answers to all questions either. As a matter of fact, there are two schools of thought dealing with *mühleme* in Turkish and neighboring languages: Brinzeu (1945) gives a richly documented account of the phenomenon and identifies an Oriental source from where it spread over adjacent territories. In this scenario, the ultimate source is either Persian or Arabic, whose properties were adopted first by Turkish and then carried over to other languages of the Ottoman empire. *Mühleme* cannot be an inheritance from Common Turkic as the phenomenon is unknown in peripheral Turkic languages such as Yakut. Tietze (1963) counters Brinzeu’s too confident hypothesis by way of proving that the pattern was rare in Old Arabic though more frequent in Persian. He believes they are independent parallel developments which are only loosely interconnected, if at all. This conclusion is unsatisfactory because





Map 1. The areal diffusion of a pattern of echo-word formation.

it does not take into account the fact that some of the examples adduced by students of *mühleme* outside Turkish stem from heavily Turkicized varieties, namely Bulgarian in Istanbul, Kurdish in East Anatolia and the non-native Russian of Karachay-speaking women. Furthermore, if *mühleme* cannot be of Arabic origin, then its being attested as a productive pattern in Lebanese Arabic calls for an explanation. Language contact with Turkish is historically the most convincing one. In addition, certain patterns of partial reduplication-*cum*-variation which are typical of Turkish adjectives (Müller 2004) are also attested in Modern Eastern Armenian (Minassian 1980).<sup>6</sup> It remains doubtful as to how these many independent developments can occur in neighboring languages without any historical tie between them. Note however that /m/ is only rarely used as initial consonant of genuine Turkic words in Turkish. This infrequency of initial /m/ has not changed much since Old Turkic (Müller 2004: 195–196). The Turkish-Persian relationship needs to be looked at more closely. Some of the Caucasian cases of TRCV might turn out to be copied from Persian. I try to capture the

6. Turkish has a pattern of partial reduplication which consists of the copy of the body of the initial syllable plus the insertion of an additional coda consonant whose default realization is /p/ (Müller 2004). In Armenian, the same pattern is employed with /s/ and /p'/ as candidates for the additional coda consonant (Minassian 1980: 63).

areal network of contact relations in Map I.<sup>7</sup> The arrows on the map indicate only one possible pathway of language contacts.

Before I move on to another region where TRCV is an areal feature, a glance at two problematic cases in Europe is in order. From various publications (Thun 1963; Bzdęga 1965), we know that, in a variety of languages outside the above area, there are lexicalized cases which look like *mühleme*. However, German *Techtelmechtel* 'love affair' and English *hurly-burly* and the like are occasional formations which do not give rise to whole series of similar formations with a common semantico-functional bond. Moreover, there is no such word as *\*Techtel* in German to which a separate meaning could be assigned and thus one of the basic requirements of *mühleme* and TR in general is not met. Nevertheless, the existence of examples of this kind in languages which have no areal ties to the Turkish zone of influence suggests that it is possible that similar patterns develop in different languages independently. This is surely the case with Basque: nouns and adjectives may be subject to TR either with or without segmental variation. TRCV requires the initial segment of the copy to be the bilabial nasal /m/ – as is the rule under *mühleme*, cf. *zoko-moko* {corner}-{X} 'every corner'. The Basque pattern often has a derogatory connotation as in *gizon handi-mandi-a* 'a/the pompous man' as opposed to *gizon handi-handi-a* 'a/the great/tall man' with *handi* meaning 'big' (Lafitte 1998). It is impossible to connect the Basque evidence directly to the data from Turkish and its neighbors as there is no historical link between the northern shores of the Iberian Peninsula and the former Ottoman Empire. For the time being, we must accept that something strikingly similar has developed independently in two geographically separated regions.

In contrast to the far-off Basque case, Hungarian is a language which has had its share of Turkish influence (perhaps not only) during its more recent history. It is therefore not surprising that Hungarian is reported as also employing TRCV as in *tarka-barka* {colorful}-{X} 'very colourful'. There is no need for Hungarian to have been exposed to Turkish influence. However, the Hungarian patterns of TRCV differ from the ones of *mühleme* in one important aspect, namely that in Hungarian, the bilabial nasal /m/ does not have the monopoly nor does it play the most prominent role as replacement of the initial segment of the copy. The consonants which are typically used in Hungarian share the feature labial (= /p/,

---

7. The data in the map stem from the following sources: Abkhaz (Bruening 1997), Albanian (Grannes 1973), Arabic (Jiha 1964), Armenian (Minassian 1980), Azeri (Müller 2004), Bulgarian (Grannes 1973), Georgian (Tamar Khizanishvili p.c.), Greek (Lukas Pietsch p.c.), Hungarian (Dezső et al. 1961), Kalmyk (Benzing 1985), Karachay (Grannes 1978), Kurdish (Paul 1998), Lezgian (Haspelmath 1993), Ossetic (Abaev 1964), Persian (Mahootian 1997), Rumanian (Grannes 1973), Russian (Nataliya Levkovych p.c.), Turkish (Müller 2004).

/b/, /f/, /v/, /m/; Dezső et al. 1961:431). The majority of these do not form part of the inventory of the major choices in Turkish and other languages which employ *mühleme* or variants thereof.

The difficulty in interpreting the Hungarian evidence notwithstanding, one thing is clear: beyond the northern borders of southeast Europe *mühleme* and similar patterns become by far less frequent and lose their systematicity. What we find are isolated lexicalized cases which outwardly resemble *mühleme* but do not obey the same principles of morphological formation. In-between the Basque country and the Balkans, there is also a huge gap in the geo-linguistic distribution of the phenomenon. Scholars like Tietze (1963) invoke universal principles which allow for TRCV to develop in any language. However, this potential universality cannot explain why the phenomenon is frequent exactly in those languages which have been in close contact with each other, whereas it is, at best, marginal outside this area. With the exception of Basque, I therefore assume that *mühleme* is indeed a pattern which has diffused from a Turkic center to a non-Turkic periphery predominantly within the former Ottoman empire. In addition to direct copies from Anatolian Turkic, the patterns may have additionally been diffused from other Turkic languages spoken in the Caucasian and the Eurasian regions. All of the languages under scrutiny have TR without variation and TR with variation. Other Circum-Mediterranean languages only allow for proper TR. The *mühleme*-kind of TR thus qualifies as the marked member of the class of phenomena of TR.

### 4.3 Hindi-Vindi

Much the same holds for the languages of South Asia studied by Abbi (1992). TR is attested from Pakistan down to the Indonesian archipelago. However, Abbi (1992:96–99) identifies an areal focus in North India where languages of different genetic backgrounds are characterized by high frequency rates for TR. TRCV is also very common in this focal area and in the languages spoken nearby. There is evidence for this phenomenon in Indo-Aryan, Mon-Khmer, Tibeto-Burman, Dravidian, Austronesian and Dai. The general principles of TRCV – including semantic and pragmatic connotations – are in line with what is familiar from Turkish *mühleme*. However, the details yield a relatively varied picture. First of all, the segmental variation on the copy involves a variety of options. The copy has an obligatory initial /v/ in Hindi, /š/ in Punjabi, /t/ in Bengali, /gi/ in Tamil, etc. Additionally, the latter case shows that the substitution process may also affect the entire body of the initial syllable, cf. Hindi *nam vam* ‘names and stuff’, *khun vun* ‘blood and stuff’, *tras vras* ‘sorrow and the like’, Tamil *puli gili* ‘tigers and stuff’, *pəRəm giRəm* ‘fruit and stuff’, *uppi gippi* ‘salt and stuff’, Telugu *puvu givu*

‘flowers and stuff’, *pre:ma gi:ma* ‘love and stuff’, *a:ta gi:ta* ‘pretentious behavior’, etc. The situation is as intricate as the one described in the previous subsection: there are considerable formal differences as to the phonological specifications applying to the initial segment(s) of the copy – alongside largely identical rules for the word-classes subjected to TRCV and very similar semantic and pragmatic interpretations of the constructions. Genetically unrelated neighboring languages share the feature of echo-word formation, whereas further outside this region, languages do not participate in the isogloss. Since the qualities of the initial segment(s) of the copy differ from language to language, it is difficult to determine from where the phenomenon was diffused, if at all.

The discussion of the possible diachrony of reduplicative patterns on the Indic subcontinent in Abbi (1992: 159–168) postulates a certain chronology of mutual influences between languages of all affiliations in India. Presumably, Austro-Asiatic languages already had this feature before Indo-Aryan speaking people moved to India. Abbi (1992: 160) specifically assumes that “Old Iranian instances of Reduplicated Structures with vowel suffixation, vowel interchange, vowel interfixation and echo formations appear to be influences of Austro-Asiatic languages perhaps spoken then in that region or of a language of which we have no traces so far”, i.e. Abbi invokes an adstratal or substratal transfer of the feature into Indo-Aryan languages. Admittedly, it remains doubtful whether one should assume an extinct substrate language or assume that, in the distant past, representatives of Austro-Asiatic were spoken on the western fringes of the sub-continent where they subsequently gave way to Indo-Aryan and Dravidian languages. More importantly however, Abbi (1992: 152–155) discusses evidence from modern and ancient Iranian languages drawn from Heston (1980). To her mind, the Iranian data show that TRCV was relatively well established in Middle Persian times, but that these cases do not necessarily constitute old autochthonous formations. Abbi (1992: 153) claims that in modern Iranian languages, reduplication of all kinds is restricted to “examples of the language universals type”. However, she evidently did not identify the *mühleme*-like cases reported for Modern Persian, Kurdish and Ossetic (all members of her sample!). Furthermore, she completely ignores the existence of the Circum-Mediterranean area where not only echo-word formation is attested but proper TR abounds. Moreover, she does not take into account the fact that TR was firmly established already in the oldest Afro-Asiatic languages and also occurs in Sumerian and other languages of the Ancient Orient (Skoda 1982: 30–31; Stolz & Sansò *forthc.*). In light of these facts and the evidence adduced by Rubino (2005) and Fabricius (1998), the South Asian map in Abbi (1992: 165) provides only half of the truth. As to proper TR, South Asia connects neatly to the surrounding areas. Only if TRCV is taken into account does an area

emerge within South Asia which stands out from at least some of the neighboring regions – but might also turn out to intersect with the Circum-Mediterranean one to yield another macro-area, or so it seems.

The lesson we learn from this is that there are still too many empirical gaps and uncertainties to allow for definitive conclusions. For proper TR, we still need not only a thorough investigation of the geo-linguistic distribution of the pattern alone but also a large-scale cross-linguistic comparison of the functional domains and the word-classes which are subject to TR. The areal co-existence of a device does not necessarily require that identical functions are fulfilled by the same construction type in different languages. If function does not follow form in this case, what does that mean for a contact-oriented interpretation? Furthermore, TRCV has been shown to be the more promising candidate for an areal-linguistic study. Yet, irrespective of the relatively rich literature on the topic, the systematic study of the phenomenon is still in its infancy.

## 5. On the ways of borrowing

Given that TRCV is a likely candidate for diffusion in linguistic areas in various parts of the world, one might ask why this pattern is so successful in language-contact situations. Discounting social factors such as prestige of the donor language, what is there structurally and/or cognitively that makes TRCV attractive for borrowing?

First of all, the languages in which it is known that *mühleme*-like constructions and similar patterns have been introduced, are typologically diverse. This diversity applies to various phonological, morphological and syntactic parameters (including alignment). Thus, there seems to be no major structural constraint on the borrowability of TRCV into the system of a given recipient language. Of course, this observation also extends to all possible donor languages. The phenomenon is readily integrated into agglutinating, fusional, introflective, polysynthetic, and mildly isolating languages; it occurs in ergative and accusative languages; it is attested in VO and OV languages, etc. The phenomenon is reported for head-marking, dependent-marking, double-marking languages alike – as well as in inconsistent ones. The recipient languages display syllable structures of varying degrees of complexity ranging from relatively simple to highly complex. Similar to proper TR, TRCV is compatible with each and every language type. This compatibility translates into easy borrowability – but what is it exactly that renders TRCV “easy”?

For the sake of the argument, I assume Turkish to be the ultimate source of *mühleme*-like constructions in other languages in its vicinity. I start the

Trivially, syllables with CV-bodies are common-place in practically all of the languages which adopted Turkish *mühleme*. In contrast to Turkish, however, the majority of these recipient languages have a much richer inventory of syllable structures. Among other things, their syllable heads are not obligatorily simple. The bulk of the evidence consists of TRCV involving (C)V-bodies. For some of the recipient languages, there is a constraint which restricts the applicability of *mühleme*-like operations to words with simple or empty word-initial syllable heads (e.g. Georgian, Armenian). Where the pattern is in conflict with syllable-structure constraints, either the application of *mühleme* is blocked or the appropriate syllable structure is created on the copy. The latter is the case in Abkhaz, cf. (9) (Müller 2004: 324–325).

- (9) Abkhaz (Bruening 1997:291)
- (9.1) *čə'-k'* 'horse'                      *čə'-k'-m ə'-k'* 'horses and so on'
- (9.2) *awropla'n-k'* 'aeroplane'      *awropla'n-k' mawropla'n-k'*  
'aeroplanes and so on'

8. Müller (2004) assumes a glottal plosive in head position of putative vowel-initial syllables and thus denies the existence of naked syllables in Turkish.

9. For an isolated violation of this rule, cf. Müller (2004: 52 and 57).

(9.3) *kʷtáR'-k'* 'egg'    *kʷtáR'-k' mʷtáR'-k'* 'eggs and so on'

(9.4) *bná -k'* 'forest'    *bná -k' mʷbná -k'* 'forests and so on'

The four examples illustrate different realisations of *mühleme* in Abkhaz. In (9.1), the process is as expected because the single consonant in the syllable head of the original is replaced by /m/ in the copy. (9.2) does not pose any problems as the prothetic /m/ on the copy occupies the predicted position and thereby creates a head for an erstwhile naked syllable. In (9.3), /m/ replaces the initial labialized velar ejective /kʷ/ – superficially, this substitution operates on a complex syllable head. However, the initial segment of both original and copy is itself part of a weight-less secondary syllable (whose nucleus is a reduced vowel with varying quality). This is in line with general laws of Abkhaz syllable phonology which disfavor complex syllable heads. However, the original in (9.4) already contains a complex syllable head. The addition of /m/ alone would increase the complexity even more and this explains why vowel epenthesis applies. Bruening (1997:293) uses this example to defend the idea that we are dealing with prefixation in lieu of internal modification by way of segmental substitution. This analysis depends heavily on how the complexity of the syllable heads is interpreted. I consider (9.3) a case of simple-head substitution as the labialisation of the initial segment of the original may also be understood as the epenthesis of an u-coloured epenthetic vowel. Only (9.4) contains a full-blown consonantal cluster (an illicit head structure according to Bruening). Since the head is complex, the usual kind of *mühleme*-process has lost its preferred domain and thus an additional syllable has to be created. In Abkhaz, TRCV operates on simple syllabic heads only, and in this it perfectly reflects the Turkish situation – with the difference that Abkhaz allows for syllable structures unknown in Turkish.

However, languages do not always have to follow the Turkish pattern closely. Grannes (1973:38–39) provides a list of examples from Bulgarian which contains several patterns which are ruled out by Turkish syllable-structure constraints. Apart from the fact that Bulgarian and Turkish versions of *mühleme* are blocked for words beginning with /m/, Bulgarian has to handle complex syllable heads relatively often. In (10), I survey those patterns for which Turkish lacks equivalents.

(10) Bulgarian (Grannes 1973:38–39)

- |        |                |           |                         |                            |
|--------|----------------|-----------|-------------------------|----------------------------|
| (10.1) | <i>riza</i>    | 'shirt'   | <i>riza-mriza</i>       | 'any old shirt'            |
| (10.2) | <i>skandal</i> | 'scandal' | <i>skandal-mskandal</i> | 'scandals and such things' |
| (10.3) | <i>skandal</i> | 'scandal' | <i>skandal-mandal</i>   | 'scandal and such things'  |
| (10.4) | <i>snjag</i>   | 'snow'    | <i>snjag-mnjag</i>      | 'snow and stuff'           |
| (10.5) | <i>strach</i>  | 'fear'    | <i>strach-mrach</i>     | 'fear and stuff'           |



In (10.1), the original has a simple syllable head – the rhotic /r/ – and nevertheless this /r/ is not subject to the substitution rule which covers all other instances of syllables with an empty or simple head. The head of the copy is made more complex by the prothetic /m/. No substitution applies. (10.2–3) are two competing realizations of *mühleme*. In (10.2), m-prothesis applies, rendering the already complex syllable head even more complex.<sup>10</sup> Again no substitution applies. However, in (10.3), /m/ replaces the entire syllable head of the original and thus ousts a bi-consonantal cluster. In (10.4), only the onset of the consonantal cluster is replaced by the bilabial nasal whereas the rightmost member of the cluster remains in place. Thus, the complexity of the syllable head is identical for original and copy. This is different in (10.5) because /m/ replaces two members of a tripartite cluster leaving only the rightmost segment unaffected. The common bond between (10.1) and (10.4–5) is the quality of the rightmost consonant: it always belongs to the class of sonorants and is thus relatively close to vowels on the sonority scale. This suggests the following interpretation. The prothetic /m/ deletes all consonants from the syllable head which have a lower sonority than the bilabial nasal itself. Those with equal or higher sonority are kept unless they are the only fillers of the original syllable head. The two possibilities for replacement or preservation of the initial rhotic perhaps mark a transitory stage. This is a rather complex rule that cannot be deduced from the Turkish pattern.

However, one should not jump to the conclusion that what fails to follow identical rules cannot be connected historically via language contact. I assume that *mühleme* diffused from Turkish and then developed differently in the individual recipient languages. In addition to the (C)V-bodies of syllables, the donor language and the recipient languages also share the phonemic status of the default consonant on the copy: /m/ is nothing exotic nor phonologically marked and therefore easy to handle.<sup>11</sup> Step 1 is characterized by the wholesale take-over of ready-made Turkish expressions which already contained *mühleme*. The general transparency of the pattern allowed for an easy application to non-Turkish words (Step 2). In this phase, the autochthonous words to which *mühleme* is extended correspond to the Turkish syllable structure, meaning they have empty or simple syllabic heads. Many languages do not go beyond this stage and thus their version of *mühleme* remains restricted to words whose syllable structure conforms to the one of the donor language. Step 3 shows how some languages (e.g. Abkhaz and

10. According to Grannes (1973:38), this rule is not compulsory as other r-initial words obey the substitution rule: *režiš'ori-mežiš'ori* 'film directors and so on'.

11. The universally unmarked status of /m/ is a leitmotif in Müller (2004). The preference for the bilabial nasal as initial consonant of the copy might be understood as an instance of the emergence of the unmarked (Downing 2005:90–92).



Bulgarian) dissociate the applicability of *mühleme* from the existence of (C)V-bodies and generalize TRCV. The rules which develop language-internally differ. The speakers have the choice to re-interpret the erstwhile Turkish rules in several ways. They may focus on onsets by way of always replacing the leftmost consonantal segment independent of the degree of complexity of the syllable head. They may equally well opt for a rule whose scope is the entire syllable head which is then replaced en bloc. A third possibility takes the phonological qualities of segments as a criterion. A fourth possibility restricts full substitution to simple heads and requires different strategies for complex heads.

What remains constant, irrespective of the individual diachronic developments, is the prominence of the left margin. The substitution rules of Turkish and all putative recipient languages involve the leftmost position of the phonological strings of original and copy. Since the leftmost segment is open to re-interpretation as onset, head or part thereof, it comes as no surprise that the results of language-internal developments differ considerably. The ease with which the Turkish pattern was adopted in language-contact situations becomes less of a riddle if we take into account its openness to re-interpretation: for a start, every language has (C)V-bodies to which the Turkish pattern can be directly applied; from then onwards, the minds of the speakers are free to re-interpret the pattern in order to use it productively with non-Turkish words.

Apart from these phonological aspects, there are other even more important factors qualifying echo-word formations of the above kind for contact-induced transfer. Among these factors, the pragmatic component stands out. The *mühleme*-construction signals the attitude of the speaker towards the referents/events involved in or qualified by the *mühleme*-construction. The construction is evaluative in nature. In the extant literature on the topic, the semantics of echo-word formation is often described as deprecativ and/or list-like. Therefore, the copy is most often translated by circumscriptions such as *and the like*, etc. This list-like property of the construction makes it, at the same time, inherently plural independent of number marking on its constituents. For instance Turkish *kitap kitap* 'books and the like' contains two word-forms unmarked for number. *Mitap* then stands for referents which are similar though not identical to those of *kitap*. In this way, these similar referents are treated in a sweeping fashion as they are not deemed important enough to be identified specifically. They are not worth being mentioned – and this assumed unworthiness is already an instance of deprecation. In combination with the original *kitap*, this deprecativ effect extends over the entire construction because it is implied that it does not make a difference whether we talk about "real" books or pseudo-books – the predication applies

in any case. This negative connotation is reported for practically all instances of *mühleme* in non-Turkic languages and also for TRCV in India. In the majority of cases, deprecation seems to be the main function of these constructions. In language contact, pragmatic functions which reflect the speaker's attitude towards referents and events are cognitively more salient than say, purely grammatical functions such as number marking or the like. Thus, their transfer is relatively easy to achieve.

The pragmatic value of the *mühleme*-construction is accompanied by a high degree of semiotic transparency which also adds to the borrowability of the construction. Among groups of youngsters, adherents to certain segments of the music scene in Berlin, *mühleme* has begun to spread into the native German of group-members with a non-Turkic linguistic background. Phrases like *Hip-Hop Mip-Hop – wer will das denn noch hören?* 'Hip-Hop Mip-Hop – who wants to listen to that anymore?' uttered by ethnic Germans have been overheard repeatedly (Reyhan Şahin p.c.). The young German Hip-Hop addicts have copied the *mühleme*-pattern from their Turkish peers. This transfer is made possible because the pattern instantiates a peculiar kind of iconicity (Mayerthaler 1977; Stolz 2007). Given that proper TR often associates with an increase of a property expressed by the original, TRCV indicates via the failure to copy all segments of the original faithfully that what is expressed by the original is not fully represented by the copy – which thus is literally a bad copy and invites the association with incompleteness, sloppiness, "not the real thing", etc.

Space restrictions do not allow me to look more closely at three additional factors which make *mühleme* especially borrowable. These factors are interconnected: the rhyming relation between original and copy, the prosodic/metrical properties (the *mühleme*-construction as a whole always has a parisyllabic structure) and the ludic character (*mühleme* and TRCV in general have something in common with language games which require the systematic modification of existing phonological strings in order to achieve a certain aesthetic or funny effect). Moreover, *mühleme* is a phenomenon from the fuzzy boundary between compounding, affixation and syntagms. This unclear status is favorable to borrowing as *mühleme* is compatible with a variety of structures into which it has to be integrated.<sup>12</sup>

---

12. On the borrowing scale (Thomason 2001:70–71), the copying of *mühleme*-constructions could be allocated between the stages 1 and 2 (incipient contacts with only small groups of bilingual speakers).

## 6. Conclusions

We observe that proper TR and its variant TRCV are differently suited for a contact-minded interpretation. TR is attested on a global scale. Its presence alone in a given pair of neighboring languages is not unambiguously indicative of either universality or contact-induced transfer. Very often what appears to be an instance of the unexpected presence of TR in language X turns out to be also marginally possible in some relative of X. Since structural constraints do not seem to determine the use or non-use of TR, the phenomenon under scrutiny might be activated in any language. To prove that the presence of TR in language X is the product of language contact, one needs evidence of a different kind. In addition to the simple observation that the pattern is attested in the language, we have to identify the word-classes which are subject to TR. This identification procedure has to go along with the determination of the functions for which TR is employed. Wherever the range of word-classes and the range of functions differ from what is universally expected, it is likely that language contact is an explanation for the observed facts. The same applies to differences in both type and token frequency. TR may very well be marginally attested in all relatives of X. However, only in X does the frequency with which the phenomenon occurs reach statistical significance (and thus converge with the high frequency of TR in the neighboring language Y). Owing to the almost general lack of information about the type and token frequency of the phenomenon, the investigator will have to work with actual texts. Cross-linguistic investigations based on comparative textual analyses is an approach which is only just taking shape in typology and universals research.

Thus, a lot needs to be done before anything definitive can be said about the role of TR in language-contact situations. This is different with echo-word formation. The patterns of TRCV are attested in sizeable areas. However, they are by no means universal. As far as we know, the types Yiddish-Shmiddish, Turki-Murki and Hindi-Vindi are not attested in Sub-Saharan Africa, Oceania, Australia, East Asia, the Americas nor in Pidgins and Creoles. The one exception that is mentioned in the literature (Clements 2003: 198–199) is Korlai, a Portuguese-based Creole spoken in India whose pattern has been borrowed recently via adstratal contacts with Marathi. This is a late contact phenomenon – a kind of Indicisation of Korlai Portuguese – and not a language-internal development of the Creole. Everywhere else, Portuguese-based Creoles lack this feature although their systems are replete with proper TR (de Silva Jayasuriya 2003). As a matter of fact, TR is ubiquitous among Creoles whereas TRCV is only attested exceptionally. Proper TR is potentially universal whereas TRCV is not. Furthermore, TRCV is also

functionally more restricted than proper TR. All these factors taken together suggest that wherever TRCV occurs in, at best, loosely related languages spoken in the same area this geo-linguistic distribution is the result of diffusion via language contact. To verify this hypothesis, future studies have to adduce as much evidence as possible including diachronic evidence. At present, our knowledge of TRCV is still fragmentary. I hope that this study has shown that it is worthwhile studying echo-word formation and related phenomena in a cross-linguistic perspective.

For students of language contact, my discussion of TR and TRCV suggests that one should not shy away from questioning the universal status of a given construction or category. Even where it can be safely assumed that a given phenomenon is universally attested, this does not automatically mean that it is attested with the same frequency or in the same range of contexts/functions. For universals researchers, on the other hand, the above findings show that universal properties may have a variable spell-out from language to language or from area to area. Thus, the concept of language universal must allow for a certain flexibility as to its “local” implementation/realisation. In language contact, languages might influence one another in such a way that universals become more or less frequent. Perhaps, a new vista at universals could take into account the degree to which a universal is realized in a given language.

## Abbreviations

ACC = accusative; ALL = allative; CONJ = conjunctive; DAT = dative; DEF = definite; DET = determiner; ABS = absolutive; FUT = future; HAB = habitual; M = masculine; NEG = negation; NOM = nominative; O = object; PART = Participle; PAST = past tense; PL = plural; SG = singular; SUB = subordinator; TR(CV) = total reduplication(-cum-variation).

## Sources

- |         |  |
|---------|--|
| Basque  | Antoine de Saint-Exupéry [translator: Iñaki Sipiri]. 1989. <i>Printze Txikia</i> . Donostia: Txertoa Argitaldaria. |
| Greek   | Antoine de Saint-Exupéry [translator: Thanases Theodorakes]. s.a. <i>O mikros Pringkepas</i> . Athena: Pella.      |
| Italian | Antoine de Saint-Exupéry [translator: Nini Bompiani Bregoli]. 1994. <i>Il Piccolo Principe</i> . Milano: Bompiani. |

## References

- Abaev, Vladimir I. 1964. *A Grammatical Sketch of Ossetic*. Bloomington IN: Indiana University.
- Abbi, Anvita. 1992. *Reduplication in South Asian Languages*. New Dehli: Allied Publishers.
- Benzing, Johannes. 1985. *Kalmückische Grammatik zum Nachschlagen*. Wiesbaden: Harrassowitz.
- Berg, Thomas. 2001. *Linguistic Structure and Change. An Explanation from Language Processing*. Oxford: OUP.
- den Besten, Hans, Luijks, Carla & Roberge, Paul. 2003. Reduplication in Afrikaans. In *Twice as Meaningful. Reduplication in Pidgins, Creoles and other Contact Languages*, S. Kouwenberg (ed.), 271–288. London: Battlebridge.
- Bollée, Annegret. 1978. Reduplikation und Iteration in den romanischen Sprachen. *Archiv für das Studium der neueren Sprachen und Literaturen* 215: 318–336.
- Botha, Rudolf P. 1988. *Form and Meaning in Word-Formation: A Study of Afrikaans Reduplication*. Cambridge: CUP.
- Brînzeu, Félicien. 1945. Un procédé oriental qui exprime l'approximation ou l'emphase: la reduplication à m-initial. *Garp Filolojileri Dergisi* 361: 35–55.
- Bruening, B. 1997. Abkhaz Mabkhaz. *MIT Working Papers in Linguistics* 30: 291–329.
- Burkhardt, Dagmar. 1985. Zu den Doppelwörtern in Balkansprachen. Semantische und pragmatische Aspekte der Reduplikation. *Zeitschrift für Balkanologie* 21: 121–135.
- Bzdęga, Andrzej Z. 1965. *Reduplizierte Wortbildungen im Deutschen*. Poznań: Polskiej Akademii Nauk.
- Calder, George. 1980. *A Gaelic Grammar*. Glasgow Gairm.
- Capidan, Theodor. 1932. *Aromânii. Dialectul aromân. Studiu lingvistic*. Bucureşti: Imprimeria Națională.
- Clements, J. Clancy. 2003. An a-morphous morphological account of the reduplication processes in three Asian-Portuguese Creoles. In *Twice as Meaningful. Reduplication in Pidgins, Creoles and other Contact Languages*, S. Kouwenberg (ed.), 193–202. London: Battlebridge.
- Dezső, Pais et al. 1961. *A mai magyar rendszere. Leíró nyelvtan. I. Kötet: Bevezetés – Hangtan – Szótan*. Budapest: Akadémiai Kiadó.
- Downing, L. J. 2005. The emergence of the marked: Tone in some African reduplicative systems. In *Studies on Reduplication*, B. Hurch (ed.), 89–110. Berlin: Mouton de Gruyter.
- Dressler, Wolfgang U. 1985. *Morphonology. The Dynamics of Derivation*. Ann Arbor MI: Karoma.
- Fabricius, Anne H. 1998. *A Comparative Survey of Reduplication in Australian Languages*. München: LINCOM Europa.
- Fähnrich, Heinz. 1994. *Grammatik der altgeorgischen Sprache*. Hamburg: Buske.
- Gil, David. 2005. From repetition to reduplication in Riau Indonesian. In *Studies on Reduplication*, B. Hurch (ed.), 31–64. Berlin: Mouton de Gruyter.
- Godel, Robert. 1945. Formes et emplois du redoublement en turc et en arménien moderne. *Cahiers Ferdinand de Saussure* 5, 5–16.
- Grannes, Alf. 1973. *Zajčiki-majčiki; turki-murki* – An example of linguistic interference between the Turkic language Karachai (Northern Caucasus) and Russian, in the spoken Russian of Karachai women. *Orbis* 22(2): 526–534.

- Grannes, Alf. 1978. Le redoublement turc a m-initial en bulgare. *Linguistique balkanique* 21 (2): 37–50.
- Haspelmath, Martin. 1993. *A Grammar of Lezgian*. Berlin: Mouton de Gruyter.
- Heston, W. L. 1980. Some areal features: Indian or Indo-Iranian. *International Journal of Dravidian Linguistics* 9(1): 141–157.
- Hetzer, Armin. 2001. *Sephardisch, Judeo-Español, Djudezmo. Einführung in die Umgangssprache der südosteuropäischen Juden*. Wiesbaden: Harrassowitz.
- Horcajada Diezma, Bautista & Sánchez-Prieto Borja, Pedro. 1999. La reduplicación distributiva del numeral y el arabismo morfosintáctico en el romance hispánico medieval. *Zeitschrift für romanische Philologie* 115: 280–299.
- Inkelas, Sharon & Zoll, Sheryl. 2005. *Reduplication. Doubling in Morphology*. Cambridge: CUP.
- Jiha, Michel. 1964. *Der arabische Dialekt von Bišmizzīn. Volkstümliche Texte aus einem libanesischen Dorf mit Grundzügen der Laut- und Formenlehre*. Wiesbaden: Steiner.
- Johanson, Lars. 2002. *Structural Factors in Turkic Language Contacts*. Richmond MD: Curzon.
- Koptjevskaja-Tamm, Maria & Wälchli, Bernhard. 2001. The Circum-Baltic languages: An areal-typological approach. In *Circum-Baltic languages*. Vol. 2: *Grammar and typology*, Ö. Dahl & M. Koptjevskaja-Tamm (eds), 615–750. Amsterdam: John Benjamins.
- Kouwenberg, Silvia (ed.). 2003. *Twice as Meaningful. Reduplication in Pidgins, Creoles and other Contact Languages*. London: Battlebridge.
- Lafitte, Pierre. 1998. *Grammaire basque (Navarro-Labourdin littéraire)*. Donostia: Elkarlanean.
- Leroy, Maurice. 1986. Le redoublement comme procédé de formation nominale en arménien classique. In *La place de l'arménien dans les language indo-européennes*, M. Leroy & F. Mawet (eds), 62–75. Louvain: Peeters.
- Lewis, Geoffrey L. 1967. *Turkish Grammar*. Oxford: Clarendon.
- Lindström, Jan. 1999. *Vackert, vackert! Syntaktisk reduplikation i svenskan*. Helsingfors: Svenska litteratursällskapet i Finland.
- Mahootian, Shahrzad. 1997. *Persian*. London: Routledge.
- Mayerthaler, Willi. 1977. *Studien zur theoretischen und französischen Morphologie. Reduplikation, Echowörter, morphologische Natürlichkeit, Haplogie, Produktivität, Regeltelescopung, paradigmatischer Ausgleich*. Tübingen: Niemeyer.
- Minassian, Martiros. 1980. *Grammaire d'arménien oriental*. Delmar: Caravan.
- Moravcsik, Edith. 1978. Reduplicative constructions. In *Universals of Human Language*. Vol. 3: *Word structure*, J. H. Greenberg et al. (eds), 297–334. Stanford CA: Stanford University Press.
- Mörth, Karlheinz. 1997. *Die Kardinalzahlwörter von eins bis zehn in den neuarabischen Dialekten*. Wien: WUV-Universitätsverlag.
- Müller, Hans-Georg. 2004. *Reduplikationen im Türkischen. Morphonologische Untersuchungen*. Wiesbaden: Harrassowitz.
- Paul, Ludwig. 1998. *Zazaki. Grammatik und Versuch einer Dialektologie*. Wiesbaden: Ludwig Reichert.
- Plähn, Jürgen. 1987. Xujnja-mujnja i tomu podobnoje. *Russian Linguistics* 11: 37–41.
- Rainer, Franz. 1998. La reduplication française du type *filille* d'un point de vue diachronique. In *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza*. Palermo 18–24 settembre 1995. Sezione 1: *Grammatica storica delle lingue romanze*, G. Ruffino (ed.), 279–290. Tübingen: Max Niemeyer.

- Rohlf, Gerhard. 1967. *Grammatica storica della lingua italiana e dei suoi dialetti. Morfologia*. Torino: Einaudi.
- Rubino, Carl. 2005. Reduplication. In *The World Atlas of Language Structures*, M. Haspelmath et al. (eds), 114–117. Oxford: OUP.
- Sgroi, Salvatore C. 1986. *Interferenze fonologiche, morfo-sintattiche e lessicali fra l'arabo e il siciliano*. Palermo: Centro di Studi Filologici e Linguistici Siciliani.
- de Silva Jayasuriya, S. 2003. Reduplication in Indo-Portuguese, Malayo-Portuguese and Sino-Portuguese. In *Twice as Meaningful. Reduplication in Pidgins, Creoles and other Contact Languages*, S. Kouwenberg (ed.), 185–191. London: Battlebridge.
- Skoda, Francoise. 1982. *Le redoublement expressif: un universal linguistique. Analyse du procédé en grec ancien et en d'autres langues*. Paris: SELAF.
- Stolz, Thomas. 2002. Crosscurrents – the Mediterranean region as a potential linguistic area. In *Der Mittelmeerraum – Brücke oder Grenze*, C. Masala (ed.), 52–73. Baden-Baden: Nomos.
- Stolz, Thomas. 2004. A new Mediterraneanism: Word iteration in an areal perspective. A pilot-study. *Mediterranean Language Review* 15: 1–47.
- Stolz, Thomas. 2006. (Wort-)Iteration: (k)eine universelle Konstruktion. In *Konstruktionsgrammatik. Von der Anwendung zur Theorie*, K. Fischer & A. Stefanowitsch (eds), 105–132. Tübingen: Stauffenburg.
- Stolz, Thomas. 2007. Re: duplication. Iconic vs. counter-iconic principles (and their areal correlates). In *Europe and the Mediterranean as Linguistic Areas. Convergencies and Typological Perspectives*, P. Ramat & E. Roma (eds), 317–350. Amsterdam: John Benjamins.
- Stolz, Thomas & Sansò, Andrea. Forthc. The Mediterranean area revisited. Word-iteration as a potential Mediterraneanism. *Orbis*.
- Thomas, Peter W. 1996. *Gramadeg y Gymraeg*. Caerdydd: Gwasg Prifysgol Cymru.
- Thomason, Sarah G. 2001. *Language Contact: An Introduction*. Washington DC: Georgetown University Press.
- Thun, Nils. 1963. *Reduplicative Words in English. A Study of Formations of the Types tick-tick, hurly-burly and shilly-shally*. Lund: Carl Bloms Boktryckeri.
- Tietze, Andreas. 1963. Die Wiederholung mit verändertem Wortbeginn im Türkischen. *Ural-Altaische Jahrbücher* 25: 92–108.
- Viereck, Wolfgang, Viereck, Karin & Ramisch, Heinrich. 2002. *Dtv-Atlas Englische Sprache*. München: Deutscher Taschenbuch Verlag.
- Wagner, Max L. 1957. Die Iteration im Sardischen. In *Syntactica et Stilistica. Festschrift für Ernst Gamillscheg zum 70. Geburtstag*, G. Reichenkorn et al. (eds), 611–624. Tübingen: Max Niemeyer.
- Weinreich, Uriel. 1953. *Languages in Contact. Findings and Problems*. The Hague: Mouton.
- Wierzbicka, Anna. 1986. Italian reduplication: Cross-cultural pragmatics and illocutionary semantics. *Linguistics* 24: 287–315.
- Wiese, Richard. 1990. Über die Interaktion von Morphologie und Phonologie – Reduplikation im Deutschen. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 43(1): 603–624.

### PART III

## Diachrony





# Variability within the French interrogative system

## A diachronic perspective<sup>\*</sup>

Martin Elsig

University of Hamburg, Research Center on Multilingualism

The French interrogative system is a fecund source of variation. Several different variants of forming yes/no- and wh-questions coexist and compete in usage. Diachronically, a replacement has taken place of variants involving subject-verb inversion by variants featuring the non-inverted SVO word order. An empirical evaluation of interrogative tokens from late Middle to early Classical French suggests that the underlying structural change is both a loss of verb movement to C<sup>o</sup> and a subsequent satisfaction of the wh-criterion at the TP-level, and not any longer at the CP-level. As a result of this change, several diastratic, diaphasic and diatopic varieties of French have emerged with each having a particular grammatical system, which are (at least partly) in close contact with each other. In spite of this contact situation, the varieties do not seem to exert mutual influence on one another which would affect and alter the underlying structure of their respective interrogative systems.

**Keywords:** syntax, language change, linguistic variation, Varbrul, French, interrogatives, subject-verb inversion, yes/no-questions, wh-questions, verb movement

---

<sup>\*</sup> This study has been carried out as part of the research project “Multilingualism as cause and effect of language change: Historical syntax of Romance languages”, directed by Jürgen M. Meisel. This project is one of currently 18 funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) within the Collaborative Research Center on Multilingualism, established at the University of Hamburg. I wish to thank Jürgen M. Meisel, Esther Rinke, Matthias Bonnesen, Susanne Rieckborn, Claudia Stöber and the editors Peter Siemund and Noemi Kintana, as well as two anonymous reviewers for their helpful comments. All remaining errors are mine.

## 1. Introduction

This chapter focuses on a diachronic word order change in the French interrogative system. One of the most conspicuous facts is the replacement in the Middle, Renaissance and Classical French periods of constructions featuring subject-verb inversion (VS) by equivalent constructions with subject-verb-object (SVO) word order (cf. Roberts 1993; Vance 1997). I will discuss the question as to how this and the related changes in the variable system of interrogatives may structurally best be accounted for.

Taking into consideration the different variants of *wh*- and *yes/no*-questions which have emerged or vanished during the evolution from Old French to contemporary colloquial and Standard French, it will be argued that these changes suggest three major evolutionary stages: In a first step, Old French exhibited a system in which the *wh*-criterion (Rizzi 1996) was met at the CP-level. The inflected verb and the *wh*-word both moved to the CP. Changes in the usage of interrogative variants between the 15th and the 17th centuries indicate the first structural change whereby the verb no longer moved to C°, but only as far as T°. I will propose that this is also the state of affairs in contemporary Standard French. In a last and final stage, observable in contemporary colloquial French, the *wh*-word only targets SpecTP, so that the CP-level is not involved any more in the formation of interrogative sentences.

The diachronic evolution has led to a situation in which several grammatical options (or variants) coexist within the French language, some as grammaticalized relics of formerly productive structural mechanisms, others as innovative forms. Some of these variants, however, are representative of different stylistic registers or regional varieties of the language. The result is a heterogeneous system, composed of a diversity of diachronic and synchronic subsystems. The allocation of certain question variants to specific varieties of French suggests that they are indicative of distinctive underlying structural interrogative systems which in spite of being in close contact with each other do not show any contact induced alterations.

## 2. Interrogative variants

The French language offers a multitude of interrogative variants some of which are nowadays extinct and others prolific in usage. The variants may be distinguished according to whether they occur in *wh*- or *yes/no*-questions and according to whether the subject of the interrogative sentence is nominal or pronominal. This

Table 1. Overview of the interrogative variants in French.

	Yes/no		Wh	
	Pronominal subject	Nominal subject	Pronominal subject	Nominal subject
VS	Pronominal inversion	Simple Inversion (†)	Pronominal inversion	Simple Inversion (†)
		Free Inversion (†)		Free Inversion (†)
				Stylistic inversion
S <sup>NP</sup> VS <sup>pron</sup>	Complex inversion ( <i>indistinguishable from clitic left dislocation</i> )		Complex inversion	
SV	Interrogative morpheme <i>-ti/-tu</i>			
	Interrogative marker <i>est-ce que</i>			
	Intonation questions		Wh- <i>in situ</i>	
			Wh-fronting	

(Grey background color: The variant is inexistent in this context. †: This variant is extinct from contemporary varieties of French)

distinction is useful, since not all variants may occur in each and every of these four contexts. Table 1 provides an overview of the variants referred to throughout this chapter.

Simple inversion:

- (1) a. N=*est pas ceste robe assez longue?*  
 NEG=IS NEG this dress enough long  
 ‘Isn’t this dress long enough?’  
 (Cnn.531.51)
- b. *Comment seroit la sagesse presente restaurée?*  
 how would the wisdom present(ADJ) be.restored  
 ‘How would the present wisdom be restored?’  
 (Rab.266)

Free inversion:

- (2) a. [...] *en luy est renouvelée l=antique palintocie*  
 in him is renewed the=ancient compensation  
*des Megariens, et la palingenesie de Democritus?*  
 of.the Megariens and the palingenesis of Democritus  
*Erreur!*  
 Error

'In him is renewed the ancient compensation of the Megariens and the palingenesis of Democritus? No!'

(Rab.416)

- b. *D=où peut donc provenir ce bizarre transport?*  
from=where can thus come this strange humor  
'From where might such a strange humor thus come?'  
(Mol.55.576)

Stylistic inversion:

- (3) *Où pouvait être alors la reine des clartés?*  
where could(IPFV) be then the queen of.the clearnesses  
'Where could then be the queen of the clearnesses?'  
(Corn.515.303)

Pronominal inversion:

- (4) a. *Suis je chievre?*  
am I goat  
'Am I goat?'  
(Far.1302)
- b. *Que mangeront ilz ?*  
what will.eat they  
'What will they eat?'  
(Cnn.369.142)

Complex inversion:<sup>1</sup>

- (5) a. *Ce divertissement n=aura=t=il point de fin ?*  
this diversion NEG=will.have=/t/=it NEG of end  
'Will this diversion never come to an end?'  
(Corn.559.1339)
- b. *Oui; mais pourquoi chacun n=en*  
yes but why everyone NEG=therefrom  
*fait=il pas de même [...]*  
does=he NEG of same  
'Yes, but why doesn't everyone do it the same way?'  
(Mol.181.440)

Interrogative morpheme -ti/-tu

- (6) *Je devons=t=il là nous frotter?*  
I have=/t/=Q there ourselves rub  
'Do I have to rub myself there?'

---

1. Unlike in wh-questions, Complex inversion in yes/no-questions cannot be distinguished from left dislocation of the DP-subject into a topicalized position which is resumed by a post-verbal subject clitic (cf. Roberts 1993: 167).

(Jouin, Nicolas (1764). *Le vrai recueil des sarcelles, mémoires, notes et anecdotes intéressantes sur la conduite de l'Archevêque de Paris & de quelques autres Prélats François: le philotanus, et le porte-feuille du diable; ouvrage absolument nécessaire à ceux qui veulent prendre une juste idée des maux que l'Eglise a soufferts pendant le règne de la ci-devant soi-disant Société de Jesus*, A Amsterdam: Aux dépens de la Compagnie, page 287)

Interrogative marker *est-ce que*

- (7) a. *Est-ce qu=on n=en meurt point ?*  
 Q=one NEG=therefrom dies NEG  
 'Doesn't one die therefrom?'  
 (Mol.107.496)
- b. *comment esse que l=en t=appelle?*  
 how Q PRT=one you(ACC.2SG)=call  
 'How does one call you?'  
 (Far.1137)

Intonation question

- (8) *Donc si ton espérance à la fin n=est déçue,*  
 so if your hope at the end NEG=is deceived  
*Ces deux amours auront une pareille issue?*  
 these two loves will.have a similar way.out  
 'So, if your hope is finally not deceived, will these two loves have a similar way out?'  
 (Corn.42.799)

Wh-*in situ*

- (9) *Pierre a parlé à qui ?*  
 Pierre has spoken to whom  
 'To whom has Pierre spoken?'  
 (Zubizarreta 2003, example (2))

Wh-fronting

- (10) *Pourquoy Nabuzardan, maistre cuisinier du roy*  
 why Nabuzardan master cook of.the king  
*Nabugodonosor, feut entre tous aultres capitaines*  
 Nabugodonosor was among all other captains  
*esleu pour assieger et ruiner Hierusalem?*  
 chosen to besiege and ruin Jerusalem  
 'Why was Nabuzardan, master cook of the king Nabugodonosor, among all other captains chosen to besiege and ruin Jerusalem?'  
 (Rab.165)

The variants featuring a postverbal subject are displayed in the upper part (examples (1) through (5)). The existence of Pronominal inversion dates back to the Old French period (cf. Roberts 1993:117–123). In contemporary French, it is generally assumed to show a significant decrease from colloquial varieties and to be increasingly restricted to a formal style (cf. Coveney 2002:190).<sup>2</sup> It displays a superficial similarity with Simple inversion: The two variants exhibit a subject immediately to the right of the inflected verb (and hence to the left of participles and other nonfinite verbs). Contrary to Pronominal inversion, Simple inversion vanished from the system a long time ago (in the 16th century, according to Roberts 1993:190). Complex inversion manifests a postverbal subject pronoun just like Pronominal inversion. In addition, it contains a coreferential preverbal subject noun. Roberts (1993:169) asserts that this variant was introduced into the French interrogative system by the 1450s. The use of the postverbal interrogative particle *-ti* (or *-tu* in Québec) is closely related to Complex inversion: There is a consensus in the literature that this particle is the outcome of a grammaticalization of the postverbal third person subject pronoun into an interrogative morpheme on the finite verb which is exempt from all referential features (cf. Picard 1992; Roberts 1993:220–224). Therefore, the use of *-ti/-tu* is not restricted any more to preverbal third person subject nouns (as in Complex inversion), but it also extends to all other persons and hence to subject pronouns as well (cf. (6)). According to Foulet (1921), *-ti* emerged towards the end of the 16th century.

Both Free (cf. (2))<sup>3</sup> and Stylistic inversion (cf. (3)) feature a postverbal subject-DP which follows nonfinite verbs. According to Roberts (1993:219), Stylistic inversion emerged in the early 17th century shortly after Free inversion had disappeared. Contrary to the latter, it is excluded from yes/no-questions and incompatible with lexical direct objects (cf. Kayne 1972; Kayne & Pollock 1978, 2001).

The lower five rows in Table 1 show variants with a preverbal subject. *Est-ce que* is a grammaticalized interrogative marker, similar to *-ti/-tu*, but situated in a clause-initial position in yes/no-questions, and in a position immediately

---

2. I use the term ‘Standard French’ in order to indicate the variety of French which is promoted by normative pressure (e.g. the educational system) and which is mainly used in written discourse and formal speech. I assume that Standard French is a variety which may productively be used and which features different grammatical properties than today’s vernacular French (cf. also de Wind 1995:144).

3. One of the reviewers pointed out that the term ‘Free inversion’ is infelicitous in so far as it implies that the inverted occurrence of the subject is unconstrained, contrary to fact (cf. Belletti 2004). I use this term, following Roberts (1993:24), only as a means to distinguish it from the other interrogative variants featuring postverbal subjects, but not in order to qualify the constraints possibly operating on this variant.

	Old Fr.	Middle Fr.						Classical Fr.			Modern Fr.
	1350	1400	1450	1500	1550	1600	1650	1700	1750		
Simple Inv.											
Free Inv.											
Stylistic Inv.											
Complex Inv.											
Pronominal Inv.											
- <i>ti/-tu</i>											
<i>Est-ce que</i>											
Intonation		(rhetorical questions)									
Wh- <i>in situ</i>				?							
Wh- <i>fronting</i>				?							

Figure 1. Diachronic overview of the interrogative variants in French.

following the fronted *wh*-constituent in *wh*-questions respectively. The appearance of *est-ce que* dates back to the 16th century, according to Foulet (1921:264–268) and Marchello-Nizia (1997:217), among others.

Intonation questions have always been an option in French. However, it is well known that their use has progressively gained ground during the diachronic development, yielding majority status in contemporary vernacular French (cf. Pohl 1965; Terry 1970; Söll 1971; Behnstedt 1973; among others). *Wh*-fronting and *wh-in situ* resemble intonation questions in so far as they exhibit the canonical subject-verb-object word order without making use of any interrogative particle (such as *est-ce que* or *-ti/-tu*). In *wh-in situ*, the interrogative word remains in its base-generated position, whereas in *wh*-fronting, it targets a clause-initial position without triggering the inversion of the subject.<sup>4</sup>

Figure 1 summarizes the diachronic evolution of the French interrogative variants, as proposed by the literature.

The statements in the literature about the chronological steps of the French interrogative system (which are visualized in Figure 1) suggest that at some point during the 15th or 16th century, some structural change caused the replacement of VS-variants by SV-variants. Instead of disappearing, most of the VS-variants have survived by passing into a rather formal style of (oral or written) discourse. The contemporary vernacular is characterized by its nearly complete restriction to interrogative SV-variants. In the following sections, I will propose that the diachronic variability may best be accounted for by assuming that the finite verb moved to  $C^0$  in the more ancient varieties, while it moves to  $T^0$  in the more recent

4. I use the term ‘*wh*-fronting’ solely to refer to the uninverted *wh*-SVO word order (cf. Zuckerman 2001:94).



ones.<sup>5</sup> The diaphasic variability between formal and casual styles indicates that Standard French and colloquial French make use of two different grammatical options: While both of them feature the finite verb in T<sup>0</sup>, the former realizes the wh-word in SpecCP and the latter in SpecTP. The diaphasic split turns out to be a direct effect of the diachronic evolution and is related to the more conservative character of Standard French. Before further developing this line of reasoning, I will present the empirical data.

### 3. Empirical evaluation of the data

Six sources of near-colloquial literature and plays from the 15th to the 17th century have been evaluated. These are displayed in Table 2 below.

These texts cover a time period from the Middle French to the early Classical French period. All matrix (yes/no- and wh-) interrogative sentences have been extracted from these sources. Table 3 shows the distribution of interrogative variants across the literary sources.

#### 3.1 Yes/no-questions

Turning our attention first to yes/no-questions (Table 3, upper part), the table shows that there is a division between the three texts from the 15th and 16th century, on the one hand, and the three texts from the 17th century, on the other hand: Pronominal inversion decreases in usage from 84% in the former to 64% in the latter (these and the following percentages being a mean value of the ones provided in Table 3). At the same time, the possibility of inverting a subject-DP completely disappears (except for one token in Molière). Complex inversion only features a small increase from 5% up to slightly less than 9%. *Est-ce que* made its appearance in the 17th century (abstracting away from one token in Rabelais). Most importantly, however, the use of intonation questions rises from less than 9% to more than 25%. In other words, at some point during the 16th century, postverbal DP-subjects vanish from yes/no-questions,

---

5. The account which will be outlined in the following sections of this chapter makes use of the traditional approach which presupposes three structural layers, the complementizer phrase (CP), the tense phrase (TP) and the verbal phrase (VP). This shall not be understood as a stand against a more articulated structure. As a reviewer pointed out, it might absolutely be possible to apply a split-CP system to the phenomena studied here (cf. Rizzi 1997). I refrained from doing so, since the empirical evaluation of the data and the distribution of variants do not necessarily force the recourse to several projections in the left periphery.

Table 2. Primary literary sources.

Author	Year	Full reference
(Cnn) Anonymous	1456–1461	<i>Les cent nouvelles nouvelles</i> , édition critique par Franklin P. Sweetser (1966), Série: Textes littéraires français, Geneva: Librairie Droz.
(Far.) Anonymous	1465	<i>Farce de maître Pierre Pathelin</i> (1489), Paris: Pierre Levet, Bibliothèque Nationale de France.
(Rab.) François Rabelais (*1493, Chinon/Indre-et-Loire – †1553, Paris)	1532–1552	<i>Œuvres de Rabelais</i> . Seule édition conforme aux derniers textes, revue par l'auteur avec les variantes de toutes les éditions originales, des notes et un glossaire, 2 tomes (1858). Paris: P. Jannet ( <i>Gargantua et Pantagruel</i> )
(Corn.) Pierre Corneille (*1606, Rouen – †1684, Paris)	1629–1637	<i>Théâtre complet de Corneille</i> . Texte établi sur l'édition de 1682, avec les principales variantes, une introduction, des notices, des notes et un glossaire, par Maurice Rat, tome premier (1960), Paris: Éditions Garnier Frères. ( <i>Mélite</i> , 1629; <i>La galerie du palais</i> , 1634; <i>L'illusion comique</i> , 1635; <i>La place royale</i> , 1637)
(Agr.C) Anonymous	1649–1651	<i>Agréables conférences de deux paysans de Saint-Ouen et de Montmorency sur les affaires du temps</i> (1649–1651), édition critique par Frédéric Deloffre (1999), Geneva: Slatkine Reprints.
(Mol.) Molière (*1622, Paris – †1673, Paris)	1645/50– 1663	<i>Molière. Œuvres complètes</i> (1962), Paris: Éditions du Seuil. ( <i>La jalousie du barbouillé</i> , 1645/50; <i>Le médecin volant</i> , 1655; <i>L'étourdi ou les contretemps</i> , 1655; <i>Dépit amoureux</i> , 1656; <i>Les précieuses ridicules</i> , 1659; <i>Sganarelle ou le cocu imaginaire</i> , 1660; <i>Dom Garcie de Navarre ou le prince jaloux</i> , 1661; <i>L'école des maris</i> , 1661; <i>Les fâcheux</i> , 1661; <i>L'école des femmes</i> , 1662; <i>La critique de l'école des femmes</i> , 1663)

*est-ce que* emerges instead, and intonation questions compensate for the decrease of Pronominal inversion (as does Complex inversion, but to a much lesser extent).

In order to verify whether these observations are accompanied by revelatory conditioning patterns of the contexts of usage of these variants, a variable rule analysis has been carried out with the most important variants of yes/no-questions (i.e. Pronominal inversion, intonation questions and Complex inversion).<sup>6</sup>

6. The software used for this purpose was GoldVarb X (cf. Sankoff 1988; Rand & Sankoff 1990; Sankoff et al. 2005).

Table 3. Distribution of interrogative variants among the literary data sources.

yes/no-questions											
data source	Pronominal inversion		Simple, Free, Stylistic inversion		Complex inversion		intonation questions		<i>est-ce que</i>		Σ
	%	N	%	N	%	N	%	N	%	N	
Cnn	87.6	156	2.2	4	4.5	8	5.6	10			178
Far.	87.3	96	0.9	1	2.7	3	9.1	10			110
Rab.	81.9	354	2.3	10	6.0	26	9.5	41	0.2	1	432
Corn.	65.2	234			9.2	33	25.6	92			359
Agr.C	60.8	31			9.8	5	25.5	13	3.9	2	51
Mol.	63.9	371	0.2	1	8.3	48	25.6	149	2.1	12	581
Σ	72.6	1242	0.9	16	7.2	123	18.4	315	0.9	15	1711

wh-questions											
data source	Pronominal inversion		Simple, Free, Stylistic inversion		Complex inversion		wh-fronting		<i>est-ce que</i>		Σ
	%	N	%	N	%	N	%	N	%	N	
Cnn	72.2	127	26.1	46			0.6	1	1.1	2	176
Far.	60.3	38	30.2	19			3.2	2	6.3	4	63
Rab.	70.4	243	25.8	89			2.3	8	1.4	5	345
Corn.	75.5	117	21.9	34	1.3	2			1.3	2	155
Agr.C	84.4	38	4.4	2			2.2	1	8.9	4	45
Mol.	78.5	252	15.9	51	2.8	9	0.3	1	2.5	8	321
Σ	73.8	815	21.8	241	1.0	11	1.2	13	2.3	25	1105

(Tokens of wh-questions whose subject is the wh-word are not displayed)

I selected those independent variables which turned out to have a significant influence on French yes/no-questions in Elsig & Poplack (2006).

Table 4 indeed confirms at the level of statistical significance the observation that Pronominal inversion is the favored variant in the 15th and 16th century texts, while intonation questions and Complex inversion prevail in the 17th century. The conditioning patterns further reveal that Pronominal inversion is favored in a number of contexts, where the other variants are significantly less likely to occur: The subject pronouns *je* and *on* favor Pronominal inversion and disfavor intonation questions (cf. (11)); lexical verbs occurring with a high frequency among the interrogative tokens favor Pronominal inversion, but disfavor intonation questions (cf. (12)); finally, cognitive verbs represent a favorable context for Pronominal inversion, but not for Complex inversion (cf. (13)).

**Table 4.** Variable rule analysis of 15th to 17th century yes/no-questions (FW = factor weight).<sup>7</sup>

Variants	Pronominal inversion			Intonation questions			Complex inversion		
	FW	%	Total N	FW	%	Total N	FW	%	Total N
Corrected mean, %, N	.75	75.3	1242/1650	.18	18.4	315/1711	.46	46.0	121/263
<b>Subject identity</b>									
<i>je</i>	<b>.70</b>	87.5	184	<b>.38</b>	11.4	184			
<i>on</i>	<b>.69</b>	83.1	65	<b>.30</b>	10.8	65			
<i>ce</i>	<b>.59</b>	83.5	164	<b>.51</b>	15.2	164			
<i>nous</i>	<b>.56</b>	81.0	42	<b>.57</b>	19.0	42			
<i>tu, vous</i>	<b>.51</b>	77.7	813	<b>.60</b>	21.5	813			
<i>il(s), elle(s)</i>	<b>.32</b>	58.6	382	<b>.29</b>	8.9	382	100.0		121
NP				<b>.93</b>	73.8	61			
range	38			64					
<b>Verb frequency<sup>8</sup></b>									
very frequent	<b>.54</b>	78.2	1252	<b>.46</b>	15.9	1288	[ ]	49.4	174
frequent	<b>.50</b>	74.5	55	<b>.46</b>	16.1	56	[ ]	45.5	11
average frequency	<b>.31</b>	61.8	68	<b>.73</b>	27.1	70	[ ]	23.8	21
rare	<b>.40</b>	68.6	172	<b>.58</b>	25.5	184	[ ]	48.1	27
singleton	<b>.32</b>	60.2	103	<b>.67</b>	31.0	113	[ ]	40.0	30
range	22			21					
<b>cognitive verbs<sup>9</sup></b>									
cognitive verbs	<b>.56</b>	80.4	445	[ ]	18.5	448	<b>.24</b>	20.0	15
other verbs	<b>.48</b>	73.4	1205	[ ]	18.4	1263	<b>.52</b>	47.6	248
range	8						28		
<b>Literary source</b>									
Cnn	<b>.73</b>	90.2	173	<b>.22</b>	5.6	178	<b>.45</b>	44.4	18
Far.	<b>.74</b>	89.7	107	<b>.32</b>	9.1	110	<b>.17</b>	15.8	19
Rab.	<b>.63</b>	84.1	421	<b>.33</b>	9.5	432	<b>.39</b>	33.3	78
Corn.	<b>.41</b>	69.0	339	<b>.63</b>	25.6	359	<b>.58</b>	54.1	61
Agr.C	<b>.34</b>	60.8	51	<b>.61</b>	25.5	51	<b>.64</b>	55.6	9
Mol.	<b>.35</b>	66.4	559	<b>.67</b>	25.6	581	<b>.64</b>	59.0	78
range	40			45			47		

(Not significant: (1) *number of syllables* of the inflected verb, (2) *polarity* of the interrogative sentence)

7. 61 yes/no-questions with a nominal subject have been removed from the variable rule analysis of Pronominal inversion, since this variant requires the subject to be pronominal. All non-third person contexts have been removed from the variable rule analysis of Complex inversion, since this variant requires the subject to be third person.

8. The lexical verb's frequency was determined based on the number of occurrences among the extracted interrogative tokens (once = singleton; 2–4 = rare; 5–10 = average frequency; 11–19 = frequent; ≥ 20 = very frequent).

9. All those verbs which refer to some cognitive state or activity were classified as 'cognitive verbs' (such as *trouver*, *penser*, *vouloir* etc.).

## Pronominal inversion

- (11) a. *Tenez, dit elle, fays je bien*  
hold.on says she do I well  
*ce qu'il vous plaist ?*  
what you(ACC.2PL) pleases  
'Hold on, says she, am I doing well what pleases you?'  
(Cnn.420.62)

## Intonation questions

- b. *Moi, j=aurais de l=amour*  
me I=would.have of the=love  
*pour ta chienne de face ?*  
for your female.dog of face  
'Would I feel love for your nasty face?'  
(Mol.95.1421)

## Pronominal inversion

- (12) a. *Me voulés vous tenir pour beste?*  
me want you hold for silly  
'Do you regard me as silly?'  
(Far.1513)

## Intonation questions

- b. *Quoi ! tu veux te sauver*  
what you want you(REFL) save(INF)  
*à l=autre bord sans moi ?*  
at the=other side without me  
'What! Do you want to run away to the other side without me?'  
(Corn.62.1341)

## Pronominal inversion

- (13) a. *Sçaiz tu pas bien que la fin*  
Know you NEG well that the end  
*du monde approche?*  
of.the world comes.near  
'Don't you know that the end of the world comes near?'  
(Rab.451)

## Complex inversion

- b. *L=amour sait=il pas l=art d=aiguiser les esprits ?*  
The=love knows=it NEG the=art to=stimulate the spirits  
'Doesn't love know the art to stimulate the spirits?'  
(Mol.187.919)

These patterns of conditioning show that Pronominal inversion is characterized by its preferential use in rather specific and delimited linguistic and lexical contexts and that intonation questions and Complex inversion function as generalists. The frequency effect, in particular, illustrates this tendency. It complies with the often made assertion that Pronominal inversion, while gradually disappearing as a productive variant from spoken varieties of French, still survives in a number of lexicalized contexts (cf. for instance Pohl 1965; Behnstedt 1973; Dewaele 1999; Coveney 2002). The favoring influence of frequent verbs could be interpreted as one of the earliest steps of such a lexicalization process.

### 3.2 Wh-questions

The distribution of wh-variants (Table 3, lower part) does not display as clear tendencies of a diachronic evolution as the yes/no-variants did. The proportion of wh-questions featuring Pronominal inversion shows considerable fluctuations between 60% and 85%, depending on the author but not on the time.<sup>10</sup> The same may be said for the inversion of a subject-DP. Wh-fronting and *est-ce que* do not show any signs of emergence or increase. They have been present in the data sources from the earliest one on, however at a very low rate not exceeding 4% taken together. The patterns of wh-variants nonetheless reveal three interesting findings: First, the inversion of a subject-DP is much more frequent than in yes/no-questions (22% vs. 1%). Second, Complex inversion shows emergence only in the 17th century, contrary to yes/no-questions where it has been present from the 15th century on. Third, wh-fronting is no more than a minor variant (1%), contrary to its counterpart in yes/no-questions (i.e. intonation questions which account for 18%).<sup>11</sup>

All in all, the wh-questions extracted from the literary sources show a large (and almost categorical) proportion of subject-verb inversion (96%)<sup>12</sup> without revealing any decrease over time, contrary to yes/no-questions in which intonation questions (and *est-ce que* to a much lesser extent) gradually gain ground. Yet, the emergence of Complex inversion in 17th century wh-questions may be interpreted

10. The observation that Pronominal inversion accounts for 74% and Simple, Free and Stylistic inversion for 22% of *all* wh-questions shall not curtail the near categoricity of inversion. If only those 858 wh-questions are considered whose subject is a pronoun, this pronoun is inverted in 95% of the cases (N = 815). If only those 247 wh-questions are considered whose subject is a DP, this DP is inverted in 97.6% of the cases (N = 241).

11. There were no instances of wh-*in situ* in the 15th to 17th century data.

12. This high proportion would make a variable rule analysis of wh-questions meaningless.

as being symptomatic of a new grammatical interrogative system which provides an additional preverbal subject position for the DP which was absent from earlier French. Besides, in spite of not showing a chronological increase, the use of *est-ce que* in wh-questions may be taken as a further indicator for this new system: *Est-ce que* is quantitatively above the average in those texts which approximate oral speech (*Farce de maître Pathelin, Agréables conférences*).

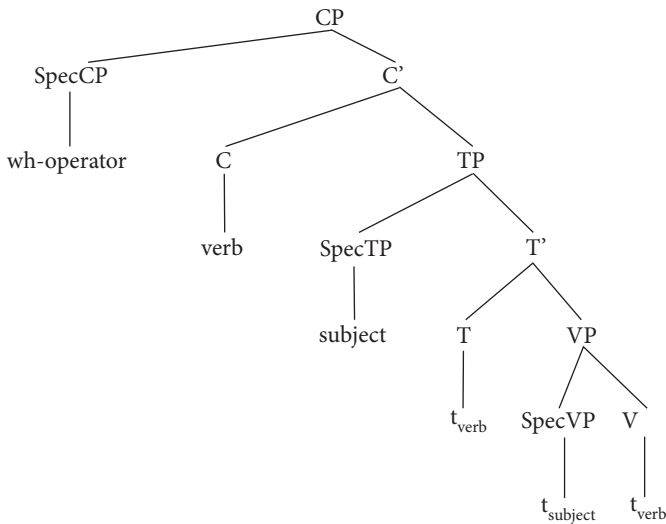
Before turning to the theoretical implications of the results, it should be mentioned that the above evaluation of the data does not distinguish between the different kinds of subject-DP inversion. Simple inversion can only be distinguished from Free and Stylistic inversion, if a nonfinite verb appears in the sentence. In the vast majority of tokens, however, this was not the case. Hence, it was impossible to determine whether the subject-DP appears in a position typical for Free and Stylistic inversion (and thus following the nonfinite verb) or in a position indicating the application of Simple inversion (and thus preceding it).

#### 4. The French interrogative system before the sixteenth century

The empirical evaluation of the data in Section 3 has shown that the inversion of the subject (be it in yes/no- or in wh-questions, with a pronominal or a nominal subject) is the default case in French interrogatives between the 15th and the 17th century. Other variants exhibiting preverbal subjects only occur sporadically. Many authors attribute this near categoricity of inversion to a supposed V2 character of Old French, whereby the inflected verb always moves to C<sup>0</sup> (cf. Adams 1987; Roberts 1993; Vance 1997; among others). The corresponding structure is illustrated in Figure 2.

As is usually the case in V2-languages (such as German), the verb occurs in the second structural position (C<sup>0</sup>), following the first constituent (SpecCP). In wh-questions, the first position is occupied by the wh-word; in yes/no-questions, it is supposedly occupied by the interrogative operator. In the two cases, the subject remains in a postverbal position (SpecTP).

With regard to interrogatives, Kaiser (1996, 2002) asserts that instead of invoking a V2-rule, V<sup>0</sup>-to-C<sup>0</sup> movement may rather be accounted for by means of Rizzi's (1996) wh-criterion whereby an interrogative head (i.e. the inflected verb in C<sup>0</sup>) must be in a specifier-head agreement configuration with an interrogative operator (in SpecCP) and vice versa. In addition, the structure in Figure 2 does not account for the variation between Simple inversion and Stylistic/Free inversion, also challenging the V2-approach. V2-languages do not generally allow for the latter two variants. A possible solution to this problem is provided by the analyses of Alexiadou & Anagnostopoulou (1998) and Barbosa (2001). Old French



**Figure 2.** Structure of the Old French interrogative system according to the V2-hypothesis.<sup>13</sup>

licensed null subjects, just as other contemporary Romance languages do. The authors argue that in Romance null subject languages the argument position of the subject is not in SpecTP, but in its *in situ*-position (SpecVP). This accounts for their postverbal localization in Free inversion. If they occur preverbally, they are either topicalized into a TP-adjoined position or they are A'-moved into SpecTP as contrastively focused elements (Barbosa 2001:39).<sup>14</sup> The syntactic structure associated with such a system is shown in Figure 3.

Since Old French was a null subject language and since it allowed for two structural positions of the subject, contrary to V2-languages, I conclude that the structure in Figure 4 may also be applied to this linguistic variety.

13. Trees were created using phpSyntaxTree (<http://ironcreek.net/phpsyntaxtree/>).

14. Cardinaletti (2004) challenges the view according to which preverbal subjects in null subject languages occur in a topicalized position (contrary to their counterparts in non-null subject languages). The author instead proposes a unified account for preverbal subjects in both types of languages which distinguishes between an upper SubjP hosting strong subjects and a lower AgrSP hosting weak subjects (in the sense of Cardinaletti & Starke 1999), the two projections being located within the IP-field. However, the author agrees with Alexiadou & Anagnostopoulou (1998) and Barbosa (2001) with regard to the possibility in null subject languages of stranding the subject-DP in its base-generated thematic position in SpecVP (Cardinaletti 2004: 117). Since the account proposed in this chapter does not investigate further the exact status of pre- and postverbal subject-DPs (e.g. their information-structural functions) and rather depends on the mere possibility of licensing the subject-DP in two structurally distinct positions, I will continue applying Barbosa's (2001) analysis.



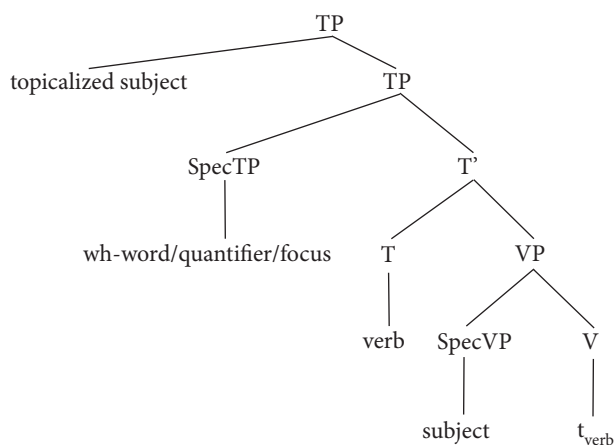


Figure 3. Structure of contemporary Romance null subject languages, according to Alexiadou & Anagnostopoulou (1998) and Barbosa (2001).

Yet, there remains an important difference between contemporary Romance null subject languages and Old French: While in the former the two subject positions discussed above, i.e. the topicalized and the *in situ* position, occur pre- and postverbally respectively (as shown in Figure 3), they occur postverbally in the latter (Simple vs. Free/Stylistic inversion). The absence of Simple inversion from contemporary Romance null subject languages has been taken as evidence against the involvement of the CP-layer in interrogatives (Barbosa 2001). Conversely, one may argue that its presence in Old French indicates that the verb actually moves to  $C^0$ . The structure I propose for the French interrogative system from before the sixteenth century is based on the above analysis of contemporary Romance null subject languages and incorporates verb movement to  $C^0$  (cf. Figure 4).

The structure proposed in Figure 4 makes available two subject positions, the VP-internal *in situ* position (yielding Free/Stylistic inversion) and the topicalized adjunction to TP.<sup>15</sup> The crucial difference between this account and the alternative V2-based proposal in Figure 2 is the status of SpecTP. I argue, following Alexiadou & Anagnostopoulou (1998) and Barbosa (2001), that SpecTP is an  $A'$ -position.

15. Postverbal pronominal subjects already had a clitic status in Old French and hence did never appear *in situ* (cf. Vance 1989: 70ff. and Roberts 1993: 117–123).

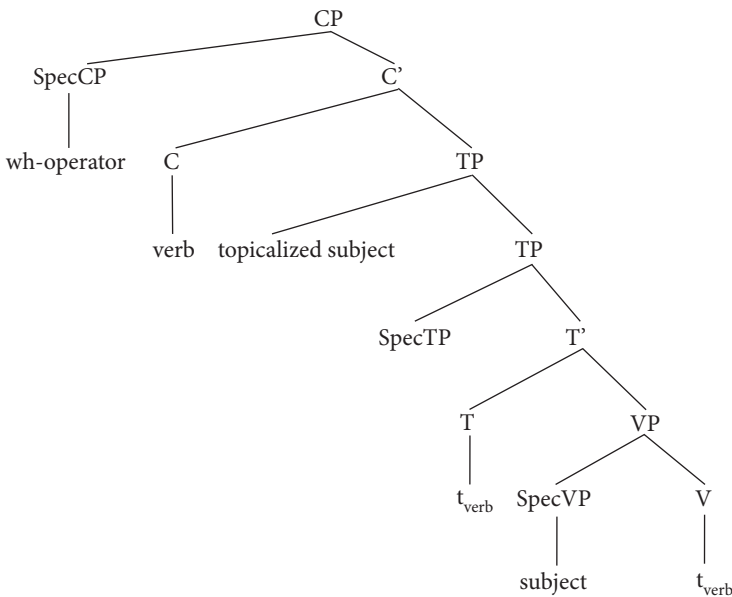


Figure 4. Structure of the French interrogative system until the sixteenth century.

## 5. Structural changes affecting the Middle and Renaissance French interrogative system

The Middle and Renaissance French periods feature the emergence of two interrogative newcomers: Complex inversion and *est-ce que*, as the evaluation of the interrogative data confirms (with the exception that *wh-est-ce que* was present throughout the time periods studied here). At the same time, Simple inversion disappeared (cf. Roberts 1993:189). Pronominal inversion was unaffected by these changes and continued being used.

These changes have caused a controversy in the literature about whether the inflected verb henceforth still moves to  $C^0$  (cf. Rizzi & Roberts 1989; Roberts 1993) or whether it remains in a lower position such as  $T^0$  (e.g. Noonan 1989; Drijkoningen 1990; Hulk 1993; de Wind 1995). I interpret them as a direct consequence of a structural change whereby verb movement to  $C^0$  has been lost. The syntactic structure accounting for the Classical French interrogative system is displayed in Figure 5.

The crucial difference between the two interrogative systems displayed in Figure 4 and Figure 5 are the two structural positions preceding  $T^0$ : In the older system, subjects could move into a TP-adjoined position while SpecTP was an  $A'$ -position analogous to the one in other Romance null subject languages (cf.

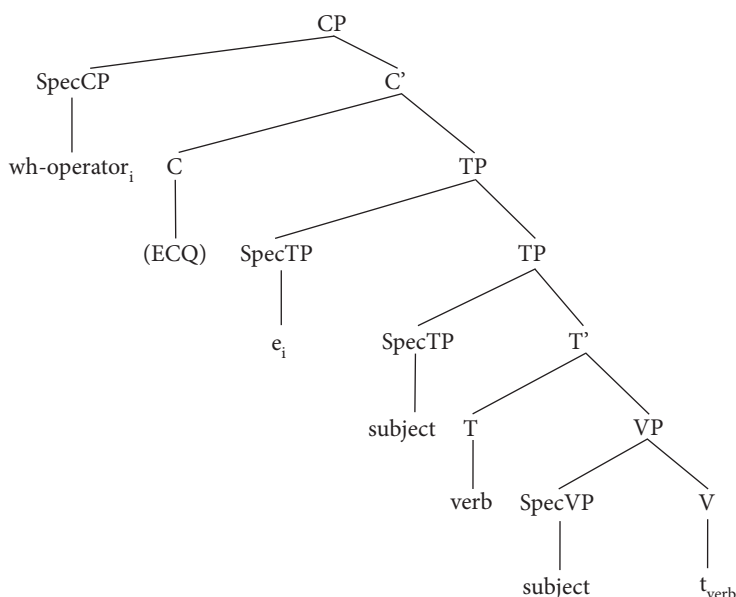


Figure 5. Structure of the French interrogative system after the sixteenth century.

Barbosa 2001). In the more recent interrogative system, the (non-*in situ*) subject has been reanalyzed as occupying the (inner) A'-specifier of TP, while the TP-adjoined position has been reanalyzed as a second (outer) specifier of TP hosting a variable bound by the wh-operator in SpecCP. This is in line with Roberts (1993: 166–177) who cites some empirical evidence that the preverbal subject-DP in Complex inversion is derived (through reanalysis) from a formerly topicalized and left-dislocated subject.

The assumption that the preverbal position of the subject-DP in Complex inversion is not an A-position is strengthened by the occurrence of Free inversion featuring a VP-internal subject. Simple inversion is excluded according to Figure 5, since the verb (or the auxiliary) henceforth never reaches a position from where it would precede the (non-*in situ*) subject-DP.

The possible realization of the interrogative marker *est-ce que* in C° constitutes further evidence that the inflected verb has not moved to this position. The wh-criterion is fulfilled by specifier-head agreement between *est-ce que* in C° and the wh-operator in SpecCP.

The presence of the *est-ce que*-marker has a certain effect on the position of the subject with regard to the inflected verb: Whereas postverbal subject-DPs are possible in combination with this marker (cf. (14)), postverbal subject pronouns are categorically excluded. They necessarily occur in preverbal position. Pronominal and Complex inversion are only possible when the *est-ce que*-marker is

absent. However, even in this case they are not obligatory. Wh-fronting is another possibility (both with nominal and pronominal subjects, cf. (10)).

- (14) *qu=est ce que veut dire ce fol?*  
 what=Q wants to say this fool  
 'What does this fool want to say?'  
 (Rab.202)

I propose that the reason for this pattern may be found in the wh-criterion and in the place where the interrogative feature has its origin. Rizzi (1996) provides evidence that the source of the interrogative feature is the inflectional head, T<sup>0</sup> (or I<sup>0</sup> in his terminology). In fact, the very existence of the postverbal interrogative marker *-ti/-tu* (cf. (6)) indicates that this is indeed the case. In the French interrogative system after the 16th century, i.e. after Complex inversion and *est-ce que* had emerged and Simple inversion had disappeared, the interrogative feature could apparently be checked either at the TP- or at the CP-level. In the first case, a postverbal subject pronoun was needed as a phonological clitic to the verb in T<sup>0</sup> in order to phonetically realize this feature. In the second case, the marker *est-ce que* assumed this role. The mutual incompatibility of simultaneously having the two (inversion of a pronoun and *est-ce que*) is then due to the fact that the interrogative feature may be checked and deleted only once, either by the inverted pronoun or by *est-ce que*. Such an analysis entails however that in Pronominal and Complex inversion, the feature checking relation between the interrogative feature in T<sup>0</sup> and the wh-operator in SpecCP is non-local, which would be a situation prohibited by the wh-criterion. Rizzi (1996) himself introduced the notion of *dynamic agreement* for such cases: The wh-operator in SpecCP may exceptionally endow C<sup>0</sup> with the [wh]-feature. Instead of invoking this mechanism, I rather suggest that in both cases (Complex/Pronominal inversion and *est-ce que*) the wh-criterion is fulfilled at a local level. When the inflectional head hosts a postverbal pronoun, the uninterpretable interrogative feature of this head is checked against the interpretable operator feature provided by the wh-operator in SpecCP and passed down to the variable it binds in the outer SpecTP (which was formerly the adjoined position hosting the subject-DP in Simple inversion). The wh-criterion is hence fulfilled by means of specifier-head agreement between the outer SpecTP and T<sup>0</sup>. In the alternative case (i.e. the insertion of the question marker *est-ce que*), the uninterpretable interrogative feature moves to C<sup>0</sup> and is checked against the operator feature in SpecCP.

The assumption that the inflectional projection licenses two specifiers the inner one of which contains a subject-DP checking its strong [N]-feature against the inflectional head and the outer one of which contains a co-indexed expletive pronoun goes back to Chomsky (1995, Chapter 4). The similarity of this proposal

with French Complex inversion which also features two subjects, a DP and an inverted pronoun (which has been attributed the status of an expletive by Kayne 1983) is obvious. It does not come as a surprise then that the inverted pronoun in Complex inversion is in some way connected to the outer SpecTP (namely by means of the interrogative feature checking relation).

Most notably, the structural change leading from the Old French interrogative system displayed in Figure 4 to the more recent one displayed in Figure 5 had an important impact on the status of the postverbal subject pronoun. Its position in Pronominal inversion was not automatically obtained any more by verb movement to C°. Instead, the pronoun had henceforth to incorporate into T° in order to derive Pronominal inversion. At the surface, Pronominal inversion before and after the change looked the same. This might have prevented its sudden disappearance. Meisel (2001) discusses the possibility that a diachronic structural change (such as the loss of verb movement to C° proposed here) may lead to a reinterpretation of certain word orders during first language acquisition. This could have been the case here as well: The pronoun underwent several steps of grammaticalization. It first changed from a phonological clitic (base-generated as a maximal projection in SpecTP and phonologically cliticizing to the inflected verb) into a syntactic clitic (base-generated in the inflectional head position), cf. Sportiche (1998). Furthermore, it shows signs of turning from an external argument of the verb into a functional agreement marker expressing interrogation. In the literature (e.g. Picard 1992), the postverbal interrogative morpheme *-ti/-tu* has been claimed to be the end result of this development whereby the inverted subject pronoun has gradually lost its referential features and turned into a morphological affix to the verb only present when coinciding with the interrogative feature in T°. The observation that Complex inversion contains two subjects, the preverbal subject-DP and the postverbal pronoun, gives rise to the assumption that the pronoun first and foremost serves as a phonetic expression of the interrogative feature in T°. This has been proposed by Noonan (1989) and Friedemann (1997). In other words, as soon as verb movement did not target C° any more but only T°, and as soon as a preverbal subject position became available, there has been a considerable grammaticalization pressure on the postverbal pronoun progressively depriving it of its referential features and turning it into an interrogative morpheme. Hilbert (this volume) reports a comparable case in Indian and Singapore English interrogatives where the copular verb *be* has formed a unit with the *wh*-word (yielding the unanalyzed *what's* and *how's* and giving rise to the use of a second copular verb in the same sentence).

Another change reportedly occurred in the late 16th or in the early 17th century whereby Free inversion disappeared and Stylistic inversion emerged (cf. Roberts 1993: 219). In Section 4, I suggested that Alexiadou & Anagnostopoulou's

(1998) and Barbosa's (2001) account of Romance null subject languages extends to Old French as well. Free inversion was then equivalent to the kind of free inversion observable in all Romance null subject languages. Rinke & Meisel (2007) relate the possibility to strand the subject-DP in its base-generated *in situ*-position to the property common to all Romance null subject languages (including Old French) to express focus in this position. The loss of Free inversion then is a natural consequence of the loss of this focus position in French. Elsig (2007) reports that in contemporary colloquial French, Stylistic inversion is largely restricted to predicative contexts which may receive a small-clause analysis (following Stowell 1978) and whose subject is hence an internal argument of the copular verb. Under this perspective, contemporary Stylistic inversion is a relic of the former null subject property in being restricted to contexts which lack an external argument (i.e. a referential null subject).

The variable use of interrogative forms in Modern Standard French, i.e. the formal spoken and written variety of French, suggests that the structure illustrated in Figure 5 accounts for this variety as well. It essentially features the same range of variants: VS (i.e. Complex, Pronominal and Stylistic inversion) variants are preferred over SV variants (i.e. intonation questions, *est-ce que*). Simple inversion is ungrammatical (due to the loss of T<sup>0</sup>-to-C<sup>0</sup> verb movement). *-Ti/-tu* has never been accepted in the standard variety. I suppose that normative influences have constituted a counterpressure with regard to the grammaticalization of the post-verbal pronoun into a functional interrogative marker. Pronominal and Complex inversion are therefore still the outcome of a productive means of inverting the pronoun, contrary to Modern vernacular French.

## 6. The interrogative system of contemporary vernacular French

Numerous empirical studies of contemporary varieties of colloquial French have shown that most of the inversion variants have disappeared and been replaced by their non-inverted counterparts (e.g. Terry 1970; Behnstedt 1973; Coveney 2002; Elsig & Poplack 2006): Simple inversion is ungrammatical (not only in spoken French, but also in normative Standard French). Complex inversion and Pronominal inversion are perceived as highly formal and are rarely found in spoken discourse (e.g. Coveney 2002). Only Stylistic inversion still seems to be the default construction in interrogatives featuring a subject-DP, as Bonnesen & Meisel (2005) state. The bulk of interrogative tokens of contemporary spoken French is made up of yes/no-intonation questions, *wh*-fronting and *wh-in situ* (cf. Pohl 1965; Behnstedt 1973; Söll 1983; Coveney 2002) with *est-ce que* and *-ti/-tu* ranging behind as minor contenders in yes/no-contexts (the latter being increasingly

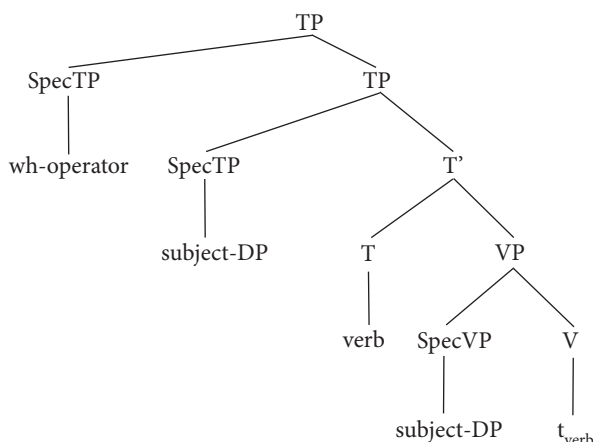


Figure 6. Structure of the interrogative system of contemporary colloquial French.

restricted in usage to rural speech or certain regional varieties such as Québec French, and *est-ce que* being reserved for a formal style, cf. Behnstedt 1973; Fox 1989; Elsig & Poplack 2006).

I consider these observations as an indication that the colloquial varieties of Modern French have undergone a further structural change whereby not only the movement of the inflected verb to  $C^0$  has been lost, but also the movement of the wh-phrase to SpecCP (cf. also Hulk 1993). The structural system I propose for contemporary colloquial French is displayed in Figure 6.

The interrogative system provides two positions for the subject-DP, a first one in the inner SpecTP and a second one in the subject's *in situ* position, SpecVP. The subject-DP may occur in either one of them (yielding wh-fronting, wh-*in situ* and intonation questions when moving to SpecTP and Stylistic inversion when staying *in situ*).<sup>16</sup>

Pronominal subjects are located within the inflectional head,  $T^0$ , due to their status as clitics and inflectional agreement markers hosted by the verb. In the literature, the grammaticalization of pronominal subjects from XPs into agreement markers and ultimately into morphological affixes to the verb has been extensively discussed (cf. Roberge 1986; Kaiser & Meisel 1991; de Wind 1995; Friedemann 1997; Cardinaletti & Starke 1999), one of the main arguments being the often made

16. By assuming that the subject-DP in Stylistic inversion is localized in SpecVP, I follow Déprez (1989, 1990), Valois & Dupuis (1992), de Wind (1995), Friedemann (1997), Barbosa (2001), and Alexiadou & Anagnostopoulou (2001, 2007) among others. This assumption is not uncontested. Kayne & Pollock (2001) argue that the subject-DP in Stylistic inversion has vacated its *in situ* position and has been moved to the left periphery.

observation of subject doubling in colloquial French (cf. Sankoff 1982; Thibault 1983; Kaiser 1994; Nadasdi 1995, 2001; Auger 1998; Nagy et al. 2003).

The main argument advocated here is that in 15th to 17th century French, interrogative variants showing the SV word order have outranged and replaced variants showing subject-verb inversion due to the loss of verb movement to C°. With the verb in T°, postverbal pronouns have progressively turned into interrogative morphemes losing their referential features (as in Complex inversion and *-ti/-tu*). Whereas the interrogative system from the 16th century onwards (up to today's Standard French) allowed for both Complex and Pronominal inversion (and hence the realization of a postverbal interrogative marker) in wh-questions, this option has been lost in contemporary vernacular French which lacks the two variants.

The structure displayed in Figure 6 contains two TP-specifiers, just like the structure proposed for the Classical French interrogative system, the only difference being that the outer SpecTP does not host the variable of the wh-operator any more, but the operator itself.

This structural difference accounts for the observation that Modern vernacular Québec French never features postverbal subject pronouns or interrogative morphemes (*-tu*) in wh-questions (cf. Fox 1989; Auger 1996; Elsig & Poplack 2006; Elsig 2007). This restriction may be explained by stating that the interrogative feature is phonetically realized only once, either by the postverbal interrogative marker (i.e. the inverted pronoun or *-ti/-tu*) or by the preverbal wh-phrase.

The assumption that in contemporary colloquial French the wh-phrase is localized in the outer SpecTP (and not in SpecCP) receives further evidence from *est-ce que*-questions: According to Elsig & Poplack (2006), the use of this interrogative marker in yes/no-questions (cf. (7)a.) correlates with a formal style (indicating the recourse to Standard French). In this context, the use of *est-ce que* implies the involvement of C° (as in Figure 5).

In wh-questions, however, *est-ce que* is much more widespread and does not at all coincide with formality. Phonological assimilation processes (such as in *où ce qui est ...*, cf. Druetta 2002, 2003; Elsig 2007: 141) indicate that *est-ce que* has cliticized to the wh-word in SpecTP forming a structural unit with it.

## 7. Conclusion

In this chapter, I have presented arguments in favor of a diachronic structural change in French interrogatives resulting in two different grammatical systems, namely Standard French and colloquial French.



From the Old French period up to the 16th century, the inflected verb in interrogatives moved to  $C^0$  in order to satisfy the wh-criterion against the wh-operator in SpecCP. The possibility for the subject-DP to remain *in situ* (Free inversion) suggests that this diachronic variety of French was very similar to other contemporary Romance null subject languages (cf. Rinke & Meisel 2007).

The changes in the Middle and Renaissance French period (i.e. the disappearance of Simple inversion and the emergence of Complex inversion and of *est-ce que*) were accounted for by assuming a loss of  $T^0$ -to- $C^0$  movement of the inflected verb. The TP-adjoined position hosting subjects (in Simple inversion) was reinterpreted as an outer specifier of this projection hosting a variable bound by the wh-operator in SpecCP. The wh-criterion was fulfilled either at the TP-level (accounting for Complex and Pronominal inversion) or at the CP-level (accounting for *est-ce que*). The postverbal position of the pronoun was no longer obtained by moving the verb across this pronoun to  $C^0$ , but by incorporation of the pronoun from SpecTP into  $T^0$ . As a consequence, Pronominal inversion was structurally reanalyzed and the pronoun progressively changed its status into an interrogative morpheme. The results in Tables 3 and 4 conform to this structural analysis.

As for the contemporary varieties of French, I suggested an analysis of Standard French along the same lines as the post-sixteenth century interrogative system. Normative influences have prevented the grammaticalization processes observable in today's vernacular French from happening. In the colloquial variety, however, subject-verb inversion has vanished as a productive application of an inversion rule. Postverbal occurrences of subject pronouns have been reanalyzed as interrogative particles (e.g. the grammaticalization of Complex inversion into *-ti/-tu*). The asymmetrical behavior of *est-ce que* with regard to yes/no- and wh-questions suggests that the CP-layer has been lost as the projection hosting the wh-operator.

The diachronic analysis of the French interrogative system proposed here has shown that a single structural language-internal change (namely the loss of verb movement to  $C^0$ ) may result in important alterations of the variable system, even leading to a split into different (diaphasic and diatopic) grammatical systems. In other words, I have argued in favor of an account whereby the variation in the interrogative system of contemporary varieties of French may be related to a parametrical change which occurred in the late Middle French period and which did not depend on a language contact scenario. The subsequent divergence of the (interrogative) system into different diastratic and diaphasic varieties (colloquial vs. Standard French), and the split into distinct diatopic varieties (Québec vs. European French), is the result of an interplay between natural tendencies of language change (such as the grammaticalization of postverbal subject pronouns into interrogative morphemes) and the influence of prescriptive and normative pressures.

The latter ones are responsible for the maintenance of inversion variants (Pronominal, Complex, Stylistic inversion) in the Standard variety which is reflected in written and literary French. The virtual absence of subject-verb inversion from colloquial French shows that although many speakers have access to the two distinct grammars of Standard and colloquial French and may apply them in the respective contexts, the normatively prescribed system remains without influence on the underlying structure of the vernacular. To put it differently, even though the vernacular is in continuous contact with the prescribed Standard, its evolution proceeds independently and in response to universal mechanisms of language change.

## Abbreviations

2PL = second person plural; 2SG = second person singular; ACC = accusative; ADJ = adjective; INF = infinitive; IPFV = imperfective; NEG = negation marker; PRT = particle; Q = question particle/marker; REFL = reflexive.

## References

- Adams, Marianne. 1987. Old French, Null Subjects and Verb Second Phenomena. PhD dissertation, UCLA.
- Alexiadou, Artemis & Anagnostopoulou, Elena. 1998. Parametrizing AGR: Word order, V movement and EPP-checking. *Natural Language and Linguistic Theory* 16(3): 491–539.
- Alexiadou, Artemis & Anagnostopoulou, Elena. 2001. The subject-in-situ generalization and the role of case in driving computations. *Linguistic Inquiry* 32(2): 193–231.
- Alexiadou, Artemis & Anagnostopoulou, Elena. 2007. The subject-in-situ generalization revisited. In *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax and Semantics*, U. Sauerland & H.-M. Gärtner (eds), 31–60. Berlin: Mouton de Gruyter.
- Auger, Julie. 1996. Subject-clitic inversion in Romance: A morphological analysis. In *Aspects of Romance Linguistics. Selected Papers from the Linguistic Symposium on Romance Languages XXIV*, March 10–13, 1994, C. Parodi, C. Quicoli, M. Saltarelli & M. L. Zubizarreta (eds), 23–40. Washington DC: Georgetown University Press.
- Auger, Julie. 1998. Le redoublement des sujets en français informel québécois: Une approche variationiste. *Canadian Journal of Linguistics* 43: 37–63.
- Barbosa, Pilar. 2001. On inversion in wh-questions in Romance. In *Subject Inversion in Romance and the Theory of Universal Grammar*, A.C.J. Hulk & J.-Y. Pollock (eds), 20–59. Oxford: OUP.
- Behnstedt, Peter. 1973. *Viens-tu? Est-ce que tu viens? Tu viens? Formen und Strukturen des direkten Fragesatzes im Französischen*. Tübingen: Narr.

- Belletti, Adriana. 2004. Aspects of the low IP area. In *The Structure of CP and IP*, L. Rizzi (ed.), 16–51. Oxford: OUP.
- Bonnesen, Matthias & Meisel, Jürgen. 2005. Die 'Subjekt-Verb Inversion' in Interrogativkonstruktionen des gesprochenen Französischen: Zum Problem der syntaktischen Variation. In *Deutsche Romanistik – generativ*, G. Kaiser (ed.), 31–48. Tübingen: Narr.
- Cardinaletti, Anna. 2004. Toward a cartography of subject positions. In *The Structure of CP and IP*, L. Rizzi (ed.), 115–165. Oxford: OUP.
- Cardinaletti, Anna & Starke, Michal. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In *Clitics in the Languages of Europe*, H.v. Riemsdijk (ed.), 145–233. Berlin: Mouton de Gruyter.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge MA: The MIT Press.
- Coveney, Aidan. 2002. *Variability in Spoken French. A Sociolinguistic Study of Interrogation and Negation*. Bristol & Portland OR: Elm Bank.
- Déprez, Viviane. 1989. Stylistic inversion and verb movement. In *Proceedings of the Fifth Eastern States Conference on Linguistics*, J. Powers & K.d. Jong (eds), 71–82. The Ohio State University.
- Déprez, Viviane. 1990. Two ways of moving the verb in French. *MIT Working Papers in Linguistics* 13: 47–85.
- Dewaele, Jean-Marc. 1999. Word order variation in French interrogative structures. *ITL Review of Applied Linguistics* 125–126: 161–180.
- Drijkoningen, Franck. 1990. Functional heads and the unification of French word order. *Probus* 2(3): 291–320.
- Druetta, Ruggero. 2002. *Qu'est-ce tu fais?* État d'avancement de la grammaticalisation de *est-ce que*. Première partie. *Linguæ &* 2/2002: 67–88.
- Druetta, Ruggero. 2003. *Qu'est-ce tu fais?* État d'avancement de la grammaticalisation de *est-ce que*: Deuxième partie. *Linguæ &* 1/2003: 21–35.
- Elsig, Martin. 2007. Diachronic Aspects of Syntactic Variation in Québec French Interrogatives. PhD dissertation, University of Hamburg.
- Elsig, Martin & Poplack, Shana. 2006. Transplanted dialects and language change: Question formation in Québec. *University of Pennsylvania Working Papers in Linguistics* 12(2): 77–90.
- Foulet, Lucien. 1921. Comment ont évolué les formes de l'interrogation. *Romania* 47: 243–348.
- Fox, Cynthia A. 1989. Syntactic Variation and Interrogative Structures in Québécois. PhD dissertation, Indiana University.
- Friedemann, Marc-Ariel. 1997. *Sujets syntaxiques. Positions, inversions et pro*. Bern: Lang.
- Hulk, Aafke C.J. 1993. Residual V2 and the licensing of functional features. *Probus* 5: 127–154.
- Kaiser, Georg. 1994. More about INFlection and agreement: The acquisition of clitic pronouns in French. In *Bilingual First Language Acquisition. French and German Grammatical Development*, J. M. Meisel (ed.), 131–159. Amsterdam: John Benjamins.
- Kaiser, Georg A. 1996. V2 or not V2? Subject-verb inversion in Old and Modern French interrogatives. In *Language Change and Generative Grammar*, E. Brandner & G. Ferraresi (eds), 168–190. Opladen: Westdeutscher Verlag.
- Kaiser, Georg A. 2002. *Verbstellung und Verbstellungswandel in den romanischen Sprachen*. Tübingen: Max Niemeyer Verlag.
- Kaiser, Georg A. & Meisel, Jürgen M. 1991. Subjekte und Null-Subjekte im Französischen. In 'DET, COMP und INFL'. *Zur Syntax funktionaler Kategorien und grammatischer Funk-*

- tionen [Linguistische Arbeiten 263], S. Olsen & G. Fanselow (eds), 110–136. Tübingen: Niemeyer.
- Kayne, Richard. 1972. Subject inversion in French interrogatives. In *Generative Studies in Romance Languages*, J. Casagrande & B. Saciuk (eds), 70–126. Rowley MA: Newbury House.
- Kayne, Richard. 1983. Chains, categories external to S, and French complex inversion. *Natural Language and Linguistic Theory* 1: 107–139.
- Kayne, Richard & Pollock, Jean-Yves. 1978. Stylistic inversion, successive cyclicity and move-NP in French. *Linguistic Inquiry* 9: 595–621.
- Kayne, Richard & Pollock, Jean-Yves. 2001. New thoughts on stylistic inversion. In *Subject Inversion in Romance and the Theory of Universal Grammar*, A.C.J. Hulk & J.-Y. Pollock (eds), 107–162. Oxford: OUP.
- Marchello-Nizia, Christiane. 1997. *La langue française aux XIVème et au XVème siècles*. Paris: Nathan.
- Meisel, Jürgen M. 2001. From Bilingual Language Acquisition to Theories of Diachronic Change. *Working Papers in Multilingualism (University of Hamburg)* 30: 1–29.
- Nadasdi, Terry. 1995. Subject NP doubling, matching, and minority French. *Language Variation and Change* 7: 1–14.
- Nadasdi, Terry. 2001. *Variation grammaticale et langue minoritaire: Le cas des pronoms clitiques en français ontarien*. München: Lincom.
- Nagy, Naomi, Blondeau, Hélène & Auger, Julie. 2003. Second language acquisition and ‘real’ French: An investigation of subject doubling in the French of Montreal Anglophones. *Language Variation and Change* 15: 73–103.
- Noonan, Máire. 1989. Operator licencing and the case of French interrogatives. In *Proceedings of the Eighth West Coast Conference on Formal Linguistics*, J. Fee & K. Hunt (eds), 315–330. Stanford CA: CSLI Publications.
- Picard, Marc. 1992. Aspects synchroniques et diachroniques du *tu* interrogatif en québécois. *Revue québécoise de linguistique* 21(2): 65–74.
- Pohl, Jacques. 1965. Observations sur les formes d’interrogation dans la langue parlée et dans la langue écrite non-littéraire. In *Linguistique et Philologie Romanes. Actes du Xe Congrès International de Linguistique et Philologie Romanes*, G. Straka (ed.), 501–516. Paris: Klincksieck.
- Rand, David & Sankoff, David. 1990. *GoldVarb. Version 2. A Variable Rule Application for the Macintosh*, Centre de recherches mathématiques, Montréal: Université de Montréal.
- Rinke, Esther & Meisel, Jürgen M. 2007. On the Syntax of Subject-Verb Inversion and the Role of Information Structure in the History of French. Ms, University of Hamburg.
- Rizzi, Luigi. 1996. Residual verb second and the wh-criterion. In *Parameters and Functional Heads. Essays in Comparative Syntax*, A. Belletti & L. Rizzi (eds), 63–90. Oxford: OUP.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of Grammar. A Handbook in Generative Grammar*, L. Haegeman (ed.), 281–337. Dordrecht: Kluwer.
- Rizzi, Luigi & Roberts, Ian. 1989. Complex inversion in French. *Probus* 1: 1–30.
- Roberge, Yves. 1986. The Syntactic Recoverability of Null Arguments. PhD dissertation, University of British Columbia.
- Roberts, Ian. 1993. *Verbs and Diachronic Syntax. A Comparative History of English and French*. Dordrecht: Kluwer.
- Sankoff, David. 1988. Variable rules. In *Sociolinguistics: An International Handbook of the Science of Language and Society*, Vol. 2, U. Ammon, N. Dittmar & K. J. Mattheier (eds), 984–987. Berlin: Mouton de Gruyter.

- Sankoff, David, Tagliamonte, Sali & Smith, Eric. 2005. *Goldvarb X: A Variable Rule Application for Macintosh and Windows*. Toronto: Department of Linguistics, University of Toronto.
- Sankoff, Gillian. 1982. Usage linguistique et grammaticalisation: Les clitiques sujets en français. In *Die Soziolinguistik in romanischsprachigen Ländern*, N. Dittmar & B. Schlieben-Lange (eds), 81–85. Tübingen: Narr.
- Söll, Ludwig. 1971. Der neufranzösische direkte Fragesatz in einem Korpus der Kindersprache. In *Sprache und Geschichte: Festschrift für Harri Meier*, E. Coseriu & W.-D. Stempel (eds), 178–168. München: Fink.
- Söll, Ludwig. 1983. L'interrogation directe dans un corpus de langage enfantin. In *Études de grammaire française descriptive*, F.J. Hausmann (ed.), 45–54. Heidelberg: Groos.
- Sportiche, Dominique. 1998. Subject clitics in French and Romance Complex inversion and clitic doubling. In *Partitions and Atoms of Clause Structure: Subjects, Agreement, Case and Clitics*, D. Sportiche (ed.), 308–341. London: Routledge.
- Stowell, Tim. 1978. What was there before there was there? *Proceedings of Chicago Linguistic Society* 14: 458–471. (Chicago Linguistic Society, University of Chicago, Chicago, Illinois.)
- Terry, Robert M. 1970. *Contemporary French Interrogative Structures*. Montréal: Éditions Cosmos.
- Thibault, Pierrette. 1983. Équivalence et grammaticalisation. PhD dissertation, Université de Montréal.
- Valois, Daniel & Dupuis, Fernande. 1992. On the status of (verbal) traces in French: The case of stylistic inversion. In *Romance Languages and Modern Linguistic Theory*, P. Hirschbühler & K. Koerner (eds), 325–338. Amsterdam: John Benjamins.
- Vance, Barbara. 1989. Null Subjects and Syntactic Change in Medieval French. PhD dissertation, Cornell University.
- Vance, Barbara. 1997. *Syntactic Change in Medieval French: Verb-Second and Null-Subjects*. Dordrecht: Kluwer.
- de Wind, Maarten. 1995. Inversion in French. PhD dissertation, University of Groningen.
- Zubizarreta, Maria Luisa. 2003. Intervention effects in the French wh-in-situ construction: Syntax or interpretation? In *A Romance Perspective on Language Knowledge and Use. Selected Papers from the 31st Linguistic Symposium on Romance Languages*, R. Nuñez-Cedeño, L. López & R. Camero (eds), 359–379. Amsterdam: John Benjamins.
- Zuckerman, Shalom. 2001. The Acquisition of Optional Movement. PhD dissertation. University of Groningen.

# Verb-late word order in Old Swedish subordinate clauses

Loan, *Ausbau* phenomenon, or both?\*

Steffen Höder and Ludger Zeevaert

University of Hamburg, Research Center on Multilingualism

Subordinate clause word order in Old Swedish differs from both Old Norse and Modern Swedish. Language-internal approaches as well as explanations based on language contact have been put forward in the literature as an explanation of this development. The authors argue that hypotaxis in Swedish emerged neither as a simple loan of syntactic structures nor as a purely internal phenomenon, but as the result of the adaptation of new Latin-based text types which made a functional differentiation of syntactic means necessary. In this paper, hypotaxis is seen as an *Ausbau* phenomenon (in the sense of Kloss) in a multilingual setting induced by language contact but resulting in new and independent structures.

**Keywords:** syntactic change, word order, language contact, Old Swedish, Latin, Middle Low German, *Ausbau*, subordination

## 1. Introduction

One of the most controversial issues in historical linguistics is the question of whether language change is primarily motivated by language-internal factors, such as functional principles, or rather by language-external factors, such as language contact. However, internal and external motivations need not be mutually

---

\* This paper is based on a talk given by the authors at the international colloquium on *language contact and contact languages* in Hamburg (Germany), 6–8 July 2006. It is part of the project *Scandinavian syntax in a multilingual setting* at the Collaborative Research Center on Multilingualism hosted at the University of Hamburg and funded by the German Research Foundation (DFG). We wish to thank the editors and two anonymous reviewers for their helpful comments. All remaining errors are, of course, our own.

exclusive, but can also interact with each other. In this paper, we will argue in favor of an explanation of a particular language change phenomenon in Old Swedish that takes both types of factors into account in a complex contact situation.

### 1.1 Typological classification of the North Germanic languages

Traditionally, the North Germanic languages are classified into two main groups, on the one hand West Scandinavian including Norwegian, Faroese and Icelandic, and on the other hand East Scandinavian with Danish and Swedish (and, additionally, extinct languages in each branch). More recently, this genetically based model has been replaced by a synchronic classification that divides Modern North Germanic into two subgroups, Insular Scandinavian (Icelandic and Faroese) and Continental Scandinavian (Danish, Norwegian, and Swedish; cf. Figure 1).

This classification is essentially based on a number of salient typological differences, such as the existence of complex nominal and verbal inflectional morphology in Insular Scandinavian compared to the much simpler inflectional paradigms in the Continental Scandinavian languages. Such differences along with divergences in the lexicon are usually explained by language change as a consequence of the intensive contact between Continental Scandinavian and (mainly) Low German in the Late Middle Ages and in Early Modern Times.<sup>1</sup>

Even syntactic differences between Insular and Continental Scandinavian have been explained as a result of contact-induced language change. In this paper, we will be discussing a particular word order phenomenon found in Old Swedish subordinate clauses, viz. the verb-late word order.<sup>2</sup> This word order variant has often been characterized as a syntactic loan from Low German or Latin. Such an explanation, though, seems in a way to be quite simplistic and therefore unsatisfactory. Thus, we will outline a rather tentative alternative explanation based on several studies carried out on a balanced corpus of Old Swedish texts.<sup>3</sup>

---

1. For a comprehensive outline see Norde (1997).

2. Cf. Vennemann (1984). *Verb-late* designates word order patterns where the finite verb is to be found not in the second position (as in declarative main clauses in most Germanic languages), but in a later position in the clause.

3. The corpus covers all diachronic stages and different text types in the Old Swedish period (legal and administrative texts as well as religious and secular prose). It has been assembled and (partly) annotated within the research project *Scandinavian syntax in a multilingual setting* at the Collaborative Research Center on Multilingualism in Hamburg (Germany) according to the standards set by the Medieval Nordic Text Archive (Menota), so that existing texts or corpora using the same scheme can easily be added. While primarily aiming at digitalizing (and lemmatizing) a large corpus of Old Nordic manuscripts, the Menota project ([www.menota.org](http://www.menota.org))



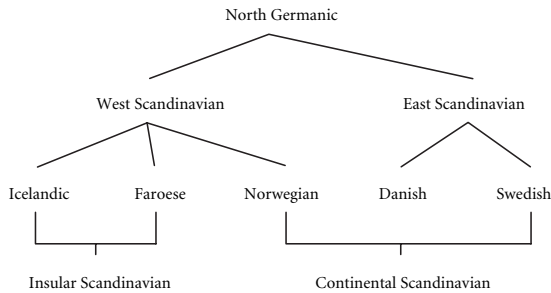


Figure 1. Genetic and typological classification of the North Germanic Languages.

### 1.2 Languages in Late Medieval Sweden

In the Late Middle Ages, i.e. roughly from the 13th to the 15th century, the Swedish-speaking area was characterized by a triglossic distribution of three co-existing languages (cf. Figure 2), a situation which has been discussed extensively, among others, by Braunmüller (2005a, 2007). On the one hand, the vernacular Old Swedish was used as a Low variety, primarily spoken as a language of proximity (cf. Koch & Oesterreicher 1994). On the other hand, we are confronted with two High varieties that were used as languages of distance: Like in Central Europe, Medieval Latin was spoken as a *lingua franca*, and more importantly served as a written language for special purposes in religious domains, e.g. as the language of the Bible, the liturgy and other religious texts. In addition, it was also used in secular literature as well as in administrative contexts. Middle Low German served as an additional *lingua franca* within the Hanseatic area where it was predominantly used as a trading language, but also for administrative purposes, especially in Swedish towns. Furthermore, it was spoken as a vernacular by the economically and socially influential German minority.

Within the same period, Old Swedish began to develop into a literary language, with larger-scale text production in Swedish (written in the Latin alphabet) starting in the early 13th century and eventually replacing the Runic epigraphy in the vernacular language. This implicates that the language took over domains in which it previously had not been used. As a consequence, Swedish also underwent considerable changes within its language system. This process of language

---

also provides a detailed standard for the encoding of medieval Nordic texts in XML (Extensible Markup Language). In order to maximize compatibility with other frameworks, it is based on the more general scheme defined by the Text Encoding Initiative (TEI; cf. [www.tei-c.org](http://www.tei-c.org)) which is widely used in corpus linguistics.



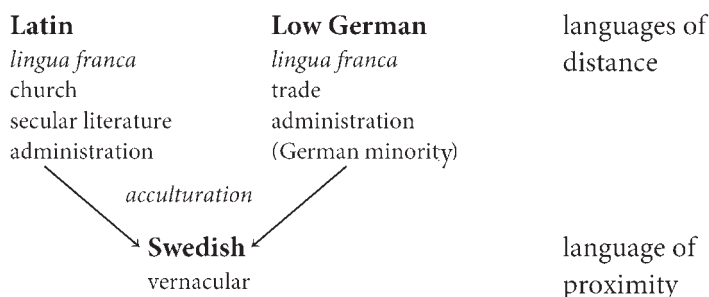


Figure 2. Triglossia and language *Ausbau* in Late Medieval Sweden.

*Ausbau* (in the sense of Kloss 1967, 1978)<sup>4</sup> was by no means an autonomous, monolingual process, but in contrast highly dependent on the influence of the established High varieties, namely Latin and Low German (*acculturation* in Koch & Oesterreicher's 1994:592 terms): Lexical and grammatical features were transferred from these two donor (or 'acculturating') languages to the recipient ('acculturated') language. This relationship is, of course, particularly evident in the context of changes which took place in the Old Swedish lexicon. These changes can be illustrated by early loanwords from Latin (predominantly in the religious domain, e.g. *predika* 'preach' < Latin *praedicare*, *mässa* 'Mass' < Latin *missa*) and by countless loanwords from Low German, which can be found in almost any lexical field, such as *borghamästare* 'mayor' < Low German *borgermēster* (cf. Modern Swedish *borgmästare*), *bruka* 'use' < Low German *brüken*, *kloker* 'wise' < Low German *klōk* (cf. Modern Swedish *klok*), and, incidentally, *sprak* 'language, speech' < Low German *sprāke* (cf. Modern Swedish *språk*). Furthermore, the Old Swedish derivational morphology is to a large extent borrowed from Low German.<sup>5</sup>

4. "The term *Ausbausprache* may be defined as 'language by development'. Languages belonging in this category are recognized as such because of having been shaped or reshaped, molded or remolded – as the case may be – in order to become a standardized tool of literary expression. We might say that an *Ausbausprache* is called a language by virtue of its having been reshaped, i.e., by virtue of its 'reshapedness' if there were such a word" (Kloss 1967:29).

5. The unstressed prefixes that were lost in the Common Scandinavian period (6th to 11th century) were partly replaced by Middle Low German affixes (e.g. *be-*, *ent-*, *er-*, *ge-*). Along with the suffixes of Middle Low German origin, they belong to the most productive derivational morphemes of Swedish (e.g. *-inna*, *-ska*, *-het*, *-ing*, *-bar*, *-aktig*, *-ig*). Comp. Wessén (1967:20ff.), Haugen (1976:221).

## 2. Word order in main and subordinate clauses

### 2.1 A diachronic and contrastive sketch

It has been claimed that the basic word order of the oldest attested North Germanic language, Ancient Nordic, was (subject –) object – verb (cf. Antonsen 1975: 24; Faarlund 2001: 1707). Yet, a thorough analysis of the entire Ancient Nordic textual corpus – i.e. of all known and syntactically analysable Runic inscriptions written in the so-called Older Futhork – reveals that the empirical evidence supporting this claim is by no means sufficient, and that there is most likely no conclusive answer to the question of the basic word order of Ancient Nordic (cf. Braunmüller 2005b: 22f.; Þórhallur Eyþórsson 2001: 46ff.). In addition, subordinate clauses are found to be extremely rare and difficult to identify.<sup>6</sup> Therefore, it cannot be decided on empirical grounds whether or not there was any syntactic differentiation between main and subordinate clauses in Ancient Nordic.

Word order in the oldest North Germanic literary language, classical Old (West) Norse, can generally be said to be relatively free. Declarative main clauses and subordinate clauses show basically identical word order patterns, with verb-second being the dominant one (Faarlund 2004: 191ff.). This is illustrated in examples (1) and (2):<sup>7</sup>

- (1) *Diplomatarium Norvegicum* (cf. Faarlund 2004: 191)

*hann snerisk siðan til trúar*  
 he turned.REFL then to faith.GEN  
 'He then turned to (the Christian) faith.'  
 X V<sub>fin</sub> ...

- (2) *Heimskringla* (cf. Faarlund 2004: 251)

*ef hann var eigi þinn bróðir*  
 if he was not your.NOM brother.NOM  
 'If he was not your brother.'  
 SUBR X V<sub>fin</sub> ...

6. The corpus used by Braunmüller (2005b) contains only two possible instances of subordinate clauses, which, in addition, are nearly identical to each other (cf. Krause and Jankuhn 1966, inscriptions 96, *sā þat bariutþ*, and 97, *sār þat brytr*, '(he) who breaks this').

7. The glossing generally conforms to the Leipzig Glossing Rules (<http://www.eva.mpg.de/lingua/resources/glossing-rules.php> [26 February 2008]; cf. also Croft 2003: xix–xxiii). A list of the abbreviations used is included at the end of the paper.

However, main clauses and conditional subordinate clauses lacking a lexical subordinator<sup>8</sup> may show verb-initial word order as well, as in examples (3) and (4):

- (3) *Finnboga saga* (cf. Faarlund 2004: 192)

*þótti* *monnum* *þat* *líkligt*, *at* ...  
 seemed men.DAT it.NOM likely.N.NOM that  
 'It seemed likely to the men that ...'  
 V<sub>fin</sub> ...

- (4) *Heimskringla* (cf. Faarlund 2004: 252)

*hefði* *þá* *verit* *þetta* *boðit*, *þá* ...  
 had.SBJV then been this.N.NOM offered.N, then  
 'If this had been offered at that time, then ...'  
 V<sub>fin</sub> ...

Modern Insular Scandinavian has not changed significantly in this respect, as verb-second remains the most prominent word order in all clause types (cf. Höskuldur Práinsson 2006: 25, 254). In Modern Continental Scandinavian, however, main and subordinate clauses show slightly different word order patterns: Although the basic word order in both clause types is verb-second, subordinate clauses are marked syntactically by the relative position of the finite verb and, if present, the so-called sentence adverbial. The different patterns are illustrated by the position of the finite verb and the negative particle *inte* in the following Modern Swedish examples (5) and (6):

- (5) Sverige *spelade inte* bra i VM  
 Sweden played not good in World.Cup  
 'Sweden didn't play well in the World Cup.'  
 ... V<sub>fin</sub> NEG ...

- (6) Alla *tyckte att* Sverige *inte spelade* bra i VM  
 all thought COMP Sweden not played good in World.Cup  
 'Everyone thought that Sweden didn't play well in the World Cup.'  
 ... NEG V<sub>fin</sub> ...

In main clauses the finite verb precedes the sentence adverbial, whereas in subordinate clauses this word order is reversed, normally leaving the finite verb in third position, not counting the subordinator except when it is used as a relativizer in subject function. In Scandinavian literature this word order is usually referred to

8. This clause type is not dealt with throughout the remaining part of the paper. Unless otherwise indicated, the term subordinate clause should be understood as a clause that is introduced by a lexical subordinator, such as a subjunction or a relative or interrogative pronoun. Likewise, the term main clause is shorthand for declarative main clause.

as the *AF order* (adverbial before finite verb), as opposed to the *FA order* (finite verb before adverbial) in main clauses (cf. SAG 2000: 5ff.).<sup>9</sup>

## 2.2 Old Swedish

Interestingly, subordinate clause word order in Old Swedish differs strikingly from both Old Norse and Modern Swedish. In general, verb-second can be assumed to be the dominant word order pattern in main and subordinate clauses, as in Old Norse (cf. Wessén 1956: 187ff., 304ff.). Verb-initial word order is a less frequent – in Late Old Swedish only marginal – alternative pattern. Both patterns are exemplified in (7)–(10):

(7) BU4.42 (p. 83)

Gwz vini **skulu** vara sua som dör  
 God's friends shall be so as door  
 'God's friends shall be as a door.'  
 X V<sub>fin</sub> ...

(8) BU4.41 (p. 82)

Tho **foruarna** jak thänna aff twem thingom  
 though warn I this.ACC of two.DAT things.DAT  
 'Yet I warn him of two things.'  
 X V<sub>fin</sub> ...

(9) Pent (p. 35)

**Raadhir** iak hwariom crisnom manne ...  
 advise I every.M.DAT Christian.M.DAT man.DAT  
 'I advise every Christian man ...'  
 V<sub>fin</sub> ...

(10) ÄVgl 12.5 (p. 54)

... at han ær fandær þiuvær  
 COMP he is true.M.NOM thief.M.NOM  
 '... that he really is the thief.'  
 SUBR X<sub>1</sub> V<sub>fin</sub> X<sub>2</sub>

There are very few exceptions from these dominant patterns in main clauses, whereas word order in subordinate clauses varies greatly. Besides the prevailing

9. To be more precise, these are the prototypical word order patterns for subordinate and main clauses in Modern Swedish. There are certain (marginal) types of both main and subordinate clauses that do not conform to these patterns (cf. SAG 2000: 467, 676). For a typological sketch of Modern Swedish word order see Holmer (2006).

verb-second word order, the finite verb can occur in virtually any position after the second one, resulting in what can be called the *verb-late* word order, a salient innovative pattern in Old Swedish subordinate clauses.<sup>10</sup> This word order variant, as described by Wessén (1956: 306ff.) and recently discussed in great detail by Zeevaert (2006), is illustrated in examples (11)–(13):

(11) *Pent* (p. 133)

... *som iak nw giorde*

REL I now did

‘... that I did now.’

SUBR X<sub>1</sub> X<sub>2</sub> V<sub>fin</sub>

(12) *BU7.3* (p. 140)

... *hwikom iach ødhmywkelegha lyddhe j allom bwdordhom*

RPRN I humbly obeyed in all.DAT precepts.DAT

‘... whom I humbly obeyed in all precepts.’

SUBR X<sub>1</sub> X<sub>2</sub> V<sub>fin</sub> X<sub>3</sub>

(13) *Dipl 8822* (p. 493)

... *som jac idher til kirkio stwfwō gifwit hafdhe*

REL I you.PL.DAT to church.GEN room.GEN given had

‘... that I had given you as a church room.’

SUBR X<sub>1</sub> X<sub>2</sub> X<sub>3</sub> V<sub>inf</sub> V<sub>fin</sub>

Different word order patterns often occur within the same text and can even appear within one single sentence as shown in example (14):

(14) *BU4.41* (p. 82)

... *at han skal ey faa thz han kötlica älskir*

COMP he shall not get RPRN he carnally loves<sup>11</sup>

‘... that he shall not get what he carnally loves.’

SUBR X<sub>1</sub> V<sub>fin</sub> X<sub>2</sub> V<sub>inf</sub> – SUBR Y<sub>1</sub> Y<sub>2</sub> V<sub>fin</sub>

Zeevaert (2006), in his study of our corpus, shows that the variation in Old Swedish subordinate clause word order increases from the 14th to the 16th century, with clauses showing the verb-late pattern becoming more frequent towards the end of that period. This diachronic development seems to support the claim by,

10. For the term *verb-late* cf. note 3.

11. In this context, *thet* (graphically *thz*) is analysed as a relative pronoun and, consequently, as a subordinator. A more traditional analysis would refer to it as a demonstrative followed by an unIntroduced (zero-marked) relative clause. With respect to word order, however, this difference is irrelevant.

among others, Pettersson (1988), who suggests that the Modern Swedish adverbial-finite word order found in subordinate clauses actually originates from a reduction and restriction of a more general word order pattern in Old Swedish (but see 3.1). Further investigations carried out on the corpus suggest that also the length of subordinate clauses exhibiting verb-late word order increases over time, meaning that, as a result, the finite verb can actually occur in the fourth or even a later position within the clause.

### 3. Proposed explanations

It has to be explained why a distinct word order pattern in subordinate clauses actually evolved in Old Swedish. Several – in our view, unsatisfactory – explanations have been proposed and elaborated in the literature, which we feel should be at least briefly discussed.

#### 3.1 Language-internal innovation

It has been claimed that the verb-late pattern in Old Swedish was a language-internal innovation that emerged independently of language contact. Pettersson (1988) considers the older so-called *wedge construction* ('kil-konstruktionen') a highly important factor in the evolution of the adverbial-finite word order, i.e. a word order pattern in which other constituents can take the position of the subject before the finite verb in clauses without an overt subject.<sup>12</sup> Interestingly, however, she does not link this particular construction to all types of verb-late patterns in Old Swedish (with the finite verb in the fourth or a later position), which she considers to be of German origin. Unstressed pronouns and adverbs in preverbal position have also been said to have played a key role in the diachronic development of verb-late word order (cf. Åkerlund 1943).

---

12. Pettersson (1988:159), following Larsson (1931), assumes that the basic word order in Early Old Swedish subordinate clauses was verb-second, with the subject usually in preverbal position ([...] *at alle änglane hafðho förft natwrligt wit*, 'COMP all angels.NOM.DEF had.PL first natural.N knowledge', Pent 155). However, in relative clauses with the relativizer in subject function ([...] *fom nw war för jaght* 'REL now was before said.N, Pent 159) as well as in other instances of subordinate clauses without an (overt) subject, other constituents could take the preverbal position as well (*wedge construction*). Pettersson is able to show that negations are the constituents to be found most frequently as wedges in subordinate clauses of this type. Her conclusion from this is that the order negation – finite verb was perceived as a typical feature of subordinate clauses. However, in her presentation the loss of the verb-second constraint in Late Old Swedish and Modern Swedish subordinate clauses remains less clear.

Regardless of whether such language-internal factors did or did not actually play a role, the question still remains as to why the Insular Scandinavian languages did not develop at least similar word order patterns – since, after all, the initial structural conditions were identical within North Germanic.<sup>13</sup>

Since it is undisputed that the intensive language contact with Latin and Low German had an enormous impact on other parts of Continental Scandinavian grammar, it is at least reasonable to consider some form of contact-induced change as a possible explanation for this syntactic innovation as well.

### 3.2 Syntactic loan

In the older linguistic literature, verb-late word order has often been explained as a simple syntactic loan. While some researchers argue in favor of Low German as the source language (e.g. Larsson 1931; Lindblad 1943), others regard Latin as being the actual origin of this pattern (e.g. Wenning 1930). In order to judge the plausibility of these different assumptions, firstly, the type of language contact involved and, secondly, the word order of the respective languages have to be examined from a modern linguistic perspective.

#### 3.2.1 *Middle Low German*

The contact between Old Swedish and Middle Low German was very intensive and led to a considerable amount of speakers becoming bilingual. According to contact typological research, this type of contact situation is deemed to be a necessary factor in contact-induced change (cf. Thomason 2001:1640f.). Nettle (1999:138) argues that

Word order [...] is extremely prone to areal convergence. The chief vector of this appears to be bilingualism. The processing and parsing habits of the bilingual brain lead to the word order of one language interfering with that of another wherever there is a substantial number of people speaking both languages.

---

13. Following Christoffersen (2002:185) fundamental differences between Old West and Old East Nordic word order patterns are not attested. Instances of verb-late constructions can be found in Early Modern Icelandic texts (e.g. Bible translations) but have to be regarded as translation interferences. In Oddur Gottskálkssons Icelandic translation of the New Testament from 1540, the final position of the finite verb is quite a usual word order pattern (Jón Helgason 1999:168). One of the examples Jón gives is Matthew 13,12: *sua at þeir eigi með augum Siae | og eyrum heyre* ('so COMP they.M.NOM NEG with eyes.DAT see.SBJV.3PL, and ears.DAT hear.SBJV.3PL').

In this respect, Low German certainly qualifies as a possible source language for verb-late word order. However, the – admittedly somewhat out-of-date – Middle Low German reference grammars (e.g. Lasch 1974 [1914]) are all but clear with respect to word order. Härd (2000: 1460f.) acknowledges at least some degree of variation in the word order of subordinate clauses. Furthermore, recent diachronic studies carried out on Middle Low German syntax (e.g. Rösler 1997; Braunmüller 1998; Tophinke forth.) reveal that the subordinate clause word order in fact varies to a fairly high degree. Thus, although being a typical syntactic feature of modern German, verb-final order – in contrast to expectations – does not seem to be the basic pattern in subordinate clauses in Middle Low German. In addition, the development towards verb-final word order in Middle Low German subordinate clauses is a very slow and gradual process that took place at a later stage in time than the corresponding evolution of the verb-late word order pattern in Old Swedish. Verb-final word order was not obligatory until the end of the 16th century. On top of that, one could expect its manifestation in Low German texts, at that time, to be influenced by the language shift towards High German. These facts make Low German seem a less plausible candidate for being the source of verb-late word order in Old Swedish.

### 3.2.2 *Medieval Latin*

As for Medieval Latin, the picture is not clear either. On the one hand, one has to point out that even if it is certain that Latin had some impact on the development of Old Swedish, the contact between the two languages can by no means be regarded as being intensive in the same sense as the contact with Low German. This is basically due to the following two reasons: Firstly, Latin was predominantly a written language, and in addition it was used almost exclusively in formal contexts, not in everyday life situations. Secondly, there never was a high number of Swedish-Latin bilinguals. Therefore, the contact situation does not provide the conditions that are necessary for contact-induced word order convergence (cf. 3.2.1). But on the other hand, this argument holds only when taking the entire Swedish population into consideration, the majority of whom were, of course, illiterate and did not have any competence in Latin at all. However, within the elite group of the authors, translators, scribes, and of course readers of Old Swedish literature in the Late Middle Ages, the percentage rate of – at least receptively – bilingual people must have been quite high, since the vast amount of all formal education took place in Latin (cf. Öberg 1994; Braunmüller 2007).

Latin is typically said to have subject – object – verb as its basic word order, with verb-final word order often supposed to be the most frequent pattern as well. However, this general view seems to be flawed in that it is mainly based on the



most well-known variety of Latin, viz. literary Classical Latin. But in fact, verb-final word order is neither obligatory nor most frequent in any non-classical – i.e. archaic, non-literary, or medieval – variety of Latin, where the word order in both main and subordinate clauses seems to be governed mainly by communicative and pragmatic principles rather than grammatical rules (cf. Panhuis 1982; Pinkster 1991; Kiesler 2006:65ff.). On the other hand, however, it has to be pointed out that verb-final word order is far more frequent in subordinate clauses than elsewhere (cf. Raible 1992: 328; Chirita 2003: 182ff.).

#### 4. Subordination as an *Ausbau* phenomenon

Apparently, the evolution of the verb-late word order in Old Swedish subordinate clauses can neither be satisfactorily explained as a language-internal development nor as a simple syntactic loan from Low German or Latin. Therefore, we wish to present a different approach that takes its point of departure in the notion of subordination itself but also takes the type of language contact situation into account.

##### 4.1 Coordination vs. subordination

Traditionally, the distinction between coordination and subordination is regarded as a dichotomy, meaning that each clause is analyzed as either a main or a subordinate clause. However, research on grammaticalization has shown that connective and deictic elements introducing main clauses quite frequently develop into various kinds of subordinators (cf. Heine & Kuteva 2002: 106f., 115f.).<sup>14</sup> This in turn implies a diachronic stage in which the status of such connective elements is ambiguous and cannot be captured by the traditional either-or-distinction. Moreover, typological research has shown that a cross-linguistically viable notion of subordination is very difficult to apply. Thus, it has been proposed to refer to coordination and subordination as constituting a *continuum* of different clause linkage types (cf. Cristofaro 2003: 22ff.) which are distinguished by means of various lexical, morphological, or syntactic subordination markers. This approach seems to be not only more reasonable from a general typological perspective, but also more appropriate in the case of the description and explanation of the very

---

14. Comp. Modern High German *weil* ‘because’ < Old High German *dia wīla so* ‘the time COMP’, Modern English *that* ‘COMP’ < Old English *thæt* ‘DEM.N’, Modern Swedish *då* ‘when, as’ < Old Swedish *þa* ‘then’.

process in which the distinction between main and subordinate clauses is being established or changed in a language system.

Applying this view to the case of Old Swedish leads to a classification of subordinate clauses into different types and degrees of subordination. Although functional definitions of subordination have been both proposed and applied successfully in cross-linguistic typological studies in the past (cf. Cristofaro 2003: 29ff.), the analysis of Old Swedish data can also be based on a syntactic classification that takes into account distinct types of overt subordination marking and varying degrees of overt syntactic integration.<sup>15</sup> In combination with a balanced corpus, such a classification additionally enables an analysis of the diachronic development of different subordination strategies (including verb-late word order) in various types of texts. For such an analysis, we propose the following set of criteria:

1. Syntactic integration:

- a. Embedding of the subordinate clause in the matrix clause as in example (15), as opposed to a mere linear sequence of clauses as in (16) (brackets indicate clause structure):

(15) BU7.8 (p. 161)

*iak [som föddhe sannan gwdh] bær vithne ...*  
 I REL bore true.M.ACC god bear witness  
 'I, who bore the true God, bear witness ...'

- b. Occurrence of the subordinate clause as an obligatory argument of some element (typically a verb) within the matrix clause without any additional expletive element, as shown in example (16) in which the subordinate clause functions as the subject of the matrix clause:

(16) BU4.49 (p. 92)

*thy.at thäkkiare är gudhi [at ey sighis mäsä ...*  
 for pleasant.CMPR is god.DAT COMP not say.SBJV.REFL Mass  
 'For it pleases God more that no Mass is held ...'

- c. Topicalization of the subordinate clause within the matrix clause with the finite verb in second position (i.e. immediately following the subordinate clause) as illustrated in example (17). The latter two criteria are crucial arguments in favor of a clear syntactic rank shift of the subordinate clause ('desententialization', 'constituent behavior'):

---

15. In addition, such an operational definition of subordination (and its subtypes) can be directly implemented in order to identify subordinate clauses in a large corpus by means of automatic search tools. An abstract or traditional concept would be of little use in this respect.

(17) BU7.15 (p. 188)

[Sedhan thet war giorth] Satte the ...  
 since that.N was done.N set they  
 ‘When that had been done, they put ...’

## 2. Subordination marking:

- a. By semantically empty subjunctions, either an all-purpose complementizer (typically *at* ‘that’) or an all-purpose relativizer (such as *sum*). Subjunctions of this type are both the earliest attested and most frequent ones in Old Swedish texts, and are functionally unambiguous. They occur in non-integrated as well as in embedded and topicalized clauses, and can introduce clauses that function as obligatory arguments.
- b. By semantically filled (morphologically simplex) subjunctions, e.g. *māpan* ‘while’, *om* ‘if’, or *sīpan* ‘since’. These subjunctions also show little functional ambiguity as subordinators, even though they can also belong to other word classes (typically adverbs and/or prepositions). However, they are far less frequent than their semantically empty counterparts and typically do not occur in clauses that are obligatory arguments.
- c. By relative or interrogative pronouns, e.g. *hvilkin* ‘who, which’, or similar adverbs, such as *hvar* ‘where’. These subordinators are particularly frequent in Late Old Swedish texts and have to be considered the most recently evolved among the lexical subordinators.
- d. By morphologically complex subjunctions, possibly unverbated or phonetically reduced, e.g. *för þy at* (cf. 4.2). These elements are highly frequent in Late Old Swedish texts. Strikingly, clauses marked by these subordinators hardly ever meet any of the syntactic integration criteria listed above, i.e. they are typically non-integrated.<sup>16</sup>
- e. Verb-late word order, which occurs in all possible types of subordinate clauses.

## 4.2 Old Swedish text production

In general, language contact from a historical point of view is often still understood and described as an abstract process operating on the level of two or more language systems interacting with each other. However, a realistic explanation of language change has to be more specific with respect to the level on which the interaction actually takes place (cf. Braunmüller 2004). In other words, we have

---

16. The fact that subordinate clauses of this type normally are not syntactically integrated in the matrix clause makes word order the crucial criterion for deciding on whether these morphologically complex connective elements can actually be regarded as subordinators or not.

to take into account the specific individual and social factors of language contact, such as the speakers' multilingualism and the network structures within the speaker communities, as exercised in contemporary synchronic sociolinguistic research. It would, for instance, be a lot more plausible to refer to Latin as a possible source for a change in the Old Swedish language system if one could show that this change originated from a socially influential group of bilingual speakers or was a specific feature of their particular variety of Old Swedish.

This is highly relevant because the language system we find attested in the Old Swedish sources can in no way be regarded as being representative of the whole language. On the contrary, given the sociohistorical and sociolinguistic information available, the Old Swedish texts have to be considered merely as a sample of the elaborated written variety of the highly educated and (to some extent) multilingual cultural elite of that time (cf. Wollin 2003). Thus, features of this variety need not be typical for Old Swedish as a whole, but instead may be considered as being specific to this one form of the written language. This is particularly important if a feature is not preserved in spoken varieties of the modern language (e.g. in any of the traditional dialects), as is the case with verb-late word order in Swedish.<sup>17</sup>

Thus, we propose to view the development of the verb-late word order in Old Swedish subordinate clauses in the context of the adaptation of continental European – i.e. Latin-based – text types (cf. Lönnroth 2005) in the course of the language *Ausbau* in the Late Middle Ages. The Old Swedish literary production is highly dominated by translations from Latin sources (cf. Wollin 2002), especially within the increasingly important genre of religious prose. While earlier texts are rather free paraphrases, later translators aim at producing texts in the vernacular that are formally equivalent to the foreign originals.<sup>18</sup> This process results in a kind of imitation of the more than often complex *hypotactic* sentence

---

17. This assumption is supported by the evidence presented by Jørgensen (1978: 63). From his analysis of the word order in a corpus of written and spoken Swedish it becomes clear that the adverbial-finite word order in Modern Swedish subordinate clauses is primarily a feature of the written language while being far from obligatory in spoken (colloquial) Swedish. In the spoken part of the corpus only 37% of the conjunctive clauses containing a sentence adverbial and a finite verb exhibit the adverbial-finite order obligatory in Standard Swedish, whereas in the written part of the corpus this is true for 93% of the conjunctive clauses.

18. Due to this process, translation interferences in the syntax of the target language are to be expected and do indeed occur. For example, infinitival and participial constructions as found in Latin are frequently copied in Late Old Swedish texts. However, interference in word order, although occasionally proposed as a possible explanation for the occurrence and frequency of verb-late word order particularly in Late Old Swedish texts, can hardly be observed. The innovative pattern is clearly productive, i.e. it occurs independently of foreign originals.

structure within the translations. This effect is illustrated in the Late Old Swedish translation in example (18) of the Latin sentence given in (19) compared to the characteristic paratactic sentence structure of the Early Old Swedish example in (20) (*italic face* indicates a lexical subordinator):

- (18) *BU7.10* (p. 164f.)

<sup>MC</sup>[Hør nw the tingh <sup>RC</sup>[*som* iach vill oppenbara tig]<sub>RC</sub> aff them ærchebi-  
scopenom / <sup>RC</sup>[*hwlkin som* sagde <sup>SC</sup>[*ath* han ville geffwa allom prestomen  
loff <sup>INF</sup>[*ath* leffwa j kothliko hyonelaghy]<sub>INF</sub> <sup>SC</sup>[*om* han ware pawe]<sub>SC</sub>]<sub>SC</sub>  
<sup>PART</sup>[Tænkiande ok throandhe <sup>INF</sup>[thet vara gwdhy tækkelikare / æn <sup>SC</sup>[*ath*  
klærkane leffde swa løsluka <sup>RC</sup>[*som* the nw leffwa]<sub>SC</sub>]<sub>SC</sub>]<sub>INF</sub>]<sub>PART</sub>]<sub>RC</sub>]<sub>MC</sub>

- (19) *Rev 7.10.3-4* (p. 137)

<sup>MC</sup>[audi nunc ea, <sup>RC</sup>[*que* tibi manifestare volo]<sub>RC</sub>, videlicet de illo archiepis-  
copo, <sup>RC</sup>[*qui* dixit, <sup>SC</sup>[*quod* <sup>SC</sup>[*si* esset papa]<sub>SC</sub>, ipse daret licenciam omni-  
bus clericis et presbiteris <sup>GER</sup>[matrimonium contrahendi carnaliter]<sub>GER</sub>]<sub>SC</sub>  
<sup>PART</sup>[cogitans et credens <sup>SC</sup>[*quod* hoc esset acceptabilius Deo, quam <sup>SC</sup>[*quod*  
clerici viuerent dissolute <sup>SC</sup>[*sicut* modo viuunt]<sub>SC</sub>]<sub>SC</sub>]<sub>PART</sub>]<sub>RC</sub>]<sub>MC</sub>  
‘Hear now the things that I want to reveal to you, concerning the archbishop  
who said that if he were pope, he would give permission to all clerics and  
priests to contract carnal marriages, thinking and believing that this would  
be more acceptable to God than that the clerics live dissolutely as they live  
now.’

- (20) *ÄVgl 2.5* (p. 13)

<sup>MC</sup>[Dræpær maþar fvænþkan man. eller fmalenþkæn innan konongþrikis.  
man eigh væftgöþkan]<sub>MC</sub> <sup>MC</sup>[böte firi atta örtogher. ok þrettan markær. ok  
ænga ætar bot.]<sub>MC</sub>  
‘A man slays a Swedish man or a man from Småland, a man from the king-  
dom, not a West Geat. He shall be fined eight *örtugar* and thirteen marks  
and no weregild.’

In order to establish a formal equivalence between the translations and their originals and at the same time to follow the hypotactic structure of these texts, Old Swedish writers increasingly make use of formal means in order to differentiate between coordinated and subordinated structures. Thus, subordinated structures are overtly marked. On the one hand, this is achieved by means of newly emerging lexical subordinators, mainly subjunctions resulting from the grammaticalization of connective constructions (often including univerbation and phonetic reduction of prepositional phrases and complementizers, such as *för þy at* ‘for this that’ > *för, þy at, þyt, þy* ‘for, because’ or *äptir þy at* ‘after this that’ > *äpte at, äpte*

‘after, when’)<sup>19</sup> and inflected relative pronouns (such as *hvilkin*, *pän* ‘who, which’, as opposed to the older uninflected relative particles such as *sum*, *pär* ‘that’; see 4.1). On the other hand, a particular word order pattern, verb-late, is grammaticized as a syntactic subordination marker, i.e. an optional strategy of unambiguous subordination marking.

Thus, the emergence of an unambiguous (syntactic) subordination marker in Old Swedish can possibly be explained as a contact-induced innovation within the Latin-based *Ausbau* process, resulting in the replication (cf. Heine & Kuteva 2005; Heine [this volume]) of the strict distinction between coordination and subordination found in Latin. This approach does not, however, account for the origin of the actual word order pattern that is grammaticized. If this pattern itself is actually a loan from one of the contact languages, we would argue that Latin cannot be rejected out of hand as the less likely candidate of the two. It is quite possible that Old Swedish writers with at least a receptive competence in Latin have adopted a word order variant that was relatively frequent in Latin subordinate clauses and that was therefore perceived as a salient pattern in comparison to the main clause word order even though the syntactic distinction between the two clause types was not obligatory in Latin.<sup>20</sup>

As already mentioned in 3.2.1, for chronological reasons Middle Low German, despite of its undisputed influence on the Swedish lexicon and morphology, is a less probable factor in this development. However, (High and Low) German may have had a supporting influence.<sup>21</sup>

## 5. Concluding remarks

Is verb-late word order as found in Old Swedish subordinate clauses a syntactic loan, an *Ausbau* phenomenon, or both? In this paper we have tried to demonstrate that the traditional (or ready-made) dichotomy between internal and external motivations for language change is misleading in this case, and that we therefore have to follow a different approach in order to explain such a phenomenon in a

---

19. The phonetic and especially prosodic changes involved in this process cannot be discussed in detail here.

20. Interestingly, this would imply that Old Swedish speakers replicated a primarily pragmatically motivated word order variant in Latin as an (optional) subordination marker in Old Swedish. This process is partially similar to the phenomenon which Heine (this volume) terms ‘pragmatic unmarking’ in that it involves the grammaticization of a pragmatically marked use pattern into a syntactic pattern.

21. Comp. Wessén (1956: 333).

complex contact situation. On the one hand, the evolution of an optional subordinate clause word order in Old Swedish is likely to be a contact-induced innovation. However, on the other hand, even if the actual word order can be regarded as a loan from another language, this development must also be looked at in the context of language *Ausbau* in a multilingual setting. In this context, the most important factors are Latin-script literacy, the adaptation of Latin-influenced text types, and the increasing use of hypotactic structures in formally equivalent translations and original Swedish texts, leading to the emergence of new lexical subordinators and other subordination strategies. In order to fully understand this process and to further analyze the linguistic material, it is necessary to specify the notion of 'language contact' (in terms of individual or social multilingualism), to reflect on the language of the Old Swedish sources (which represents an elaborated written variety), and at the same time to adopt a differentiated view on subordination as a gradual and evolving phenomenon.

## Abbreviations

ACC = accusative; CMPR = comparative; COMP = complementizer; DAT = dative; DEF = definite; GEN = genitive; GER = gerundial construction; IMP = imperative; INF = infinitival construction; M = masculine; MC = main clause; N = neuter; NEG = negation; NOM = nominative; PART = participial construction; PL = plural; RC = relative clause; REL = relative particle; REFL = reflexive; RPRN = relative pronoun; SBJV = subjunctive; SC = subordinate clause; SUBR = subordinator; SUP = supine; V<sub>fin</sub> = finite verb; V<sub>inf</sub> = infinite verb; X, Y = any constituent.

## Sources

- ÄVgl = Äldre Västgötalagen. Collin, H. S. and Schlyter, C. J. (eds). 1827. *Westgöta-Lagen* [Samling av Sveriges gamla lagar 1]. Stockholm: Haeggström.
- BU4 = St. Bridget's revelations, 4th book (Swedish): Klemming, G. E. (ed.). 1860. *Heliga Birgittas uppenbarelser*, Vol. 2 [Samlingar utgivna av Svenska fornskriftsällskapet 1.14.2]. Stockholm: Norstedt.
- BU7 = St. Bridget's revelations, 7th book (Swedish): Lindell, Inger (ed.). 2000. *Heliga Birgittas uppenbarelser bok 7 efter Cod. Ups. C 61. Diplomatarisk utgåva med kommenterande inledning* [Samlingar utgivna av Svenska fornskriftsällskapet 1.84]. Uppsala: Fornskriftsällskapet.
- Dipl = Diplomatarium: Riksarkivet (ed.). 2002. *Svenskt diplomatarium*, Vol. 10.3: 1374 juli – 1375. Stockholm: Elander.

- Pent* = Pentateuch paraphrase: Thorell, Olof (ed.). 1959. *Fem Moseböcker på svenska enligt Cod. Holm. A 1* [Samlingar utgivna av Svenska fornskriftsällskapet 1.60]. Uppsala: Almqvist & Wiksell.
- Rev 7* = St. Bridget's revelations, 7th book (Latin): Bergh, Birger (ed.). 1967. *Den Heliga Birgittas Revelaciones. Bok VII* [Samlingar utgivna av Svenska fornskriftsällskapet 2.7.7]. Uppsala: Almqvist & Wiksell.

## References

- Åkerlund, Walter. 1943. Om det finita verbets plats i den fornsvenska bisatsen. *Arkiv för nordisk filologi* 57: 1–67.
- Antonsen, Elmer. 1975. *A Concise Grammar of the Older Runic Inscriptions* [Sprachstrukturen A 3]. Tübingen: Niemeyer.
- Braunmüller, Kurt. 1998. Wortstellungstypologische Untersuchungen zu den Kontaktsprachen der Hansezeit (Mittelniederdeutsch, Dänisch, Schwedisch). In *Historische germanische und deutsche Syntax. Akten des Internationalen Symposiums anlässlich des 100. Geburtstages von Ingerid Dal, Oslo, 27. 9.–1. 10. 1995* [Osloer Beiträge zur Germanistik 21], J. Ole Askedal (ed.), 315–334. Frankfurt: Lang.
- Braunmüller, Kurt. 2004. Plädoyer für eine Umorientierung in der Beschreibung von Sprachwandelprozessen. *NOWELE* 44: 21–49.
- Braunmüller, Kurt. 2005a. Language contacts in the Late Middle Ages and in Early Modern Times. In *The Nordic Languages. An International Handbook of the History of the North Germanic Languages*, Vol. 2 [Handbooks of Linguistics and Communication Science 22.2], O. Bandle et al. (eds), 1222–1233. Berlin: Mouton de Gruyter.
- Braunmüller, Kurt. 2005b. Variation in word order in the oldest Germanic Runic inscriptions. A case for bilingualism? *NOWELE* 46/47: 15–30.
- Braunmüller, Kurt. 2007. Receptive multilingualism in Northern Europe in the Middle Ages. A description of a scenario. In *Receptive Multilingualism. Linguistic Analyses, Language Policies and Didactic Concepts* [Hamburg Studies on Multilingualism 6], L. Zeevaert & J. D. ten Thije (eds), 25–47. Amsterdam: John Benjamins.
- Chirita, Diana. 2003. Did Latin influence German word order? Aspects of German-Latin bilingualism in the Late Middle Ages. In *Aspects of Multilingualism in European Language History* [Hamburg Studies on Multilingualism 2], K. Braunmüller & G. Ferraresi (eds), 173–200. Amsterdam: John Benjamins.
- Christoffersen, Marit. 2002. Nordic language history and research on word order. In *The Nordic Languages. An International Handbook of the History of the North Germanic Languages*, Vol. 1 [Handbooks of Linguistics and Communication Science 22.1], O. Bandle et al. (eds), 182–191. Berlin: Mouton de Gruyter.
- Cristofaro, Sonia. 2003. *Subordination*. Oxford: OUP.
- Croft, William. 2003. *Typology and Universals*. 2nd edn. Cambridge: CUP.
- Faarlund, Jan Terje. 2001. From Ancient Germanic to Modern Germanic languages. In *Language Typology and Language Universals. An International Handbook*, Vol. 2 [Handbooks



- of Linguistics and Communication Science 20.2], M. Haspelmath et al. (eds), 1706–1719. Berlin: Mouton de Gruyter.
- Faarlund, Jan Terje. 2004. *The Syntax of Old Norse. With a Survey of the Inflectional Morphology and a Complete Bibliography*. Oxford: OUP.
- Härd, John Evert. 2000. Syntax des Mittelniederdeutschen. In *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*, Vol. 2 [Handbooks of Linguistics and Communication Science 2.2], W. Besch et al. (eds), 1456–1463. 2nd edn. Berlin: Mouton de Gruyter.
- Haugen, Einar. 1976. *The Scandinavian Languages. An Introduction to their History* [The great languages]. London: Faber and Faber.
- Heine, Bernd & Kuteva, Tania. 2002. *World Lexicon of Grammaticalization*. Cambridge: CUP.
- Heine, Bernd & Kuteva, Tania. 2005. *Language Contact and Grammatical Change*. Cambridge: CUP.
- Holmer, Arthur. 2006. The place of Swedish in word order typology. *Sprachtypologie und Universalienforschung* 59: 76–102.
- Höskuldur Þráinsson. 2006. *Íslensk tunga*, Vol. 3: *Setningar. Handbók um setningafræði*. Reykjavík: Almenna bókafélagið.
- Jón Helgason. 1999. *Málið á Nýja Testamenti Odds Gottskálkssonar* [Safn Fræðafélagsins um Ísland og Íslendinga 7]. Reykjavík: Málvísindastofnun Háskóla Íslands. (Reprint of the edition Copenhagen 1929).
- Jørgensen, Nils. 1978. *Underordnade satser och fraser i talad svenska. Funktion och byggnad* [Lundastudier i nordisk språkvetenskap C 10]. Lund: Ekstrand.
- Kiesler, Reinhard. 2006. *Einführung in die Problematik des Vulgärlateins* [Romanistische Arbeitshefte 48]. Tübingen: Niemeyer.
- Kloss, Heinz. 1967. 'Abstand languages' and 'Ausbau languages'. *Anthropological linguistics* 9(7): 29–41.
- Kloss, Heinz. 1978. *Die Entwicklung neuer germanischer Kultursprachen seit 1800* [Sprache der Gegenwart 37]. 2nd edn. Düsseldorf: Schwann.
- Koch, Peter & Oesterreicher, Wulf. 1994. Schriftlichkeit und Sprache. In *Writing and its Use. An Interdisciplinary Handbook of International Research*, Vol. 1 [Handbooks of Linguistics and Communication Science 10.1], H. Günther & O. Ludwig (eds), 587–604. Berlin: de Gruyter.
- Krause, Wolfgang & Jankuhn, Herbert. 1966. *Die Runeninschriften im älteren Fupark*, Vol. 1: *Text*. Göttingen: Vandenhoeck & Ruprecht.
- Larsson, Carl. 1931. *Ordföljdsstudier över det finita verbet i de nordiska fornspråken* [Uppsala universitets årsskrift 1931.1]. Uppsala: Lundequistska bokhandeln.
- Lasch, Agathe. 1974. *Mittelniederdeutsche Grammatik* [Sammlung kurzer Grammatiken germanischer Dialekte A 9]. 2nd edn. Tübingen: Niemeyer.
- Lindblad, Gustaf. 1943. *Relativ satsfölgning i de nordiska fornspråken* [Lundastudier i nordisk språkvetenskap 1]. Lund: Gleerup.
- Lönnroth, Lars. 2005. From Old Nordic to Early Modern Nordic: The development of types of text. In *The Nordic Languages. An International Handbook of the History of the North Germanic Languages*, Vol. 2 [Handbooks of Linguistics and Communication Science 22.2], O. Bandle et al. (eds), 1189–1195. Berlin: Mouton de Gruyter.
- Nettle, Daniel. 1999. *Linguistic Diversity*. Oxford: OUP.

- Norde, Muriel. 1997. Middle Low German–Middle Scandinavian language contact and morphological simplification. *Multilingua* 16: 389–409.
- Öberg, Jan. 1994. Vem kunde latin i medeltidens Sverige? In *Medeltida skrift- och språkkultur. Nordisk medeltidsliteracy i ett diglossiskt och digrafiskt perspektiv II. Nio föreläsningar från ett symposium i Stockholm våren 1992* [Opuscula 2], I. Lindell (ed.), 213–224. Stockholm: Runica et Mediævalia.
- Panhuis, Dirk G. J. 1982. *The Communicative Perspective in the Sentence. A Study of Latin Word Order* [Studies in Language Companion Series 11]. Amsterdam: John Benjamins.
- Pettersson, Gertrud. 1988. Bisatsordföljden i svenskan eller Varifrån kommer BIFF-regeln? *Arkiv för nordisk filologi* 103: 157–180.
- Pinkster, Harm. 1991. Evidence for SVO in Latin? In *Latin and the Romance Languages in the Early Middle Ages*, R. Wright (ed.), 69–82. London: Routledge.
- Raible, Wolfgang. 1992. The pitfalls of subordination: Subject and object clauses between Latin and Romance. In *Historical Philology. Greek, Latin and Romance. Papers in Honor of Oswald Szemerényi II* [Current Issues in Linguistic Theory 87], B. Brogyanyi & R. Lipp (eds), 299–337. Amsterdam: John Benjamins.
- Rösler, Irmtraud. 1997. *Satz – Text – Sprachhandeln. Syntaktische Normen der mittelniederdeutschen Sprache und ihre soziofunktionalen Determinanten* [Sprachgeschichte 15]. Heidelberg: Winter.
- SAG = Teleman, Ulf, Hellberg, Staffan & Andersson, Erik. 2000. *Svenska Akademiens grammatik*, Vol. 4: *Satser och meningar*. Stockholm: Norstedts Ordbok.
- Thomason, Sarah Grey. 2001. Contact-induced typological change. In *Language Typology and Language Universals. An International Handbook*, Vol. 2 [Handbooks of Linguistics and Communication Science 20.2], M. Haspelmath et al. (eds), 1640–1648. Berlin: Mouton de Gruyter.
- Þórhallur Eyþórsson. 2001. The syntax of verbs in Early Runic. *Working Papers in Scandinavian Syntax* 67: 1–55.
- Tophinke, Doris. Forth. Vom Vorlesetext zum Lesetext: Zur Syntax mittelniederdeutscher Rechtsverordnungen im Spätmittelalter. In *Oberfläche und Performanz. Untersuchungen zur Sprache als dynamische Gestalt*, A. Linke & H. Feilke (eds). Tübingen: Niemeyer.
- Vennemann, Theo. 1984. Verb-second, verb late, and the brace construction. Comments on some papers. In *Historical Syntax* [Trends in Linguistics. Studies and Monographs 23], J. Fisiak (ed.), 627–636. Berlin: Mouton de Gruyter.
- Wenning, E. Alf. 1930. *Studier över ordföljden i fornsvenskan. Predikatets bestämningar i äldre och yngre fornsvenskan*. Lund: Lindstedt.
- Wessén, Elias. 1956. *Svensk språkhistoria*, Vol. 3: *Grundlinjer till en historisk syntax* [Stockholm Studies in Scandinavian Philology 19]. Stockholm: Almqvist & Wiksell.
- Wessén, Elias. 1967. *Om det tyska inflytandet på svenskt språk under medeltiden* [Skrifter utgivna av Nämnden för svensk språkvård 12]. 3rd edn. Stockholm: Norstedt.
- Wollin, Lars. 2002. Nordic language history and the history of translation II: Swedish. In *The Nordic languages. An International Handbook of the History of the North Germanic Languages*, Vol. 1 [Handbook of Linguistics and Communication Science 22.1], O. Bandle et al. (eds), 511–518. Berlin: Mouton de Gruyter.

- Wollin, Lars. 2003. Swedish and Swedish. On the origin of diglossia and social variation in the Swedish language. In *Aspects of Multilingualism in European Language History* [Hamburg Studies on Multilingualism 2], K. Braunmüller & G. Ferraresi (eds), 145–171. Amsterdam: John Benjamins.
- Zeevaert, Ludger. 2006. Variation und kontaktinduzierter Wandel im Altschwedischen. *Working Papers in Multilingualism* 74, Series B. University of Hamburg: Sonderforschungsbereich 538 Mehrsprachigkeit.

# Contact-induced phonological changes in the Catalan spoken in Barcelona\*

Conxita Lleó<sup>a,b</sup>, Susana Cortés<sup>a</sup> and Ariadna Benet<sup>a</sup>

University of Hamburg, Research Center on Multilingualism<sup>a</sup> /

Department of Romance Languages<sup>b</sup>

On the basis of elicited data, this article focuses on the purported on-going change of the Catalan language spoken in Barcelona, especially as regards the pronunciation of the two open mid vowels /ɛ/ and /ɔ/ and the central vowel [ə]. The data have been collected from three generations, i.e., children, youngsters and adults, in two districts (i.e., Gràcia and Nou Barris) differing on the degree of presence of Spanish. Percentages of target-like pronunciation for these vowels have been quantified. Whereas adults do not exhibit any noticeable difference across districts, the two younger generations lead to statistically different results, with considerably higher percentages in Gràcia, the traditionally Catalan district, than in Nou Barris. These results are discussed in the light of internal and external factors for language change, as production seems to be less dependent from the language spoken in the family than from the language dominant in the district, at school, and within the peer group.

**Keywords:** allomorphy, allophony, Catalan phonology, internal factors, external factors, markedness, on-going phonological change, sociolinguistic variables, variability in production, vowels

---

\* This study has been carried out as part of the research project “H6: Phonoprosodic development of Catalan in its current bilingual context”, conducted by the first author, at the Collaborative Research Center on Multilingualism. We are indebted to the Center and its sponsors, the Deutsche Forschungsgemeinschaft (German Research Foundation) and the University of Hamburg for their support. Furthermore, we would like to thank Laura Batalla, Laia Puig and Anna Schreiber for their essential help with the transcription of the data. We also thank the reviewers and both editors of this volume for their valuable suggestions that have substantially contributed to improve the article. Furthermore, this study had not been possible without the collaboration of schools, children and parents who have kindly accepted to participate; we sincerely thank them all.

## 1. Introduction

### 1.1 The Catalan language in contact with Spanish

Catalan is spoken in several regions in the east of Spain (Catalonia, Valencia, the Balearic Islands, in a part of Aragon and an area of Murcia), in Andorra, in the French territory of Roussillon (in the southeast of France) and in the Italian city of Alghero (on the island of Sardinia). All over these areas, Catalan has 7.5 million speakers and 2.5 additional million have passive command of the language (according to the Generalitat de Catalunya 2006a). Catalan has always coexisted with at least one other language (namely Spanish, French, Occitan, Italian and Sardinian), which means that bilingual speakers show dominance in Catalan or in the other language(s) to different degrees. In this article we analyze the contact of Catalan phonology with Spanish phonology. Specifically, we focus on the Catalan spoken in the capital city of Catalonia, Barcelona, because there the contact between both languages is most intense; the variety spoken there is Central Catalan.

Although Barcelona and Catalonia have historically been bilingual areas, Spanish gained strength from the 1940s onwards due to two facts. First, during Franco's dictatorship (1939–1975) Catalan was prohibited in public domains and it was therefore only used privately. Second, Catalonia received a large immigration wave coming from non-Catalan speaking areas of Spain (Andalusia, Galicia, Extremadura, etc.) mostly from the 1950s to the 1970s. Nowadays this population constitutes almost one fourth of the whole population of Catalonia (23.04%), according to the Institut d'Estadística de Catalunya (2006). Over all these years, the degree to which immigrants have learned Catalan varies considerably. For instance, it is estimated that more than a third of the second generation uses Catalan within the newly created family (Pons & Vila 2005: 71) and the other two thirds use Spanish. In fact, especially in the area of Barcelona, Spanish is used as the main language by many people. This is shown, for instance, by the fact that this area has the lowest percentage of population with Catalan as their first language:<sup>1</sup> 31.91% in 2003, whereas that same year the total percentage in Catalonia was 40.45%.

---

1. The data from the Institut d'Estadística de Catalunya (2006) show three categories: people who have Catalan as their “*first* language (the first language spoken at home)”, as their “*own* language (the language which the interviewee feels identified with)”, and as their “*usual* language (the language mostly used by the interviewee)”. People who answered with both languages, Catalan and Spanish, indistinctly were not included in the countings.

When speaking about the contact of Catalan with Spanish in Catalonia it is relevant to mention the new immigration as well. During the last decade Spain has been receiving large numbers of immigrants coming from Latin America, Maghrib, sub-Saharan African countries, Eastern Europe, etc. In 2005, in Catalonia the percentage of people born abroad was 12.56%. Besides, according to the data reported in the *Pla de ciutadania i immigració 2005–2008* (Generalitat de Catalunya 2006b: 19), the increase in the number of inhabitants in Catalonia is due to foreign immigration in more than 90%. The strongest wave of these newcomers – around 300,000 in 2005 – originates in Latin America, automatically intensifying the presence of Spanish in Barcelona. In the province of Barcelona, these Latin American newcomers, most of them Spanish speakers, correspond to 44% of the foreign population.

However, in this article we will only consider the contact of Catalan with Iberian Spanish, not with that from Latin America. We do so because the longest contact of Catalan has been with the Spanish spoken in mainland Spain. Thus, we will observe Catalan-Spanish bilingual speakers with different degrees of dominance in one or the other language, stemming from Catalan and/or Iberian Spanish speaking families. Specifically, we will consider three age groups differing mainly in the amount of Catalan they have learned at school and in the kind of input they have received in this language during their lives. Observing these three age groups will allow us to look for possible diachronic changes among the Catalan speaking population. Moreover, speakers in each age group are split up into two groups, depending on the district where they live. The two districts have been chosen because they exhibit highly different degrees of language contact with Spanish. What we want to observe within these groups is the possible influence of Spanish on part of the Catalan phonological system, specifically on the vowels /ε/, /ɔ/ and [ə].<sup>2</sup>

The present article is organized as follows. In the remainder of this section, an overview of some studies on bilingualism, which define vulnerable areas in the phonologies of two languages in contact, is provided. Section 2 describes the vocalic system of Catalan and compares it to that of Spanish. This comparison, put together with evidence of vulnerability in the phonological domain leads us to the formulation of some hypotheses as to a plausible influence of Spanish on Catalan, which will be tested in the final part of the paper. The empirical study is presented in Section 3, which is followed by the presentation of results in Section 4. Section 5 discusses these results in the light of our hypotheses, and Section 6 draws some general conclusions.

---

2. The use of [ə] instead of /ə/ is due to the phenomenon of vowel reduction: in unstressed position, the underlying non-back, non-high vowels are realized as schwa, which thus does not have a distinctive value (see Section 2.1.)

## 1.2 Some observations on permeability in the bilingual context

The last decade has witnessed a growing interest in the interaction exhibited by two grammatical systems that develop simultaneously. The phonoprosodic area has been shown to be the most vulnerable one, manifesting influence from one language onto the other at the segmental level as well as at the metrical and prosodic structure (Paradis 2001; Lleó 2002, 2006; Lleó et al. 2003). The vulnerability of the phonoprosodic realm has been explained by internal as well as by external factors. The former relate to markedness and complexity, as well as to the frequency of certain features used in the specification and mental representation of lexical items. The latter include criteria such as language dominance, as well as input frequency of certain sounds and sound patterns. Sociolinguistic criteria like prestige, gender or age may be decisive external factors in the interaction of two systems for adult bilinguals, but hardly significant in child bilingualism.

We will draw on internal criteria, in order to identify those areas that are vulnerable to the influence of the other language, as suggested in Lleó (2006: 369f). Given the various usages that some of the notions classified as internal factors have received in the literature, we will try to define our own usage of those notions, although in a cursory way, given space limitations. We consider *markedness* from a Jakobsonian perspective, by which marked entities presuppose unmarked ones in a typological, diachronic and acquisition sense. Typologically, for a marked entity to be present in a language, the corresponding unmarked entity must necessarily be contained in that same language; this entails that diachronically a language will develop a certain marked entity only if it also contains the corresponding unmarked entity; and in the realm of L1 acquisition, marked entities are acquired later than unmarked ones (Jakobson 1941). Such a notion of markedness has frequency as a consequence: unmarked entities are contained in the languages of the world more often than marked entities. *Complexity* is viewed as a psycholinguistic notion, related to allophony and to allomorphy, i.e. a phoneme comprising more than one phone is a complex category, because the child acquiring the language must figure out not only the different phonetic entities belonging to it, but must also be able to realize them in the right context. Finally, a feature that is used in the characterization of a higher number of entities is *stronger* than a feature with a reduced range of activity. We will see below that these three notions, especially markedness and complexity, are relevant for the study of Catalan vowels.

Besides the interaction that may take place between two phonologies that are acquired simultaneously, it is beyond doubt that learning the phonology of a second language takes place under the influence of the phonology of the first language. The interference of a first into a second language is a fact that has been often observed (Eckman 1987), and that may be interpreted as having fixed parameters

in the first language phonology that prevent or hinder the fixation of such parameters according to the values of L2; or as having established a constraint hierarchy, which is different from the constraint hierarchy of the second language, and which interferes with it. Within Optimality Theory, it has been proposed that learning a second language phonology involves the emergence of the unmarked, especially in those cases in which the superset language is learned as L2, L1 corresponding to the subset (Broselow et al. 1998; Broselow 2004). On the basis of what has been said above in Section 1.1, Catalan has been a second language for many speakers of Spanish, who have learned it at school or often much later, at adult age, given the massive immigration that Catalonia has experienced since the 1950s. As we will see below, the relationship between Spanish and Catalan in the vocalic space is one of subset to superset, i.e., the Catalan vocalic system comprises three more vowels than the Spanish system.

Several phoneticians have noticed that some sounds of Catalan are being modified. Recasens (1991:107), Pla Fulquet (1995), and Herrick (2006) among others have argued that in certain varieties of the Catalan spoken in Barcelona, schwa is being lost, and that the mid vowels are being reduced from four to two. Other changes are taking place, as well, but here we will concentrate on vowels.<sup>3</sup> Research on the simultaneous acquisition of two phonologies has shown that complex and marked phenomena in one language may tend to be permeated by the other language, if this is simpler and less marked (Lleó & Rakow 2005, 2006; Lleó 2006). And something parallel has been reported in the acquisition of a second language phonology: more complex and marked phenomena are learned with difficulty, require more time to be learned, and often are only learned in an incomplete manner (Eckman 1987; Broselow et al. 1998). The changes that have been identified in the Catalan of Barcelona might be due to such internal factors like complexity and markedness, and might be due to the simultaneous acquisition of two phonologies, and/or to the acquisition of a second language phonology with higher degrees of complexity and markedness. However, we do not have clarity as far as how advanced such changes are, and whether there are differences due to age and/or to population groups and city districts.

The present study tries to find out how the situation is in relation to the vowel system in two districts of Barcelona and in three age groups. Although the data reported in this article belong to a larger project that investigates present influences of Spanish onto the phonology of Catalan in general, we will focus on the

---

3. Voiced stridents or sibilants, velarized /l/, palatal /ʎ/, etc. are topics of our on-going research, which we cannot include in this paper. Besides the references mentioned, López del Castillo (1976), Veny (1978), Tuson (1987), Payrató (1988) describe properties that are typical for Barcelona Catalan, some of which are restricted to nonstandard varieties.



segmental area, and within segments, we will only present results on vowels. Consonantal as well as other prosodic areas will be treated elsewhere.

## 2. The vowel system of Catalan as a case of on-going change

### 2.1 The vowel systems of Catalan and Spanish in comparison

Generally speaking, the phonological system of Catalan has more segments than that of Spanish, and is thus more marked as regards the set of segments that it comprises. This is also the case for vowels. Moreover, Catalan has some neutralization processes, which especially in the area of vowels makes it more complex, too. In Catalan, stressed vowels have four degrees of height, mid vowels showing two opening degrees,  $\epsilon/e$  for the front and  $\text{ɔ}/o$  for the back vowels. Unstressed vowels only have two degrees of height, high  $/i/$  and  $/u/$ , and non-high schwa  $[\text{ə}]$ . Figure 1 summarizes Catalan vowels in stressed (1a) and unstressed (1b) position.<sup>4</sup>



Figure 1a. Vowels in stressed position.

Figure 1b. Vowels in unstressed position.

The unstressed vowel inventory results from a general process of neutralization of vowels in unstressed position. The correspondences between stressed and unstressed vowels are shown in Figure 2. Examples are shown in Table 1.

Whereas  $[i]$  and  $[u]$  are maintained in unstressed position, all other vowels change when unstressed, so that many words experience morphophonological variation. The vocalic system of Spanish comprises only the five cardinal vowels,  $/i/$ ,  $/e/$ ,  $/a/$ ,  $/o/$  and  $/u/$ , thus being much simpler than the Catalan one. Spanish also has morphophonological variation, for instance between single vowels and diphthongs (*tener* 'have' vs. *tiene* '(he/she) has', *poder* 'be able' vs. *puede* '(he/she)

4. This paper deals with the Eastern dialect of Barcelona, where schwa  $[\text{ə}]$  only appears in unstressed position, excluding other Catalan varieties, as e.g. the Eastern dialect of the Balearic Islands, where schwa appears both in unstressed and stressed position.

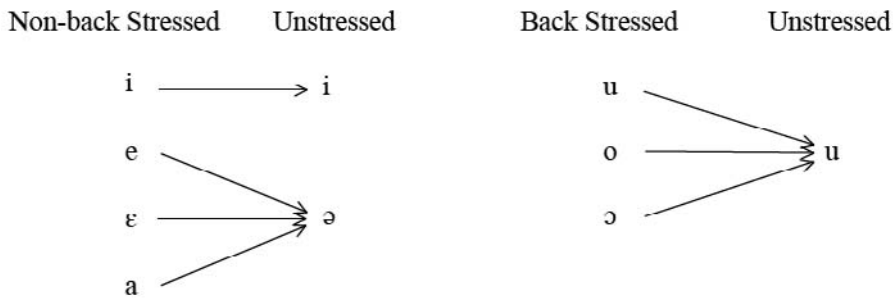


Figure 2. Neutralization of vowels in unstressed position.

Table 1. Correspondences between stressed and unstressed Catalan vowels.

Stressed		Unstressed	
d[i]t	“finger”	d[i]tet	“little finger”
fin[e]stra	“window”	fin[ə]streta	“little window”
p[ε]l	“hair”	p[ə]let	“little hair”
c[a]ma	“leg”	c[ə]meta	“little leg”
[u]ll	“eye”	[u]llet	“little eye”
b[o]ssa	“bag”	b[u]sseta	“small bag”
p[ɔ]ta	“leg”	p[u]teta	“little leg”

can’), which takes place at the lexical level, but there is no vowel reduction process as in Catalan, which applies lexically and postlexically.

According to phonological typology (Maddieson 1984), vocalic systems with two degrees of height within the mid vowels are less frequent than those with just one degree of height: in Maddieson’s survey there are 68 languages with mid vowels that have a single degree of height, but 34 languages with two degrees. As we have indicated above, we relate frequency in the languages of the world to markedness and consider the Catalan system, with two degrees of height in the mid vowels, to be more marked than the Spanish system. However, considering mid vowels individually, the open non-back vowel /ε/ appears more frequently than the closed /e/: according to Maddieson (1984), the former is attested in 116 languages and the latter in 83 languages. The same is true for the back mid vowels: /ɔ/ appears in 100 languages, and /o/ appears in 88 languages.<sup>5</sup> This would entail that the open mid vowels should be considered as less marked than the closed mid ones.

5. However, these numbers may change, once the 246 remaining vocalic systems be included in the countings. That is, there is lack of information about these 246 languages. They have 113 non-back and 133 back mid vowels, which might correspond to either degree of height, open /ε/ and /ɔ/ or closed /e/ and /o/.

Regarding schwa, on the one hand it has been proposed that [ə] is unspecified for place features, and can thus be considered as “unmarked” in the sense of unspecified. In this respect, Van Oostendorp (2000: 156) has proposed Projection Theory, according to which schwa does not have any vocalic features, but must only be specified for [-cons]; this lack of features licenses its occurrence in weak prosodic positions, as the unstressed syllable of a trochee. According to this claim, schwa appearing in a weak prosodic position would be unmarked. On the other hand, if we turn to typology, it happens to be a relatively infrequent vowel in the world’s languages. According to Maddieson (1984), it appears in 73 languages out of the 317 included in his survey, corresponding to 23%. This places schwa among marked vowels. Otherwise, it is unexplainable why being unmarked, schwa does not appear more frequently in the world’s languages. Interestingly, in a study on schwa acquisition in German, Kehoe & Lleó (2003) have found that German children have difficulties in producing this vowel. At first, they usually substitute full vowels for schwas, and it is at about age 2;6 and even later, that they begin producing schwa target-like. At the other end, the vowel /a/ is one of the most preferred vowels in the world’s languages: it appears in 88% of the languages surveyed by Maddieson (1984: 125), it is the first to be produced by children (Jakobson 1941), and the most frequent of the stressed vowels in Catalan (Rafel i Fontanals 1980: 480, 484), as well as the most frequent vowel in Spanish.

According to our definition of marked as typologically less preferred, schwa should be considered a marked vowel in Catalan, as opposed to /a/. Moreover, the process of vowel reduction in the Catalan of Barcelona implies complexity, as the production of [ə] emerges by means of the replacement of some non-back, non-high stressed vowel of a word by [ə] when unstressed (see Figure 2, above). Obviously, such a substitution makes the system complex, since it adds a vowel that otherwise does not have any further usage in the system, as it only appears in unstressed position. This situation has been characterized as allophony or as allomorphy, depending on the theoretical perspective adopted. Some phonologists consider schwa as a mere allophone of the vowels /a/, /ε/ and /e/ in Catalan (see Alarcos 1953), whereas others have attributed phonemic status to schwa (as e.g. Badia Margarit 1965; Viaplana & DeCesaris 1984). In the latter case they have then claimed that the alternation between the stressed vowels and schwa leads to allomorphy. Independently of the solution preferred, from a psycholinguistic point of view, i.e., in acquisition as well as in perception and production, the speaker is confronted with a complex category, which involves alternation of forms.

## 2.2 The permeability of the Catalan vocalic system: some general hypotheses

It has been claimed that in a situation of languages in contact, complex categories belonging to an interface (Müller & Hulk 2001) and marked categories (Lleó 2006: 396) are vulnerable to the influence of the other language, with which contact holds. On the one hand, the Catalan vocalic system, in comparison with the Spanish one, contains additional vowels, exclusively belonging to Catalan, namely, /ɛ/, /ɔ/ and [ə]. On the other hand, these vowels imply markedness, and are further involved in alternations. Given these facts, it can be predicted that Catalan speakers, being in a situation of languages in contact, in which the other language is simpler – all Catalan speakers in Barcelona are bilingual in Catalan and Spanish –, will tend to simplify the system. That is, Catalan words often have two alternating vowels, as we have seen in the previous section: stressed /ɛ/ or /e/ alternating with unstressed [ə], and stressed /ɔ/ or /o/ alternating with unstressed [u]. Moreover, the unstressed position corresponding to stressed /a/, /ɛ/ and /e/ requires a special vowel, [ə], which only appears in unstressed syllables. The vocalic system of Catalan thus exhibits a higher degree of complexity, and under the influence of Spanish it may undergo a simplification process.

Markedness, on the one hand, and complexity, on the other hand, in the form of allomorphy or allophony point to areas in the language that are permeable to the influence of another language in the bilingual context. Nonetheless, the outcome of interaction will not only be governed by such internal factors, i.e., external factors may also play a role. That is to say, the emergent phenomena may be vulnerable to change under the influence of another language that is less complex, as in this case the less complex vocalic system of Spanish in comparison to the Catalan one. However, the completion of the change will be fostered by variables such as language dominance, the degree of incidence of linguistic normalization and other sociolinguistic variables.

Considering the vocalic system of Catalan in comparison to the Spanish one, and assuming that the superset system of Catalan might be exposed to the influence of the subset system of Spanish, two hypotheses can be formulated:

- H1: The four vowels front  $\epsilon/e$  and back  $\mathfrak{o}/o$  will tend to be reduced to just two vowels, a front one and a back one. Whether these will be closer to the vowels /ɛ/ and /ɔ/ or to the vowels /e/ and /o/ will depend on whether markedness plays the main role or whether, due to external factors, the influence of Spanish is most powerful.
- H2: The schwa [ə] will tend to disappear, that is, it will tend to be produced as [ə], given that according to the acquisition data and to typological data, schwa is

a marked vowel, and it hardly has any role to play in the representation of lexical items, as it only appears in unstressed position, whereas the vowel /a/ is one of the most preferred vowels in the world's languages (see Section 2.1).

We intend to find out what influence internal and external factors have in the present variety of the Central Catalan spoken in Barcelona and, from there, we try to predict the near future of this Catalan variety. With that purpose in mind, we study three age groups: a) Children between 3 and 5 years of age (G1), when they have already acquired most of the phonological and prosodic aspects of their language, and are not literate yet, i.e., before their language competence receives the guidance of school. b) Young speakers between 19 and 23 years of age (G2), possibly students; some of whom were born in non-Catalan speaking families and have learned Catalan at school. c) Men and women between 32 and 40 years of age (G3), who are parents of the children in G1, and is thus the group that offers the input to G1; they have already passed adolescence, in which cultural and social pressure of the group is very strong, and have learned to read and write Catalan at school, although not all of them from the beginning of schooling; they were born towards the end of Franco's dictatorship, when Catalan was only starting to re-establish itself at school and in official contexts.

Geographically, this study is limited to the city of Barcelona, where we have selected two districts, Gràcia and Nou Barris, based on the different degrees of presence of the Catalan language, relatively high in Gràcia – a traditional Catalan district –, and rather low in Nou Barris – a district created in the outskirts of the city with the aim to house the immigration wave in 1950s – 1970s. The specific degree of presence of Catalan in each district is based on available data from the Institut d'Estadística de Catalunya (2006), about both origin of the population and level of linguistic competence. According to this source, on the one hand, in Gràcia 83.02% of the population can speak Catalan fluently, whereas in Nou Barris, the percentage is 61.79%. On the other hand, 71.84% of the population from Nou Barris was born in a non-Catalan speaking community and is over 54 years of age, which means that they belong to the Spanish-speaking immigration wave referred to above. Instead, in Gràcia the inhabitants born in a non-Catalan speaking community amount to 35.32%.

### 2.3 Specific hypotheses about on-going phonological change of the vowel system

The two general hypotheses formulated above as H1 and H2 should be further differentiated for the three groups of subjects. They will lead in their turn to the formulation of the following additional hypotheses:

- H3: The G3 speakers in Nou Barris will not exhibit great differences in relation to G3 in Gràcia, because all of them have acquired the language before the years in which the Catalan speaking population massively increased, and they have acquired it in a Catalan sociolinguistic environment, or in a bilingual context with predominance of Catalan. For the speakers of Nou Barris, this implies that they acquired the language before they moved to the district, which was created in the sixties. Some of them might have even been raised in Gràcia. Therefore, the linguistic contexts of the two groups, G3 from Gràcia and G3 from Nou Barris, during infancy and youth were not significantly different.
- H4: If subjects in G2 are assumed to have always lived in the district where they live at present, it is to be expected that G2 from Gràcia and G2 from Nou Barris will build two different constellations, depending on the language of their parents. Those with Spanish speaking parents will show a stronger influence of Spanish onto their Catalan than those whose parents are Catalan speaking. Moreover, it is to be expected that the input they receive is not limited to that at home; this means that G2 in Nou Barris is exposed to more Spanish than G2 in Gràcia, and thus their variety of Catalan will exhibit some of the changes formulated in H1 and H2: reduction of the two degrees of opening in the mid vowels and substitution of /a/ for [ə], that is, G2 from Nou Barris will use less of the “Catalan” vowels, /ɛ/, /ɔ/ and [ə].
- H5: G1 children receive the input of G3, so that, if it is true that there are no differences between both G3 groups, we should not find large differences between G1 in Gràcia and G1 in Nou Barris. But since the sociolinguistic context is different in the two districts, in the same manner that we expected some differences between the two G2 groups, we also expect to find some differences between the two G1s, under the assumption that language development is not only determined by parental input, but also by the language spoken by peers at school and in the street. This makes us expect a greater continuity between G1 and the language of G3 in Gràcia than in Nou Barris; G1’s production in Nou Barris will be getting further apart from the language of their parents, so that they will exhibit some of the changes formulated in H1 and H2: reduction of the two degrees of opening in the mid vowels and substitution of [ə] by /a/, that is, they will use less of the “Catalan” vowels, /ɛ/, /ɔ/ and [ə].

### 3. Methodology

In this section, we present the methodology used in order to obtain the data for this study. It is worth mentioning that the recording sessions included tasks which

aimed at collecting further data not discussed in the present article. Here, we only describe the data with regard to vowel production.

### 3.1 Subjects

The subjects in this study belonged to the three different age groups introduced above:

- G1: children aged between 3 and 5
- G2: young adults aged between 19 and 23
- G3: adults aged between 32 and 40

Data were collected from speakers in all three age groups in two different districts: Gràcia and Nou Barris.

The speech of ten children in each district was recorded: 10 from Gràcia (8 girls and 2 boys, with a mean age of 3;10) and 10 from Nou Barris (5 girls and 5 boys, with a mean age of 4;6). The children in Gràcia attended the pre-school section in the state schools CEIP Patronat Domènech and CEIP Josep Maria Jujol. The boys and girls at Nou Barris attended the state school CEIP Gaudí and Sant Lluís, a private school. The young adults in G2 were met at sportclubs, public libraries, and youth associations in both districts: 10 from Gràcia (5 male and 5 female, with a mean age of 21) and 10 in Nou Barris (4 male and 6 female, with a mean age of 20;8). The group of adults comprised either the father or the mother of the children in G1: 10 in Gràcia (9 female and 1 male, with a mean age of 36;4) and 10 in Nou Barris (6 female and 4 male, with a mean age of 37;1).<sup>6</sup> The subjects in all three groups belong to the middle and lower-middle class. Independently of the language of their parents and level of proficiency in Catalan, they all use Catalan in their everyday life; all of them grew up in Barcelona (except for one G3 speaker from Nou Barris, who lived in Manresa for several years) and thus Central Catalan is the variety spoken by all our subjects. The following table specifies age and language spoken by both parents for each subject in each of the six groups.<sup>7</sup>

---

6. G2 and G3 include two subjects each who happen to be between one and four years beyond the age range.

7. Our population sample shows a certain bias towards female speakers. In G3, this was due to the fact that children's mothers had a greater availability than fathers to take part in the test. In G1, it is just accidental, as we selected children on the basis of the teachers' information about their linguistic background.

Table 2. Age and parents' language for each subject.

G1	Gràcia				Nou Barris			
	Subject	Age	Mother	Father	Subject	Age	Mother	Father
	APR	3	Ct	Ct	LVG	4	Ct	Ct
	EMT	3	Ct	Ct	CRS	5	Ct	Sp
	JPC	3	Ct	Ct	MJG	5	Ct	Sp
	JBC	5	Ct	Ct	MTG	5	Ct	Sp
	MBB	4	Ct	Ct	ACM	4	Sp	Ct
	GGP	4	Ct	Ct	ICC	5	Sp	Sp
	MTC	4	Ct	Sp	ACG	5	Sp	Sp
	CPL	5	Sp	Ct	CRG	5	Sp	Sp
	UPW	4	Sp	Ct	JCL	3	Sp	Sp
	MSD	4	Sp	Ct	DGS	4	Sp	Sp
G2	Gràcia				Nou Barris			
	Subject	Age	Mother	Father	Subject	Age	Mother	Father
	AGQ	22	Ct	Ct	ARJ	23	Ct	Ct
	LDB	20	Ct	Ct	AMC	19	Ct	Ct
	JFC	20	Ct	Ct	APC	19	Ct	Ct
	JAD	18	Ct	Ct	ALM	22	Ct	Ct
	DCA	23	Ct	Ct	NGE	23	Sp	Ct
	MRV	19	Ct	Ct	LSA	20	Sp	Sp
	MUV	20	Ct	Ct	IRE	19	Sp	Sp
	MBB	23	Ct	Ct	RHL	22	Sp	Sp
	ELP	22	Ct	Sp	ACM	18	Sp	Sp
	DGT	23	Sp	Sp	GFD	23	Sp	Sp
G3	Gràcia				Nou Barris			
	Subject	Age	Mother	Father	Subject	Age	Mother	Father
	MTF	33	Ct	Ct	ISB	37	Ct	Ct
	ECP	39	Ct	Ct	TGM	37	Ct	Ct
	LBR	33	Ct	Ct	MGT	39	Ct	Ct
	MCG	39	Ct	Ct	NGC	34	Ct	Ct
	IBL	38	Sp	Ct	PGB	33	Ct	Sp
	SRD	35	Sp	Sp	JCN	40	Sp	Sp
	PLC	41	Sp	Sp	CGB	33	Sp	Sp
	SWR	31	Sp	Sp	IGT	39	Sp	Sp
	EPR	39	Sp	Sp	CCR	34	Sp	Sp
	CDQ	36	Sp	Sp	FCP	45	Sp	Sp



### 3.2 Materials, procedure and equipment

In order to carry out this study, some specific materials were designed with the purpose of having a number of words containing each target vowel. Since one group of subjects, namely G1, included very young speakers (the youngest were 3 years old), the materials were elaborated keeping their abilities in mind. The two tasks used for the data elicitation were: picture naming, and answering simple questions asked by the interviewer. The same materials were used in the sessions with all subjects in order to keep the answers as similar as possible across groups. A few questions or concepts were only asked to the adult groups due to their abstractness or because the children were unlikely to know the answers.

The pictures to be named and described by the subjects were illustrations in books for children to learn basic vocabulary. The printed words on the books were hidden behind dark adhesive tape so that the written form played no role on the pronunciation of the target words, especially for the older children, who were already learning to read, and for the adults. Apart from the pictures in these books, some specific notecards with pictures found on the Internet were designed to meet our specific purposes. For example, the picture of the sun was shown in order to elicit the target word *sol* [sɒl], with a target open mid back vowel (see Appendix for the whole list of target words).

The recordings were done in several different places. The sessions with children took place in the school they attended, so that they were in a familiar environment, which they associated with the usage of Catalan. As for the children's parents (G3), the sessions were recorded either at school or at their own flat, depending on what was more convenient to them. The young adults (G2) were, in turn, recorded in study rooms at libraries in their district, in a room at youth association centers, or at their flat.

Two of the authors of this article carried out the sessions with the children. The sessions with adults were run either by both or by only one of them. The average length of these recording sessions was between 45 and 60 minutes with the children, and about 30 minutes with the adults. Sessions included not only the elicitation of the specific target words analyzed in this study, but also a casual conversation, which was carried out because subjects monitor the way they speak to a lesser extent in the spontaneous speech style. The casual conversation took place at the very beginning of the session, and subjects were prompted to talk about life in their district. In this way, their attention was focused on a sociological rather than linguistic topic.

The equipment used in the recording sessions was a Sony ECM-CS10 unidirectional lapel microphone, connected to a portable Mini-Disc Hi-MD Walkman MZ-RH10 Sony recorder.

### 3.3 Analysis

The data was auditorily analyzed by native Catalan speakers. Each and every item in the corpus was listened to and transcribed by one of the authors of this study and by another native Catalan speaker. Transcribers' agreement exhibited a reliability index of 80.62% for [ə], 80.52% for /ɔ/ and 78.08% for /ɛ/. After all productions had been transcribed, they were divided into target-like and non-target-like items. Target-like productions were those containing the vowels /ɛ/, /ɔ/ and [ə], whereas in non-target like productions these vowels had been replaced by /o/, /e/ and /a/, respectively. Subsequently, these productions were turned into percentages due to the variable number of productions by each speaker.

## 4. Results

The binary division of the transcriptions into target-like and non-target like productions has turned them into categorical data (i.e., X and no X), allowing for a nominal treatment. Therefore,  $\chi^2$  tests were run on this data. When reported, the  $\chi^2$  tests indicate the degrees of freedom followed by N – i.e., the sample size – in brackets.

### 4.1 Global results by age groups across districts

When we compare the production of each vowel by age groups in the two districts, a pattern is easily observable in the sense that both G1 and G2 show a statistically significant difference in production depending on where the speakers live. The speakers in Gràcia produce all three target vowels more often than the corresponding age groups in Nou Barris. However, the production of G3 in both Gràcia and Nou Barris displays no significant difference across districts. Significant differences between groups of the same age in different districts are marked with a star above the two bars for that age group in Figures 3, 4 and 5.

Another observable trend is that, in Gràcia, G2 is always the group with the highest target-like production followed by G1 and then G3. Only in the case of [ə]

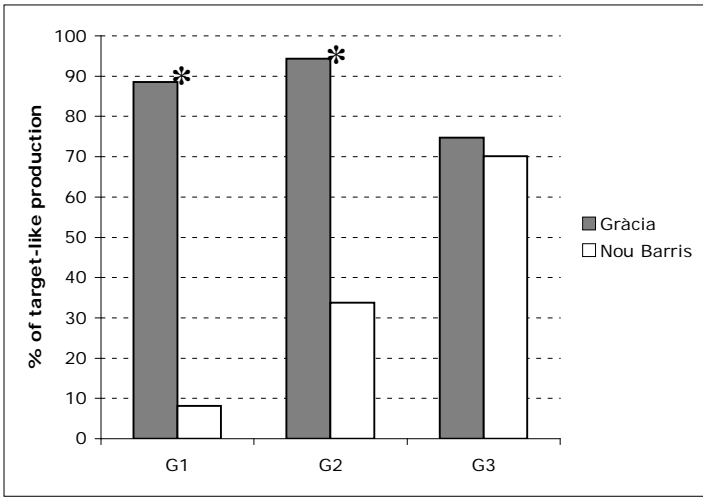


Figure 3. Percentages of /ε/ production.

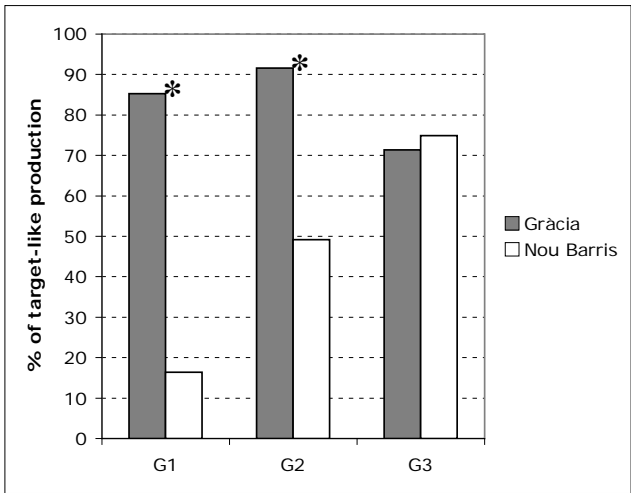


Figure 4. Percentages of /ɔ/ production.

is G1’s production lower than G3’s. In turn, the older the age group in Nou Barris is, the more target-like their production of the Catalan vowels is.

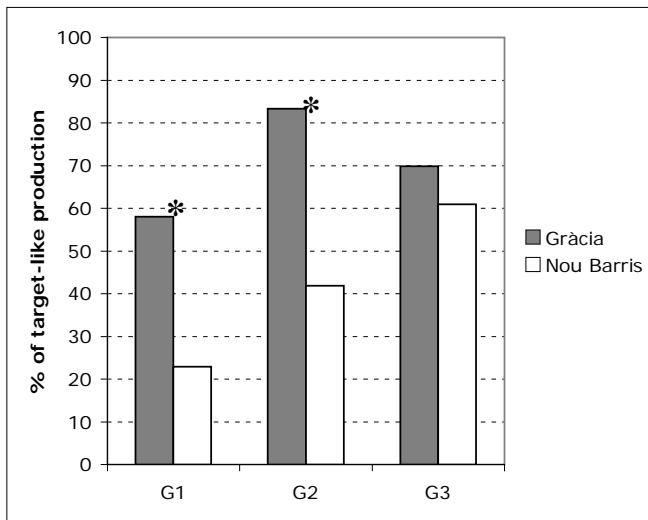


Figure 5. Percentages of [ə] production.

#### 4.2 G1 vs. G3: differences between productions by children and by their parents

The following graphs display the percentages of target-like production of / $\epsilon$ / in Gràcia and Nou Barris by G1 and G3, i.e., by the children and by their parents. Only the graphs showing the percentages of target-like production for / $\epsilon$ / are included because the trend is very similar across vowels, as regards the variability in production depending on age group and on the language(s) spoken in the family. In Figures 6 and 7, the language of the child's parents is coded under the x axis. 'CT' stands for 'Catalan' and 'SP' for 'Spanish'. From the two codes under each pair of bars in the figures, the code under the gray column labels the language spoken by the child's mother whereas the code under the white bar stands for the language spoken by the child's father. If we analyze these figures, we do not see a continuous gradation depending on the language spoken by the children's parents. If there was a correlation between the children's production and the language spoken by their parents, we should observe that the bars representing children's target-like production (in light grey) would decrease from left to right, as children from Catalan parents are placed on the left, and those with Spanish-speaking parents appear towards the right hand-side. However, percentages do not decrease from left to right.<sup>8</sup>

8. The gaps in the figures indicate that there was no target-like production in the items elicited from a given speaker.

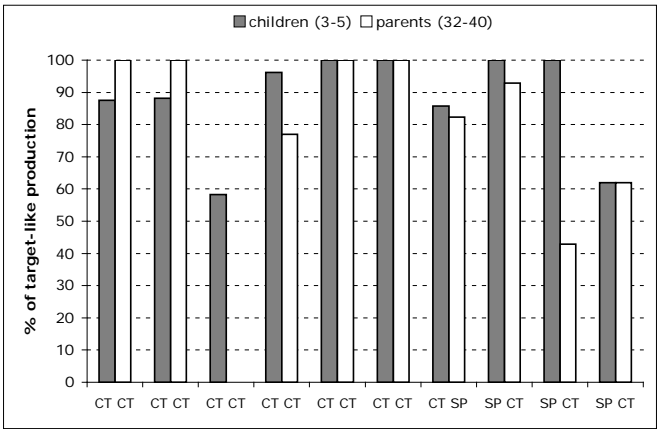


Figure 6. Percentages of /ε/ production in Gràcia by G1 and G3.

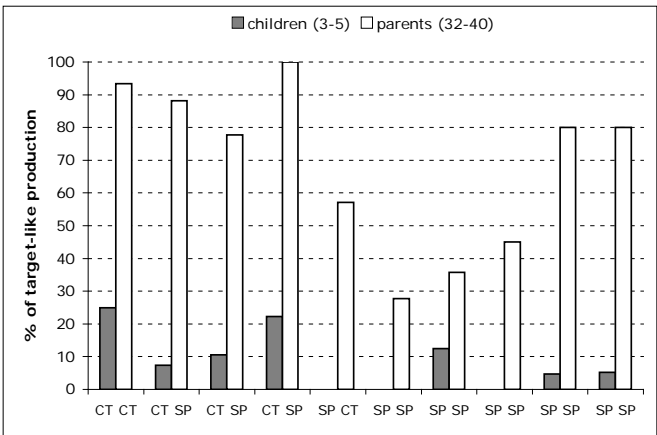


Figure 7. Percentages of /ε/ production in Nou Barris by G1 and G3.

In Nou Barris, the parents (G3) produce the three target vowels more often than their children, whereas the children in Gràcia produce two of the target vowels, /ε/ and /ɔ/, more often than their parents. However, the children in Gràcia’s G1 do not produce target [ə] more often than their parents. That is, although the values for [ə] production by the children in Gràcia are clearly higher than those for the children in Nou Barris, they do not reach the frequency they have in their parents’ productions. Such a behavior could be due to the special status of [ə], in comparison with the other vowels in the language, as schwa has no other usage than as a reduced vowel in unstressed position. At any rate, a comparison of the results of G1 and G3 in Gràcia shows that in the youngest generation, whereas the

open mid vowels are kept, [ə] seems to be drifting to /a/. The difference between the production by G1 and G3 reaches significance in all cases, with G1's target-like production being higher than G3's for /ɛ/ and /ɔ/ in Gràcia and with G1's target-like production being lower than G3's in the rest (/ɛ/ in Gràcia:  $\chi^2(1, N = 20) = 10.68$ ;  $p \leq .01$ ; /ɛ/ in Nou Barris:  $\chi^2(1, N = 20) = 147.50$ ;  $p \leq .001$ ; /ɔ/ in Gràcia:  $\chi^2(1, N = 20) = 11.62$ ;  $p \leq .001$ ; /ɔ/ in Nou Barris:  $\chi^2(1, N = 20) = 145.06$ ;  $p \leq .001$ ; [ə] in Gràcia:  $\chi^2(1, N = 20) = 9.76$ ;  $p \leq .01$ ; [ə] in Nou Barris:  $\chi^2(1, N = 20) = 91.12$ ;  $p \leq .001$ ).

#### 4.3 G2 vs. G3: comparison between the young and the older generation of adults

When comparing the production of the young adults and the older adults in Figures 3, 4 and 5, age correlates positively with percentage of target-like production in Nou Barris, but negatively in Gràcia. That is, the older the group, the higher the target-like production is in Nou Barris. But the older the group, the lower the target-like production is in Gràcia, except for [ə]. The differences in vowel production between G2 and G3 in the same district is statistically significant in all cases (/ɛ/ in Gràcia:  $\chi^2(1, N = 20) = 27.33$ ;  $p \leq .001$ ; /ɛ/ in Nou Barris:  $\chi^2(1, N = 20) = 45.89$ ;  $p \leq .001$ ; /ɔ/ in Gràcia:  $\chi^2(1, N = 20) = 26.37$ ;  $p \leq .001$ ; /ɔ/ in Nou Barris:  $\chi^2(1, N = 20) = 49.06$ ;  $p \leq .001$ ; [ə] in Gràcia:  $\chi^2(1, N = 20) = 15.20$ ;  $p \leq .001$ ; [ə] in Nou Barris:  $\chi^2(1, N = 20) = 21.71$ ;  $p \leq .001$ ). Thus, G2's vowel production is more target-like than G3's in Gràcia, whereas it is the other way round in Nou Barris.

The difference between the young adults in Gràcia and Nou Barris is statistically significant across vowels (/ɛ/:  $\chi^2(1, N = 20) = 147$ ;  $p \leq .001$ ; /ɔ/:  $\chi^2(1, N = 20) = 83.56$ ;  $p \leq .001$ ; [ə]:  $\chi^2(1, N = 20) = 110.62$ ;  $p \leq .001$ ), whereas the vowel production by the older adults in both districts is quite homogeneous, as already mentioned in Section 4.1.

#### 4.4 G1 vs. G2: the two younger generations in Barcelona

As shown in Figures 3, 4 and 5, the difference between the production of /ɛ/, /ɔ/ and [ə] across districts always reaches significance both in G1 and in G2 (/ɛ/ in Gràcia:  $\chi^2(1, N = 20) = 4.09$ ;  $p \leq .05$ ; /ɛ/ in Nou Barris:  $\chi^2(1, N = 20) = 34.86$ ;  $p \leq .001$ ; /ɔ/ in Gràcia:  $\chi^2(1, N = 20) = 4.09$ ;  $p \leq .05$ ; /ɔ/ in Nou Barris:  $\chi^2(1, N = 20) = 49.06$ ;  $p \leq .001$ ; [ə] in Gràcia:  $\chi^2(1, N = 20) = 50.41$ ;  $p \leq .001$ ; [ə] in Nou Barris:  $\chi^2(1, N = 20) = 24.59$ ;  $p \leq .001$ ). In these two age groups, production by speakers in Gràcia is more target-like than that by speakers in Nou Barris.

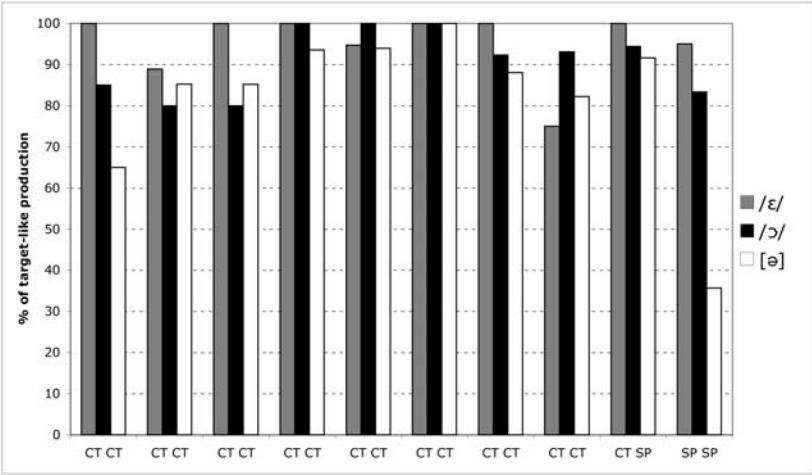


Figure 8. Production percentages of /ε/, /ɔ/ and [ə] by G2 in Gràcia.

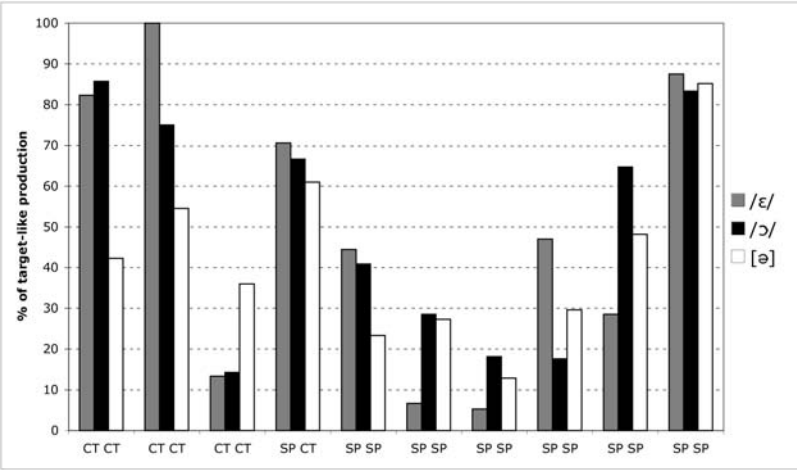


Figure 9. Production percentages of /ε/, /ɔ/ and [ə] by G2 in Nou Barris.

As for the role played by language in the family of G2 speakers, Figures 8 and 9 show that, as it was the case for G1 speakers, the language spoken by parents does not determine the percentages of target-like production of the three vowels. The bars do not decrease as one goes from left to right in the graph, which would show an effect of language in the family because subjects with Catalan-speaking parents are to the left whereas those with Spanish-speaking parents are to the right. The difference in production variance within districts should also be pointed out. Production by G2 is much more homogeneous in Gràcia than in Nou Barris.

## 5. Discussion

The results obtained partly confirm the hypotheses of this study, and allow for some predictions about possible trends in vowel production in the Catalan of Barcelona. Hypotheses H1 and H2 just foresee in a general way that open mid vowels may be replaced by closed mid vowels, and that [ə] may be replaced by /a/. This is generally the case, i.e., many instances of such substitutions have been found in the data. However, these hypotheses are too general, and they do not really show in a differentiated manner what the real linguistic situation is across districts and across generations. This has been further specified by means of three more hypotheses.

H3 states that there will be no differences between G3 in Gràcia and G3 in Nou Barris. This hypothesis has been confirmed, as no statistically significant differences in the percentages of /ɛ/, /ɔ/ and [ə] have been found between the subjects of G3 across districts. However, looking at the diagrams in some detail, we can observe three phenomena. Firstly, values hardly ever reach 100%, neither in Gràcia nor in Nou Barris, but fluctuate between 40% and 80%, reaching a mean of 72% in Gràcia and 67% in Nou Barris. This simply shows that G3, the parents of G1, do not offer a model of language equivalent to the purported standard language, in the sense that G3 already exhibits some loss of mid vowels and schwas. Secondly, production values for the three vowels in G3 are very similar in both districts, with a joint mean of 72% for /ɛ/ and /ɔ/ and a joint mean of 64% for schwa. This finding entails that the distance between the standard and the variety offered by G3 is equivalent for all three vowels, albeit slightly larger in the case of [ə]. Thirdly, although the statistical analysis does not show a statistically significant difference between the two G3s across districts, and the mean percentages are very similar in the two districts, namely 72% in Gràcia and 67% in Nou Barris, this latter district exhibits some very low individual values: one case of 28% for /ɛ/ as well as for /ɔ/, and one case of 9% for [ə]. It can thus be claimed that the loss of mid vowels and schwas is more advanced in G3 in Nou Barris, given some very low individual values.

H4 states that G2 will manifest differences in the vowel system across districts, that is, that the replacement of open mid vowels and schwas will be more advanced in Nou Barris than in Gràcia. This hypothesis is certainly confirmed, as we have seen that the G2 percentages of target-like mid vowels and schwas are significantly higher in Gràcia than in Nou Barris. This development clearly points to an influence of the environment onto the variety of language used by the subjects. That is, in Figures 8 and 9 we have seen that the fact that the language of the parents is Catalan (left hand-side of the diagrams) or Spanish (right hand-side) does not really correlate with the results of G2, neither in Gràcia nor in Nou



Barris, as the values in these figures do not decrease from left to right. What seems to indeed exert an influence on speakers' production are the language dominance relations in each district because values in Nou Barris are much lower than in Gràcia as a whole, and this correlates with the strong presence of Spanish in Nou Barris and a clear dominance of Catalan in Gràcia.

One unexpected finding is that for all three vowels, the results of G2 in Gràcia are always higher than those of G3, so that the reduction of target-like vowels that is being experienced in Nou Barris is not only compensated in Gràcia, but reversed. As it will be remembered from the Introduction, Catalan speakers of G3 were born towards the end of Franco's dictatorship and were schooled mostly in Spanish. It was after they had attended school that the language experienced a broad process of normalization, by which the language was amply used outside the family, in the government, at school, in the media, in the street, in shops, etc. It seems as if through this process of normalization the language might be maintaining some of its original phonological properties. However, this is only taking place in Gràcia, whereas in Nou Barris the language has experienced a reduction of such original phonological properties. This implies that in order to experience a positive effect through normalization, the language must reach a certain threshold of presence and usage, and this is reached in Gràcia, but not in Nou Barris.

H5 states that G1 will be different across districts, in a parallel manner as it was hypothesized for G2 by H4. In spite of the similarities between the two groups of parents (G3) across districts, G1 in Gràcia and G1 in Nou Barris exhibit statistically significant differences, which point to an on-going change in the vowel system of one of the groups of G1, namely that of Nou Barris. These findings confirm H5. In fact, and in a parallel way as we saw for G2, this finding shows that parental input cannot be the source of variation found in the two districts, as G3 is very similar in both districts, and nevertheless an on-going change is emerging in Nou Barris. It looks as if the parental input is not as important as the language of the district in general; that is, children receive the main influence from the environment where they socialize: from the day-care center, and from the school of the district. This might allow us to explain why children's productions in Gràcia maintain both degrees of opening and schwa more often than children's productions in Nou Barris, which exhibit more influence from Spanish, in spite of similar parental input in the two districts. Certainly, these results referring to G1 and similar results for G2 presented above (H4) are based on a limited sample, and will have to be verified on a larger population, in which equal numbers of subjects for each parental language condition are examined.

Moreover, in Gràcia, we find higher values for G1 than for their parents. This also points to the influence that the environment is exerting on the children's

language, beside the family; certainly school seems to play an important role in the maintenance and possible recovery of features that had experienced some reduction in G3.

If we now turn to the basic hypotheses, H1 and H2, it is clear that the four vowels /ɛ/, /e/, /ɔ/ and /o/ are being reduced to two, /e/ and /o/, and that schwa [ə] tends to disappear in one of the two districts. When stating H1 we said that because the mid vowels /ɛ/ and /ɔ/ seem to be typologically less marked, the drift of the mid vowels towards the open ones would show that internal factors are the most powerful, whereas the substitution in favor of the mid vowels /e/ and /o/ would be due to external factors, i.e., to the influence of Spanish. Our data confirm the latter case, as the vowels /e/ and /o/ are the ones that prevail. As regards H2, /a/ substituting for [ə], the results do not allow us to disentangle whether internal or external factors take the lead. That is, according both to internal factors like complexity and markedness and to external factors, in this case influence of Spanish, /a/ should be preferred, as it is a universally preferred vowel, and it is also part of the Spanish inventory, in stressed as well as in unstressed position.

However, the reduction of the vowel system is only taking place in one part of the Catalan-speaking population, i.e., in those groups of G2 and G1 that are more exposed to Spanish, because they live in a district with a large Spanish speaking population. It is possible that in a district like Nou Barris this evolution is progressing, as a decrease has been found from the older to the younger,  $G3 > G2 > G1$ . That is, there we have found a negative correlation between age and withdrawal of Catalan phonological traits: The older the subjects, the more target-like their production is, and the younger the subjects, the greater the distance to the target is. However, in Gràcia, G2 is the group with the highest target-like results for all vowels. Since youth is exposed to social factors like school, prestige, group pressure, etc., in Gràcia these factors exert a positive influence, probably due to the fact that the presence of Catalan reaches a necessary threshold there, whereas in Nou Barris they exert a negative influence, due to the strong presence of the Spanish language. G1 is not yet under the influence of social pressures related to prestige, which are very important for G2. This entails that when G1 in each district reaches the age of G2, in Gràcia they will probably keep a variety of language similar to the one G2 has now, reaching higher values than those in today's G3, but in Nou Barris the future G2 will possibly show lower values than those showed by G2 nowadays. This difference in the evolution of both G1 groups is predicted in view of the fact that the distance between G1 and G2 today is minimal in Gràcia, but rather large in Nou Barris; this is especially the case for /ɛ/ and /ɔ/.

Several studies mentioned above, like López del Castillo (1976), Tuson (1987) or Payrató (1988) have already observed some of the properties of the vowels of the Catalan of Barcelona, like those here investigated, and have attributed them

to sub-standard varieties, whereas more recent studies, like Recasens (1991), Pla Fulquet (1995) and Herrick (2006), assign these properties to the speech of Barcelona in general. Our results, which, as Veny (1978:46) would say, are based on “horizontal variation across districts [...] and vertical variation across generation levels”,<sup>9</sup> allow for more precise claims about the extension of such properties, geographical extension – across districts – as well as chronological extension – across generations. As regards differences of social level, our data cannot make more precise differentiations because, as we have mentioned in Section 3.1, our subjects belong to the middle or low-middle class. As for register, results are based on semi-formal conversations, and being on linguistic phenomena, the observer’s paradox cannot be assumed to be totally avoided. Nevertheless, we tried to exclude any type of self-correction and any tendencies to formalism on the part of the subjects, by means of focusing attention on social questions, like life in the district, and not making the linguistic aim of the study transparent. In any event, these problems only refer to G2 and G3, but do not relate to the G1 children, who are not yet exposed to pressures like linguistic correction or formality, that may influence the results within the older groups.

Turning once more to the question about the influence of internal and external factors in the present situation of Catalan, it should be recalled (see Section 1.2) that research both on simultaneous acquisition of two phonologies and on acquisition of a second language phonology has shown that complex and marked phenomena in one language (Catalan, in this case) may tend to be permeated by the other language (Spanish). From our results, we can observe that the effects of these internal factors of markedness and complexity of the Catalan phonological system play some role in those groups which are (and/or have been) more exposed to Spanish, i.e., G1 and G2 in Nou Barris. Instead, those groups whose speakers live (or have lived) in more Catalan speaking environments, especially G1 and G2 in Gràcia, maintain the Catalan vowel system, in spite of complexity and markedness. In Nou Barris, the intensive presence of Spanish in the district with its concomitant peer pressure leads G2, in particular, to the reduction of the vowel system under the influence of Spanish. It is important to keep in mind that, independently of the language spoken at home, G2 has learned Catalan at school. However, the positive effect of school on a more target-like production is hardly noticeable in G2 in Nou Barris. To sum up, we can say that markedness and complexity make Catalan phonology more permeable to the Spanish one, alongside exposure to this language. But if Catalan is the main language in the environment, these internal factors do not seem to exert any noticeable influence even at early

---

9. Our translation.

ages, as our results from G1 in Gràcia – far more target-like than those from G1 in Nou Barris – show.

Finally, it is important to note that the results reported here are based on auditory analysis, i.e., transcriptions, and that we have not carried out any acoustic analyses yet. Carrying out auditory analyses turns the data into categorical data, as mentioned in Section 4, because decisions have to be binary: a vowel is [ə] or [a], a vowel is [ɛ] or [e], a vowel is [ɔ] or [o]. We cannot deny that in more than one occasion the decision has not been easy to reach because our procedure did not allow for intermediate solutions.<sup>10</sup> That is, although we have been able to say how many times we have heard open mid vowels and schwa, we are not able to say whether there is a more or less generalized tendency to close mid vowels or to open schwas, as some phonologists have reported (Herrick 2006; Pla Fulquet 1995; Recasens 1991). This type of conclusion will require to submit our data to acoustic analysis, which we want to do in the near future.

## 6. Conclusions

The present study has shown that the Catalan spoken in Barcelona is partly merging the four mid vowels (/ɛ/, /e/, /ɔ/ and /o/) into two (/e/ and /o/), and replacing schwa by /a/. Such a trend, which is attributed to the influence of Spanish, exhibits different degrees, depending on age and on place of residence. The study has differentiated three generational groups, children (G1), young adults (G2) and adults (G3), and two districts, Nou Barris with much presence of Spanish, and Gràcia with much less presence of this language. G1 in Nou Barris is the group that shows the lowest percentages of target-like vowels, indicating that the on-going change is taking place there most forcefully. The percentages of G2 in Nou Barris, although slightly higher, are also low, making clear that the loss of the target vowels /ɛ/, /ɔ/ and [ə] is taking place in that district. At the other end, G2 in Gràcia is the group that maintains the target vowels to the greatest extent, directly followed by G1. In Gràcia, both G2 and G1 have higher values than G3, which implies that the loss of the three vowels does not seem to be taking place there. As far as adults between 32 and 40 years (G3) are concerned, they have very similar

---

10. The decisions were grounded on the intuition of the transcribers, all of them native speakers of Catalan. Using diacritics, as suggested by one of the reviewers, does not make the decisions easier or better motivated, as they are still categorical, but there are then more decision-alternatives to choose from. That is, the decision must not only be taken between /ɛ/ and /e/, but e.g. between /ɛ/, /ɛ̃/, /ɛ̂/ and /e/, which would further pose the problem of deciding where the dividing line between /ɛ/ and /e/ should be drawn.

values in the two districts, with a mean percentage of maintenance of target vowels of 72% in Gràcia and 67% in Nou Barris.

Phonological change is thus mainly taking place within the groups of the Catalan speaking population that has been and still is most exposed to language contact with Spanish from birth, that is, G1 and G2 in Nou Barris. There, the relevant factors for change are internal as well as external to the linguistic system: greater complexity and a higher degree of markedness of the Catalan vowels, on the one hand, and an intense presence of Spanish in the environment, on the other hand. Obviously, both types of factors reinforce each other. As regards G1 and G2 in Gràcia, the fact that they are able to maintain the three Catalan vowels in spite of complexity and markedness shows that external factors outrank internal factors here, as school and peer group pressure contribute to this outcome. In none of those groups is there a correlation between their language variety and the language of their parents, which is especially clear in the case of G1 compared with G3, as the subjects of the latter group are the parents of G1, and thus offer the input to G1. We attribute the high percentages of standard vowels in G1 and G2 in Gràcia to the influence of the environment where the children of G1 and the young adults of G2 socialize, i.e., school and peers rather than to the parental input they receive at home.

The most important finding in our data is the large difference found both in G1 and G2 across the two districts investigated. Whereas in Nou Barris both children and young adults have been permeable to the influence of Spanish, in Gràcia they even outdo the results of G3, which are the parents of G1, as for the maintenance of the three target vowels of Catalan. This result highlights how crucial external factors are for the fate of the language: although all three vowels involve complexity and markedness, they are either maintained or lost, depending on the interaction of languages in the environment. In Gràcia, it is clear that normalization and the influence of school has led to vowel maintenance, whereas in Nou Barris, the impact of Spanish is pushing the language into a mixed variety, with a phonology that is influenced by Spanish to a large extent, at least as far as the vowel system is concerned.

## References

- Alarcos Llorach, Emilio. 1953. Sistema fonemático del catalán. *Archivum III*: 135–146. Oviedo.
- Badia Margarit, Antoni M. 1965. Función significativa y diferencial de la vocal neutra en el catalán de Barcelona. *Revista de Filología Española* 48: 79–93.
- Broselow, Ellen. 2004. Unmarked structures and emerging rankings in second language phonology. *International Journal of Bilingualism* 8: 51–65.

- Broselow, Ellen, Chen, Su-I & Wang, Chilin. 1998. The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20: 261–280.
- Eckman, Fred R. 1987. Markedness and the contrastive analysis hypothesis. In *Interlanguage Phonology. The Acquisition of a Second Language Sound System*, G. Ioup & S. H. Weinberger (eds), 55–69. Cambridge MA: Newbury House.
- Generalitat de Catalunya. 2006a. La llengua catalana. Retrieved December 13, 2006, from <http://www.gencat.net/catalunya/cat/llengua.htm>
- Generalitat de Catalunya. 2006b. *Pla de ciutadania i immigració 2005–2008*. Retrieved July 10, 2006, from Generalitat de Catalunya, Secretaria per a la Immigració. Access: <http://www.gencat.net/benestar/societat/convivencia/immigracio/pla>
- Herrick, Dylan. 2006. Mid vowels and schwa in Eastern Catalan: five non-Barcelona dialects. In *New Perspectives on Romance Linguistics. Vol. II: Phonetics, Phonology and Dialectology*, I.-P. Y. Montreuil (ed.), 113–126. Amsterdam: John Benjamins.
- Institut d'Estadística de Catalunya (IDESCAT). 2006. Societat: llengua. Retrieved December 13, 2006, from <http://www.idescat.net/cat/societat/soclleng.html>
- Jakobson, Roman. 1941. *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell.
- Kehoe, Margaret & Lleó, Conxita. 2003. A Phonological analysis of schwa in German first language acquisition. *Canadian Journal of Linguistics* 48: 289–327.
- Lleó, Conxita. 2002. The role of markedness in the acquisition of complex prosodic structures by German-Spanish bilinguals. *International Journal of Bilingualism* 6(3): 291–313.
- Lleó, Conxita. 2006. Fenòmens evolutius i d'ús oral del català des de la finestra psicolingüística de l'adquisició bilingüe. *Actes del Tretzè Col·loqui Internacional de Llengua i Literatura Catalanes*, 361–385.
- Lleó, Conxita, Kuchenbrandt, Imme, Kehoe, Margaret & Trujillo, Cristina. 2003. Syllable final consonants in Spanish and German monolingual and bilingual acquisition. In *(In)vulnerable Domains in Multilingualism*, N. Müller (ed.), 191–220. Amsterdam: John Benjamins.
- Lleó, Conxita & Rakow, Martin. 2005. Markedness effects in voiced stop spirantization in bilingual German-Spanish children. *Proceedings of the 4th International Symposium on Bilingualism*, 1353–1371. CD-Rom. Somerville MA: Cascadilla Press.
- Lleó, Conxita & Rakow, Martin. 2006. Nasalassimilation und prosodische Hierarchie im monolingualen und bilingualen Erwerb des Spanischen und des Deutschen. In *Phonetik und Nordistik. Festschrift für Magnús Pétursson zum 65. Geburtstag*, C. El Mogharbel & K. Himstedt (eds), 95–117. Frankfurt: Wissenschaftliche Buchhandlung Theo Hector.
- López del Castillo, Lluís. 1976. *Llengua standard i nivells de llenguatge*. Barcelona: editorial Laia.
- Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge: CUP.
- Müller, Natascha & Hulk, Aafke C.J. 2001. Cross-linguistic influence in bilingual first language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition* 4 (1): 1–21.
- Paradis, Johanne. 2001. Do bilingual two-year-olds separate phonological systems? In *International Journal of Bilingualism* 5(1): 19–38.
- Payrató, Lluís. 1988. *Català col·loquial. Aspectes de l'ús corrent de la llengua catalana*. Universitat de València.
- Pla Fulquet, Joan. 1995. L'obertura de [ə] a Barcelona: el xava i altres varietats. In *La Sociolingüística de la Variació*, M.T. Turell (ed.), 139–162. Barcelona: PPU.

Pons, Eva & Vila, Francesc X. 2005. *Informe sobre la situació de la llengua catalana (2003–2004)*. Retrieved July 10, 2006, from Observatori de la Llengua Catalana. Access: <http://www.observatoridelallengua.cat>

Rafel i Fontanals, Joaquim. 1980. Dades sobre la freqüència de les unitats fonològiques en català. *Estudis Universitaris catalans* Vol. 25: 473–406.

Recasens, Daniel. 1991. *Fonètica descriptiva del català*. Barcelona: Institut d'Estudis Catalans.

Tuson, Ampar. 1987. El repertori lingüístic de la ciutat de Barcelona. In *La formazione dell'insegnante di lingue in ambiente di lingue in contatto: atti del colloquio italo-catalano 25–27 maggio 1985*, D. Hart (ed.), 63–85. Università di Roma: Bagato.

Van Oostendorp, Marc. 2000. *Phonological Projection: A Theory of Feature Content and Prosodic Structure*. Berlin: Mouton-De Gruyter.

Veny, Joan. 1978. *Els parlars catalans*. Palma de Mallorca: Moll.

Viaplana, Joaquim & DeCesaris, Janet Ann. 1984. La vocal neutra del català central: fonema o al·lòfon?. In *Estudis Gramaticals 1. Working Papers in Linguistics*, 341–380. Bellaterra: Universitat Autònoma de Barcelona.

Appendix

Target words elicited from subjects with their target pronunciation (only the target vowels in bold type in the spelling have been analyzed):

/ɛ/	/ɔ/	[ə]
cabell [kəβɛɫ]	escola [əs'kɔlə]	gegant [ʒə'ɣan]
cel ['sɛl]	flor ['flɔ]	menjador [məɲʒə'do]
jersei [ʒər'sɛj]	futbol [fub'bɔl]	papallona [pəpə'lonə]
pedreta [pə'drɛtə]	groc ['grɔk]	paper [pə'pe]
pera ['pɛrə]	nou ['nɔw]	paperera [pəpə'rerə]
pereta [pə'rɛtə]	oli ['ɔli]	pedreta [pə'drɛtə]
princesa [prin'sɛzə]	olla ['ɔlə]	pelut [pə'lut]
setze ['sɛdzə]	pilota [pi'lɔtə]	pereta [pə'rɛtə]
telèfon [tə'lɛfun]	rosa ['rɔzə]	petit [pə'tit]
vermell [bər'mɛɫ]	sol ['sɔl]	tovallola [tuβə'ɔlə]
zero ['zɛru]	taronja [tə'rɔɲʒə]	
	tovallola [tuβə'ɔlə]	

# Prepositional aspect constructions in Hiberno-English

Lukas Pietsch

University of Hamburg, Research Center on Multilingualism

Tense-aspect periphrases of the structural type *be* + preposition + gerund have been a shared feature of Irish and English for several centuries. Its best-known instantiation is the so-called *after* perfect. However, it can best be accounted for in the context of the other, similar constructions, although some of these come from a historically different origin. Based on a scenario involving reinforcement of already present constructional possibilities in English through contact with Irish, it is suggested that the syntactic productivity of the prepositional construction schema in Irish was among the crucial cognitive factors that conditioned both the replication of the *after* construction and the strengthening and preservation of its siblings with *for* and *about* in Irish English.

**Keywords:** English, Hiberno-English, Irish, tense and aspect, grammaticalisation, syntactic productivity, Construction Grammar, contact linguistics, syntactic replication

## 1. Introduction

This paper<sup>1</sup> deals with the history of a class of verbal periphrases of the type in (1), henceforth called “prepositional tense/aspect constructions”, or “prepositional constructions” for short, in Irish English (Hiberno-English) and in Irish.

---

1. This paper is based on research carried out in the *Collaborative Research Center 538 “Multilingualism”* hosted at the University of Hamburg and funded by the *German Research Foundation (DFG)*. My thanks are due to all my colleagues at the research center who have given valuable input to this research in the form of discussion and collaboration, as well as to the participants of the Workshop on “Multilingualism and Universal Principles of Linguistic Change” held at the Annual Meeting of the *Societas Linguistica Europaea* in Bremen, August 2006. I also thank David Denison, Teresa Fanego and Merja Kytö for sharing their opinions and for



- (1) a. English:  
       *be* + preposition + V-*ing*  
       b. Irish:  
       *bí* + preposition + Verbal Noun

The best-known instance of this pattern is the so-called “*after* perfect” in Hiberno-English (2a) and its equivalent in Irish (2b).

- (2) a. I am after going  
       b. *Tá mé tar éis imeacht*  
           is me after going

This, however, is only one member of a whole family of equally interesting – if less prominent – patterns, which has been quite productive in Irish and to a somewhat lesser extent in Hiberno-English. In Irish, at least the following prepositions have been attested in this role: *ag* ‘at’, *i*<sup>N</sup> ‘in’, *do* ‘to’, *ar* ‘on’, *iar*<sup>N</sup> ‘after’, *tar éis* ‘after’, *i ndiaidh* ‘after’, *chun* ‘towards’, *le* ‘with’, *ar tí* ‘on the point of’.<sup>2</sup> In English, the relevant items include *on*, *for*, *about*, *after*, and *upon*. In both languages, these constructions cover the aspecto-temporal semantic areas of simultaneity (duration/progressive), retrospectivity (perfect) and prospectivity.

This article will put forward the hypothesis that the parallel development of the English constructions following the Irish ones is at least partly due to language contact, even though the scenario is not a straightforward one. In the previous literature, only the *after* perfect has generally been recognised as a clear instance of structural transfer. The other, minor instantiations of the pattern are partly not yet very well documented historically, and – what is even more problematic from a contact-linguistic point of view – some of them have not been restricted to Ireland but are also attested in other, unrelated varieties. The pattern *be* + preposition + VP was an option whose structural foundations were already present in English, in the form of one or two minor, peripheral instantiations, when English-Irish contact was in its crucial stage. However, under contact with Irish, and enriched with new lexical instantiations such as the *after* perfect, it was apparently reinforced and gained significantly in productivity and frequency in Hiberno-English as compared with other varieties.

---

providing me with highly valuable data, and two anonymous referees for their very helpful remarks and suggestions.

2. Following a notation convention of Celtic linguistics, superscript letters in cited forms of Irish words, here and further below, denote initial consonant mutation patterns triggered on the following word. Superscript L stands for lenition, N for nasalisation (eclipsis). (See Mac Eoin 1993; Ó Siadhail 1989, for details).

Besides the purely historical interest inherent in this scenario, there is also a theoretical concern to this discussion. The contact scenario crucially involves cognitive acts of “interlingual identification” (Weinreich 1953) between abstract structural patterns in the two languages that are functionally and formally similar but not identical. This implies an approach to grammatical theory that takes the “construction” to be a central unit in the organisation of linguistic knowledge. In particular, I will argue that an approach based on Construction Grammar (Goldberg 1995, 2006) lends itself best to the description of this process, where interlingual identification seems to have operated simultaneously on the level of the more abstract constructional schema of ‘be’ + preposition + VP, and on the level of its specific, lexically filled instantiations with their individual semantic and pragmatic properties, such as the *after* perfect.

The paper is organised as follows: In Sections 2–3, I will provide an historical survey of the constructions in question. I will first give a brief outline of the situation in Irish (Section 2), based on the existing descriptive literature. This will serve as the typological backdrop to the developments in Hiberno-English. The situation in English will be the subject of Section 3. This section will present original corpus data from Hiberno-English material of the 18th and 19th centuries, as well as some comparative data from British varieties. Section 4 will then provide discussion and comparison of the data and relate them to the theoretical issue of a cognitively plausible modelling of contact-induced change.

## 2. Irish

In Irish, prepositional tense/aspect constructions – known in the grammatical literature under terms such as “aimsirí timchainteacha” (“periphrastic tenses”, Bráithre Críostaí 1960:173), “periphrastic aspectual phrases” (Ó Siadhail 1989:294–302), or “conjugaison périphrastique” (Gagnepain 1963:240) – have been a feature of the language for a long time, although most of them have not been fully grammaticalised to the same extent as in some other modern Celtic languages (Ó Corráin 1997). Table 1 lists the patterns that are or have been attested, for a brief overview. These constructions can be seen as a family linked together by a common semantic schema instantiated by each, and by a common syntactic shape – even though, it should be noted, the syntactic structure of the embedded verbal noun clause differs somewhat between constructions.<sup>3</sup>

---

3. Internal differences relate mostly to the coding of object NPs within the verbal noun clause. While in some types they are coded as genitives following the verbal noun, in others they are preposed to a position before the verbal noun, and separated from it by a particle *a*.

Table 1. Prepositional tense/aspect constructions in Irish.

Construction	Lexical source	Function
<i>ag</i> + VN	‘at’	Progressive
<i>i<sup>N</sup></i> + poss. + VN	‘in’	Progressive (unaccusatives)
<i>do<sup>L</sup></i> + poss. + VN	‘to’	Progressive (passive)
<i>a<sup>L</sup></i> + VN (< <i>do</i> )	‘to’	Progressive (in <i>wh</i> -movement environments)
<i>for/ar</i> + VN	‘on’	Durative
<i>iar<sup>N</sup> / ar<sup>N</sup></i> + VN	‘after’	Perfect
<i>tar éis</i> + VN	‘after’	Perfect
<i>i ndiaidh</i> + VN	‘after’	Perfect
<i>chun</i> + VN	‘towards’	Prospective (intentional)
<i>le</i> + VN	‘for/with’	Prospective (obligative)

2.1 *Ag* ‘at’, *do* ‘to’, and *i* ‘in’

The preposition *ag* (Old Irish *oc*) has been attested in the pattern *be at V-ing* since the earliest medieval documents of Irish (Gagnepain 1963: 49). While in Old Irish it was still only in an early phase of its grammaticalisation and not yet consistently used to denote progressive aspect (Ronan 2003), its use has steadily expanded and gained in importance throughout the history of Irish (Gagnepain 1963: 141). In Scottish Gaelic and Welsh, a similar progressive construction has all but ousted the simple tense forms in many environments (Ó Corráin 1997).

Within the progressive construction, *ag* alternates with some other prepositional items depending on the environment. In passive clauses, *ag* is replaced by *do* ‘to’. (As is generally the case in Irish, the passive itself is marked not by verb morphology, but by the presence of a resumptive possessive pronoun marking the patient of the verbal noun as coreferential with the subject).

- (3) *tá mé do mo bhualadh*  
is I<sub>i</sub> to my<sub>i</sub> beating  
‘I am being beaten.’

Additionally, in some constellations where the order between the verbal noun and its objects differs from the canonical structure, especially when the object is moved out of the verbal noun clause by means of *wh*-movement, the preposition is further weakened and replaced by *a<sup>L</sup>*, also historically a reflex of *do*. Yet another alternation occurs with a certain closed class of intransitive (unaccusative) verbs, most of them expressing physical position or state. Here, *ag* is replaced by *i<sup>N</sup>* ‘in’. (The construction then also displays a special syntax similar to that of the passives

mentioned above, with a resumptive possessive pronoun between the preposition and the verbal noun.)

- (4) *tá sé ina shuí*  
 is he<sub>i</sub> in.his<sub>i</sub> sitting  
 'he is sitting.'

## 2.2 *Iar* and *ar* 'on/after'

The historical situation with this group of prepositions is complex, owing mostly to the conflation of several source prepositions during Middle and Early Modern Irish (Gagnepain 1963: 174, 261, 287; Greene 1979; Ó Sé 2004: 191f.). In 17th-century Irish, there was one preposition *ar* 'on', itself a product of an earlier merger of two items *ar* and *for*; and another preposition *iar*<sup>N</sup> > *ar*<sup>N</sup> 'after'. Both items came to be distinguished only by the pattern of initial mutation triggered on the following word, and their functions tended to be conflated, to the point where *ar*<sup>N</sup> 'after' was eventually replaced by other items (*tar éis*, *i ndiaidh*).

*For/ar* 'on' had been an alternative to *ag* as an item expressing duration of a state or action in certain environments. Its use in such a durative function can be traced back to roughly the 14th or 15th century. In the 17th century, it occurred in passive progressives (5) and in progressives of some intransitive verbs (6).

- (5) *do chleachtadh Ceat inchinn Mheis-Geaghra*  
 was.accustomed Ceat brain M.G.GEN  
*do bheith ar iomchar aige*  
 PRT being on carrying at.him  
 'Ceat used to carry Meis Geaghra's brain with him.' [Gagnepain 1963:289]
- (6) *an roth geintlidhe do bhí ar siubhal ar dhorus an dúnaidh*  
 the wheel magic REL was on turning on gate the fort.GEN  
 'the magic wheel that was turning at the gate of the fort'  
 [Gagnepain 1963:289]

The near-homophonous (*i*)*ar*<sup>N</sup> 'after' was the original item used to form the 'after' perfect. Isolated instances of it in this function have been attested since the 12th century, but it did not become common in the written language until the 16th or 17th century (Ó Sé 2004: 187–189). Unlike in present-day Irish, where its prototypical use is to denote recent past (as is also characteristic of the Hiberno-English *after* perfect), at earlier stages it could also express more generally perfect and relative-past meanings, as in (7). It was then ousted from this wider use by the rise of a rival construction, the participial perfect of the type *tá sé déanta agam* 'I have it done' (Ó Corráin 2006: 167).

- (7) *Dia fire do Dhia fhíre, atá iarn -a gheineamhoin*  
 God true from God true, REL.is after his begetting  
*gan déanomh*  
 NEG making  
 ‘true God of true God, who was begotten, not created’  
 [1616; Uí Bheirn 2004]

The durative construction with *ar* ‘on’ and the perfect construction with *(i)ar<sup>N</sup>* ‘after’ were formally indistinguishable in some environments and were sometimes conflated. Thus (8) shows an instance where the preposition can morphologically be identified as *(i)ar<sup>N</sup>* ‘after’ but behaves semantically more like the earlier durative *ar* ‘on’, while (9) shows a sentence containing two instances that are formally identical, but of which the first can be glossed as present/durative and the second as perfect.

- (8) *bíaidh an ghrían arna dhorchughadh*  
 is.FUT the sun after.its darkening  
 ‘the sun will be darkened.’ (Lat. ‘sol obscurabitur’) [Ó Corráin 2006: 157]
- (9) *gé atá mo chorp ar marthain / táim ar scarthain*  
 although is my body on living / is.1SG after parting  
*rem anam*  
 from.my soul  
 ‘although my body lives, I have parted from my soul.’ [Ó Sé 2004: 191]

From the 17th century onwards, *(i)ar* ‘after’ became obsolete in Irish in most environments and was lexically replaced by other items, composite prepositional expressions such as *tar éis* (contracted *tréis*) and *i ndiaidh*. These also took over its role in the perfect periphrasis.

### 2.3 *Le* ‘with’, *chun* ‘towards’

Besides the progressive and perfect constructions discussed so far, Irish has for a long time had a number of future-oriented, prospective periphrases. They use the prepositions *le*, *chun*, as well as several composite prepositional items.

*Chun* ‘towards’ (from earlier *dochum*) has been attested in this function since the 15th century in written Irish, but has reportedly become restricted to Munster dialects (O’Rahilly 1932: 233). The form *le* is today identified with the lexeme *le* ‘with’. However, historically it is a reflex of yet another case of a merger with a different preposition, *fri* > *ré* > *lé*, which has gone through a semantic development from orientation in space (‘facing towards’) through adversative (‘against’), comitative (‘with’) to instrument (‘with’) meanings, as well as meanings of goal (‘to’)

and purpose ('for'). Its usage in a periphrastic construction with verbal nouns or other process nouns can be traced back to Old Irish (Gagnepain 1963:69, *Dictionary of the Irish Language*). Today *le* is used either in intransitives (10), or in passives, which in turn can either express possibility (11) or necessity/obligation (12). In many cases, its most natural translation equivalent in Standard English would be *be to V*.

- (10) *Tá sé le theacht amáireach*  
is he with/for/to coming tomorrow  
'He is to come tomorrow.'
- (11) *Níl dada le fáil*  
NEG.is nothing with/for/to getting  
'there is nothing to be got.'
- (12) *Tá an leabhar le léamh agam*  
is the book with/for/to reading at.me  
'The book is to be read by me.'

In environments where both *le* and *chun* are possible, they are reported to contrast in such a way that *le* carries implications of obligation, *chun* of intentionality (Ó Siadhail 1989:296, 293). Finally, the distinction between *ar tí* ('on the point of') and *chun* is described in Bráithre Críostaí (1960:173) as one of immediate versus non-immediate futurity.

### 3. English

We can now turn to the situation in English, both in the British English of the time of contact (i.e. mainly the 17th/18th century), and in Hiberno-English itself. The main data source for the treatment of Hiberno-English in this investigation, besides existing studies such as Bliss (1979), Filppula (1999) and McCafferty (2003, 2004, 2006), is a preliminary version of the *Hamburg Corpus of Irish English*, a collection of dialectal, subliterate English texts of Irish provenance, currently compiled at the University of Hamburg.<sup>4</sup> The material used in this study amounts to some 270,000 words of text. For varieties outside Ireland, a number of subsidiary data sources were used, in particular: the *Old Bailey Proceedings Online*, a large

---

4. The corpus consists for the most part of letters written either in the context of emigration (written to or from Ireland), or of the social and political unrest in the country (prisoners' letters, anonymous threatening letters, petitions etc.). The material collected for this purpose is complemented with some similar material already published in other sources, such as O'Farrell (1984), Fitzpatrick (1994), and Miller et al. (2003).

Table 2. Prepositional tense/aspect constructions in English.

Construction	Function
<i>on</i> + V-ing	Middle English progressive, obsolete
(will be) <i>after</i> + V-ing	Durative (?), 17th/18th cent.
(will be) <i>for</i> + V-ing <sup>†</sup>	Durative (?), 17th cent.
(will be) <i>upon</i> + V-ing <sup>†</sup>	Durative (?), 17th cent.
<i>after</i> + V-ing	Perfect
<i>about</i> + V-ing	Prospective (immediate)
<i>for</i> + V-ing	Prospective (intentional)
<i>on</i> + V-ing <sup>†</sup>	Prospective (imminent necessity)?

<sup>†</sup> Only sparse attestations; *hapax legomena*

collection of published trial records spanning the time between 1670 and 1834; the *Corpus of Late 18th Century Prose* (Van Bergen/Denison 2003,<sup>5</sup> henceforth CL18CP), which consists of letters written by and to the manager of a large rural estate in north-west England during the late 18th century; a corpus of New York newspaper advertisements, also from the late 18th century (Triggs 1996); and finally some data from the *Corpus of 19th Century English (CONCE)* (Kytö/Rudanko 2000<sup>6</sup>).

What we find in English is, again, a whole group of seemingly similar constructions (see Table 2 for an overview). They came from different sources, some of them within Standard English and some directly replicated from Irish.<sup>7</sup> Because of the heterogeneity of the sources, the contact effect did not result in an exact one-to-one correspondence between all items involved across the two languages. However, the net effect was nevertheless one of structural convergence:

5. The *Corpus of Late 18th Century Prose*, compiled by Linda van Bergen and David Denison, was created in the course of the research project “The English language of the north-west in the late Modern English period” at the University of Manchester, on the basis of material held in the John Rylands University Library of Manchester.

6. The *Corpus of Nineteenth-Century English (CONCE)* was compiled under the supervision of Merja Kytö and Juhani Rudanko at Uppsala University and University of Tampere in May 2000. My thanks go to Merja Kytö, who kindly ran several corpus searches for me and provided me with the excerpted search lists.

7. Of the constructions listed here, the Middle English progressive with *on* needs not to be considered further, as it had long become obsolete by the time of the the Irish-English contact. Hypotheses about the role of a different (British) Celtic contact effect involved in its development cannot be treated in this paper (cf. Keller 1925; Wagner 1959; Elsness 1994; Vezzosi 1996; Poppe/Mittendorf 2000; Poppe 2002, 2003; Filppula 2003).

as the outcome, we find a family of constructions in Hiberno-English that bears unmistakable points of correspondence to the Irish one.

### 3.1 Non-perfect *after* and related uses

Before dealing with the well-known *after* perfect itself, we need to discuss a group of curious use types found in earlier (17th and 18th century) literary representations of Hiberno-English, which have given rise to a good deal of discussion and some puzzlement in the literature (Bliss 1979: 302f.; Filppula 1999: 99–107; McCafferty 2003, 2004; Ó Sé 2004; Ó Corráin 2006). Here, *after*, as well as other prepositions (at least according to some isolated attestations) are employed in a function that is not easily definable but clearly distinct from the perfect function familiar from later stages of the dialect. The preposition is typically combined with other markers of non-present time, most frequently a future-oriented modal such as *will* or *would* (13). The resulting construction generally appears to have the temporal meaning expressed in Standard English by the same combination just without the prepositional element, rendering the preposition itself seemingly otiose. This use type has been dubbed the “futuric” *after* (McCafferty 2003), although it is far from obvious in what sense the preposition as such really contributes futuric meaning to the overall construction.

- (13) I expect your honour will be after doing the same this year  
 (= ‘...will be doing...?’)  
 [1800, quoted in McCafferty 2004: 140]

According to one opinion proposed by several older authors and most recently upheld to some extent by Ó Sé (2004: 243), non-perfect *after* should be seen with scepticism and should most probably be discarded as an inauthentic artefact of “stage Irish”, produced by English authors not natively familiar with the dialect. However, McCafferty (2004) and Ó Corráin (2006) argue convincingly for its authenticity, which leaves us with the question of how to analyse it semantically. Pace Hickey (2000: 100) it seems fairly clear that the construction is semantically distinct from the perfect *after* and that the semantic contribution made by the preposition is not one of perfect (relative past); thus the combinations with *will* cannot simply be analysed as future perfects (McCafferty 2004: 102). However, McCafferty’s alternative proposal, of analysing *after* as a separate futuric construction, meets with some empirical difficulties too. McCafferty (2004) develops an ingenious scenario according to which *after* underwent two parallel and partly simultaneous grammaticalisation processes, one leading to a perfect and the other to a prospective/future. This duplicity would be motivated by the “Ianus-like”



properties of the source semantics of locative *after*, in connection with different ways how time relations are conceptually mapped into the metaphorical domain of space. However, there is little or no evidence that *after*, on its own, could constitute future or prospective meaning. All the examples McCafferty quotes seem to owe their future semantics first and foremost to the presence of *will* or other elements of the grammatical context.

We are thus left with the hypothesis proposed by Bliss (1979: 302) as the most likely explanation for this curious construction. Bliss proposes that it is a borrowing from a different source preposition, the preposition *ar* ‘on’. As discussed earlier in Section 2.2, the prepositions *iar*<sup>N</sup> ‘after’ and *ar* ‘on’ had become confusable in Irish for phonological reasons and were in a state of variability. Ó Corráin (2006), building further on Bliss’ hypothesis, demonstrates that *iar*<sup>N</sup> had intruded into non-past environments and was being used in ways quite similar to the non-perfect *after* attestations in English. Ó Corráin points especially to the correspondence between examples such as (14a, b) and (15a, b).<sup>8</sup>

- (14) a. You will be after being damn’d  
       b. *beidh tú ar do fhliuchadh*  
           be.FUT you after your making.moist  
           ‘you will be made moist.’
- (15) a. I will be after being absolv’d  
       b. *biaidh mé air mo ghlanadh*  
           be.FUT I after my cleansing  
           ‘I will be cleansed.’

Both the seemingly non-perfect Irish *iar* constructions and the English non-perfect *after* constructions are very frequently used in passive forms, just like the original Irish *ar* ‘on’ construction was. Moreover, many of the uses of English *after* appear to be consistent with a reading of durativity, also corresponding to the original meaning of *ar* ‘on’.<sup>9</sup>

---

8. An anonymous reviewer questions the formal parallel observed here (which Ó Corráin calls “patently obvious”), on the grounds of the different morphosyntactic material employed by the Irish and English constructions (in particular, *being* + passive participle on the English and possessive + verbal noun on the Irish side). The parallel, however, lies in the fact that the possessive construction – as in *mo ghlanadh* ‘my cleansing’ – represents the standard way Irish marks passives in periphrases of this type; thus both the English and the Irish construction ultimately consist of an element denoting futurity, an element denoting the passive, and the preposition *ar/after*.

9. McCafferty, too, notes (2004: 138) that some examples he cannot easily classify as futuristic have “connotations of past habituality with no sense of prediction” or “iterative meanings”.

Somewhat more dubious as to their authenticity are a few attestations found in one early text edited and discussed in Bliss (1979: 124f., 303), a short satirical piece of mock Irish English of 1684 called *Bog Witticisms*, where a similar usage is found not with *after*, but other prepositions: three times with *for* (16) and once with *upon* (17).

- (16) a. Y vill be for mauking Child upon dy Body  
 b. Vee vill shet up Housh-kepin and be for livein aul togadder  
 c. Dou shaut be for sending Aunswer to vaat Y hauve sent dee
- (17) aund Y cannot be upon vaaking but the Deevil take me, Y do fall upon dream-  
 ing consharning thy shweet shelfe [Bliss 1979: 124]

Bliss notes that the source is “not a text which inspires much confidence” and that the structure may well be “the result of some error or misunderstanding”.<sup>10</sup> However, whether or not the precise usage of the two prepositions is being rendered correctly, the passage may still be of interest, at least as indirect evidence of the stereotyped outside perception of the dialect. The occurrence of as many as four tokens of this kind in a short passage of a few lines indicates that the author intended them as part of the linguistic satire, i.e. as a reference to a marked and socially stigmatised Hibernism. As such, even if the author got the precise lexical usage wrong, these examples may well attest to the existence of the overall pattern of prepositional circumlocutions.

It is not unlikely that here, too, the Irish *iar/ar* confusability may ultimately provide the explanation. As the Irish equivalents of *after* and *on* were functionally merging in these kinds of constructions, *after* could easily be replaced with *upon* in the emerging English structure. Bliss specifically explains *upon waking* as a direct calque of Irish *ar múscailt* (‘awake’, lit. ‘after waking up’): in this idiom, *ar* etymologically represents (i)ar<sup>N</sup> ‘after’, but is formally indistinguishable from *ar* ‘on’ in this phonological environment.

### 3.2 The *after* perfect

The function and history of the *after* perfect proper in Hiberno-English needs not to be discussed in great detail here, as it is sufficiently known in the literature (Filppula 1999; Hickey 2000; McCafferty 2006). In short, *be after V-ing* is one of the characteristic Irish perfect constructions, sometimes dubbed a “hot-news” perfect (Harris 1984) or “immediate perfect”. There is general agreement in the

10. A similar reservation is emphasised by an anonymous reviewer, who suggests that an isolated token like (17) “may be more an example of Saussure’s parole than of langue.”

literature that it derives from Irish.<sup>11</sup> In present-day Hiberno-English, it contrasts with a second type of perfect, the transitive so-called “medial-object perfects” (*I have my dinner eaten*), whose development has also often been discussed in relationship with a similar construction in Irish, the passival participial perfects (Filppula 1999; Ó Sé 2004; Pietsch 2005). The semantic distinction is mainly that the *after* perfect denotes an event in the immediately recent past relative to a reference time, whereas the medial-object perfect has resultative and statal uses.

Compared with the non-perfect *after* constructions of the 17th and 18th centuries discussed in the previous section, the perfect *after* construction is relatively young. McCafferty (2004: 139) demonstrates how the younger, perfect use of *after* ousted the older non-perfect (“futuristic”) use in the course of a century of variable usage, between roughly 1750 and 1850 (cf. also Filppula 1999: 104). McCafferty finds the earliest true, prototypical modern perfect token recorded in 1767 (18):

- (18) Why, friend, my master is Mr. Delamour, who is just after coming from Paris,  
.... [1767, quoted in McCafferty 2004: 139]

In modern dialect data, Filppula (1999: 101) notes an apparently higher frequency of the *after* perfect in Dublin speech as opposed to other dialects, and an “almost total absence” from the south-west; a fact possibly linked to the dialectal variation within Irish, where the south-western dialects of Munster and Connacht display less usage of the Irish *after* perfect and a relative preference for the passival participial perfect instead; see Ó Sé (2004).

Comparable data from the *Hamburg Corpus of Irish English* is not particularly conclusive, owing to small token counts. The corpus contains only seven attestations, all of them consistent with the expected recent-past semantics. The earlier non-perfect use of *after* is not attested. With the earliest perfect token recorded in 1851, the data would seem to corroborate the findings regarding a relatively late spread of the construction. It does not, however, confirm Filppula’s geographical findings, as five out of the seven tokens are indeed from Munster writers, from the dialect area that according to Filppula least favoured the construction.

---

11. Pace Kelly (1989), who is quoted in the literature as having advocated an origin of the construction in English dialects from outside Ireland, based on scattered evidence of some similar usages in late-19th century and 20th century dialects. Unfortunately I was unable to access this paper (which is discussed briefly in McCafferty 2004: 124). I thank an anonymous reviewer for drawing my attention to it.

### 3.3 *For*

Just as in Irish, the semantic domain of prospective (future-oriented) constructions has given rise to several prepositional constructions in English. The first to be discussed here is one using *for*, followed by a discussion of *about*. Both of these constructions differ from the *after* perfect in that Hiberno-English has shared them with emergent Standard English at some stage.

*Be for V-ing* was used in 17th and 18th century Standard English as a verbal prospective periphrasis, with a meaning of ‘plan to V’, ‘intend to V’. While it soon became obsolete again in Standard English, it was preserved in very much the same function as a dialectal archaism in Irish English. This usage appears to be distinct from the doubtful nonce attestation of *will be for V-ing* in the 17th century text discussed earlier in Section 3.1. Although the construction is solidly attested in some corpora, it has not made it into the *Oxford English Dictionary* (henceforth *OED*) or into compendia such as Visser (1963), where only a semantically related usage with a non-verbal NP complement is mentioned (*be for X* = ‘intend to go to X’, ‘be bound for X’; e.g. Visser 1963:I, 162). A typical example of the verbal periphrasis is (19).

- (19) M<sup>r</sup> Ashton has bought the *piece* of Land late M<sup>rs</sup>. Claytons that lyes between Laffock and m<sup>r</sup>. orrels and is for opening the Colliery [...] he is not for seling any coales unles to a perticular Flat and no money to be reciv<sup>d</sup>.  
[Th Billinge, 1788, CL18CP]

A useful overview of the diachrony of the usage of *for* in British (London-based) English can be gleaned from the *Old Bailey Proceedings*. The construction is already in existence, though rare, in the earliest parts of the corpus, between 1670 and 1700. It becomes fairly frequent during the first half of the 18th century, reaching a peak at around 1750, when it occurs in the records with a frequency well above the comparable constructions *be about to V* or *be about V-ing* (about which see next section). After 1770, it takes a very sudden decline, and by 1800 it has become decidedly marginal. The last, isolated, tokens recorded in this corpus are from 1820 and 1826. Table 3 shows the token and frequency figures across seven sub-periods.

The *for* construction is still relatively frequent in the *CL18CP*, of the late 18th century. With 13 instances in a corpus of just over 270,000 words, its raw textual frequency is somewhat higher than in the *Proceedings* (though such a direct comparison is of limited value, as raw text frequencies are obviously dependent on conditions of subject matter, genre and register); and it is also again a good deal more frequent than either the gerundial or the infinitival *about* construction within the same texts.

**Table 3.** Occurrences of *be for V-ing* in the *Old Bailey Proceedings*.

	Tokens	Tokens per million words	Word count (approx.)
1670–1699	2	1.41	1,423,000
1700–1729	30	13.09	2,292,000
1730–1759	109	13.77	7,915,000
1760–1779	74	10.50	7,049,000
1780–1799	18	1.25	14,411,000
1800–1819	9	0.67	13,462,000
1820–1834	2	0.12	16,455,000
Total	244	3.87	63,007,000

The picture is different in other corpora, however. The New York newspaper advertisements corpus, roughly contemporary with the *CL18CP*, contains not a single example. The *CONCE*, finally, has only two apparent examples, both from private letters from the first decades of the 19th century.<sup>12</sup>

- (20) a. My brother Tom looked very unwell yesterday, and I am for shipping him off to Lisbon.
- b. William is for trying for a Cure.

We may conclude that by the end of the 18th century, the *be for V-ing* construction had been effectively ousted from educated written British English. As the material from the *CL18CP* comes from a time when according to the *Proceedings* the structure seems to have already been on its decline in London speech, its continued presence in this material and its absence elsewhere probably points to a longer survival in more peripheral regional varieties.

The letters from the *Hamburg Corpus of Irish English* attest to the same kind of usage of *for*, and crucially, for its preservation all through the 19th century. Again, we must be content with much smaller token counts, due to the smaller corpus size and the overall rareness of the construction. The Hamburg corpus contains five tokens, produced by four different writers – considerably fewer than the earlier British *CL18CP*, but more than the contemporary *CONCE* in terms of

12. In identifying instances of the *be for V-ing* construction in written corpora, there is an empirical difficulty in telling them apart from the formally identical but pragmatically different modern *be for X* construction (meaning ‘be in favour of X’). The operative difference is that the latter construction pragmatically presupposes the existence of contrastive choice, dealing with something you are *for* as opposed to a presupposed alternative you are *against*. In addition, the *be for X* construction does not entail that the subject of the clause intends to carry out the action himself. The two instances quoted from *CONCE* seem to be pragmatically consistent with the verbal periphrasis.

text frequency (considering that the letters-only part of the *CONCE* and the two other corpora are of roughly equal word count.)

- (21) a. A last request I ask is to write by this ship what you are for doing [McLeeJ01, 1828]  
 b. They are for writing soon [Hammon03, 1845]  
 c. As he tells me that he is for writeing by this mail [McCanc03, 1859]  
 d. But as he told me that he was for writing to you by this mail [McCanc04, 1860]  
 e. I would like to let my mate know what I am for doing [Millik01, 1884]

By the late 19th century, the construction was apparently felt to be a characteristic marker of Irish dialectal speech. In a work on Irish folklore from the early 20th century (O'Neill 1913), it is found in a passage of a folk story purportedly "translated from the Irish" (22). The nonstandard construction is evidently used here to give the language an Irish stylistic flair:

- (22) "Upon my word, you're a fine music master," says the piper then; "but tell me where you're for bringing me." – "There's a great feast in the house of the Banshee, on the top of Croagh Patric, tonight," Says the Puca, "and I'm for bringing you there to play music and, take my word, you'll get the price for your trouble."<sup>13</sup>

Montgomery (2006:20) confirms the survival of the *for* construction in Ulster English (Ulster Scots) until the present, with attestations taken from Fenton (2000) and several literary representations of modern Ulster Scots.

- (23) a. I thought you were never for coming.  
 b. A'm for startin noo. [Montgomery 2006: 20]

It is also possible to find occasional attestations of the construction in colloquial present-day Irish English, as in the following two entries from internet chatforums and guestbooks, both written by teenagers from Northern Ireland (24a, b).

- (24) a. What are you for doing this weekend? Wana have another girly nite?<sup>14</sup>  
 b. its no problem..... what are you for doing this weekend?<sup>15</sup>

13. "The Piper and the Puca", Irish folk story, "translated from the Irish by Dr. Douglas Hyde", in O'Neill (1913: Ch.31). Online edition: <http://billhaneman.ie/IMM/IMM-XXXI.html>

14. Internet guestbook, retrieved from <http://www.bebo.com/Profile.jsp?MemberId=2029812512>, 29 January 2007.

15. Internet chat, retrieved from <http://foxtrot.xnoc.net/~monkey/forum/viewtopic.php?p=438&sid=539a5e77918b06a127ec54007c47ac92>, 29 January 2007.

There is little evidence regarding a possible continuity between this *for* construction and the nonce attestation of the *will be for V-ing* use in the 1684 Hiberno-English *Bog Witticisms*, where *for* is semantically either durative or otiose. While both are attested at roughly the same time, nothing can be said about possible influences either way. However, the very fact that the anonymous author chose to use the *will be for V-ing* form with an apparent intent of effecting a caricature of a non-standard variety, at a time when a formally so similar *be for V-ing* construction already existed in more mainstream varieties of the language, would seem to indicate that he perceived of it as distinct and unrelated.

### 3.4 *About*

Prospective constructions with *about* + VP have been common in English for a long time. The *OED* describes *be about to V*, in essentially its modern meaning, as attested from the mid-16th century. It differs from the *be for* construction semantically in denoting temporal imminence rather than planned intention.

Its grammaticalisation history, as mirrored in earlier use types reported in the *OED*, indicates that its origin lies not so much in the preposition *about*, but rather the adverb *about*, used here originally in the sense of ‘afoot’, ‘astir’ (25).

- (25) Bisi aboute þei han ben To cacchen hit with al heore miht.  
 ‘they have been busy about to catch it with all their might.’  
 [1360; *OED* v. “about”]

The verbal complement used in the *about* construction was originally a *to*-infinitive, sometimes also a *for to* phrase. The use of an *-ing* form in its stead is a later innovation, datable only to the 18th century (26). It can be interpreted as part of a long-term trend in English for an expansion of the *-ing* form at the expense of the infinitive (cf. Fanego 2004), and, in light of the discussion of the previous section, it may have partly been due to the parallel with the somewhat earlier *for* construction (whose origin is unambiguously prepositional).

- (26) He might give a satisfactory answer respecting the Person to whom he is about setting the Corn Mill [*CL18CP*, 1782, John Philips]

Again, a good diachronic summary can be gleaned from the *Old Bailey Proceedings*. Summary statistics are provided in Table 4. During the first decades of the period covered (1670–1700), and indeed still in the first decade of the 18th century, we find the older, infinitival pattern exclusively. The first two recorded instances of the *-ing* form occur, isolated as yet, in 1716 and 1726 respectively.

**Table 4.** Occurrence of *be about* V-ing and *be about to* V in the *Old Bailey Proceedings*.

	Gerund		Infinitive		Word count (approx.)	% Gerunds
	tokens	per million words	tokens	per million words		
1670–1699	0	0	11	7.7	1.423.000	0,0%
1700–1729	2	0.9	9	3.9	2.292.000	18,2%
1730–1759	24	3.0	76	9.6	7.915.000	24,0%
1760–1779	31	4.4	33	4.7	7.049.000	48,4%
1780–1799	19	1.3	77	5.3	14.411.000	19,8%
1800–1819	34	2.5	77	5.7	13.462.000	30,6%
1820–1834	60	3.7	193	11.7	16.455.000	23,7%

The gerund pattern then advances rapidly until it reaches a peak during the third quarter of the 18th century, when its frequency is almost on a par with that of the infinitive. After that, however, it very soon begins to decline again, and ends up at a frequency level of between 20% and 30% at the beginning of the 19th century, with a steady downward trend in terms of relative frequency as compared to the competing infinitive construction (even though in terms of absolute token counts the last two decades again seem to show an increase). In comparison, the *about* construction developed somewhat later than the *for* construction, and then went out of use later and somewhat less abruptly.

The data from two other late-18th century corpora at our disposal, the *CL18CP* and the New York newspaper corpus, fit in well with this picture. Each corpus (being much smaller in size than the Old Bailey collection) only has a handful of attestations, four in each case. Half of them have the gerundial form, which is just about the proportion to be expected on the basis of the Old Bailey data for that time period.

The picture changes radically in the *CONCE*, which covers language from a later period and, crucially, also from a range of registers closer to the literary standard. Here, the infinitival pattern is overwhelmingly in the majority. Besides 65 tokens of *be about to* V, found across all parts of the corpus, there is only a single instance of *be about* V-ing, dating from the 1850s. Characteristically, this instance occurs again in a trial transcript, i.e. in a portion of the corpus that represents the same kind of approximation of the spoken colloquial language as the Old Bailey collection. The witness whose speech is being recorded is a servant, presumably from the lower social ranks. Interestingly, the witness's use of the gerund echoes an immediately preceding phrase by the interrogator – but he has used the infinitive, apparently felt to be the more standard form by that time:



- (27) – Was it when you first went into the service, or when you were  
about to leave it?  
– I was just about leaving.

It appears, in short, that *be about V-ing* lived on into the mid-19th century in some varieties of the spoken language in British English, but that it was marginalised and excluded from educated standard English soon after 1800.

We can now again proceed to the Irish English data from the Hamburg corpus, for comparison. Despite a lack of conclusiveness due to overall smallish token counts, this corpus again gives some indication that the situation in Irish English differed from Standard English: the gerundial forms are more frequent, and seem to survive longer. There are 8 tokens of the gerund as opposed to only 3 of the infinitive. The gerund forms occur in writers from all geographical areas, including some of relatively high education, and crucially, they persist even into the latest stages covered by the corpus, right into the early 20th century.

- (28) a. a parcel of land on the Bury Estate which was then about being sold  
[ForreE01, 1906]  
b. Some of the estates you are about purchasing [RyanJ\_01, 1908]

These latter two examples are both written by Catholic smallholding farmers from the province of Munster. Both writers show strong indications of typical Hiberno-English dialect features otherwise, such as transitive perfects with medial object positions (of the type “I have the work done”) or the use of the present tense instead of the perfect in sentences with *for* or *since*. Tentatively, we can take these late attestations as an indication that the *about V-ing* construction survived as a dialectal archaism in Irish English, probably longer than it did in Britain and elsewhere, although, unlike in the case of *for*, I currently have no data about its possible survival into the present.

### 3.5 *On*

Finally, the Hamburg Corpus also contains a *hapax legomenon* attestation of what may have been yet another instance of the same constructional pattern: *be on V-ing*. Despite its very scarce attestation, it is interesting enough to warrant a discussion in terms of a potential further member of the group under study here. The single attested example dates from 1875 (29).

- (29) Your aunt Angess McKeer is on dying. We dont expect she will live long.  
[1875, Brenna06]<sup>16</sup>

Like the constructions with *for* and *about*, (29) seems to have a prospective, future-oriented meaning. This is important because as an adverbial subordinator, temporal *on* otherwise denotes immediate relative past (completion) rather than futurity. On this basis, one might expect *be on V-ing* to act more like a perfect construction, just like in the case of *after*. However, the intended futuric meaning of (29) is clear from the context. It apparently carries an implication not of intentionality, but rather of an impending, causally unavoidable event. The Standard English item most akin to this in form and structure would be *be close on V-ing*. Similar to the case of *for* and *after*, there seems to be little or no semantic continuity between this use and the early use of *upon* in the 1684 text discussed earlier. No hints as to the existence of a *be on V-ing* construction were found for other varieties of English. If these isolated tokens represent actual use types and not mere one-off idiosyncrasies, we would be dealing with a situation similar to that of *after* and possibly *for*: an older, 17th-century type of the pattern modal + *be* + preposition + *V-ing*, with a semantic value that is somewhat indistinct but probably related to the Irish *ar/iar* merger; and a more recent use type lacking the collocation with the modal verb, and having a clear semantic value of temporal sequence on its own.

#### 4 Comparison and discussion

We can now proceed to compare the English and the Irish data under historical and typological perspectives. In terms of historical sequence, it will be noted that the Irish construction pattern is generally older than the English one. Even though many of the Irish forms discussed above did not become common in writing until fairly late, the general pattern is attested from the earliest times in the case of *ag*, and at least from the Middle Ages onwards in the case of several of the others. In English, on the other hand, the preposition + *V-ing* patterns are attested only from the 17th century.

More interesting than the mere temporal comparison is a comparison of the functional paths of development that both languages underwent. In Irish, it will be noted that virtually all the prepositions in question are spatial in origin. The

---

16. From Fitzpatrick (1994: 409). The expression is actually spelled “undying” instead of “on dying” in the original. The emendation seems safe to make, since the negation marker *un-* was clearly not intended in the context.

items used for periphrases of temporal simultaneity (progressive and durative), i.e. *ag* 'at', *i<sup>N</sup>* 'in' and *ar* 'on', denote proximity in space. As such, they represent the universal grammaticalisation path that draws on the spatial metaphor where the boundaries of a temporal process are conceptualised as spatial entities. This path is known as a common source for progressive constructions cross-linguistically (Heine 1993). The source of *do* 'to' is similar, likewise related to local proximity, although it does not match the universal prototype for progressives quite as closely, since it denotes movement towards rather than location in a spatial entity.

In the case of the prospective constructions, the sources of *le* (< O.Ir. *fri* 'to, against') and of *chun* 'towards' are also both solidly spatial, even though in the case of *le* this semantic source has been obliterated through later developments. Both represent the common grammaticalisation path from goal-directed motion via intentionality to futurity (Bybee/Pagliuca/Perkins 1994). Even more obvious is the spatial source in the case of the modern replacement *ar tí* ('on the point of') which draws on the conceptual metaphor of the beginning of a process as the crossing of a spatial boundary.

In the case of the perfect constructions, the spatial nature is somewhat less obvious, because spatial and temporal meanings existed side by side from very early on. *Iar* 'after', although ultimately of spatial origin etymologically, had developed the temporal meaning of 'after' as one of its primary functions already by the time of Old Irish. The later replacements, *tar éis* and *i ndiaidh*, are again both spatial in origin, consisting each of locative preposition (*tar* 'across', *i* 'in') and a noun of concrete spatial meaning (*éis* 'tracks, trace'; *diaidh* < O.Ir. *déad* 'end point'). However, both had developed the purely temporal meaning of *after* as one of their functions well before they were co-opted to serve in the periphrastic perfect construction.

In English, the picture is much more mixed. The source of the standard progressive was a locative prepositional construction at least in parts (Middle English *is on huntunge* etc.), but that source had long ceased to be transparent by the time the grammaticalisation of the other, younger constructions set in. Of the two prepositional constructions that are shared marginally with Standard English, *about* and *for*, one comes from a source that is even formally quite unrelated to the *be* + prep + *V-ing* pattern. As discussed earlier, *be about V-ing* was a late replacement for a considerably older structure *be about to V*, where *about* was originally not a preposition but an adverb, and *to V* was a straightforward infinitive of purpose.

The other construction, *for*, is formally unambiguous as an instantiation of the *be* + prep + *V-ing* pattern. Its semantic source is one of destination and purpose; as such, it represents a similar grammaticalisation path as the *about* construction, namely that of volitive and obligative modality towards futurity.

Thus, in Standard Modern English, unlike in Irish, it seems to have been only this modality-related grammaticalisation path that has been productive – and only marginally so – in developing structures of the prepositional type. The spatial grammaticalisation path, active in Irish, was no longer productive in standard English, and reflexes of it are found only in those constructions peculiar to Hiberno-English that were undoubtedly directly borrowed.

To conclude, the question is: can we causally link the development of these different constructions together at all? Taking an agnostic approach, we might be content with stating that *after* (and possibly *on*) were borrowed from Irish, whereas *for* and *about* existed independently. But is the replication of *after* related to the fact that *for* and *about* were preserved along with it? Conclusive empirical proof of such a hypothesis is hardly possible in principle, but a plausibility argument in its favour, based on cognitive theories of grammatical organisation, can be attempted.

The constructions in question can be described as formal idioms. Even though their semantics is transparently motivated by their combinatorial make-up to some degree, their behaviour is not entirely predictable and their licensing as a grammatical unit must be an object of specific, arbitrary knowledge. According to a construction-based approach to grammatical organisation such as Construction Grammar (Goldberg 1995, 2006; Croft 2001), such complex patterns are assumed to possess “unit status”, that is, to be represented as holistic entities in speakers’ internalised grammatical knowledge. Crucially, this holds not only for the individual constructions with specific prepositions; it may also hold for representations of the more abstract schema instantiated by each. In this way, Construction Grammar attempts to model the internalised knowledge that speakers possess about an abstract pattern’s productivity and frequency: the more prominent a pattern is in terms of productivity (the type frequency of its instantiations) and in terms of their token frequency, the more strongly it will become entrenched in memory. Thus, in the present case, Construction Grammar predicts that speakers will form representations of an abstract schema such as *be* + preposition + gerund, if this is supported by sufficient number and frequency of instantiations. What I am suggesting now, additionally, is that such categorisation processes, in the bilingual speaker, can cross the two linguistic repertoires. Representations of perceived schemata such as the prepositional schema discussed here can then become objects of what Weinreich (1953) called “interlingual identification”, and as such, they can become channels and carriers of analogy-driven convergence and replication (cf. Bisang 2001: 188f.).

On this basis, we can now sketch the historical process we assume to have happened in Hiberno-English, roughly as follows. In Irish, at the time of contact with English, there already existed a large range of the prepositional constructions

in question, all organised around the common cognitive schema related to the locative source semantics of the various prepositions involved. English speakers only had, at most, two such constructions (with *about* and *for*), and these were sufficiently dissimilar and isolated in the system that it is doubtful whether they would have led to the establishment of a common abstract categorisation with a separate unit status. In contact with Irish, however, bilingual speakers could perceive the similarities between these existing constructions and the Irish ones, and on this basis re-categorise and assimilate them as instantiations of a common productive schema.

At about the same time, bilinguals were creating the *after* and probably the *on* construction in English, on the model of Irish. In these instances, the transfer was evidently based on a perception of both functional and formal equivalence on the level of the individual lexical items. Whether this innovation was dependent on, or aided by, a prior strengthening effect regarding the other two constructions, or whether the sequence was rather the other way round, is a question that cannot be answered on the basis of the historical data. In any case, both the constructions that were downright borrowed on the lexical level, and those constructions that already existed in English, ultimately served as input into the same new emergent category, the *be* + preposition + *V-ing* pattern in Hiberno-English. It was ultimately this abstract categorisation that was “borrowed” from Irish into English, along with the specific instantiation of the *after* perfect, and this act of abstract replication was what caused the existing *for* and *about* constructions to become strengthened and preserved in use along with their new sibling, the *after* perfect.

## References

- Bisang, Walter. 2001. Areality, grammaticalization and language typology: On the explanatory power of functional criteria and the status of Universal Grammar. In *Language Typology and Universals*, W. Bisang (ed.), 175–233. Berlin: Akademie-Verlag.
- Bliss, Alan J. 1979. *Spoken English in Ireland, 1600–1700*. Dublin: Dolmen Press.
- Bráithre Críostaí [The Christian Brothers]. 1960. *Graiméar Gaeilge na mBráithre Críostaí*. Dublin: M. H. Mac an Ghoill.
- Bybee, Joan, Pagliuca, William & Perkins, Revere. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Dictionary of the Irish Language: Based Mainly on Old and Middle Irish Materials*. 1983. Dublin: Royal Irish Academy.
- Elsness, Johann. 1994. On the progression of the progressive in Early Modern English. *Computers in English Linguistics* 18: 5–25.

- Fanego, Teresa. 2004. On reanalysis and actualization in syntactic change: The rise and development of English verbal gerunds. *Diachronica* 21: 5–55.
- Fenton, James. 2000. *The Hamely Tongue: A Personal Record of Ulster-Scots in County Antrim*. Belfast: Ullans Press.
- Filppula, Markku. 1999. *The Grammar of Irish English: Language in Hibernian style*. London: Routledge.
- Filppula, Markku. 2003. More on the English progressive and the Celtic connection. In *The Celtic Englishes III*, H. Tristram (ed.), 150–168. Heidelberg: Winter.
- Fitzpatrick, David. 1994. *Oceans of Consolation: Personal Accounts of Irish Migration to Australia*. Cork: Cork University Press.
- Gagnepain, Jean. 1963. *La syntaxe du Nom Verbal dans les langues celtiques*. Vol. I: Irlandais. Paris: Librairie C. Klincksieck.
- Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Greene, David. 1979. Perfects and perfectives in Modern Irish. *Ériu* 30: 122–141.
- Harris, John. 1984. Syntactic variation and dialect divergence. *Journal of Linguistics* 20: 303–327.
- Heine, Bernd. 1993. *Auxiliaries: Cognitive Forces and Grammaticalization*. Oxford: Oxford University Press.
- Hickey, Raymond. 2000. Models for describing aspect in Irish English. In *The Celtic Englishes II*, H. Tristram (ed.), 97–116. Heidelberg: Winter.
- Keller, Wolfgang. 1925. Keltisches im englischen Verbum. In *Anglica: Untersuchungen zur englischen Philologie 1: Sprache und Kulturgeschichte (Festschrift Alois Brandl)*, 55–66. Leipzig: Mayer & Müller.
- Kelly, Patricia. 1989. Afterthoughts on *after* + *doing*. Paper presented at the joint meeting of the Linguistics Association of Great Britain / Irish Association for Applied Linguistics, Belfast.
- Kytö, Merja & Rudanko, Juhani. 2000. *Corpus of 19th Century English (CONCE)*. University of Uppsala / University of Tampere.
- Mac Eoin, Gearoid. 1993. Irish. In *The Celtic Languages*, M. J. Ball & J. Fife (eds), 101–144. London: Routledge.
- McCafferty, Kevin. 2003. 'I'll be after telling dee de raison ...': *Be after V-ing* as a future gram in Irish English, 1601–1750. In *The Celtic Englishes III*, H. Tristram (ed.), 298–317. Heidelberg: Winter.
- McCafferty, Kevin. 2004. Innovation in language contact: *Be after V-ing* as a future gram in Irish English, 1670 to the present. *Diachronica* 21: 113–160.
- McCafferty, Kevin. 2006. *Be after V-ing* on the past grammaticalisation path: How far is it after coming? In *The Celtic Englishes IV*, H. Tristram (ed.), 130–151. Potsdam: Universitätsverlag.
- Miller, Kerby, Schrier, Arnold., Boling, Bruce & Doyle, David. 2003. *Irish Immigrants in the Land of Canaan: Letters and Memoirs from Colonial and Revolutionary America, 1675–1815*. Oxford: Oxford University Press.
- Montgomery, Michael. 2006. *From Ulster to America: The Scotch-Irish Heritage of American English*. Belfast: Ulster Historical Foundation.

- Ó Corráin, Ailbhe. 1997. On verbal aspect in Irish, with particular reference to the Progressive. In *Miscellanea Celtica in memoriam Heinrich Wagner*, S. Mac Mathúna & A. Ó Corráin (eds), 159–173. Uppsala: University of Uppsala.
- Ó Corráin, Ailbhe. 2006. On the ‘after perfect’ in Irish and Hiberno-English. In *The Celtic Englishes III*, Hildegard Tristram (ed.), 152–172. Potsdam: Universitätsverlag.
- O’Farrell, Patrick. 1984. *Letters from Irish Australia, 1825–1929*. Belfast: Ulster Historical Foundation.
- O’Neill, Francis. 1913. *Irish Minstrels and Musicians*. Chicago: Regan Printing House.
- Ó’Rahilly, Thomas F. 1932. *Irish Dialects Past and Present*. Dublin: Browne and Nolan.
- Ó Sé, Diarmuid. 2004. The *after perfect* and related constructions in Gaelic dialects. *Ériu* 43: 39–67.
- Ó Siadhail, Mícheál. 1989. *Modern Irish: Grammatical Structure and Dialectal Variation*. Cambridge: Cambridge University Press.
- Oxford English Dictionary*. 1989. 2nd edn. Oxford: Clarendon.
- Old Bailey Proceedings Online* ([www.oldbaileyonline.org](http://www.oldbaileyonline.org), accessed 20 February 2007).
- Pietsch, Lukas. 2005. Re-inventing the perfect wheel: Grammaticalisation and the Hiberno-English medial-object perfects. *Working Papers in Multilingualism, Series B*, 65. University of Hamburg: Sonderforschungsbereich 538 Mehrsprachigkeit.
- Poppe, Erich. 2002. The ‘expanded form’ in Insular Celtic and English: Some historical and comparative considerations, with special emphasis on Middle Irish. In *The Celtic Roots of English*, M. Filppula, J. Klemola & H. Pitkänen (eds), 237–270. Joensuu: Joensuu University.
- Poppe, Erich. 2003. Progress on the progressive? A report. In *The Celtic Englishes III*, H. Tristram (ed.), 65–84. Heidelberg: Winter.
- Poppe, Erich & Mittendorf, Ingo. 2000. Celtic contacts of the English progressive? In *The Celtic Englishes I*, H. Tristram (ed.), 117–145. Heidelberg: Winter.
- Ronan, Patricia. 2003. Periphrastic progressives in Old Irish. In *The Celtic Englishes III*, H. Tristram (ed.), 129–149. Heidelberg: Winter.
- Triggs, Jefferey. 1996. *New York Newspaper Advertisements and News Items 1777–1779*. Oxford: Oxford Text Archives.
- Uí Bheirín, Úna (ed.). 2004. *Corpas na Gaeilge, 1600–1882*. Dublin: Royal Irish Academy.
- Van Bergen, Linda & Denison, David. 2003. *Corpus of Late 18th Century Prose*. University of Manchester/Oxford Text Archive.
- Vezzosi, Letizia. 1996. La costruzione participio presente + verbo *essere* in anglosassone: un antesignano della forma progressiva o qualcosa di diverso? *Studi e saggi linguistici* 36: 157–210.
- Visser, Frederic T. 1963. *An Historical Syntax of the English Language*. 3 vols. Leiden: E. J. Brill.
- Wagner, Heinrich. 1959. *Das Verbum in den Sprachen der britischen Inseln*. Tübingen: Niemeyer.
- Weinreich, Uriel. 1953. *Languages in Contact: Findings and Problems*. Den Haag: Mouton.

PART IV

**Acquisition**





# Acquisition of Basque in successive bilingualism

## Data from oral storytelling\*

Margareta Almgren, Leire Beloki, Itziar Idiazabal  
and Ibon Manterola

University of the Basque Country

Two groups of five-year-old children (Basque L1 and L2), growing up in a language contact situation involving Spanish and Basque, show – based on data from oral story reconstitution – that the patterns of language development are similar in the two groups as far as story structure and use of main verb forms are concerned. The L2 subjects, who attend an educational immersion programme in Basque, also show that the density of verb forms throughout the texts and the relative absence of agreement errors are results that are very similar to those obtained by the Basque L1 children. Some differences can be observed between the two groups in the proportions of use of present and past tense forms and in the variety and amount of other verb forms. A first conclusion of the present study might be that the Basque L2 subjects offer productions which follow the adult model more closely than the L1 subjects, who tend to be more spontaneous.

**Keywords:** immersion programmes, narrative skills, oral story retelling, narrative structure and anchoring, imperfective and perfective verb forms, verb density

---

\* This research has been carried out thanks to the financial support granted by the University of the Basque Country (9/UPV 00033 130–13614/2001) and the Spanish Education and Culture Ministry (BFF2003–05196).

## 1. Introduction

This study<sup>1</sup> seeks to analyze whether five-year-old children from Spanish-speaking families acquiring Basque through immersion programmes at school from an early age show linguistic patterns similar or different from those who come from Basque-speaking families in a mixed Basque – Spanish speaking environment.

The question whether bilinguals and monolinguals follow the same pattern of development has often arisen in studies on the acquisition of syntax and the lexicon (Deuchar & Quay 2000). So far, a great number of studies on bilingual language acquisition have dealt with children acquiring two languages from birth, or bilingual first language acquisition (de Houwer 1990), as compared to their monolingual peers. However, not so much work has been done on comparing bilingual first language acquisition with successive bilingualism (Meisel 2001).

Successive bilingualism, however, is becoming more and more frequent through educational immersion programmes and in bilingual programmes for migrant population (Akinçi 2001). This fact has also brought about an increasing interest in the possible influence of L1 on L2, where concepts such as “cross-linguistic influence” and “transfer” are being interpreted not as an inevitable fusion of codes, but rather as a relief strategy or as the application of structures which are common to the two languages (Lanza 1998; Müller 1998; Müller & Hulk 2001).

These phenomena, however, should be viewed from a broader perspective on bilingual situations. The well-known educational immersion programmes in Québec (Hamers & Blanc 1983; Genesee 1987; Comeau & Genesee 2001; Hamers 2006) may have similarities with the situation of our Basque L2 subjects in the sense that they belong to the category of L1 majority language speakers who take part in immersion programmes which do not imply any risk of loss of their L1.

As pointed out on several occasions (Pfaff 1999, 2001; Romaine 1999; Backus 2004), this situation is not entirely equivalent to that of linguistic minorities as reflected in the studies carried out by Akinçi (2001) or Pfaff (1999, 2001). Nevertheless, it should not be forgotten that Basque is still a minority language in a situation which has undergone a change only during the last decades as we shall see briefly in the following.

---

1. The authors are grateful for the useful comments of the referees. However, responsibility for the views expressed in the article and for any omission or error remains with the authors.

## 2. The sociolinguistic situation of Basque

Administratively, Basque speakers belong to 3 different regions. On the French side of the Pyrenean Mountains, the Basque-speaking area belongs to the *Département des Pyrénées Atlantiques*. On the Spanish side, Navarre constitutes one community and this is where our Basque L2 data were collected. Three other Basque-speaking provinces form the Basque Autonomous Community, the most densely populated area of the three regions, where our Basque L1 data were collected.

These facts explain the contact situation between Basque and Spanish / French, where Basque's historical position has always been that of a threatened minority language. As a matter of fact, Basque was practically absent from education for centuries. It was not until after 1980 that a slow recovery began to take place, mainly through educational programmes. The amount of Basque speakers varies among the three regions, and also within each region. Most small rural areas are traditionally Basque speaking, where the oral transmission of the language has never been interrupted. In urban areas, however, the percentage of native speakers of Basque is much lower.

The Basque Autonomous Community is officially bilingual in the sense that official texts are published in both Basque and Spanish and labelling is also bilingual. However, this is not the case to the same extent in Navarre and in France. The media are still predominantly Spanish / French in all areas.

Recent data published by the Basque Statistics Office (2004) show that the major part of the population (76%) of the Basque Autonomous Community consider Spanish their mother tongue. 19% have Basque as L1, and only 5% indicate both Spanish and Basque as native languages.

As to the school system, 20% of the children attend primary and secondary education in Spanish, with Basque as a compulsory subject. Another 30% attend bilingual programmes and 50% attend school in Basque (either because it is their mother tongue, or in immersion programmes), where Spanish is introduced during the second part of the primary school period.

It should also be pointed out that in the Basque Autonomous Community part of the adult population has attended very intensive language and literacy programmes in Basque. As a result of these programmes, the proportion of teachers trained to teach in Basque rose from only 5% in 1978 to 70% in 2005.

## 3. Purpose of the present study. Hypothesis

According to our hypothesis, the children who in our circumstances become successive bilinguals at preschool age will develop their L2, Basque, following the

same patterns as their Basque L1 counterparts. The analysis deals with the acquisition of Basque as L2 in an educational immersion programme as compared to the acquisition of Basque as L1.

In both cases, our experiment was carried out in a school context, eliciting the reproduction of traditional children's stories. The activity designed for this purpose is that of oral story-telling, or re-telling, a genre familiar to preschoolers. In fact, according to some authors, most children of age 5 have acquired basic narrative skills (Bamberg 1987; Fayol 1991) and are able to reproduce the canonical narrative structure of traditional children's stories, as well as the use of verb forms which permits the distinction of narrative frame or background from foreground events, etc. (de Weck 1991; Berman & Slobin 1994).

Nevertheless, very different results can be found depending on several factors. The level of linguistic development can be as important as the age (or degree of bilingualism), the characteristics of the task and genre-variation (telling, re-telling, use of pictures, etc.), the school training, etc. For instance, using the well-known "Frog Story" (Mayer 1969), Akinçi & Kern (1998) show that Turkish – French bilinguals aged 5 are not able to reproduce the narrative superstructure. Rosat (1998) finds that four to six-year-old French-speaking children are not able to narrate autonomously (to reproduce the narrative macrostructure of an already known traditional story, such as "Snow White"). In Section 6 some of the differences between these studies and the method applied in our experiment will be analyzed.

The question whether the reproduction of narrative structures occurs in a similar way in L1 and L2 is our first point of analysis. In the second place, we will compare the trends in verb density, or the proportion of inflected verb forms in relation to the total number of words. It is expected that verb density will be higher in young children's stories, taking into account that children tend to focus more on the action or the events than on description or background comments (Bamberg 1996). Similarly, Quasthoff (1997) shows that children start to expand episodic knots around action words only after age seven.

In the third place we will go into the use of perfective and imperfective verb forms, where some differences have been attested between Basque and Spanish (Bonnotte et al. 1993). This issue brings about the question of narrative anchoring (Dolz 1990; Aksu-Koç 1994) or dominant tense, as well as variety of verb forms. The complex Basque system of tense marking by intransitive and transitive auxiliaries, incorporating morphemes of triple agreement for subjects, direct and indirect objects in singular and plural is considered difficult to acquire for non-native speakers. Consequently, it is a point of interest when analysing the process of acquisition of Basque by L2 children.

Bearing these aspects in mind, we will compare data collected in Basque and in Spanish from Spanish L1 bilinguals and in Basque from L1 children. Adult data will also be used to some extent for contrast.

#### 4. Linguistic structures analyzed

Concerning the verb cohesion of narratives, past and present tenses as well as perfective and imperfective forms play an important part in order to establish appropriate cohesive relations between background and foreground events. The languages of our study, Spanish and Basque, both distinguish perfective and imperfective aspect in past tense forms. Basque verbs mostly appear in periphrastic constructions where the main verb carries perfective and imperfective morphemes suffixed to the verb: *-tu* for perfective aspect and *-tzen* for imperfective aspect.

Tense, mood and person are marked by auxiliaries, which are different for intransitive and transitive constructions. Furthermore, Basque being an ergative language, transitive auxiliaries also include marking for direct and indirect objects, as can be seen in examples (1) to (5):

- (1) *zen* (3rd person singular): past tense for intransitive verbs.
- (2) *zuen* (3rd person singular): past tense for transitive verbs with direct object in singular.
- (3) *zitu* (3rd person singular): past tense for transitive verbs with direct object in plural.
- (4) *ziona* (3rd person singular): past tense for transitive verbs with direct and indirect objects in singular.
- (5) *zizkion* (3rd person singular): past tense for transitive verbs with direct object in plural and indirect object in singular.

The following reflects two examples of these periphrastic Basque verbs, always counted as one item in our analysis:

- (6) Perfective past tense (3rd person singular):  
*begira-tu zuen*  
 watch-PERF 3SG.PAST  
 '(He / She / It) watched.' (on one occasion)
- (7) Imperfective past tense (3rd person singular):  
*begira-tzen zuen*  
 watch-IMPERF 3SG.PAST  
 '(He / She / It) watched.' (normally, daily, now and then)

In Spanish, past tense verb forms are marked simultaneously for tense and aspect by different morphemes:

- (8) Perfective past tense (3rd person singular):  
*miró*  
watch.3SG.PAST.PERF  
'(He / She / It) watched.' (on one occasion)
- (9) Imperfective past tense (3rd person singular):  
*miraba*  
watch.3SG.PAST.IMPERF  
'(He / She / It) watched.' (normally, daily, now and then)

In Basque, a small set of frequently used verbs maintains synthetic forms where tense is marked by inflection and no specific aspectual marking appears. Semantically they are considered state verbs and grammatically they are assigned imperfective aspect.

- (10) *da / dago* (3rd person singular): present tense; permanent / temporary.  
'(He / She / It) is.'
- (11) *zen / zegoen* (3rd person singular): past tense; permanent / temporary.  
'(He / She / It) was.'
- (12) *dauka* (3rd person singular): present tense; direct object in singular.  
'(He / She / It) has.'
- (13) *zeukan* (3rd person singular): past tense; direct object in singular.  
'(He / She / It) had.'
- (14) *dauzka* (3rd person singular): present tense; direct object in plural.  
'(He / She / It) has.'
- (15) *zeuzkan* (3rd person singular): past tense; direct object in plural.  
'(He / She / It) had.'

Although languages may have parallel forms in morphology, the use that speakers make of these forms need not be identical. Spanish, French and Basque all possess perfective and imperfective past tense forms for aspect marking. However, in an experiment carried out by Bonnotte, Kaifer, Fayol & Idiazabal (1993), where university students produced short written narratives, the distribution of imperfective and perfective forms in the different parts of text turned out to be more clearly differentiated in French and Spanish, whereas in Basque the perfective verb forms were most frequent in all parts of the text (Table 1).

**Table 1.** Percentage of verb forms in relation to narrative structure (source: Bonnotte et al. 1993:93).

Language	Verb form	Setting	Unfolding	Outcome	Average
French	Imp	60	41	18	40
	Perf	30	51	76	52
Spanish	Imp	61	46	22	43
	Perf	38	53	77	56
Basque	Imp	47	37	17	34
	Perf	53	63	83	66

Whether or not this trend holds for all language uses in these three languages is a question that would require an analysis of a much broader scope than what is possible to carry out within the present study.

## 5. Narrative structure applied

Since Propp (1937, 1970), many researchers have proposed that traditional stories follow a fixed thematic structure that is considered a discursive characteristic of these texts (Adam 1992; Brigaudiot 1993, etc.) For the present study, the usual 3 phases (setting, unfolding and outcome) were extended to 5, following Adam (1992) and Dolz (1990), with the purpose of distinguishing the different phases of the narrative structure very clearly, making it easier to appreciate the change between perfective and imperfective verb forms.

Data were collected using different stories but with identical narrative structures in Basque and in Spanish. In this way we tried to avoid translation from one language to the other. The structures of both stories can be summarized as follows:

1. Setting: X / Y and his / her mother who lived in Z, were very poor / ran out of fire.
2. Onset: Mother told him / her to go and find money / fire.
3. Unfolding: The hero(ine) met three helpers and was then faced with three difficult tasks to overcome.
4. Resolution: Having solved the problems with the aid of the helpers he / she obtained the money / fire.
5. Outcome: X and his mother had money and lived happily ever after / Y and her mother had fire, could cook and lived happily ever after.



As can be seen, the stories correspond in both languages to the classical structure of children's tales with a happy end.

## 6. Method

All our data were collected in a semi-experimental classroom setting, where five to six-year-old preschool children were videotaped while listening to and re-telling a story – one in each language – based on a wordless picture-book. Before initiating the task, the children were told that their performance would be videotaped and shown to smaller children who did not know the story. These data were all transcribed, as well as all the adult's input used as a model during the story-telling sessions.

This kind of task is known to form part of preschool curricula and is often reinforced in educational immersion programmes.

In each session the adult told the story based on 12 wordless pictures representing the most important episodes to a group of 5 children. Subsequently, two children were elected at random. One of them, sitting on the teacher's chair, told the story again, with the support of the illustrations and of the adult in case of need, to the second one. After this, "the narrator" left the room, while the second child now sitting on the teacher's chair told the story to the third child, and so on, until the five children had completed the session.

After that, the adult again told the story to the next group of five children. In order not to interrupt or deviate the chain of transmission, not more than five children at the time were present during each session.

Berman & Slobin (1994) affirm that eliciting the same story makes it possible to control the content of the texts produced by children, and thus facilitates the comparison across different groups of subjects, or different languages. In this aspect our research method is similar to that used with the "Frog Story". However, our method shows some differences in relation to the "frog method". Our subjects were first asked to listen to the story and then to perform their retelling task. The "frog method" does not give any kind of model to the subjects, who just look at the pictures before telling the story.

In our opinion, adults' story telling to children does not consist in a simple input of story-contents in an audiovisual way, but it can also give some kind of input related to action competence (Schneuwly & Dolz 1998; Bronckart & Dolz 1999). Adults' story telling regarded as a model may have positive effects on children's representations or knowledge of the story-telling genre in a classroom setting. These representations deal for instance with physical aspects of the communicative action (where and when it happens, who the listeners are, etc.). They may also

be related to interactive aspects such as the social status of the speaker (the one sitting before the audience is expected to guide the interaction, in terms of a story production) or the goal of the communicative action (to tell a story to younger children via audiovisual recording).

We believe that all these communicative parameters should be kept in mind when interpreting the results of our research. In fact, some works have shown that children's narrative performance varies according to different communicative contexts. Brigauidot (1993) e.g. compares five to six-year-old children's narratives elicited from distinct tasks and concludes that the best organised and most coherent narratives are produced by children who have previously listened to the same story told by an adult.

## 7. Subjects

Two groups of children are analyzed. Children from group A are L1 speakers of Spanish having acquired Basque through immersion preschool instruction from age 2. These subjects, living in a mainly Spanish-speaking environment in Navarre, constitute the bilingual target group, and their mean age is 5:7.

Among these children, 46 subjects participated in the task, but in the subsequent analysis of the data, 6 of them were not included because they turned out to be L1 speakers of Basque and 3 of them were excluded because their narration turned into question-and-answer sessions. The remaining 37 subjects constitute group A.

Data were also collected with the same method from Basque L1 children, belonging to Basque-speaking families in a predominantly Basque-speaking area of the Basque Autonomous Community. From this group of originally 36 children, 12 were not taken into account because they turned out to be bilingual, with Spanish as one of their family languages. This leaves 24 subjects in the Basque L1 group B, who have also attended a Basque school from age 2. The mean age of this group is 5:3. Although these children had reached a certain degree of bilingualism because of the social presence of Spanish, we consider them a Basque L1 group.

## 8. Data analysis and results

In the first place, some quantitative data referring to general characteristics will be given, such as narrative structure, verb density and anchoring tense of the narratives. Secondly, a more precise analysis of verb forms and complexity will be carried out.

**Table 2.** Presence of narrative phases in stories told by children in Basque (L1 and L2) and Spanish (L1) (percentages).

Subjects	Setting	Onset	Unfolding	Resolution	Outcome
Basque L2 (Group A)	100	91.9	100	100	100
Basque L1 (Group B)	87.5	95.8	100	91.6	95.8
Spanish L1 (Group A)	97	100	100	97	97

8.1    Narrative structure

As mentioned above, we propose a five part structure (setting, onset of the plot, unfolding of the plot, resolution of the plot and outcome) to analyze our data, as reflected in Table 2. The data only take into account whether our subjects make reference or not to the five parts of the story in a way that makes them possible to identify, although they may be extremely short and not all elements (characters or actions) of each part appear.

As already mentioned, the point of these data is merely to analyze whether the narrative structure exists or not in children’s texts and in that context it should be underlined that our subjects are all able to reproduce the main parts of the story.

We want to focus on the communicative context as a crucial factor to understand these positive results. In Section 6 it was explained that children were asked to perform a retelling activity after they had listened to a story told by an adult. Furthermore, both adults’ and children’s productions were accompanied by a wordless picture-book.

As a matter of fact, the good results among both L1 and L2 speakers regarding their capacity to reproduce these narrative phases can probably be explained by the design of the experiment. The help of the images to remember the most important episodes and the fact that children are retelling and not constructing the narration are very important factors in the production of narratives as Brigaudiot (1993) claims in her work.

8.2    Verb density

Table 3 shows some general data referring to the stories produced by L1 and L2 children and the adults during their story-telling sessions. Verb density concerns the average of inflected verbs per 100 words.

**Table 3.** Number of words and inflected verbs in children's and adults' tales (average).

Subjects	N. of words/text	N. of inflected verbs	Verb density
Adult Basque	554	92	17
Child Basque L2 (Group A)	272	51	19
Child Basque L1 (Group B)	194	38	20
Adult Spanish	814	127	16
Child Spanish L1 (Group A)	395	69	18

If we consider the figures referring to verbs, the average number of inflected verbs per text used by children and adults in each language continues to be proportional to the number of words of the story; longer stories tend to contain more verbs.

On the one hand, our children follow the adult pattern in each language: the frequency of verbs is higher in Basque than in Spanish. On the other hand, adult verb density is lower than the density produced by children in both languages. Paradoxically, Basque L1 subjects produce shorter stories and fewer inflected verbs but with higher density. However, the Standard Deviation calculus also shows higher individual differences in group B than in group A (3 and 1.73 respectively).

Concerning the proportion of inflected verb forms in relation to the total number of words, it is expected to be higher in young children's stories than in those told by adults, taking into account that children tend to focus more on the action or the events than on description or background comments (Bamberg & Damrad-Frye 1991; Bamberg 1996).

Quasthoff (1997) shows that children start to expand episodic knots only after age seven and by the age of ten and fourteen all event knots are expanded. Bearing in mind that our subjects are only five to six years old, the tendency shown in Table 3 seems to confirm this assumption. In addition, the tale used as a model (told by the adult) is in fact to a great extent built up around events or episodes, which are not very much expanded.

The slightly better results in Basque L2 may partly be explained by the age difference (the mean age for group A is 5:7, whereas for group B it is 5:3). However, we consider school training as a much more significant influence than age. When interviewing the teachers, it became evident that the children in group B (Basque L1) were accustomed to the technique of telling stories to their classmates and

knew how to “play the teacher”. The Basque L2 subjects (group A), however, had less experience in carrying out such a task in the classroom, so for these children a few exercises were prepared with the aim of helping them to narrate in an autonomous way. Considering that these subjects belong to an immersion group, it may be assumed that they tend to follow given patterns rather strictly. The Basque L1 subjects (group B) also perform narrative tasks at school but they may have more linguistic resources to produce narratives that follow the given model to a lesser extent.

### 8.3 Narrative anchoring

One of the characteristics of the story-telling is the use of past tense in order to mark distance between the contents of the stories and the communicative situation. That is, past tense verb forms are one of the means by which the narrator is able to establish a narrative discourse outside the *hic et nunc*.

Anchoring can also be made in present tense as many Frog<sup>2</sup> researches have shown. Although Berman (1988) for Hebrew and Stephany (1994) for Greek show that the past is the preferred anchoring verb tense, Bamberg (1987) for German, Hickmann & Roland (1992) for French and English, and Sebastian & Slobin (1995) for Spanish have found that the anchoring is made in the present tense. Massive use of the present tense is also claimed in many studies as a trend of young children (Bamberg 1987; Berman 1988; Berman & Slobin 1995).

In the first place, as we have already mentioned, young children may have difficulties in decontextualising the story (Akinçi & Kern 1998). Furthermore, Sebastian (1989) shows that subjects younger than 3;6 tend to simply describe the contents of the illustrations in picture-based story-telling situations in the present tense and with the frequent use of progressive forms, and thus produce descriptive texts instead of narratives. It could be expected that our Basque L2 learners without a sufficient degree of linguistic maturity would reproduce this kind of pattern.

However, our subjects are able to anchor their texts in the past tense in L1 as well as in L2, almost without exception. This feature is particularly reflected in the setting of the stories, as shown in Table 4.

The occasional appearance of the present tense in Spanish does not correspond to narrative anchoring or to descriptive elements, but is due to inserted

---

2. The criteria used by Berman (1988:484) to identify the anchoring tense is the percentage of forms in the text “...as 75% or more occurrences of either present or past tense verb forms out of the total finite verbs in the text”. Sebastián (1989) establishes the 80% criterion.

**Table 4.** Narrative anchoring in the setting sequence (percentages).

	<b>Basque L2 (Group A)</b>	<b>Basque L1 (Group B)</b>	<b>Spanish L1 (Group A)</b>
Past tenses	100	100	96.4
Present tense	0	0	2.4
Others*	0	0	1.2

\* The category "Others" refers to subjunctive, future, conditional, present perfect and past perfect forms (also in subsequent tables).

dialogues. In this context, it should be stressed that at age five and with 2 or 3 years of schooling, these children are familiar with some of the main characteristics of this narrative genre.

#### 8.4 Tense and aspect related to narrative sequences

Although the percentage of forms can be significant to figure out the children's mode of production, not all parts of the narrations are similar. Discourse studies claim that narratives have different sequences with particular verb form exigencies (Bronckart et al. 1985; Dolz 1990; Bronckart 1996, etc.). When children are able to reproduce the complete frame of the story they use verb forms consistently. Past tense forms are used for the anchoring of the events and present tense forms appear mainly in dialogue parts.

As pointed out in Section 8.1, in some cases of our corpus the narrative phases are so extremely short that an analysis of perfective and imperfective verb forms would not throw any light on inter-linguistic differences. This is particularly true in the onset, resolution and outcome phases, which are not included in our analysis for this reason.

The anchoring value of tense in the setting of the story has been reflected in Table 4. When it comes to aspect, as shown in Table 5, imperfective forms are on the whole far more frequent than perfective ones in both languages in the setting sequence. There are, however, some important inter-linguistic differences. In Spanish, the imperfective forms are clearly dominant (94%). In Basque L1 these forms constitute 60 %, very similar to 57.7% in Basque L2.

This indicates in the first place that the Basque L2 subjects follow the same pattern as Basque L1 children. And in the second place, both groups show a similar trend to that found in Bonnotte et al. (1993): perfective forms are more frequent in Basque and in both cases clearly different from the use in Spanish (Table 1). But

**Table 5.** Use of tense and aspect in setting and unfolding sequences (percentages).

	Setting			Unfolding		
	Basque L2 (Group A)	Basque L1 (Group B)	Spanish L1 (Group A)	Basque L2 (Group A)	Basque L1 (Group B)	Spanish L1 (Group A)
Past perfective	42.2	40	2.4	74.2	81.4	50.6
Past imperfect	57.7	60	94	6.3	6.7	5.1
Present	0	0	2.4	14.8	8.4	35.4
Others	0	0	1.2	4.4	3.5	8.9

the inter-linguistic difference in the use of both aspect and tense is also particularly clear in the unfolding phase, which contains the main body of the story.

Data from the setting and unfolding sequences clearly show that although there are slight differences in the distribution of tense and aspect use between L1 and L2 speakers of Basque, the main difference is inter-linguistic. Again, past perfective forms are clearly predominant in Basque L1 and L2. In Spanish, past perfective forms are less frequent than in Basque, whereas present tense forms are more frequently used in Spanish than in Basque. However, L2 speakers of Basque produce more present tense forms than L1 speakers. In this context, it should be pointed out that the fairly high frequency of present tense forms in Spanish is due to the appearance of inserted dialogues. Some examples of this are given in Section 8.7 (examples (17), (18)). Inserted dialogues are also slightly more frequent in Basque L2 than in Basque L1 (example (16) in Section 8.7).

8.5    Verb agreement

As pointed out in Section 4, the complex Basque verb system with triple agreement marking for subjects, direct and indirect objects, as well as for number can be a source of difficulty for non-native speakers. This is the reason why we have analyzed possible errors in Basque L2 productions as compared to Basque L1. However, as Table 6 shows, agreement errors occur in both L1 and L2.

**Table 6.** Agreement errors in verbs (Basque L1 and Basque L2 children).

	Total number of verbs	% of verbs with agreement errors	% of texts with agreement errors
Basque L2 (Group A) (n=37)	1898	4.37 (n=83)	83.8% (n=31/37)
Basque L1 (Group B) (n=24)	900	4.66 (n=42)	66.7% (n=16/24)

In the first place, the low percentage of errors in both groups is remarkable (4.66 % in L1 and 4.37 % in L2). In the second place, the proportion of agreement errors is similar in both groups, even if the absolute number of verbs is much higher in the L2 group (it should be remembered that there are thirteen more subjects in this group). Finally, the percentage of texts with agreement errors is significantly lower for L1 speakers than for L2 children (66.7% vs. 83.8%, respectively). From a comparative point of view, these data could be interpreted in ways of intra-group homogeneity and heterogeneity. More L2 subjects (83.8 %) commit fewer errors, a fact that seems to indicate less inter-individual differences, whereas fewer L1 speakers (66.7%) commit more errors, which in turn seems to indicate more inter-individual differences.

## 8.6 Variety

The inter-linguistic differences shown in Table 5 are also apparent in Table 7, where the distribution of verb forms in all narrative sequences is summed up. As we have already pointed out, the higher percentage of present tense forms in Spanish is due to a more frequent use of inserted dialogues.

The percentage of other forms (as indicated in Table 7) is markedly higher in Spanish than in Basque. This fact is interesting since it reflects that the main difference produced is inter-linguistic: the subjects of group A do not reproduce the same pattern in their L2, Basque, as in their L1, Spanish. The percentages show more similarity between Basque L1 and L2 patterns than between Spanish L1 and Basque L2 in the same subjects.

## 8.7 Verb tense maintenance throughout the narration

Discourse analysis has revealed that each genre tends to hold to a basic or dominant tense, as e.g. the relation *imparfait* / *passé simple* in written narratives in French. It follows that if unexpected tense switching occurs, this produces problems in

**Table 7.** All verb forms in percentages (past, present and other forms).

	Past perfective-imperfective	Present	Others
Basque L2 (Group A)	76.4	17.5	6.1
Basque L1 (Group B)	87.2	8.4	4.4
Spanish L1 (Group A)	59.03	26.72	14.25



**Table 8.** Maintenance and non-maintenance of past tense anchoring in the narration (percentages).

	Narratives with dominant past tense	Narratives with tense mixing
Basque L2 (Group A)	92	8
Basque L1 (Group B)	71	29
Spanish L1 (Group A)	78	22

the cohesion of the text (Dolz 1990). In this context, Table 8 shows the extent to which our subjects are able to maintain the dominant tense of their texts.

As can be observed, our subjects are able to maintain the anchoring tense to a very high extent. In fact, the L2 speakers of Basque seem to be most systematic in this respect (92%). These same children are less systematic in Spanish L1 (78%), and the Basque L1 speakers are those who seem to pay less heed to this question (71%).

The following are a few examples of this kind of tense switching: in most cases (examples (16), (17), (18)), the switching occurs just before a dialogue is introduced. In some cases, however, it is more difficult to find an explanation, as in example (19).

- (16) *esan du/ atera azeria/*  
say AUX.TRAN.3.SG.PRES come.out fox-the  
*eta azeria atera zen/*  
and fox-the come.out AUX.INTR.3SG.PAST  
*hurrengo egunean/ e etorri zen*  
next day e come AUX.INTR.3SG.PAST  
'(He) says come out, fox! and the fox came out the next day e he came.'  
(Basque L2)
- (17) *y luego fue Centellita a casa*  
and then go (3SG.PAST) Centellita to house  
*de la bruja del fuego y le dice*  
of the witch of-the fire and her (IO) say 3SG.PRES.  
*hay tres mil pájaros*  
be IMPERS.PRES. three thousand birds  
*en el árbol y le dijo*  
in the tree and her (IO) say 3SG.PAST  
'and then Centellita went to the fire-witch's house and says to her: there are  
three thousand birds in the tree – and she said to her.' (Spanish L1)

- (18) *le dijo a casa*  
 him (IO) say 3SG.PAST to house  
*de la bruja de fuego le dice*  
 of the witch of fire her (IO) say 3SG.PRES  
*la la bruja- y la hada le dice*  
 the the wit- and the fairy her (IO) say 3SG.PRES  
 ‘(She) said to him – to the fire-witch’s house the the wi- tells her – and the fairy tells her.’ (Spanish L1)
- (19) *jun tzan ba-barrura/*  
 go AUX.INTR.3SG.PAST in-into  
*ta gero han ba/ zaldiengana jun*  
 and then there well horses-to-the go  
*da eta kris kras jun tzan*  
 AUX.INTR.3.SG.PRES and yum-yum go AUX.INTR.3SG.PAST  
 ‘(He) went in-into and then in there well he goes to the horses and yum-yum (he) went.’ (Basque L1)

Sometimes children correct themselves spontaneously. These corrections concern tense (20, 21) or mode (22) but not agreement. This fact can be interpreted as a feature of discourse awareness.

- (20) *esan zion*  
 say IO.AUX.TRANS.3SG.PAST  
*ea nahi baduzu*  
 if want AUX.TRANS.2SG.PRES  
*nahi bazuen e: berekin etorri*  
 want AUX.TRANS.3SG.PAST e him-with come  
 ‘(He) said to him if (you) want to if (he) wanted to e come with him.’  
 (Basque L2)
- (21) *eta esan zuen erregeak*  
 and say AUX.TRANS.3SG.PAST king-the (s)  
*eraman dutene*  
 take AUX.TRANS.3PL.PRES(-whe-)  
*eraman zutenean berriro/*  
 take AUX.TRANS.3PL.PAST (-when) again  
 ‘and the king said when take- when (they) took (him) back again.’  
 (Spanish L1)
- (22) *que si quería ayuda que le*  
 that if want 3SG.PAST help that her (IO)

*llamaba                    que    le                    llamara                    a    ella*  
 call 3SG.PAST.IND    that    her (IO)    call 3SG.PAST.SUBJ    to    her  
 ‘that if (she) wanted help that (she) called –should call her.’ (Spanish L1)

Curiously, no examples of self corrections have been found in Basque L1.

## 9. Summary and conclusions

The main hypothesis of our study postulating that young successive bilinguals acquiring Basque follow the same linguistic and discourse patterns as those who acquire Basque as L1 seems to be partly confirmed by our results. In the same sense, the design of our experiment has allowed us to observe that five to six-year-old children are able to reproduce stories which follow fairly closely the given genre model. This conclusion, however, can only be considered valid if certain conditions such as the communicative context of the task and the school training facilitate this reproduction.

In effect, all our subjects are able to structure their narratives coherently. The five phases are present in almost all stories, with slight variations. In this aspect, our L2 speakers of Basque are quite as competent as Basque L1 subjects.

The contrast between imperfective and perfective verb forms for background information and foreground events is produced by all our subjects, and furthermore with patterns that correspond to inter-linguistic differences attested among adult L1 speakers of Basque and Spanish (Table 1).

Verb density is higher in our productions by children than in adult productions. Again, inter-linguistic differences are visible: it is lower in Spanish and higher in both L1 and L2 Basque. It should also be pointed out that there is more variety of verb forms in Spanish L1 than in both Basque L1 and L2. An interesting result can also be observed concerning verb agreement. Basque L2 productions contain no more agreement errors than Basque L1 productions.

These results seem to indicate that our young Basque L2 subjects are developing their second language following patterns that are similar to Basque L1 children, and also different from their L1, Spanish. It would seem that the educational immersion programme they are taking part in has had positive effects. On the whole, the situation of the language contact these subjects are immersed in seems to lead to fairly well balanced bilingual language development.

Our Basque L2 subjects show more consistent maintenance of past tense anchoring in Basque than in Spanish, but also more than Basque L1 subjects, who appear to be more spontaneous in their story-telling. This consistency might be interpreted as a more rigid adherence to the adult model of narration, but also as

a positive effect of immersion teaching programmes, where story-telling is often practiced. The special help given to these children (Section 8.2) may also have reinforced certain reproductive patterns.

In Basque L2 more examples of inserted dialogue can be found than in Basque L1. This fact might explain why narratives are shorter in Basque L1. Shorter narratives do not necessarily reflect lower narrative competence, taking into account that more condensed versions may be found in L1 productions, a fact that has also been pointed out by Juvonen et al. (1989). A first analysis of our material shows that our L2 subjects tend to reproduce the given model to a higher extent than the L1 subjects, who are able to reformulate the text converting direct speech into reported sentences. This fact has been illustrated by example (20) in Section 8.7.

These are features that open up a promising field for future research on successive bilingualism.

## Abbreviations

AUX = auxiliary; IMPERF = imperfective; IMPERS = impersonal; IND = indicative; INTR = intransitive; IO = indirect object; PERF = perfective; PL = plural; PRES = present; SG = singular; SUBJ = subjunctive; TRAN = transitive.

## References

- Adam, Jean-Michel. 1992. *Les textes: types et prototypes*. Paris: Nathan.
- Akınç, Mehmet Ali. 2001. *Développement des compétences narratives des enfants bilingues turc-français en France âgés de 5 à 10 ans* [Studies in Language Acquisition 03]. München: Lincom.
- Akınç, Mehmet Ali & Kern, Sophie. 1998. Développement de la temporalité chez des enfants monolingues et bilingues. *BCILL 99 Temps et Discours*: 237–255.
- Aksu-Koç, Ayhan. 1994. Development of linguistic forms: Turkish. In *Relating Events in Narrative: A Crosslinguistic Developmental Study*, R. A. Berman & D. I. Slobin (eds), (329–385). Hillsdale NJ: Lawrence Erlbaum.
- Backus, Ad. 2004. Turkish as an immigrant language in Europe. In *The Handbook of Bilingualism*, T. Bhatia & W. Ritchie (eds), 689–724. Oxford: Blackwell.
- Bamberg, Michael. 1987. *The Acquisition of Narratives. Learning to Use Language*. Berlin: Mouton de Gruyter.
- Bamberg, Michael. 1996. Perspective and agency in the construal of narrative events. In *Proceedings of the 20th Annual Boston Conference on Language Development* Vol. 1, A. Stringfellow, D. Cahana-Amitay, E. Hughes & A. Zukowski (eds), 30–39. Somerville MA: Cascadia Press.

- Bamberg, Michael & Damrad-Frye, Robin. 1991. On the ability to provide evaluative comments: Further explorations of children's narrative competencies. *Journal of Child Language* 18: 689–710.
- Basque Statistics Office (2004): [www.eustat.es](http://www.eustat.es)
- Berman, Ruth A. 1988. On the ability to relate events in narrative. *Discourse Process* 11: 469–497.
- Berman, Ruth A. & Slobin, Dan I. (eds). 1994. *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale NJ: Lawrence Erlbaum.
- Bonnotte, Isabelle, Kaifer, Anton, Fayol, Michel & Idiazabal, Itziar. 1993. Rôle des types de procès et du co-texte dans l'emploi des formes verbales de la narration. *Langue Française* 97: 81–101.
- Brigaudiot, Mireille. 1993. Quelques remarques à propos du récit et des images à l'école maternelle. *Repères* 7: 7–21.
- Bronckart, Jean-Paul & Dolz, Joaquim. 1999. La notion de compétence: quelle pertinence pour l'étude de l'apprentissage des actions langagières? *Raisons Educatives* 2: 27–44.
- Bronckart, Jean-Paul, Bain, Daniel, Schneuwly, Bernard, Davaud, Clairette, Pasquier, Auguste. 1985. *Le fonctionnement des discours. Un modèle psychologique et une méthode d'analyse*. Paris: Delachaux et Niestlé.
- Bronckart, Jean-Paul. 1996. *Activité langagière, textes et discours. Pour un interactionisme socio-discursif*. Paris: Delachaux et Niestlé.
- Comeau, Liane & Genesee, Fred. 2001. Bilingual children's repair strategies during dyadic communication. In *Trends in Bilingual Acquisition*, J. Cenoz & F. Genesee (eds), 231–256. Amsterdam: John Benjamins.
- Deuchar, Margaret & Quay, Suzanne. 2000. *Bilingual Acquisition. Theoretical Implications of a Case Study*. Oxford: OUP.
- Dolz, Joaquim. 1990. *Catégorie verbale et action langagière. Le fonctionnement des temps du verbe dans les textes écrits des enfants catalans*. PhD dissertation, Université de Genève.
- Fayol, Michel. 1991. Stories. A psycholinguistic and ontogenetic approach to the acquisition of narrative abilities. In *Language Bases...Discourse Bases*, G. Piéaut-Le Bonniec & M. Dolitsky (eds), 229–244. Amsterdam: John Benjamins.
- Genesee, Fred. 1987. *Learning through Two Languages: Studies of Immersion and Bilingual Education*. Cambridge MA: Newbury House.
- Hamers, Josiane. 2006. Bilingüismo, multilingüismo y desarrollo mental. In *Palabras y mundos. Informe sobre las lenguas del mundo*, F. Martí, P. Ortega, I. Idiazabal, A. Barreña., P. Juaristi, C. Junyent, B. Uranga & E. Amorrortu (eds), 63–69. Barcelona: Icaria.
- Hamers, Josiane & Blanc, Michel. 1983. *Bilingüisme et bilinguisme*. Bruxelles: Pierre Mardaga éditeur.
- Hickmann, Maya & Roland, Françoise. 1992. Déterminants sémantiques et pragmatiques dans l'acquisition du temps et de l'aspect: Étude comparative du Français, de l'Allemand et de l'Anglais. Lyon: Réseau Européen de Laboratoires sur l'Acquisition des Langues.
- de Houwer, Annick. 1990. *The Acquisition of Two Languages from Birth: A Case Study*. Cambridge: CUP.
- Juvonen, Päivi, Lindberg, Inger & Viberg, Ake. 1989. Narrative skills in two languages. *Scandinavian Working Papers on Bilingualism* 8: 49–69.
- Lanza, Elizabeth. 1998. Cross-linguistic influence, input and the young bilingual child. Commentary to N. Müller: *Transfer in bilingual first language acquisition*. *Bilingualism: Language and Cognition* 1 (3): 181–182.

- Mayer, Mercer. 1969. *Frog, where are you?* New York NY: Dial Books for Young Readers.
- Meisel, Jürgen M. 2001. The simultaneous acquisition of two first languages: Early differentiation and subsequent development of grammars. In *Trends in Bilingual Acquisition*, J. Cenoz & F. Genesee (eds), 11–42. Amsterdam: John Benjamins.
- Müller, Natascha. 1998. Transfer in bilingual first language acquisition. *Bilingualism: Language and Cognition* 1: 151–192.
- Müller, Natascha & Hulk, Aafke C.J. 2001. Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition* 4 (1): 1–21.
- Pfaff, Carol W. 1999. Changing patterns of language mixing in a bilingual child. In *Bilingualism and Migration* [Studies on Language Acquisition 14], G. Extra & L. Verhoeven (eds), 97–121. Berlin: Mouton de Gruyter.
- Pfaff, Carol W. 2001. The development of co-constructed narratives by Turkish children in Germany. In *Narrative Development in a Multilingual Context* [Studies in Bilingualism 23], L. Verhoeven & S. Strömquist (eds), 153–188. Amsterdam: John Benjamins.
- Propp, Vladimir. 1937/1970. *Morphologie du conte*. Paris: Seuil.
- Quasthoff, Uta M. 1997. An interactive approach to narrative development. In *Narrative Development: Six Approaches*, M. Bamberg (ed.), 51–83. Hillsdale NJ: Lawrence Erlbaum Associates.
- Romaine, Suzanne. 1999. Early bilingual development from elite to folk. In *Bilingualism and Migration* [Studies on Language Acquisition 14], G. Extra & L. Verhoeven (eds), 61–74. Berlin: Mouton de Gruyter.
- Rosat, Marie-Claude. 1998. Comparaison des stratégies discursives d'étayage dans un conte et un récit d'expériences oraux. *Travaux Neuchatelois de linguistique* 29: 29–47.
- Schneuwly, Bernard & Dolz, Joaquim. 1998. *Pour un enseignement de l'oral. Initiation aux genres formels à l'école*. Paris: ESF.
- Sebastián, Eugenia. 1989. *Tiempo y aspecto verbal en el lenguaje infantil*. PhD dissertation, Universidad Autónoma de Madrid.
- Sebastián, Eugenia & Slobin, Dan I. 1995. Más allá del aquí y el ahora: El desarrollo de los marcadores temporales en el discurso narrativo en español. *Substratum* 5: 41–48.
- Stephany, Ursula. 1994. Tense and aspect in narrative child discourse. A comparison of Greek and French. Paper presented in First Lisbon Meeting on Child language, June 14–18.
- de Weck, Geneviève. 1991. *La cohésion dans les textes d'enfants*. Paris: Delachaux et Niestlé.



# Interrogative inversion in non-standard varieties of English

Michaela Hilbert  
University of Hamburg

This article investigates interrogative inversion in two varieties of English, Indian English and Singapore English. In both main and embedded clause interrogatives these varieties deviate from the standard in that they allow uninverted main clause and inverted embedded clause interrogatives. Both features occur in a range of other varieties as well; for some, substrate influence has been proposed as an explanation, which, however, cannot be applied to all varieties displaying the features in question. Alternatively, phenomena that occur across a range of historically unrelated varieties have been associated with a potential status as “vernacular universals” or “Angloversals”, an approach which so far has failed to provide a more precise explanation as to the origin of such “universal” tendencies. The present study will use an innovative approach comparing the factors governing the occurrence of inversion, rather than its mere presence or absence. It will show that only if these determining factors are parallel in the varieties under consideration, potential universal status can be claimed, and located in a specific linguistic domain, in this case in second language acquisition.

**Keywords:** interrogative constructions, inversion, embedded interrogatives, Indian English, Singapore English, L2 varieties, language shift, group shift, formulaic language

## 1. Introduction

Interrogative constructions are a field of syntax in which varieties of English deviate from Standard English in a wide range of features, the most important being the absence of inversion in main clause interrogatives, the presence of inversion in embedded interrogatives, the non-initial position of the interrogative word, the use of invariable interrogative tags and the development of interrogative particles.



Most of these phenomena characterize the “New Englishes” more than the traditional dialects of Britain. This is why they have been discussed for the most part under two perspectives: Approaches within the framework of contact linguistics have primarily studied single varieties, and have looked for origins in either the “superstrate” language (English), or the various “substrate” languages involved in the contact. More recent approaches have taken into account the fact that some of these features characterize a whole range of different varieties, and have hypothesized potential universal status for them, ultimately linking variation in vernaculars with cross-linguistic generalisations and functional typology (e.g. Kortmann & Szmrecsanyi 2004).

Either perspective usually misses the advantages of the other. Explanations claiming substrate influence to be the dominant source of a specific feature cannot neglect the fact that it might characterize varieties where no such explanation is possible. Irish influence, for instance, has been considered the source of embedded inversion in Irish English (Filppula 2004); it cannot, however, account for embedded inversion in Indian English. Approaches taking the universal perspective, on the other hand, often fail to look at the occurrence and non-occurrence of a specific phenomenon in more detail, neglecting the possibility that what looks the same on the surface might be governed by different determinants and thus might be based on different sources.

This article investigates interrogative inversion in two Asian contact varieties of English, Indian English and Singapore English. It will show that what is described as “lack of inversion” and “extension of inversion to embedded interrogatives” does not adequately describe the use of interrogative inversion in these two varieties of English. It will appeal for going beyond describing the occurrence of specific features and claiming potential universal (or other) status; instead it will appeal for finding the factors determining their occurrence or non-occurrence in order to actually *explain* their development. For this purpose, the precise domain in which a potentially universal tendency or mechanism seems to hold, and therefore a more precise notion of its source, needs to be determined.

I will proceed as follows: I will first give an overview of non-standard features in interrogatives and the explanations given in the recent literature (Chapter 1), and their theoretical status in the field of contact linguistics (Chapter 2). The quantitative analysis of interrogative inversion in Indian English and Singapore English based on the corresponding spoken sections of the ICE corpora (Chapters 3 and 4) will be compared to data from individual second language acquisition (Chapter 5). The results will be analyzed with a view to their implications for the further study of contact vernaculars and the quest for universals in non-standard varieties (Chapter 6).

## 2. Interrogative constructions in non-standard varieties of English

Varieties of English deviate from Standard English interrogative clauses in a wide range of features. These features include:

- lack of inversion in main clause interrogatives
  - (1) Why so many are being killed?  
(Pakistani English; Mahboob 2004: 1051)
- inversion in embedded interrogatives
  - (2) I asked her could I go with her.  
(African-American Vernacular English; Wolfram 2004: 334)
- non-initial position of the interrogative word
  - (3) They are going where?  
(Malaysian English; Baskaran 2004: 1078)
- the use of invariable interrogative tags
  - (4) You like that, isn't it?  
(Nigerian English, Alo & Mesthrie 2004: 817)
- the development of interrogative particles
  - (5) *á* in Jamaican Creole; *ngi* in Kriol

Lack of inversion in polar main clause interrogatives is a feature that can be found in an overwhelming majority of the so-called New Englishes, that is, in L2 varieties such as Indian English (emerging through the process of language shift, see next chapter), in pidgins such as Nigerian Pidgin English, and creoles such as Jamaican Creole. In polar interrogatives, lack of inversion is also common in (colloquial) traditional British and American dialects.

In post-creole continua, inversion begins to be used in the upper mesolect, presupposing that the periphrastic auxiliary *do* is available in the respective variety. Periphrastic *do* is reportedly being introduced in mesolectal Belize English, typically without agreement marker (Escure 2004: 539). In most creole and post-creole varieties there are constructions that superficially look like inversion, e.g.

- (6) Is wat? (Jamaican Creole)  
'What is it?'
- (7) Did him give you what you a look for? (British Creole)
- (8) Is they come yet? (African-American Vernacular English)

Cassidy (1971) interprets the Jamaican Creole example as the absence of expletive *it* rather than as an incidence of inversion. The AAVE example is likewise not interpreted as inversion, but analyzed as involving the initial question particle *is* followed by declarative word order. Similarly, Sebba (2004: 205) considers the *did him* in the British Creole example as a memorised invariable string rather than an incidence of “real” inversion.

All the varieties that do not use inversion in main clause interrogatives have intonation as the primary means of distinguishing their polar interrogatives from declaratives. Varieties that use intonation in this way have rising intonation in polar interrogatives and tags, and falling intonation in constituent interrogatives. If a variety has distinctive grammaticalised intonation patterns, it has them for polar interrogatives. In Jamaican Creole, for instance, intonation in polar interrogatives is a series of downsteps after a high (“final cadence”), which is also used in relative constructions.

Inversion in embedded interrogatives is not as pervasive as in main clause interrogatives; still there is a range of different varieties displaying the feature. It is generally attested for in the following varieties of English (based on Kortmann et al. 2004):

- |                                       |                                |
|---------------------------------------|--------------------------------|
| – Irish English                       | – Cameroon English             |
| – Scottish English                    | – East African English         |
| – Welsh English                       | – Black South African English  |
| – Newfoundland English                | – Indian South African English |
| – Colloquial American English         | – Indian English               |
| – Appalachian English                 | – Singapore English            |
| – Gullah                              | – Malaysian English            |
| – African-American Vernacular English | – Aboriginal English           |

The list includes L1 varieties as well as L2 and post-creole varieties, so that the feature can be said to exist beyond borders of region and type of variety.

Embedded inversion can occur in both polar and constituent interrogatives. It does so in all varieties of Irish English (Filppula 2004: 93f.). The pervasiveness of the phenomenon in present-day Irish English depends on the region (it occurs more frequently in the south and west), and on the type of interrogative (it is more frequent in polar than in constituent interrogatives).

Henry (1995), writing about Belfast English, also notes a preference for the construction in polar interrogatives. She observes that the inverted form is even considered ungrammatical in constituent questions by some speakers. Belfast English also allows, in contrast to other Hibernian English (HibE) dialects, the use of *that* without inversion, as in *I wonder which dish that they picked*. Inversion

is also possible in extracted clauses (*Who did John hope would he see?*), where it occurs with verbs that do not normally take *wh*-complements.

Embedded inversion is a widespread feature throughout spoken Scottish English, also in more formal contexts, and currently finds its way into written texts, including not only letters but also newspapers (Miller 1993). The phenomenon is also mentioned for Tyneside English (Beal 1993). Filppula (2004) finds the construction only rarely in his corpus of southwestern British dialects and in the Survey of English Dialects (for the north and West Midlands), but he avoids a generalisation on the basis of his incomplete data.

The fact that these features (and in fact quite a few others) occur in several varieties, across the borders of region and type of variety, has given rise to hypotheses claiming that they are potential universals of certain domains of language. Firstly, Mair (2003) tentatively introduces the notion of “Angloversals”, as joint tendencies in postcolonial varieties which are not linked historically or genetically (Mair 2003: 84). As to the explanation of potential “Angloversals”, Mair suggests learner strategies of non-native speakers as a likely source. Within this concept, both lack of inversion and embedded inversion might serve as potential candidates, as both dominantly characterize L2 varieties, although especially the former are by no means restricted to them.

Secondly, Chambers (2004) introduces the concept of “vernacular universals”, in his terms

a small number of phonological and grammatical processes [which] recur in vernaculars wherever they are spoken. This conclusion follows from the observation that, no matter where in the world the vernaculars are spoken [...] these features inevitably occur. (Chambers 2004: 128)

As to the explanations of these universal features, Chambers claims the only potential source to be “natural structural linguistic developments [...] natural outgrowths, so to speak, of the language faculty, that is, the species-specific bio-program that allows (indeed requires) normal human beings to become *homo loquens*” (Chambers 2004: 128). Although lack of inversion in interrogative clauses is not mentioned by Chambers in his list of potential universal candidates, it would probably be a good one, considering the fact that hardly any other non-standard phenomenon characterizes the various groups of English varieties so consistently.

Chambers’ concept falls short, however, of providing any approach to *predict* the occurrence of potential universals in a given variety, considering that even the most pervasive phenomena in varieties of English do not occur in every one of them. As Kortmann & Smrecsanyi (2004: 1192ff.) point out in their overview of non-standard morphosyntactic features, the type of variety, based on its genesis,

seems to govern the occurrence of specific features more than region or status, so that we will expect variation largely to be based on the historical evolution of specific groups of varieties. Actual explanations must be grounded in specific domains in which universal status is claimed to hold, such as language change, language acquisition or language contact (Siemund 2008).

### 3. Varieties of English from the perspective of contact linguistics

#### 3.1 Thomason/Kaufman's classification of language contact scenarios

A central field of classifying English varieties according to their genesis, and thus of predicting the major determinants of their development, is the field of contact linguistics. Its central aim is to predict the outcome of specific language contact situations. A major model is the one by Thomason/Kaufman (1988), who claim that social factors rather than linguistic ones primarily determine the outcome of language contact: "it is the social context, not the structure of the languages involved, that determines the direction and the degree of interference" (Thomason/Kaufman 1988: 19).

Based on this initial hypothesis, Thomason/Kaufman (1988) distinguish three major types of language contact situation: language maintenance, language shift and language creation. The first, language maintenance, applies to situations in which speakers keep their native language from generation to generation. Observable changes in the languages come gradually and are primarily due to internal processes, and to smaller degrees due to the language contact(s) in question. Contact-induced changes are introduced by native speakers, not the speakers they are in contact with (Thomason/Kaufman 1988: 37).

Contact-induced changes brought about by language shift are predicted to affect phonology and grammar primarily rather than the lexicon (Thomason/Kaufman 1988: 38f.). The fact that these changes are introduced by non-native speakers implies that many of them are associated with phenomena of imperfect (group) learning, or "learner errors". The linguistic results of language shift are subsumed as *substratum interference*, "a subtype of interference that results from imperfect learning during a process of language shift" (Thomason/Kaufman 1988: 38). Various other terms have been suggested for deviations from the target language in the speech of shifting speakers, such as *interference* (whether modified as *substratum*, *superstratum*, or *adstratum interference*), *retention* and *transfer*.

The third scenario, language creation, involves the creation of a new language which is clearly distinct from the languages involved in the contact. The result

**Table 1.** Contact-induced change in English varieties.

<b>Language maintenance</b> <i>L1 varieties</i>	<b>Language shift</b> <i>L2 and shift varieties</i>	<b>Language creation</b> <i>Pidgins and creoles</i>
British English	Irish English	Jamaican Creole
American English	Indian English	Nigerian Pidgin English
Australian English	Singapore English	etc.
New Zealand English	Malaysian English	
South African English	Nigerian English	
etc.	Ghanaian English	
	Jamaican English	
	Cameroon English	
	etc.	
	??? African-American English ???	

is either a mixed language, including features of both or all languages involved without one being dominant on the whole; or pidgins and creoles, which might also involve features of all languages involved, but which are more clearly distinct from the latter due to a greater amount of features which originate in internal processes independent from the source languages. The result of these processes is a language which is considerably distinct from the languages originally involved in the contact, so that “[t]hese outcomes involve such extreme restructuring that they cannot be considered cases of either maintenance or shift in the strict sense of those terms” (Winford 2003: 18).

Thomason/Kaufman’s distinction between language maintenance, language shift and language creation would classify varieties of English into three categories, according to the origin of *contact-induced* changes in the respective variety (Table 1).

The status of African-American Vernacular English (AAVE) among English varieties has been debated for a long time. The most common claims are that AAVE is either a dialect of English like other regional dialects; or that AAVE is a creole, or is originally based on one. In Thomason/Kaufman’s model, AAVE would definitely not be classified as an L1 variety, due to the fact that some kind of shift has taken place at the beginning of its history. Thus, the question to be answered along these lines is rather whether AAVE has developed more along the L2- or along the creole line.

As the varieties studied in this article are both L2 varieties, the determinants of language shift according to Thomason and Kaufman will be outlined in more detail here.

3.2 Varieties of English as the result of language shift

In order to make hypotheses about the outcomes of language contact in L2 varieties of English within the framework of Thomason/Kaufman's (1988) model, two separate groups of linguistic factors have to be considered, which potentially influence the results of language shift. These are

- factors related to (individual) L2 acquisition;
- factors related to group shift.

The interplay of the two factors is displayed in Figure 1.

Principles of (individual) second language acquisition are obviously a primary source of information in Thomason/Kaufman's (1988) model of language shift, in as much as they are introduced by non-native speakers into the affected language: "To a considerable extent these learners' errors are directly comparable to shift-induced language change" (Thomason/Kaufman 1988: 145). Similarly, the study of the linguistic results of language shift provides insights into the processes and developmental stages of individual second language acquisition. Developmental stages in second language acquisition are transitory and thus difficult to study; the results of language shift, on the other hand, have been claimed to reflect an intermediate stage of learner development: "Fossilized intermediate or advanced stages in second language acquisition, such as [...] Irish English, provide better opportunities for the study of systematic effects of first-language interference" (Thomason/Kaufman 1988: 145f.).

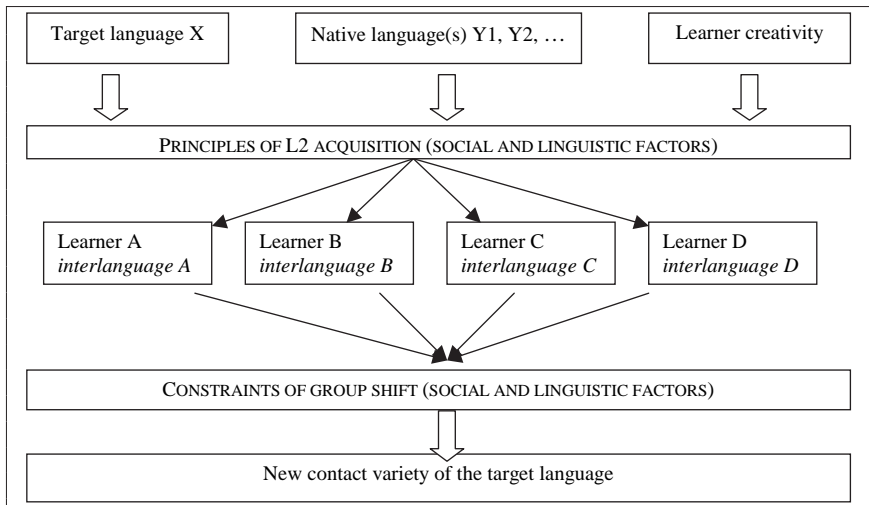


Figure 1. Determinants of group language shift.

Based on the assumption that second language acquisition lies at the heart of the emergence of an English contact variety, it has to be conceptualised why and how these varieties display “fossilised” stages of individual second language acquisition. Selinker’s (1987) original concept of fossilisation primarily refers to individual L2 acquisition and defines it as “linguistic items, rules, and subsystems which speakers of a particular NL [native language; MH] will tend to keep in their IL [interlanguage] relative to a particular TL [target language] no matter what the age of the learner or amount of explanation and instruction he receives in the TL” (Selinker 1987: 17). Explanations of why features of such intermediate subsystems become stabilised in a whole shifting speech community are mostly based on the fact that during the acquisition process, the target of learners shifts from native speakers of the target language to early bilinguals in the non-native speech community:

Often too, the shift can take place over several generations, and involve successive stages of bilingualism among different sections of the shifting group. As a result, over time, the target for new learners would have been second language varieties of the TL spoken by bilinguals. (Winford 2003: 245)

This fossilisation of early acquisition stages implies the significance of establishing developmental stages in L2 acquisition, in order to predict shift-induced changes in L2 varieties of English:

It is important to investigate these stages if we are to fully understand the role of TL input, reduction, L1 “transfer” or retention, and internal change in the learning process, as well as the principles that constrain it. These developmental stages, as well as the processes and constraints that guide them, are (also) very relevant to the emergence of contact vernaculars such as pidgins, creoles, and the “indigenized” varieties of a TL that result from group SLA or shift. (Winford 2003: 221)

Linguistic subsystems for which developmental stages are relatively clear and to some extent universal, are negation, relative clause formation, and general word order. The general development in all these fields is claimed to be that “a single invariant rule is applied across the board in the first stage, followed by gradual acquisition of more specific rules in later stages” (Winford 2003: 221).

To sum up, this model of the emergence of English contact vernaculars predicts that the variation we find in varieties involving language shift reflects individual L2 acquisition phenomena to the extent that developmental stages of the latter should be evident in the synchronic variation displayed in the former. In order to test this hypothesis, interrogative inversion in two contact varieties of English, Indian English and Singapore English, will be outlined and compared to the development of interrogative inversion found in individual L2 acquisition.



## 4. Interrogative inversion in Indian English and Singapore English

### 4.1 Inversion in Indian English: a “mirror image”?

Indian English, given that what we look at is vernacular speech, is commonly described as having no inversion of subject and verb in main clause interrogatives, both in polar and constituent ones (cf. e.g. Bhatt 2004: 1020). The same descriptions usually also comment on the presence of inversion in embedded clauses and express surprise at this phenomenon which seems to turn the Standard English rule completely around:

The simple empirical generalization that emerges [...] is that in vernacular IndE [Indian English] inversion is restricted to embedded questions; it does not apply in matrix questions. The question formation strategy in vernacular IndE is the mirror image of that of StIndE [Standard Indian English; and Standard English in general; MH], where inversion is restricted to matrix contexts.

The fact that direct *wh*-questions in IndE do not invert is not mysterious; StInd/Br/AmE [Standard Indian English/British English/American English] questions with the question phrase *how come* [...] do not invert either. (Bhatt 2004: 1020)

Two aspects are of importance here: firstly, the seeming mirror image of main clause and embedded interrogatives, which seems implausible considering that for many varieties embedded inversion is claimed to be the result of overgeneralisation (mostly in the sense of hypercorrection) of the main clause rule; secondly, the question of the source of uninverted main clause interrogatives arises, which Bhatt (2004) ascribes to the overgeneralisation of structures like *how come*, but which does not seem plausible given the rarity of the structure. The problem remaining within these two approaches is that both presuppose the existence of different rules (non-inversion in the main clause, inversion in the embedded clause), resulting in the paradox that the two are incompatible.

In the following section, the use of inversion in both main and embedded interrogatives will be studied in detail, using the spoken sections of the ICE (International Corpus of English)-India corpus. Data from Singapore English, which is generally described to display very similar properties as Indian English, will be directly compared, in order to see whether the same determinants govern the variation.

The discussion will be largely restricted to constituent interrogatives as they are unambiguously retrievable from the corpus by means of the interrogative pronouns. The analysis of polar interrogatives in Indian English is based only on the mark-up in ICE-India, which has question marks in part of the spoken section of the corpus, a mark-up which is not used in the other ICE corpora. Main clause constituent interrogatives are selected independently of this mark-up.

#### 4.2 Main clause polar interrogatives (Indian English)

Lack of inversion occurs in about one third of polar interrogatives in ICE-India (Table 2). This number might be surprising given that main clause polar interrogatives without inversion are perfectly acceptable in spoken British English, so that the number of such uninverted main clause polar interrogatives could have been expected to be higher in Indian English than it actually is.

The interrogative clauses with standard-like inversion show a preference as to the verb type which is inverted. Most inversions occur with forms of *be*, among which an overwhelming portion consists of *is it* and *are you*. Results for the two other groups of verbs, modal auxiliaries and full verbs, are about roughly the same (Table 3).

Compared to spoken British English as covered by ICE-GB (Table 3), the prominence of *be* in Indian English is only slight and might as well result from the frequency of forms of *be* in the corpus as a whole. Frequency, however, is not the governing factor. Figure 2 shows that of all occurrences of the respective verb types, *be* is indeed the verb that is most likely to be inverted, with the modal auxiliaries following. Full verbs including *do*-periphrasis are considerably less likely to be inverted.

Breaking down the inverted clauses, those with full verbs and *do*-periphrasis show a clear clustering around the two initial phrases *do you* and *did you*, which together amount to 80 percent. If we add *don't you* to this group of frequent question-introducing chunks, cases of inversion that do not involve such frequent chunks constitute only twelve percent of all inverted polar interrogatives (Table 4). Examples from the corpus include:

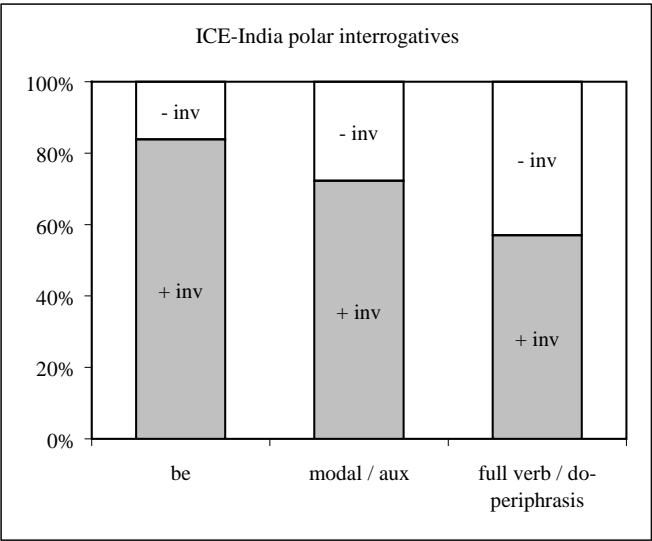
- (9) Do you add a bit of flour later? (ICE-India-S1A-007)

**Table 2.** Inversion in main clause polar interrogatives in ICE-India.

	Tokens	%
+ inv	474	63
– inv	261	37

**Table 3.** Basic verb type in standard-like inverted polar interrogatives.

+ Inv	ICE-India %	ICE-GB %
<i>be</i>	38	33
modals, <i>have/has</i>	30	30
full verb	32	32



**Figure 2.** Percentage of inverted and uninverted polar interrogatives within basic verb type (ICE-India).

- (10) Did you buy sarees?

(ICE-India-S1A-031)
- (11) Don't you think so?

(ICE-India-S1A-032)

The hypothesis that inversion with full verbs and *do*-periphrasis is largely based on fixed chunks like *did you* is supported by the fact that there are several examples of the type [did] + [subject] + [V-ed] in the corpora, in which this “question-introducing” chunk is followed by main clause word order and structure. The same analysis could, of course, also hold for *do you*, which can also be analyzed as [do] + [main clause], only that in this case the strategy is not detectable as no double marking can occur.

- (12) Did he continued as a teacher?

(ICE-India-S1A-035)

**Table 4.** Inversion with *do*-periphrasis in polar interrogatives (ICE-India).

+ Inv	%
<i>do you</i>	57
<i>did you</i>	24
<i>do non-you</i>	10
<i>don't you</i>	6
<i>did non-you</i>	2

### 4.3 Constituent interrogatives (Indian English and Singapore English)

Constituent interrogatives in Indian English display more inversion than polar interrogatives (Table 5). This might again be surprising, given that inversion can be seen as redundant when it comes to strategies of marking a clause as interrogative, with the interrogative word in initial position fulfilling this function much more clearly.

The analysis of inversion in constituent interrogatives shows roughly the same picture as in polar interrogatives, though much clearer. Again, inversion with *be* is more common than with modals and *have*, and with *do* for full verbs. Here, however, the relative number of forms of *be* is much higher than for polar interrogatives (Table 6).

The results of Singapore and Indian English are absolutely parallel, and taken together they support the conclusions drawn from main clause polar interrogatives. Inversion is again largely dependent on *be*, and full verbs with *do*-periphrasis constitute about one third of the sample.

Again, however, the difference from British English is slight and only quantitative in nature. As for polar interrogatives, the prominence of *be* becomes apparent once the distribution within each type of verb is focused on (Figures 3 and 4). In both Indian and Singapore English *be* occurs more often in inverted clauses than modal auxiliaries and full verbs with *do*-periphrasis. The differences between the three verb types are clearer in Indian than in Singapore English, which might be interpreted as a greater degree of fossilisation of the pattern in the former variety.

The fact that the prominence of *be* in inverted constituent interrogatives is not due to its overall frequency is shown in Table 10. If frequency was the governing factor, inverted and uninverted interrogatives should display the same distribution of

**Table 5.** Inversion in polar and constituent interrogatives (ICE-India).

	Polar interrogatives		Constituent interrogatives	
	tokens	%	tokens	%
+ inv	474	63	540	81
– inv	261	37	76	11

**Table 6.** Basic verb type in standard-like inverted constituent interrogatives (in %).

+ Inv	ICE-India %	ICE-Singapore %	ICE-GB %
<i>be</i>	57	62	55
full verb	36	30	33
modals, <i>have/has</i>	7	9	12

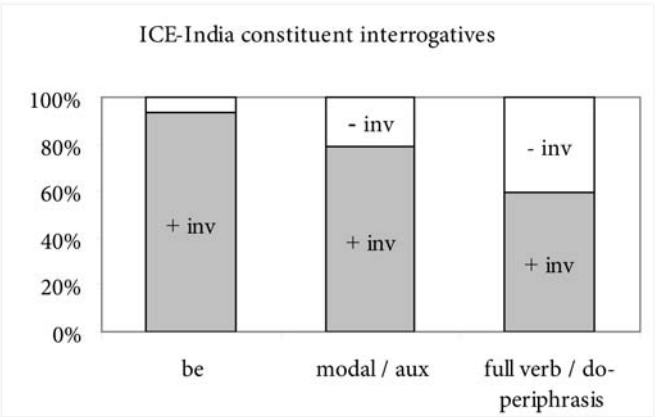


Figure 3. Inverted and uninverted constituent interrogatives within basic verb type (ICE-India).

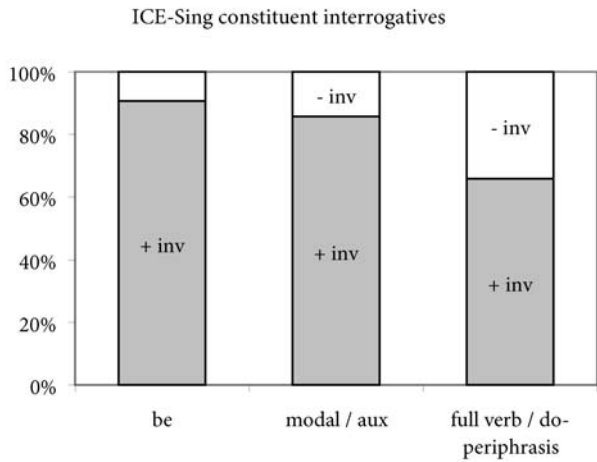


Figure 4. Inverted and uninverted constituent interrogatives within basic verb type (ICE-Singapore).

verb types. The reverse order of verb types, however, is present, with full verbs and *do*-periphrasis constituting more than 70 percent in uninverted clauses (Table 7).

In most of these uninverted interrogatives involving *be*, *be* functions as a present progressive auxiliary, a phenomenon which also occurs in Singapore English, but which is hard to explain at this stage of research.

Breaking down inversion with *do*-periphrasis for full verbs again reveals the same picture as for polar interrogatives. In ICE-India, *do you* and *did you* amount to more than 80% of standard-like inverted clauses; if we include *don't you* and

**Table 7.** Lack of inversion in constituent interrogatives (in %).

– Inv	ICE-India	ICE-Singapore
full verb	72	65
be	17	30
modals, <i>have</i>	11	5

**Table 8.** Inversion with *do*-periphrasis in constituent interrogatives (in %).

+ Inv	ICE-India %	ICE-Singapore %
<i>do you</i>	64	25
<i>did you</i>	21	13
<i>do non-you</i>	8	13
<i>don't you</i>	3	30
<i>didn't you</i>	2	7
<i>did non-you</i>	1	11

*didn't you*, which are frequent chunks in *why*-interrogatives, the numbers add up to 90%, leaving a mere ten percent for inversions not relying on frequent question-introducing chunks, which is the same as for polar interrogatives. In ICE-Singapore the rate for *did you* is considerably lower, but *don't you* is more pervasive, on the other hand, so that what can be regarded as fixed chunks also add up to about 75 % (Table 8). This is especially true for *why*-interrogatives, where *why don't you* and *why didn't you* constitute about one half of the total number of main clause *why*-interrogatives.

As for polar interrogatives, overgeneralisations can also be found with constituent interrogatives, again involving *did* + main clause:

- (13) When did it happen? (ICE-Singapore-S1A-093)

## 5. Embedded constituent interrogatives

The majority of embedded interrogatives in both corpora are uninverted as in Standard English (Table 9).

Inversion in embedded interrogatives is heavily restricted as to the verb involved in the interrogative clause. Inversion almost exclusively occurs with forms of *be* in both varieties (Table 10).

The “other” category comprises all verbs which are represented in only one instance each, such as *did* + full verb, *does* + full verb, and some modals. All of these occur with *how* in Singapore English.

**Table 9.** Inversion in embedded constituent interrogatives (in %).

	ICE-India	ICE-Singapore
+ inv	18	14
– inv	82	86

**Table 10.** Inverted verbs in embedded constituent interrogatives (in %).

+ Inv	ICE-India	ICE-Singapore
<i>is</i>	49	43
<i>'s</i>	44	33
<i>are</i>	2	11
<i>would</i>	–	3
<i>why don't</i>	2	3
<i>was</i>	3	2
other	–	4

These results cast strong doubts on the idea that Indian English has inversion in embedded interrogative clauses. Rather, it suggests that the forms of *be* are closely associated with the interrogative word itself, so that they are produced in direct sequence to it. The conclusion that the inversion of subject and verb in these clauses consists of a fusion of interrogative word and *be*, rather than an extension of the rule for main clause interrogatives, is supported by examples involving overgeneralisations, in which the form of *be*, mostly the clitised variant, follows the interrogative word, with the full non-inverted main clause following:

(14) So at that time I didn't know what's was indication and after that was when  
I met this guy uh in River Valley swimming pool. [ICE-India-S1A-065]

(15) Don't understand how's it is coming up. [ICE-Singapore S1A-053]

To sum up, “inversion” in embedded interrogatives in Indian English and Singapore English is largely restricted to forms of *be*, mostly in the form of *is*, *'s* and *are*. This reflects the inversion patterns found in main clause interrogatives. Also most inversion other than with *be* involves a large proportion of other fixed chunks such as *do you* and *did you*, as well as *why don't* and *why didn't* for *why*-clauses.

The presence of inversion in Indian English and Singapore English is thus based on the availability of frequent and fixed chunks, rather than the application of a productive rule. Inversion in embedded interrogatives is not a case of overgeneralisation of the main clause word order to embedded clauses, but rather again of the availability of fixed chunks that are used to introduce an embedded interrogative clause. The overwhelming majority of cases that do not fit into these

patterns remain uninverted in both main clause and embedded interrogatives, and in both the varieties under consideration.

The close parallels in the internal variation in the two varieties suggests a common explanation of where these comparable tendencies come from. Supporting the hypothesis outlined in Chapter 2, the tendencies observed for the two L2 varieties under study are largely paralleled by data from individual second language acquisition.

## 6. Individual L2 acquisition of English interrogatives

### 6.1 Polar interrogatives

The individual second language acquisition of English interrogatives has been studied extensively. Most of the seminal studies of the 1970s focus on one child or a small number of children who learn a second language while living in the native speech community of the target language. Fewer studies involve adult learners; these usually focus on the imperfectness or incompleteness of the learning process, and sometimes compare their results to those of pidginisation processes. The field has become increasingly heterogeneous, involving classroom and university teaching in addition to native speaker environment, and a greater diversity of languages, involving for instance Arabic and Chinese in addition to Indo-European languages like Spanish and French. Due to this heterogeneity, I will restrict my survey to those tendencies that appear to be common in all studies, independent of the learning context and first languages involved.

The very first polar questions produced by learners are usually one-word and two-word utterances, usually without a verb, such as *me? Like that?* Among these early elliptic utterances, Tiphine (1983:225), whose study involves four children with L1 French, finds one productive pattern for her subjects: a subjectless clause with initial *is*:

(16) Is me?

(17) Is my turn?

The first non-elliptic polar questions produced by learners of English are most often statements with main clause word order (i.e. without inversion) and rising intonation (e.g. Butterworth & Hatch 1978; Cancino 1978; Huang & Hatch 1978; Ravem 1978; Felix 1981; Ellis 1984; Brines 1990). Uninverted polar interrogatives seem to be produced by all learners of English, often but not necessarily before inverted ones (cf. Cancino 1978:219).



Early inversions are usually not analyzed as evidence for the acquisition of the inversion rule by the respective authors. Three patterns of such early inversion can be distinguished: inversion with the copula *be*, inversion with frequent modal verbs, and inversion involving a full verb and *do*-periphrasis. Firstly, inversion with the copula *be* is present very early for the great majority of subjects. Tiphine (1983: 225) finds inversions with *be* before all other instances of inversion in polar interrogatives, such as

(18) Are you?

She considers them formulaic for three reasons: Firstly, because they occur much earlier than any other instances of inverted interrogatives; secondly because these early instances never occur uninverted; and thirdly because subject-verb inversion, notably in the form of “is that”, also occurs in declarative clauses produced by her subjects (Tiphine 1983). Likewise, Ellis (1984), in a study of ten to thirteen-year-old subjects living in Britain (L1s Punjabi or Portuguese), stresses that *be* is the only verb for which inversion becomes a frequent strategy in the early stages of acquisition: “There are few examples of inverted Yes-No questions and those that occur are almost entirely restricted to BE” (Ellis 1984: 36).

Secondly, other early target-like inversions are also consistently inversions involving an auxiliary present in the respective declarative clause (e.g. Adams 1978). Central in this development are frequent modal verbs such as *can*. Tiphine (1983: 256) finds that *can* is always correctly inverted in her study. The process of stabilising inversion in polar interrogatives first occurs as an extension to other modals before it is extended to full verbs (cf. Adams 1978: 285f.; Tiphine 1983: 231).

Thirdly, inverted polar interrogatives involving a full verb and *do*-periphrasis appear late in most studies. Early instances of *do*-periphrasis often involve declarative clauses with a present tense main verb preceded by *do*, as in

(19) You live in a big house.      Do you live in a big house?

Shapira (1978: 251) finds that the inverted polar interrogatives produced by her subject (Zoila, 22, L1 Spanish) exclusively consist of *do you* + VP. Overall, the percentage of this construction is 21%, as against 79% percent declaratives plus rising intonation (Shapira 1978: 251). The fact that this construction is neither extended to other persons nor to other tenses suggests that it is a formulaic expression rather than a productive strategy (Shapira 1978: 252).

Adams (1978: 286) shows that this use of *do you* is later paralleled by chunks involving *does* and *did*, although Brines (1990: 244) finds that his 16 subjects (14–15 years old, complete beginners in English, L1 Spanish) generally tend to

simplify *does* in favour of *do*. There is evidence that chunks involving *did* are not only produced as imitations, but are also generalised to some sort of interrogative particle. A case in point is the frequent double tense marking when *do*-periphrasis is present (Adams 1978:286):

(20) Did you saw [...]

As to the kinds of full verbs involved, early instances of *do*-periphrasis are followed by frequent main verbs like *have* and *know* (e.g. Adams 1978:285). The fact that these are high-frequency verbs in spoken discourse (*does she have*, *do you know*), it is again likely that they are formulaic rather than early instances of an acquired rule.

To sum up, the results surveyed in this chapter suggest that in the acquisition of English polar interrogatives certain preferences of imitation and rule formation can be found. Firstly, imitation involves both full interrogative clauses (e.g. *How are you?*) and memorised chunks followed by variable sentence elements (e.g. *What's + NP*). The earliest and most common combinations are

*is + NP*  
*are you*  
*can you*  
*do you have, do you know*  
*do you + VP; did you + VP*

Secondly, the learners attempt to find and apply productive strategies of constructing interrogative clauses (here called 'rule formation'). As mentioned above, authors have often found it difficult to determine whether a learner's utterance is based on imitation or the application of a productive rule. Evidence of rule formation and application, however, might be provided by cases of overgeneralisation, i.e. the production of non-target like utterances. These have been reported for

*is + NP / declarative clause*  
*do + declarative clause (involving an auxiliary)*  
*did + declarative clause (in the past tense)*

If it is assumed that the application of rules involves cases of overgeneralisation, it has to be concluded that the inversion rule itself, i.e. the inversion of subject and verb, is not learned as such, but that target-like inversion develops adding more and more chunks of the pattern [(invariable) question marker] + declarative clause. This hypothesis is confirmed by the fact that in contrast to the latter three structures (*is + NP/declarative clause*, *do/did + declarative clause*), inversion is hardly ever overgeneralised to full verbs (i.e. without *do*-periphrasis). This

would result in sentences involving the full verb in initial position followed by the subject. This phenomenon, however, is very rarely found. Ravem (1978: 151) reports it for one of his subject's (Rune, age 6;5, L1 Norwegian) early polar interrogatives:

(21) Know you?

Ravem (1987) attributes this to the child's L1 Norwegian, in which full verb inversion is grammatical (and required), so that this is not analyzed as an overgeneralisation of the inversion rule but as L1 interference. In fact, most cases of inversion with full verbs are reported for learners whose L1s allow this structure, e.g. Arabic (cf. Mukattash 1981 for constituent interrogatives).

## 6.2 Constituent interrogatives

As in polar interrogatives, lack of inversion is a highly frequent feature in the early L2 acquisition of English constituent interrogatives. Early instances of inversion are, again, often analyzed as instances of memorised chunks rather than a productive use of inversion. Thus, the first seemingly inverted constituent interrogatives involve initial *where's* and *what's* (e.g. Huang & Hatch 1978: 130).

(22) Where's pen?

According to Huang & Hatch (1978: 129f.), there is no stage in which their subject (Paul, age 5, L1 Taiwanese) fronts the interrogative word only, neither is there a following stage in which he inverts subject and verb, nor a subsequent stage in which he contracts the copula. Rather, the copula is considered to be learned as part of the initial interrogative word, as the separation of the interrogative word and the copula appears only later (Huang & Hatch 1978: 129f.):

(23) Where is ball?

The copula *is* is inverted in Cancino's (1978) study in all constituent interrogatives from the beginning on (Cancino 1978: 220). Other early target-like instances involving the copula *be* that are analyzed as formulaic questions by the respective authors include:

(24) What's your name? (Butterworth & Hatch 1978: 239)

(25) What time is it, please? (Butterworth & Hatch 1978: 239)

(26) How are you? (Shapira 1978: 252)

(27) How old are you? (Butterworth & Hatch 1978: 239)

(28) How is NP? (Shapira 1978:252)

(29) How do you say X? (Butterworth & Hatch 1978:239)

Apart from clearly memorised elements, Adams (1978:286f.) claims *be* to be the first verb that is productively inverted, earlier than modal verbs. She argues that inversion with *be* is actually acquired, because it is extended to embedded constituent interrogatives. Tiphine (1983:243) comes to the same conclusion. Interestingly, she adds that this is especially true for *what* and *where*. This would, however, parallel the early and highly frequent occurrence of *what's* and *where's* found in questions as mentioned by Huang & Hatch (1978), as outlined above, and would again point to a memorised chunk interpretation rather than to instances of productive rule and application.

Overgeneralisations with *be* occur in Mukattash (1981:324), which are ascribed to the chronological priority given to inversion with copular *be* in classroom teaching:

(30) Why are the boys ran home?

As in polar interrogatives, the second group of early interrogative inversion involves frequent modal verbs. Again, they typically involve modal *can*, before other auxiliaries (e.g. Hatch 1974). *Do*-periphrasis usually occurs relatively late in the acquisition constituent interrogatives. There are early instances of negated *do*-periphrasis, Ravem (1978) doubts, however, that these can be considered productive, but rather that they are learned “as an unsegmented negative particle” (Ravem 1978:151).

Other frequent chunks again involve *do you*: Tiphine (1983:247) concludes that the high frequency of correctly used *what*-clauses with *do*-periphrasis is due to formulaic expressions involving *what do/does*, similar to *what is / what's* in other studies. The same occurs later for *what did you*. There is some transference of the pattern to *why did you*. In addition to these cases of imitation, Tiphine's (1983:250) subjects use *what do you* as an all-purpose question marker. Other examples of overgeneralisation include:

(31) Why does the tail are too short? (Tiphine 1983:254)

(32) Where do you was? (Tiphine 1983:254)

(33) Where does/do Mrs. Smith was cooking? (Mukattash 1981:323)

(34) Who did want to leave? (Mukattash 1981:323)

In these examples, *do* is overgeneralised to contexts in which the inverting auxiliary verb would actually be present, either the copula *be*, as in (31) and (32) or auxiliary *be*, as in (33). In contexts of classroom or university teaching “[i]t seems

that some students think that the auxiliary *DO* is always required in the formation of *Wh*-questions regardless of the type of verb incorporated in the declarative sentence” (Mukattash 1981:323). (34) illustrates the extension of *do*-periphrasis to a context in which inversion is not required. “What is involved [...] seems to be an instance of ‘false analogy’ or ‘ignorance of rule restriction’ [...] where students are accustomed to using the morpheme *DO* in the absence of an auxiliary (Mukattash 1981:323). Similar instances of overgeneralisation can be found in Ravem (1978).

To sum this up, the acquisition of constituent interrogatives seems to develop very similarly to the acquisition of polar interrogatives. Again we can find evidence of imitation and rule formation. Frequently imitated chunks involve

*where’s / what’s + NP*  
*what do / what does + declarative clause*

Evidence of rule formation can be found in overgeneralisations that favour the following elements as initial question markers:

*interrogative pronoun + do you*  
*interrogative pronoun + do + declarative clause*

Cases in which the subject is inverted against a full verb are again rare. Butterworth and Hatch (1978) cite the following examples:

- (35) Who like you more?
- (36) What city like you more?
- (37) What time come by you San Francisco?

They attribute these instances to L1 interference from Spanish, rather than to an overgeneralization of inversion. Mukattash (1981:320) comes to the same conclusion for his Arab subjects (cf. also Adams 1978:287; Butterworth & Hatch 1978:241).

### 6.3 Embedded interrogatives

The acquisition of embedded interrogatives is not studied as extensively as main clause interrogatives. Most studies confirm a developmental sequence first established by Cazden et al. (1975). This model predicts that the acquisition of target-like embedded interrogatives is closely related to the learners’ development of main clause interrogatives. The crucial element is the differentiation between the presence of inversion in target-like main clause interrogatives, and its absence in embedded interrogatives (Cazden et al. 1975:38):

Stage I – Undifferentiation: learner did not distinguish between simple and embedded *wh*-questions.

- a. Uninverted: Both simple and embedded *wh*-questions were *uninverted*.  
     simple: *What you study?*  
     embedded: *That's what I do with my pillow.*
- b. Variable inversion: Simple *wh*-questions were sometimes inverted, sometimes not.  
     inverted: *How can you say it?*  
     uninverted: *Where you get that?*
- c. Generalization: increasing inversion in *wh*-questions with inversion being extended to embedded questions.  
     simple: *How can I kiss her if I don't even know her name?*  
     embedded: *I know where are you going.*

Stage II – Differentiation: Learner distinguished between simple and embedded *wh*-questions.

- simple: *Where do you live?*  
     embedded: *I don't know what he had.*

Stage I of the model, which claims that learners do not differentiate between main and embedded clause word order, predicts that phenomena which characterize main clause interrogatives should also be found in embedded interrogatives in the learners' language. The data from Indian and Singapore English show that this does not only apply to word order, i.e. the presence or absence of inversion, as predicted by Cazden et al. (1975), but also to the variation related to verb type and the occurrence of frequent chunks: The preferences found for main clause interrogatives, the prominence of *be* especially in its cliticised variant, as well as the use of strings like *what's*, *where's* and *why don't you*, have also been shown for embedded interrogatives.

## 6.4 Formulaic language

The data from both the varieties under discussion and individual language acquisition suggest that memorised "chunks", i.e. frequent strings or combinations of words, are crucial in the acquisition of English interrogative clauses, especially in initial position as productive question markers. The occurrence of such chunks in early learner language has often been documented in the SLA literature. Their role and even the criteria by which to define such chunks are debated, however, and commonly based on the authors' intuitions.

To begin with, the labels vary, including “formulaic language”, “language formulas”, “ready-made chunks”, “unanalyzed wholes”, “lexical phrases”, and “strings” (cf. Peters 1983; Nattinger & DeCarrico 1992; Weinert 1995). Still, most authors agree with respect to the criteria they use for characterising formulaic language. These recurring criteria include phonological coherence, frequency, invariance of form, overextension and idiosyncratic/inappropriate use (cf. Peters 1983; Weinert 1995).

Irrespective of the methodological difficulties concerning the extent to which formulas feed into rule formation, the copula *be*, modals, and other auxiliaries are central and frequent elements in the formation of formula, especially for negation and interrogation: “These verb types may be favourable contexts because they are limited in number and some combinations may be learnt as formulas” (Weinert 1995: 189). Weinert stresses frequency as a crucial factor in the formation of language formula (cf. Bolander 1989); other authors refer to the chronology of teaching (Mukattash 1981).

The major role of formulaic language claimed by these authors can be supported by the frequent occurrence of such chunks in both the English varieties under discussion and the overview of individual L2 acquisition. The data support the claim that firstly copular *be* and frequent modals are likely to feed into language formulas; and secondly that limited number of forms and overall frequency are factors that favour combinations like *what’s* and *where’s*. Thus, both in terms of imitation as well as rule formation, as defined above, the English varieties under consideration parallel the data found in individual second language acquisition. This observation yields important implications for the study of L2 and shift varieties, as well as the cross-variety approach to English vernaculars supported in this article.

## **7. Conclusion: implications for the study of contact vernaculars and “vernacular universals”**

The present study has shown that there are clear parallels between individual second language acquisition and the characteristics of English varieties emerging through language shift (referring to Thomason/Kaufman’s use of the term), to the extent that developmental phenomena of individual L2 acquisition seem to be reflected in the synchronic variation within the English varieties under consideration. This might support the hypothesis that learner development can become fossilised in the context of group second language acquisition. What has been shown for Indian English and Singapore English indeed looks like an intermediate stage in the learners’ acquisition of the target language. The link between the two language domains would then be based on the fact that the principles and

constraints applying to individual and to group L2 acquisition could generally be considered identical or at least strongly related (e.g. Winford 2003: 247).

Despite the obvious relation between individual L2 acquisition and group shift in Thomason/Kaufman's (1988) concept, there have been authors casting doubt on the comparability of the two acquisition contexts. Sridhar & Sridhar (1986: 5ff.), for instance, list five major differences between the two contexts: the goal of acquisition, input, functions of the acquired language, role of the learner's first language, and the motivation to learn the target language. They conclude that:

What is needed is a reevaluation of the applicability of SLA theories to the particular circumstances in which IVEs [indigenized varieties of English] are acquired, rather than simply legitimizing IVE learners as 'normal' second-language-acquirers according to currently fashionable models. (Sridhar & Sridhar 1986: 4)

It turns out, however, that in terms of the actual linguistic results, group shift contexts like the ones leading to varieties like Indian English and Singapore English, are mainly characterized by the additional linguistic factor of fossilisation. Social factors, on the other hand, such as the ones mentioned by Sridhar & Sridhar (1986) seem to have an influence on quantity rather than quality, that is the quantitative presence of interlanguage features and the distance or nearness of the L2 variety grammar to the original target language. These properties have been related to social factors such as the demographics of the groups involved, the extent of interaction between the groups and within the group of 'learners', the length of contact, power relationships and attitudes (cf. Winford 2003: 236). Winford (2003) finds that shared social characteristics surrounding the emergence of several English L2 varieties cluster around a relative distance between the native speakers of the target language and its learners, resulting in the L2 variety becoming an early target for other learners in the community:

Since native speakers of the original TL are typically in a small minority and have higher social status, only the more educated and elite sections of the community have access to native TL models. [...] The fact that most members of the community interact primarily among themselves rather than with native speakers of the TL means that the contact variety of the TL itself becomes the primary target of learning. (Winford 2003: 252)

Irrespective of the debate whether individual and group second language acquisition can use the same models and base their hypotheses on the same foundations – a debate which cannot and was not attempted to be solved here – generalisations in the field of second language acquisition seem to provide a first domain in which potential "universal status" of certain non-standard phenomena of English varieties can be located. The analysis of the factors determining the



occurrence and non-occurrence of inversion in the two varieties under discussion, as well as their comparison with individual language acquisition has provided a first solution to the problem of the incompatibility of preceding explanations of inversion in the two varieties. It has been shown that the acquisition of inversion cannot be undoubtedly considered as the acquisition of an “inversion rule”, and consequently that the use of embedded inversion does not necessarily need to be the extension of such a rule to embedded interrogatives. Thus, claims about the presence or absence of “inversion” and the “extension of inversion to embedded clauses” in varieties of English are severely challenged by these findings, and require detailed further investigation in other varieties.

As to the occurrence of embedded inversion in other varieties of English, the importance of investigating the determinants of its occurrence is evident. Inversion in embedded interrogatives is not restricted to the “New Englishes”, but also occurs in Irish English, Welsh English and Scottish English. Along the traditional lines of substrate and superstrate, embedded inversion has been traced both to influence from Irish and to the retention of a formerly common structure in earlier English (e.g. Filppula 2004). Taking up the recent insight that embedded inversion is a phenomenon not restricted to Irish-substrate contact vernaculars, and that it is a frequent phenomenon in learner English, Filppula defends his Irish-substrate argument:

Despite their merits, these accounts fail to explain the geographical distribution of embedded inversion among the dialects of English spoken in the British Isles, and more specifically, its prominence in the western, north-western and northern varieties such as IrE, ScE, and WeE. Thus, it is hard to escape the conclusion that the Celtic substrate languages have had some role in promoting the use of embedded inversion in the said varieties. (Filppula 2004: 95)

Here, a closer analysis of the form in which embedded inversion occurs is necessary in order to solve such uncertainties. If inversion in embedded interrogatives in Irish English exhibits the same determinants (majority of *be*, then frequent auxiliaries and certain chunks involving forms of *do*) as “learner Englishes” such as Indian English and Singapore English, the learner explanation definitely has to be confirmed. Only if Irish English displays markedly different tendencies must we take into account that the phenomenon might have different sources in different varieties and types of varieties of English. For Indian and Singapore English, however, it has clearly been shown that lack of inversion and embedded inversion can only be explained on the basis of, and within the domain of second language acquisition.

This article has shown that claims for universal status in vernacular data can only be claimed to hold in a specific domain, in this case in second language

acquisition, combined with concepts of language contact. Thus, the status of joint tendencies is crucially dependent on the domain in which they are rooted (such as language acquisition). Siemund (2008), discussing various concepts of universals, concludes that “universals research can only be seriously taken advantage of if the specific properties of vernacular data are taken into consideration”. Only this approach makes sure that superficial correspondences go hand in hand with a common *explanation* of their occurrence.

## References

- Adams, Marilyn. 1978. Methodology for examining second language acquisition. In *Second Language Acquisition. A Book of Readings*, E. Marcussen Hatch (ed.), 277–296. Rowley MA: Newbury House.
- Alo, M. A. & Mesthrie, Rajend. 2004. Nigerian English: Morphology and syntax. In *A Handbook of Varieties of English*, B. Kortmann & E. Schneider, in collab. with K. Burridge, R. Mesthrie, C. Upton (eds), 813–828. Berlin: Mouton de Gruyter.
- Baskaran, Loga. 2004. Malaysian English: Morphology and syntax. In *A Handbook of Varieties of English*, B. Kortmann & E. Schneider, in collab. with K. Burridge, R. Mesthrie, C. Upton (eds), 1073–1088. Berlin: Mouton de Gruyter.
- Beal, Joan. 1993. The grammar of Tyneside and Northumbrian English. In *Real English. The Grammar of English Dialects in the British Isles*, J. Milroy & L. Milroy (eds), 187–213. London: Longman.
- Bhatt, Rakesh M. 2004. Indian English: Syntax. In *A Handbook of Varieties of English*, B. Kortmann & E. Schneider, in collab. with K. Burridge, R. Mesthrie, C. Upton (eds), 1016–1030. Berlin: Mouton de Gruyter.
- Bolander, Maria. 1989. Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish. In *Bilingualism across the Lifespan. Aspects of Acquisition, Maturity and Loss*, K. Hyltenstam & L. Obler (eds), 73–86. Cambridge: CUP.
- Brines, Vivenc. 1990. Learning yes-no questions under different academic conditions. *International Review of Applied Linguistics in Language Teaching* 28(3): 234–247.
- Butterworth, Guy & Hatch, Evelyn. 1978. A Spanish-speaking adolescents's acquisition of English syntax. In *Second Language Acquisition. A Book of Readings*, E. Hatch (ed.), 231–245. Rowley MA: Newbury House.
- Cancino, Herlinda. 1978. The acquisition of English negatives and interrogatives by native Spanish speakers. In *Second Language Acquisition. A Book of Readings*, E. Hatch (ed.), 207–230. Rowley MA: Newbury House.
- Cassidy, Frederic G. 1971. Tracing the pidgin element in Jamaican Creole. In *Pidginization and Creolization of Languages*, D. Hymes (ed.), 203–221. Cambridge: CUP.
- Cazden, Courtney B., Cancino, Herlinda, Rosansky, Ellen J. & Schumann, John H. 1975. *Second Language Acquisition Sequences in Children, Adolescents and Adults*. Final Report, US Department of Health, Education and Welfare.
- Chambers, J. K. 2004. Dynamic typology and vernacular universals. In *Dialectology Meets Typology: Dialect Grammar from a Cross-Linguistic Perspective*, B. Kortmann & E. Schneider,

- in collab. with K. Burrige, R. Mesthrie, C. Upton (eds), 127–145. Berlin: Mouton de Gruyter.
- Ellis, Rod. 1984. Formulaic speech in early classroom second language development. In *On TESOL '83. The Question of Control*, J. Handscombe, R. Orem & B. Taylor (eds), 53–65. Washington DC: Teachers of English to speakers of other languages.
- Escure, Geneviève. 2004. Belize and other central American varieties: Morphology and syntax. In *A Handbook of Varieties of English*, B. Kortmann & E. Schneider, in collab. with K. Burrige, R. Mesthrie, C. Upton (eds), 517–546. Berlin: Mouton de Gruyter.
- Felix, Sascha. 1981. The effect of formal instruction on second language acquisition. *Language Learning* 31(1): 87–112.
- Filppula, Markku. 2004. Irish English: Morphology and syntax. In *A Handbook of Varieties of English*, B. Kortmann & E. Schneider, in collab. with K. Burrige, R. Mesthrie, C. Upton (eds), 73–101. Berlin: Mouton de Gruyter.
- Hatch, Evelyn. 1974. Second language learning-universals? *Working Papers on Bilingualism* 3: 1–17.
- Henry, Alison. 1995. *Belfast English and Standard English*. Oxford: OUP.
- Huang, Joseph & Hatch, Evelyn. 1978. A Chinese child's acquisition of English. In *Second Language Acquisition. A Book of Readings*, E. Hatch (ed.), 118–131. Rowley MA: Newbury House.
- Kortmann, Bernd & Szmrecsanyi, Benedikt. 2004. Global synopsis: Morphological and syntactic variation in English. In *A Handbook of Varieties of English*, B. Kortmann & E. Schneider, in collab. with K. Burrige, R. Mesthrie, C. Upton (eds), 1142–1202. Berlin: Mouton de Gruyter.
- Mahboob, Ahmar. 2004. Pakistani English: An overview of its syntax, morphology, and lexis. In *A Handbook of Varieties of English*, B. Kortmann & E. Schneider, in collab. with K. Burrige, R. Mesthrie, C. Upton (eds), 1045–1056. Berlin: Mouton de Gruyter.
- Mair, Christian. 2003. Kreolismen und verbales Identitätsmanagement im geschriebenen jamaikanischen Englisch. In *Zwischen Ausgrenzung und Hybridisierung*, E. Vogel, A. Napp & W. Luttermann (eds), 79–96. Würzburg: Ergon.
- Miller, Jim. 1993. The grammar of Scottish English. In *Real English. The Grammar of English Dialects in the British Isles*, J. Milroy & L. Milroy (eds), 99–138. London: Longman.
- Mukattash, Lewis. 1981. Wh-questions in English: A problem for Arab students. In *IRAL, International Review of Applied Linguistics in Language Teaching* 19(4): 317–332.
- Nattinger, James R. & DeCarrico, Jeanette. 1992. *Lexical Phrases and Language Teaching*. Oxford: OUP.
- Peters, Ann. 1983. *The Units of Language Acquisition*. Cambridge: CUP.
- Ravem, Roar. 1978. Two Norwegian children's acquisition of English syntax. In *Second Language Acquisition. A Book of Readings*, E. Hatch (ed.), 148–154. Rowley MA: Newbury House.
- Sebba, Mark. 2004. British Creole: Morphology and syntax. In *A Handbook of Varieties of English*, B. Kortmann & E. Schneider, in collab. with K. Burrige, R. Mesthrie, C. Upton (eds), 813–828. Berlin: Mouton de Gruyter.
- Selinker, Larry. 1987. Interlanguage. In *Interlanguage Studies*, D. Nehls (ed.), 11–33. Heidelberg: Julius Groos.
- Shapira, Rina G. 1978. The non-learning of English: Case study of an adult. In *Second Language Acquisition. A Book of Readings*, E. Hatch (ed.), 246–255. Rowley MA: Newbury House.

- Siemund, Peter. 2008. Linguistic universals and vernacular data. In *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*, M. Filppula, J. Klemola & H. Paulasto (eds). London: Routledge.
- Sridhar, Kamal K. & Sridhar, Shikaripur N. 1986. Bridging the paradigm gap; second language acquisition theory and indigenised varieties of English. In *World Englishes* 5(1): 3–14.
- Thomason, Sarah & Kaufman, Terence. 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley CA: University of California Press.
- Tiphine, Ursula. 1983. *The Acquisition of English Statements and Interrogatives by French-Speaking Children*. Kiel: Englisches Seminar der Universität.
- Weinert, Regina. 1995. The role of formulaic language in second language acquisition. *Applied Linguistics* 16(2): 180–205.
- Winford, Donald. 2003. *An Introduction to Contact Linguistics*. [Language in Society 33]. Oxford: Blackwell.
- Wolfram, Walt. 2004. Urban African American Vernacular English: Morphology and syntax. In *A Handbook of Varieties of English*, B. Kortmann & E. Schneider, in collab. with K. Burridge, R. Mesthrie, C. Upton (eds), 319–340. Berlin: Mouton de Gruyter.



PART V

**Translation**



# Linguistic variation through language contact in translation<sup>\*</sup>

Nicole Baumgarten and Demet Özçetin

University of Southern Denmark, Sønderborg / University of Hamburg,  
Research Center on Multilingualism

Translations are classic instances of language contact. This paper examines translational language contact between English and German in the genre of written business communication, where translations from English into German seem to act as a conduit through which Anglophone communicative styles enter German text production. By the example of the first person plural pronouns *we* and *wir* we show how source text induced language variation surfaces in the German register and how it shapes the interaction between the writer and the reader both in German translations and non-translated German texts.

**Keywords:** translation, writer-reader interaction, deixis, register variation, business communication

## 1. Introduction

In his article “Übersetzungen ins Deutsche und ihre Bedeutung für die deutsche Sprachgeschichte” Koller (2000) suggests that in the history of the German language, translations from other languages have always played an important part in the shaping of the grammatical system and preferences in communicative styles. He points out that the total of written texts (*Sprachwirklichkeit* Koller 2000: 113) in German has always consisted of both original text production and translations, without the two ever having been clearly marked off from each other. From a

---

<sup>\*</sup> The study reported on in this article was carried out within the project *Covert Translation* (PI Juliane House) which is part of the Research Center on Multilingualism at the University of Hamburg. The Center is funded by the German Science Foundation (DFG). For an overview of the project’s work see e.g. Baumgarten (2007), House (2006), Böttger (2006).



historical perspective, translations into German have always resulted in adaptations of the German linguistic system in the sense that stylistic and systematic properties of the source text's language were incorporated into German. Koller's main argument is that communication across language boundaries challenges the target language to provide equivalent – or at least adequate – linguistic means for the expression of communicative tasks, which may be to various degrees new to the receiving culture. The target language can deal with this linguistic confrontation in two ways, each of them making a different contribution to the process of language variation and change: Either linguistic features of the source language are introduced into the target language as new linguistic features, or existent usage norms for particular linguistic features in the target language are redefined to fit a new communicative task.

Koller distinguishes two ways of triggering this kind of language variation and change in the target language: 'Innovations on the level of the language system' (*Systeminnovationen*) and 'innovations on the level of norms of usage or stylistic innovation' (*Norm-/Stilinnovationen*). It is the stylistic/norm innovations which are of most interest in the present context. They are divided further into qualitative and quantitative norm innovations. With qualitative norm innovations the first occurrence of a particular linguistic item, feature or structure in a particular use can be traced back to a translation. This is where its proliferation in native target language use arguably stems from. In the case of a quantitative norm innovation, a particular linguistic item, feature or structure occurs both in translations and original texts in the target language. However, its occurrence in translations is triggered by a similar structure in their source texts and is, on the whole, found significantly more frequently in translations as compared to original texts produced in the target language.

Not surprisingly, surface phenomena such as usage and stylistic preferences (i.e. *Norm-/Stilinnovationen*) tend to be more susceptible to change than the underlying linguistic system (i.e. *Systeminnovationen*). At the same time, it is almost impossible to verify the existence of 'qualitative norm innovations'. The definition of 'quantitative norm innovations', however, corresponds to the view of translation-induced language change followed in this article: The frequent translation of source text structures by grammatical, but less used linguistic structures of the target language can, over time (through sheer frequency), marginalize other linguistic means used for the particular communicative function in the target language, and it may eventually override prevailing norms of usage in translation and original text production in the target language. The result of this linguistic variation would be a change in communicative preferences. These 'new' communicative preferences would subsequently become established as prescriptive stylistic

principles, and would hence be perceived as 'normal', appropriate language choice on the part of contemporary native speakers.

The present research attempts to trace and describe such a shift in communicative preferences in the context of English-German translation. The article presents an investigation of the use of speaker deictic elements – or speaker pronouns – in English and German translated and non-translated business communication as they occur, for example, in letters to shareholders. The focus will be on the use of first person plural pronouns in subject position and their co-occurrence patterns both on sentence and on discourse level. Below are some examples from English (E) and translated (GT) and non-translated German (G) written business communication.

- (1) *We strive for excellence in everything we do by being open to new ideas and better ways of working and not being afraid to take risks.* (E)
- (2) *Wir glauben, dass es wichtig ist, in einem Unternehmensbericht die Zukunft und die Vergangenheit gleichermaßen zu betrachten.* (GT)  
*[We believe that it is important, in an annual report, to deal with both the future and the past in equal measure.]*  
 (Here and throughout all translations in square brackets are our own)
- (3) *Aber selbst wenn wir alle Voraussetzungen schaffen, um bald wieder an die gewohnten Verhältnisse anschließen zu können, tun sich allbekannte Defizite nur um so deutlicher wieder auf.* (G)  
*[But even if we create the conditions to enable us to return soon to the usual state, universally known shortcomings only show more clearly.]*

We will show that even though English and German have formally and functionally equivalent speaker deictic means, i.e. direct translation is systematically possible in virtually every instance, it is the presence of language and culture-specific communicative conventions which seems to govern the use of speaker deictic elements in German translations from English and non-translated German texts. It will be suggested, however, that some of the differences in the use of speaker deictic elements in German translations of English texts and non-translated German texts can be explained as source text induced variation in the German register, i.e., linguistic variation in the target language through English-German language contact in translation.

The following section provides a brief review of the differences between English and German communicative conventions as these are the measure by which we assess a possible influence of English on German through translational contact. Section 3 gives a brief overview of the research into the functions of first person personal pronouns in written texts. Section 4 introduces the corpus and

presents the results of the investigation into speaker deictic elements in business communication, and Section 5 is a brief conclusion.

2. English and German communicative conventions

Converging evidence from a variety of English-German contrastive pragmatic studies points to a general pattern of differences in the communicative behavior of native speakers of English and of German (cf. e.g. Byrnes 1986; Clyne 1987, 1994; Kotthoff 1989; Doherty 1996, 2003; House 1996, 2003, 2006). This pattern can be displayed in tabular form (Figure 1).

These dimensions are clines rather than clear-cut dichotomies with absolute values. They reflect tendencies in language use rather than categorical distinctions. It is possible to say, however, that across a variety of communicative situations native speakers of German tend to realize linguistic features associated with the values listed on the right. Native speakers of English, in the same situations, prefer linguistic structures associated with the values listed on the left. In other words, German language use (spoken and written) can be said to be often more direct, taking little heed of the addressees in terms of raising their attention and keeping them interested; it is more content-oriented, more explicit in terms of referential explicitness and denotational precision and less likely to employ verbal routines than English language use in comparable situations. Another way of categorizing the differences is that, in German, on the whole, a ‘transactional’ style focusing on the content of the message is frequently preferred, whereas in English, speakers tend to prefer an ‘interactional’, addressee-focused style.

It is assumed that the different preferences for communicative styles are, at least partly, the surface-level expression of underlying language-systematic differences between English and German as they have been described, for example, by Hawkins (1986), Doherty (1995) and Fabricius-Hansen (1999). Nevertheless, within the confines of language-systematic differences a considerable amount

English		German
<i>Indirectness</i>	↔	<i>Directness</i>
<i>Orientation towards other</i>	↔	<i>Orientation towards self</i>
<i>Orientation towards persons</i>	↔	<i>Orientation towards content</i>
<i>Implicitness</i>	↔	<i>Explicitness</i>
<i>Use of verbal routines</i>	↔	<i>Ad-hoc-formulation</i>

Figure 1. Dimensions of communicative preferences between German and English (House 1996).

of stylistic variation seems possible. An influence of English on German, then, would manifest itself in (for the most part quantitative) changes in the use of certain linguistic items and grammatical structures in German texts which result in a shift towards the more Anglophone communicative style. This would be especially easy in cases in which both linguistic systems offer equivalent linguistic means which can be translated from English into German without a shift in semantic and pragmatic meaning and without consequences for the surrounding syntactic structure.

### 3. Speaker-hearer deictic elements: Overview of existing research

The use of personal pronouns, i.e., their function in text and discourse, has lately been frequently investigated in English text and register linguistics. Most of the research carried out has focused on written academic discourse and on the differences in the use of personal pronouns in different scientific disciplines (e.g. Bernhardt 1985; Vassileva 1998; Kuo 1999; Hyland 2001, 2002; Harwood 2005a, b). Within the group of personal pronouns, a distinction is usually made between speaker-deictic elements (*I, we, me, mine, my, us, our, ours*) and hearer-deictic elements (*you, yours*). Within the speaker-deictic elements we can furthermore distinguish between exclusive and inclusive uses of the first person plural pronouns (*we, us, our, ours*). First person personal pronouns in subject position are especially conspicuous in written texts because their use is highly constrained by rhetorical/communicative conventions and consequently their occurrence is highly register-specific. In the four English genres analyzed in Biber et al. (1999) – conversation, fiction, news writing, written academic discourse – spoken language shows the highest frequency of speaker deictic elements and academic discourse the lowest. Similar corpus-based investigations for German do not exist, but qualitative and theoretical studies suggest a possibly even more distinct writer eradication in German scientific writing (Bungarten 1981; von Polenz 1981; Kretzenbacher 1994, 1998; Drescher 2003). Nevertheless, speaker deictic elements do occur in both English and German texts, and when they occur, they fulfill a special function in the text because they refer directly to the reader and the writer. They ‘enact’, as it were, an interaction between the writer and the reader where no ‘real’, immediate interaction takes place. This imitation of face-to-face interaction in written, ‘distance’, communication raises the question of why a writer would want to have writer-reader interaction in a text at all, when text producers in written communication – as Tannen (1982), Chafe (1984), and others (e.g. Biber 1988) have observed – usually choose linguistic items which express a detachment from the communicative event in which they are involved,

thus reflecting the spatio-temporal separation of the contexts of text production and reception.

For any writer it is essential to keep this interaction between the reader and the writer within the limits of the rhetorical conventions of the genre, text type, or register. For academic discourse, a number of communicative functions of the use of speaker deictic elements have been suggested. Hyland (2002), for example, argues that through the use of first person pronouns in subject position “writers gain credibility by projecting an identity invested with individual authority, displaying confidence in their evaluations and commitment to their ideas” (p. 1091). Kuo (1999) claims that the use of pronouns in scientific articles serves to encode the writers’ perception of their own role in research and their relationship with the intended readers of the text, and Harwood (2005a) argues that *I* and *we* serve to create a tenor of self-promotion in scientific articles, which helps to publicize both the research and the person of the writer. On a more conceptual level of description, Hyland (2005) describes the use of personal pronouns as one of the resources for what he calls the “intersubjective positioning” of the writer and the reader in written texts. In this framework, personal pronouns are relevant for the expression of both “stance” and “engagement” – the two central dimensions of “intersubjective positioning”. Stance-taking encompasses the writers’ way of expressing attitudes, conveying judgments, opinions and commitments. Engagement refers to the way in which writers align themselves with the readers by focusing their attention, acknowledging their presence and thus including them as discourse participants. In sum, first person personal pronouns serve to personalize claims and beliefs, to stress the originality and the ownership of the work and the ideas presented, to align the reader with the writers’ perspective and to express solidarity with the readership.

It follows from the above that personal pronouns are not randomly scattered throughout a text. They display specific co-occurrence patterns with other linguistic elements – e.g. with modalizing and evaluating expressions and, as will be shown in Section 4, certain semantic verb types. They also occur at specific junctures in the text, which fulfill specific discourse functions with respect to the progression of the argument and the positioning of the writer and the reader in the text. Summarizing the evidence from the investigations carried out by Bernhardt (1985), Vassileva (1998) and Kuo (1999), we can say that first person personal pronouns occur with the following discourse functions:

1. Explaining what was done
2. Proposing a theory/approach
3. Stating a goal/purpose
4. Showing results/findings

5. Comparing approaches/viewpoints
6. Giving reasons
7. Outlining a procedure/methodology
8. Referring to one's own research
9. Disputing other researchers' findings
10. Acknowledging assistance
11. Justifying a proposition
12. Hedging a proposition
13. Assuming shared knowledge/goals/beliefs
14. Seeking agreement/cooperation
15. Showing commitment to research
16. Indicating necessity
17. Expressing wish/expectation
18. Discussing implications for reader behavior

The list shows that speaker deictic elements occur in both predominantly ideational (content-oriented) discourse contexts (1–10) and interpersonal-pragmatic discourse contexts (11–18), which relate to the participant framework (Goffman 1981; Schiffrin 1987) and the communicative situation. The use of speaker deictic elements thus apparently serves an overall orientation of the discourse towards persons because it encompasses both the content-related and the interpersonally-related aspects of the communicative event in the text. The functional description of first person personal pronouns in subject position suggests that personal pronouns are the linguistic expression of both an orientation towards persons in the presentation of information and an orientation towards the other (the reader) in the communicative event. As noted earlier, these are characteristics of English communicative styles rather than of German ones.

Some investigations indeed point towards cross-linguistic and cross-cultural differences in the way the lexicalization of the writer and the reader in the text is conventionalized. For example, in her analysis of English and Finnish academic writing Mauranen (1993) finds that in English articles written by English native speakers, more linguistic elements are used to actively guide and orient the reader around the information than in comparable texts by Finnish L2 writers of English. In her view, the English texts are, on the whole, more reader-oriented, which coincides with a more explicit presence of the writer in the text who usually appears in subject position and navigates the reader towards the relevant information.<sup>1</sup>

---

1. Comp. also Ädel (2006) who finds differences in the frequency of the use of metadiscourse between American and British English student writing.

In her investigation of the use of *I* and *we* and their formal correspondences in German, French, Bulgarian and Russian academic discourse, Vassileva (1998) reports on cross-linguistic commonalities and differences. For example, first person pronouns in subject position are used to express conclusive statements and the introduction of procedures in all the languages investigated, whereas the reference to previous personal experience is much more common in English and German than in French, Bulgarian and Russian. She assumes that the similarities between the English and the German uses of the speaker pronouns is the result of an influence of English scientific writing styles on German communicative conventions, which has not (yet) affected the French, Bulgarian and Russian academic writing communities.

As pointed out above, this description of the use of first person personal pronouns in written communication is the result of the investigation of academic discourse. There are virtually no analyses of other written genres. In this article we will present a synchronic description of the use of speaker-deictic elements in a corpus of business communication – a genre which, up to now, has never been investigated in terms of the role speaker-deictic elements play in it.

## **4. Speaker pronouns in business communication**

### **4.1 Database and methodology**

The study was carried out on an English-German translation and comparable text corpus of written business communication. The texts were sampled from the annual reports of internationally operating companies and comprise letters to shareholders and a few mission statements. Annual reports are part of so-called ‘external business communication’, i.e., they are addressed to investors and people considering an investment while at the same time serving as information material for any other persons interested in the company. Especially letters to shareholders and mission statements are characterized by non-specialist language: they ostensibly appear to target the individual private investor even though many companies are in fact in large parts owned by other companies and not immediately by individual persons. Thus, these two text types generally do not require any in-depth knowledge of economics and business operations.

The database used for the investigation comprises the following three datasets:

1. English texts from the years 1999–2002 (21222 words).
2. The German translations of these texts (21808 words).
3. German original texts from the years 1999–2002 (26091 words).

For the present investigation, all occurrences of first person plural pronouns in subject position and their translational structures plus an extended context of their occurrence (5 preceding, 5 following sentences) were extracted from the corpus. Each occurrence of a speaker deictic element (and its translational structure) was then annotated manually for its co-occurrence with the following features: finiteness, tense, voice, transitivity, verb and sentence mood, aspect, modality, clause complexing, position in sentence and text, agentivity, thematic markedness, cohesion, metadiscourse, referential continuity and discourse functions. In the present context transitivity choices and the co-occurrence of the first person pronouns with discourse functions will be focused on. The goal of the annotation is to be able to deduce

- a. whether or not formally corresponding linguistic items are found in the same linguistic contexts in both languages;
- b. whether or not formally corresponding linguistic items are used for the same communicative purpose in both languages.

Only when these questions are answered can we assess the nature of language contact occurring in translation in terms of source text induced linguistic variation in the target language register.

## 4.2 Results of the corpus study

We will analyze the data from four different perspectives. First, we will look at the frequency counts for *we* and *wir*. Secondly, we will focus on the co-occurrence of the pronouns with semantic verb types. Thirdly, we will present the discourse functions fulfilled by the sentences in which *we* and *wir* occur, before finally analyzing the translation relation.

### 4.2.1 Frequencies

In Table 1 we illustrate the frequency of speaker-deictic elements in English texts (EO), their German translations (GT) and German original texts (GO).

Contrary to our expectations of a higher frequency of *we* in the English texts than of *wir* in the German translations and the German original texts, the data in fact show more occurrences of *wir* in the German original texts than of *we* in



**Table 1.** Frequency of *we* and *wir* (normalized on the basis of 10000 words; raw frequencies in brackets).

	EO		GT		GO	
<i>we/wir</i>	181.4	(385)	183.8	(401)	218.4	(570)

the English texts and *wir* in the German translations. If we count the first person pronouns in subject position as markers of a person-orientation in the presentation of the subject matter, the German original texts clearly surpass the remaining two data sets. This result neither coincides with the differences between the communicative preferences in German and English as described in Section 2, nor does it match the findings described in the literature on German annual reports and letters to shareholders (Bolten et al. 1996; Gazdar & Kirchhoff 2001; Baumgarten et al. 2004; Böttger 2004, 2006; Bührig & House 2004).<sup>2</sup>

The English texts and the German translations are almost identical with regard to the frequency of first person plural pronouns. Apparently, the English texts constrain the use of *wir* in the German translations. As will be shown in Section 4.2.4 below, this is not the result of straightforward one-to-one translation of *we* into *wir*.

4.2.2 *Process types*

In order to characterize the use of *we* and *wir* in the English and German texts more precisely, the co-occurrence of the pronouns with verbs was analyzed. The verbs were classified according to Halliday’s (1994) classification of process types. Halliday distinguishes between six process types which model the linguistic expression of human experience along the general processes of “doing”, “sensing” and “being”. The processes and the participant in subject position are part of the “experiential center” of the clause (Halliday & Matthiessen 2004: 176). In Table 2 we provide a summary of the process types; in Table 3 we illustrate the co-occurrences of *we* and *wir* with the different process types in English texts, their German translations and German original texts.

Despite the slight differences between the frequencies of *we* and *wir*, the distribution across process types shows remarkable similarities, with the majority of verbs belonging to the material category in all three data sets. In co-occurrence with *we* and *wir*, material verbs express activities carried out or events taken part in by the writer. The second most frequent category is the category of mental verbs which, in co-occurrence with personal pronouns, denote mental activities and states of mind of the writer. The relational verbs construe the writer

2. All these studies are qualitative case studies.

Table 2. Process types.

Process type	Description	Example
Material	Doing, acting Happening, being created Creating, changing	do
Behavioral	Behaving	look
Mental	Seeing (perception) Feeling (emotion, desideration) Thinking (cognition)	see like think
Verbal	Saying	say
Relational	Having attribute Having identity Symbolizing	be have
Existential	Existing	be

Table 3. Co-occurrences of *we* and *wir* with process types (percent).

	EO (N = 385)	GT (N = 401)	GO (N = 570)
Material	56.2	58.6	60.6
Mental	20.7	18.3	22.5
Relational	17.8	17.3	10.8
Verbal	5.0	5.8	5.9
Behavioral	0.3	1.0	0.2
Existential	0.0	0.0	0.0

as the carrier of attributes or as having a particular identity, while verbs of saying denote the writer's communication activities in the form of self-quotations or reported speech. Behavioral and existential verbs occur either very rarely or not at all.

Halliday & Mathiessen (2004) distinguish between four subtypes of mental processes, which characterize the main types of mental activities participants can engage in: cognition, desideration, emotion and perception. In Table 4 we present the co-occurrences of *we* and *wir* with these subtypes in our data.

The distribution of the mental process subtypes shows a clear difference between the English texts and the German translations on the one side and the German original texts on the other side. In the English texts and the German translations, mental verbs predominantly express cognitive meanings, whereas in the German original texts mental verbs are primarily used to express desiderative meanings (often realized by *wollen* 'want'). Thus, although mental verbs occur with a comparable frequency in all three data sets, the English texts and their

**Table 4.** Co-occurrences of *we* and *wir* with mental process subtypes (percent).

	EO (N = 79)	GT (N = 73)	GO (N = 127)
Cognition	70.1	64.4	29.8
Desideration	22.1	23.3	67.8
Emotion	3.9	9.6	0.7
Perception	3.9	2.7	1.7

German translations reveal a markedly different distribution of the pronouns across the mental verb subtypes compared with the German original texts. The writer in the German original texts is therefore associated with distinctly different verbal meanings.

To further characterize the contexts in which *we* and *wir* occur, we also analyzed the discourse functions which the sentences with the pronoun in subject position fulfill with regard to the progression of the argument and the positioning of the writer in the text.

**4.2.3 Discourse functions**

As described in Section 3, first person personal pronouns fulfill a number of discourse functions in written academic discourse. Following Kuo (1999), for the analysis of our corpus, we defined ‘discourse function’ as the function that a sentence containing a first person personal plural pronoun in subject position performs in the immediate context of its occurrence in the text. The analysis of our data yielded the set of discourse functions for business texts presented in Table 5.

The discourse functions found in the business texts partly overlap with those found in academic discourse (cf. Section 3). However, the present data feature a number of additional discourse functions (shaded in Table 5) which can be regarded as being typical of written business communication or, more generally, of genres designed to appeal to an educated lay audience. As already mentioned, in contrast to peer-to-peer expert communication, the text producer, in this genre, cannot be sure of the reader’s knowledge base and attitudes towards the subject matters presented. We can assume that, in order to secure the transmission of the intended meanings, it might be necessary to connect more areas of subject matter presentation (encompassing ideational, interpersonal and discourse organizational aspects) with the writer as the explicit source of the information.

The table shows that, overall, the English texts and their German translations show similar co-occurrences of first person pronouns and discourse functions, whereas the German original texts show a somewhat different profile. DESCRIBING STATE OF AFFAIRS and EXPLAINING WHAT WAS DONE are the two most frequent

**Table 5.** Discourse functions (percent). Shaded: additional discourse functions apparently primarily occurring in business texts.

	EO (N = 385)	GT (N = 401)	GO (N = 570)
Describing state of affairs	34.0	33.4	22.1
Explaining what was done	13.5	13.2	20.9
Expressing commitment/obligation*	6.2	6.2	–
Stating goal/purpose	3.6	6.5	19.5
Expressing feelings	2.9	2.7	–
Admitting difficulties	0.2	–	2.1
Expressing request	–	–	0.7
Explaining results	9.1	8.5	7.2
Showing commitment (to corporate goals)	5.4	5.2	3.1
Evaluating achievements	4.4	3.0	6.0
Expressing opinion	3.9	3.0	3.7
Expressing prediction	5.4	4.2	3.0
Thanking	1.8	2.0	2.8
Expressing condition	1.0	1.5	0.5
Expressing expectation of results	4.4	6.2	4.7
Pre-organizing the discourse	1.3	1.5	1.4
Metalanguage	0.5	1.0	0.3
Expressing wish	0.3	0.2	0.2
Unclear	2.1	1.7	1.8
<b>Total</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

\* We consider it necessary to distinguish between EXPRESSING COMMITMENT / OBLIGATION using modal verbs like *must* / *have to*... and SHOWING COMMITMENT (TO CORPORATE GOALS) using verbs like *strive* / *pledge to be* / *be dedicated to*...

types of discourse functions in the data (cf. the examples (4) and (5) below). This is not surprising, as letters to shareholders typically describe states of affairs and explain the company's operations during the past business year. The English texts and their German translations focus on personalized descriptions, while in the German original texts the pronoun is used to focus backward and personalize explanations.

#### DESCRIBING STATE OF AFFAIRS:

- (4) Our portfolio of brands and products, and our record of innovation, are unmatched. We have 60 new models in the pipeline. We lead in the luxury car segments. We defined the minivan, and will continue to do so.

## EXPLAINING WHAT WAS DONE:

- (5) Im April 2002 haben wir die restlichen Anteile an Temic übernommen und damit einen wichtigen Schritt zur vollständigen Integration in den Continental-Konzern getan. [...] Im Oktober 2002 haben wir mit der Moscow Tyre Plant (MTP) einen Vertrag zur Gründung eines Unternehmens für die Produktion und den Vertrieb von Pkw- und Leicht-Lkw-Reifen in Moskau geschlossen, an dem Continental die Mehrheit hält. [*In April 2002, we took over the remaining shares in Temic and thereby made an important step towards the full integration into the Continental Group.* [...] *In October 2002, we signed a contract with the Moscow Tyre Plant (MTP) for the foundation of a new company for the production and distribution of automobile and lightweight truck tires in Moscow in which Continental holds the majority of shares.*]

STATING GOAL/PURPOSE is more frequent in the German original texts than in the English texts and their German translations, which explains the high frequency of the *Desideration* mental process type in the German original texts presented above (cf. Table 4). (6) below is a good example for this.

- (6) *Wir wollen unseren Kunden weltweit Spitzenleistungen, unseren Mitarbeitern zukunftsfähige Arbeitsplätze und unseren Aktionären kontinuierliche Wertsteigerung bieten.*  
[*We want to offer our customers worldwide top class services, our colleagues sustainable workplaces and our shareholders continuous value accretion.*]

The pronounced preference for statements of goals and purposes in the German texts can be regarded as a confirmation of House's findings (e.g. 1996, 2003, 2006), which point to a difference in German and English discourse moves that express explicit announcements of topics or plans and giving reasons. The discourse functions ADMITTING DIFFICULTIES and EXPRESSING A REQUEST do not occur at all in the German translations and only once in the English texts. On the other hand, EXPRESSING COMMITMENT/OBLIGATION and EXPRESSING FEELINGS only occur in the English texts and the German translations, and not in the German original texts. Thus, the high frequency of *wir* in German original texts (cf. Table 1 in Section 4.2.1) is not the result of additional discourse functions in co-occurrence with *wir*, but rather that of a more frequent realization of statements of goals and purposes, the admission of difficulties, the expression of requests and thanks, and the evaluation of achievements featuring this pronoun.

The fact that the discourse functions EXPRESSING COMMITMENT/OBLIGATION and EXPRESSING FEELINGS do not occur at all in German original texts, corresponds to the findings made by Böttger (2006). She posits that in the US

economy, employees are generally expected to identify strongly with their company, so that American business communication often contains expressions of self-commitment, and thus tends to be more ‘emotional’ than German business communication, which focuses almost solely on the objective transmission of information. Expressions of self-commitment are generally avoided in German business communication – presumably in order to avoid the possibility that they might be understood as commands to the addressees.

On the whole, the hypothesis that German translations are influenced by their English source texts and English communicative conventions is confirmed by this analysis. The German translations differ from the German original texts, such that the German genre of business communication encompasses texts which linguistically resemble the English sister genre. The similarities between English texts and their German translations may lead us to assume that all occurrences of *we* are translated into the formally corresponding *wir*. In the following, however, we will show that this is not the case.

#### 4.2.4 Translations of *we*

In our data, the first person plural pronoun *we* is translated into German as shown in Table 6. In 76.8% of the cases *we* has been translated into German as the formally corresponding *wir*, leading us to the question of what happened to the rest. Why has the translator not in every case taken the easy way of transferring the *we* of the source text into *wir* in the target text? In 21.3% of the cases *we* is translated into a formally non-corresponding structure in German. Two major reasons can be distinguished: the use of impersonal expressions and the omission of the clause with *we* in subject position.

**4.2.4.1 Impersonal expressions.** The largest group of impersonal expressions which appear in the translations in place of the source text’s first person plural pronoun may be explained by different preferences for ‘congruent realization’ and ‘metaphorical realization’ in English and German, respectively. Following systemic functional theory, meanings can be expressed either congruently, i.e., a prototypical and unmarked construction that “speakers recognize as typical patterns of

**Table 6.** Translations of *we* (percent).

<i>Wir</i>	76.8
Impersonal expressions*	10.4
No equivalent structure / omission	10.4
<i>Uns</i> (‘us’)	0.5
Deleted sentences	1.8

\* NP, passive, *man* (‘one’)

wording” (Halliday 1994:343), or in a ‘grammatical metaphor’, “where the typical pattern has not been used and the [...] writer has chosen to say things differently” (Halliday 1994:343). In other words, in a congruent realization things are expressed by nouns, properties by adjectives and actions by verbs as displayed in example (7a):

- (7) a. *She was so beautiful that he could not speak.*

Example (7b) expresses the same meanings by means of grammatical metaphor: Properties are expressed by nouns and actions by adjectives.

- b. *Her beauty left him speechless.*

According to Halliday (1994), grammatical metaphor is characteristic of adult discourse and written language, while congruent realizations are typical of spoken language. It follows, first, that congruent realizations in written texts can evoke impressions of the use of spoken language, and, secondly, that the structure *we/wir in subject position + verb* with which we are concerned here can be seen as such a congruent realization reminiscent of spoken discourse in written texts.

(8) is an example from our data in which the congruent realization of the English text is rendered as a grammatical metaphor in German.

- (8) a. There is a great desire at all levels of Baker Hughes to effectively serve our customers and to benefit our stockholders as we enter a period of increasing demand in the oil field services business.
- b. Ich kann feststellen, dass auf allen Ebenen bei Baker Hughes große Anstrengungen unternommen werden, um angesichts steigender Anforderungen an Dienstleistungsunternehmen im Bereich der Erdölförderung Höchstleistungen im Interesse unserer Kunden und Aktionäre zu erbringen.  
[*I find that on all levels of Baker Hughes there are great endeavors to effectively serve our customers and to benefit our stockholders in view of increasing demand in the oil field services business.*]

Through the use of a grammatical metaphor, the German translation in (8b) conveys the information in a more compact and lexically dense manner by packing more lexical items into the clause. In addition, the preposition *angesichts* (*in view of*) in the German translation has a stronger causal meaning component than the predominantly temporal *as* in the English source text. In virtually every case in which the source text's *we* is not realized as *wir* in the translation, an equivalent

congruent translation realization would have been possible. Compare the constructed congruent version in (8c):

- (8) c. (...) Auf allen Ebenen bei Baker Hughes werden große Anstrengungen unternommen, um unseren Kunden effektiv zu dienen und unseren Aktionären Gewinne zu bringen, da wir in eine Zeit der steigenden Anforderungen an Dienstleistungsunternehmen im Bereich der Erdölförderung eintreten.  
 [On all levels of Baker Hughes there are great endeavors to effectively serve our customers and to benefit our stockholders as we enter a period of increasing demand in the oil field services business.]

The use of the grammatical metaphor in (8b) renders the German translation more general and, despite the *ich* (I) in the matrix clause, distinctly content-oriented as there is no linguistic item that signals exactly who is going to be facing the increasing demand in the oil field services business.

The second group of impersonal expressions in the translations are passive constructions. Consider example (9).

- (9) a. We continued to develop our management team, leveraging the depth of talent within the Ecolab ranks. In March 2002, we promoted four of our key leaders:  
 b. Im Berichtsjahr haben wir durch optimales Einsetzen unserer talentierten Mitarbeiter innerhalb von Ecolab unser Management-Team weiter gestärkt. Im März 2002 wurden vier unserer wichtigsten Führungsmitglieder befördert.  
 [In March 2002, four of our key leaders were promoted:]

We may hypothesize that the passive is used to avoid repetition of the structure *temporal adverbial + haben wir* (we have), which already occurs in the previous sentence. This assumption is given support by German stylistic conventions according to which such repetitions should be avoided (cf. Möller 1958; Seibicke 1969; Besch 1989). However, the passive may also have been used to downplay the writer's personal involvement in the development of the careers of the employees. This downplaying of personal involvement also matches findings suggesting that German texts tend to be less oriented towards persons in the sense that the focus is placed on the process rather than on the human being responsible for the process (Weinrich 2005: 166). Both the use of the passive and the condensation of information through metaphorization are typical features of specialized discourse. As such, the impersonal expressions in the German translations appear to lift the discourse to a higher stylistic level.



**4.2.4.2 Omissions.** The second main category of formally non-corresponding translations of *we* is the category of omissions. Omissions usually occur in cases in which the pronoun *we* in the source text is part of a verbal routine or a routine-like expression functioning as a discourse marker, as in (10).

- (10) a. At Microsoft, software is our passion, the reason we're excited to come to work every morning. As we look to the future, our commitment to advance the frontiers of software technology offers extraordinary opportunities for consumers, businesses and our company.
- b. Basis für diese Entwicklungen war die zur Umsetzung der neuen Technologien erforderliche Software. Dabei war und ist es Ziel von Microsoft, die Softwaretechnologien kontinuierlich zu verbessern und Endkunden sowie Unternehmen optimierte Produkte mit erweiterten Funktionen zur Verfügung zu stellen.
- [*The basis for these advances was the software needed to realize the new technologies. Thereby, it was and still is the goal of Microsoft to improve software technologies and offer consumers and businesses optimized products with advanced functions.*]

The routine-like expression in the English text cannot be translated into German without a loss in idiomaticity and the decomposition of the fixed expression because no corresponding expression exists in German. In addition, verbal routines functioning as discourse markers are especially associated with spoken discourse (Aijmer 1996). Their occurrence in written texts thus evokes meanings of 'orality' and colloquiality (Coulmas 1981; Wierzbicka 1991; Aijmer 1996; House 1996, 2003) and a particular kind of rhetorical strategy which could be unwanted in German translations (Bühridg & House 2004).

The previous discussion of the formally non-corresponding translations of *we* suggests that in some cases the person-orientation expressed by the personal pronouns is overridden by the expression of the 'typically German' impersonal and content-oriented presentation of the subject matter. The instances of formally non-corresponding translations appear to be unsystematic, i.e., the data provide no clues as to which circumstances in the source text trigger the choice of a formally non-corresponding item in the translation.

A cursory analysis of a sample of 100 randomly selected occurrences of *we* translated by *wir* in our data revealed that in 60% of the cases, the translation closely imitates the source text's syntactic structure. In the remaining 40%, the sentence structure of the translation deviates considerably from its source text structure, while retaining the formally corresponding *wir* in subject position. We can therefore assume that the instances of one-to-one *we*-to-*wir* translation cannot be regarded as straightforward imitations of the source texts (so-called

‘translationese’, cf. e.g. Baroni & Bernardini 2005), because translators often preserve the pronoun in subject position despite creatively rearranging the remaining part of the sentence structure. For the 21.3% of formally non-corresponding translations discussed above, this means that other reasons than simply aiming to avoid the pronoun in German must in fact motivate the use of impersonal constructions and omissions. One possibility, we suggest, is the avoidance of an overly ‘oral’ and non-elaborated written style.

Lastly, we will briefly address the case of occurrences of *wir* in the German translations which appear without an equivalent *we* in the English original texts.

#### 4.2.5 Occurrences of *wir* in the German translation without a corresponding *we* in the English original text

Occurrences of first person personal pronouns in the German translations in excess of what is licensed by their source texts are triggered by impersonal expressions and verbal routines in the English texts. Impersonal expressions are avoided for grammatical reasons, i.e., semantic and syntactic incompatibilities between the two languages. As suggested in the previous section, verbal routines seem to be avoided in order to evade meanings of ‘orality’ and colloquiality in the texts.

The vast amount of cases in which we find *wir* in the translation but no respective pronoun in the source text are triggered by inanimate noun phrases in subject position in the source text. In contrast to English, German has semantic restrictions with respect to the selection of inanimate and non-conscious entities as subjects in agentive roles (cf. Hawkins 1986). However, the selectional restrictions in German apply only to a certain degree. Some inanimate, non-conscious subjects are possible in German, although they are usually considered as stylistically marked (i.e. inappropriate) choices. In example (11) the use of an inanimate subject in the German translation is avoided by means of the use of *wir*.

- (11) a. In closing, I express my appreciation to the dedicated employees who have stuck with Baker Hughes through the downturn, made sacrifices, and who are now bringing the company into recovery. Their hard work has kept our operations focused on delivery value to our customer.
- b. Abschließend möchte ich den hochmotivierten Mitarbeitern, die in schweren Zeiten zu Baker Hughes gehalten haben, die Opfer gebracht haben, und die nun das Unternehmen wieder voranbringen, meine Anerkennung aussprechen. Durch ihre harte Arbeit konnten wir uns darauf konzentrieren, unseren Kunden eine hohe Wertschöpfung zu sichern.
- [Through their hard work we were able to focus on delivery value to our customers].

In example (12) syntactic differences in the realization of participle clauses between English and German lead to the use of a *wir*-clause in the German translation.

- (12) a. In contrast, we expect a gradual pick-up in economic activity in Europe, except in the United Kingdom. In Japan, further slow recovery is expected, with continued faster recovery in South East Asia. The outlook for Latin America varies widely.
- b. In Europa erwarten wir, abgesehen von Großbritannien, eine Verstärkung des Wachstums. Für Japan wird eine weitere langsame Erholung der Wirtschaft erwartet. Für Südostasien gehen wir von einer anhaltenden schnelleren Erholung aus. Die Prognosen für Lateinamerika ergeben kein einheitliches Bild.  
[*In Japan, further slow recovery is expected. In South East Asia, we assume a continued faster recovery*].

Example (12) is of one of the very few cases in the data in which an analogous translation would be grammatically unacceptable. Compare (12c).

- c. \*Für Japan wird eine weitere langsame Erholung der Wirtschaft erwartet, mit einer anhaltenden schnelleren Erholung in Südostasien.  
[*In Japan, further slow recovery is expected, with continued faster recovery in South East Asia*].

Finally, example (13) illustrates the omission of a verbal routine which has no direct equivalent in German and – as discussed above – is therefore deleted in order to avoid ‘oral’ and colloquial meanings within the text.

- (13) a. We drove and will drive the consolidation process in the world-wide automotive industry, we will seek further world-wide growth in the automotive business and lead this development into the future. At the end of the day, our job as leaders is to anticipate the future and to make it happen to the best advantage of you, the shareholder.
- b. Wir gestalten den Konsolidierungsprozeß in der Automobilindustrie. Wir werden ihn weiter anführen und wir werden die sich daraus ergebenden Wachstumschancen für uns nutzen. Sie können von uns erwarten, dass wir die Zukunft vorwegnehmen und sie zu Ihrem Vorteil gestalten.  
[*You can expect us to anticipate the future and to make it happen to your best advantage*].

## 5. Conclusion

Our comparative analysis of the use of the first person plural pronouns *we* and *wir* in English business texts and German translated and non-translated business texts yielded the following results: The English texts and their German translations show similarities (1) in the frequency of *we* and *wir*, (2) in the distribution of process types and (3) in the distribution of discourse functions. In all three areas, the German original texts differ from both the English texts and the German translations. Consequently, we may suggest that the German translations of the English texts introduce variation in the target language register. In the German translations, sentences featuring first personal plural pronouns in subject position collocate with a different set of mental process types than found in comparable German original texts, and they co-occur with a different set of discourse functions. This means, in the translations, the first person plural pronoun *wir* does not occur as frequently in some contexts as would be expected by a reader familiar with German original texts (e.g. with respect to *Desideration*-type mental processes, the discourse functions EXPLAINING WHAT WAS DONE and STATING A GOAL/PURPOSE). In addition, the first person plural pronoun occurs in contexts in which it might not be expected at all (e.g. with the discourse function EXPRESSING COMMITMENT/OBLIGATION). The differences in the co-occurrence patterns of *wir* in German translations and *wir* in German original texts are not trivial because the meanings that are established by the co-occurrences result in different constructions of the writer 'persona' in the texts. This, in turn, means that the readers of the translations are addressed by writers who – at least slightly – differ from the writers of German original texts in what they think, say and do, i.e., in the way they are present in the texts as the source of information. Consequently, the German genre 'business texts' cannot be regarded as a homogenous one, but rather as a genre which encompasses some 'foreign' aspects, resulting from English communicative norms.

Ideally, diachronic analyses, supporting the present synchronic investigation, will show whether the discourse functions present in the English texts and the German translations are taken up in German original text production. However, it is likely that no large scale translation corpora will be available due to the fact that in written business communication translations from English into German are becoming increasingly uncommon (cf. House 2004). This in itself may be taken as a sign of the extent to which the English language influences German business communication.

## References

- Ädel, Annelie. 2006. *Metadiscourse in L1 and L2 English*. Amsterdam: John Benjamins.
- Aijmer, Karin. 1996. *Conversational Routines. Convention and Creativity*. London: Longman.
- Baroni, Marco & Bernardini, Silvia. 2005. *A New Approach to the Study of Translationese: Machine-Learning the Difference between Original and Translated Text*. Oxford: OUP.
- Baumgarten, Nicole. 2007. Converging conventions?: Macrosyntactic conjunction with English *and* and German *und*. In *TEXT & TALK* 27 (2): 139–170.
- Baumgarten, Nicole, House, Juliane & Probst, Julia. 2004. English as *lingua franca* in covert translation processes. *The Translator* 10(1): 83–108.
- Bernhardt, Stephen A. 1985. The writer and the reader of the scientific text. *Journal of Technical Writing and Communication* 15(2): 163–174.
- Besch, Elmar. 1989. *Wiederholung und Variation: Untersuchungen ihrer stilistischen Funktionen in der deutschen Gegenwartssprache*. Frankfurt: Lang.
- Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: CUP.
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. 1999. *Grammar of Spoken and Written English*. London: Longman.
- Bolten, Jürgen, Dathe, Marion, Kirchmeyer, Susanne, Roennau, Marc, Witchalls, Peter & Ziebell-Drabo, Sabine. 1996. Interkulturalität, Interlingualität und Standardisierung bei der Öffentlichkeitsarbeit von Unternehmen. In *Fachliche Textsorten*, H. Kalverkämper & K. Baumann (eds), 389–425. Tübingen: Narr.
- Böttger, Claudia. 2004. Genre-mixing in business communication. In *Multilingual Communication*, J. House & J. Rehbein (eds), 115–129. Amsterdam: John Benjamins.
- Böttger, Claudia. 2006. *Lost in Translation? Analysing the Role of English as the Lingua Franca of Multilingual Business Communication*. PhD dissertation, University of Hamburg.
- Bühlig, Kristin & House, Juliane. 2004. Connectivity in translation: Transitions from orality to literacy. In *Multilingual Communication*, J. House & J. Rehbein (eds), 87–114. Amsterdam: John Benjamins.
- Bungarten, Theo. 1981. Wissenschaft, Sprache und Gesellschaft. In *Wissenschaftssprache. Beiträge zur Methodologie, theoretischen Fundierung und Deskription*, T. Bungarten (ed.), 14–53. München: W. Fink.
- Byrnes, Heidi. 1986. Interactional style in German and English conversations. *Text* 6: 89–106.
- Chafe, Wallace. 1984. Integration and involvement in speaking, writing, and oral literature. In *Spoken and Written Language*, D. Tannen (ed.), 35–53. Norwood NJ: Ablex.
- Clyne, Michael. 1987. Cultural differences in the organization of academic texts. *Journal of Pragmatics* 11: 101–124.
- Clyne, Michael. 1994. *Intercultural Communication at Work. Cultural Values in Discourse*. Cambridge: CUP.
- Coulmas, Florian (ed.). 1981. *Conversational Routine: Explorations in Standardized Communication Situations and Prepatterned Speech*. The Hague: Mouton.
- Doherty, Monika. 1995. Prinzipien und Parameter als Grundlagen einer allgemeinen Theorie der vergleichenden Stilistik. In *Stilfragen*, G. Stickel (ed.), 181–197. Berlin: Mouton de Gruyter.
- Doherty, Monika. 1996. Passive perspectives: Different preferences in English and German – A result of parameterized processing. *Linguistics* 34(3): 591–643.

- Doherty, Monika. 2003. Parameterized beginnings of sentences in English and German. *Across Languages and Cultures* 4(1): 19–51.
- Drescher, Martina. 2003. Sprache der Wissenschaft, Sprache der Vernunft? – Zum affektleeren Stil in der Wissenschaft. In *Gruppenstile. Zur sprachlichen Inszenierung sozialer Zugehörigkeit*, S. Habscheid & U. Fix (eds), 53–79. Frankfurt: Lang.
- Fabricius-Hansen, Cathrine. 1999. Information packaging and translation: Aspects of translational sentence splitting (German-English/Norwegian). In *Sprachspezifische Aspekte der Informationsverteilung*, M. Doherty (ed.), 175–214. Berlin: Akademie Verlag.
- Gazdar, Kaevan & Kirchhoff, Klaus Rainer. 2001. *Geschäftsbericht ohne Fehl und Tadel*. Neuwied-Kriftel: Luchterhand.
- Goffman, Erving. 1981. *Footing. Forms of Talk*. Oxford: Blackwell.
- Halliday, M. A. K. 1994. *An Introduction to Functional Grammar*. London: Arnold.
- Halliday, M. A. K. & Matthiessen, Christian. 2004. *An Introduction to Functional Grammar*. London: Arnold.
- Harwood, Nigel. 2005a. 'Nowhere has anyone attempted ... In this article I aim to do just that': A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of Pragmatics* 37: 1207–1231.
- Harwood, Nigel. 2005b. 'We do not seem to have a theory ... The theory we present here attempts to fill this gap': Inclusive and exclusive pronouns in academic writing. *Applied Linguistics* 26(3): 343–375.
- Hawkins, John A. 1986. *A Comparative Typology of English and German: Unifying the Contrasts*. Austin TX: University of Texas Press.
- House, Juliane. 1996. Contrastive discourse analysis and misunderstanding: The case of German and English. In *Contrastive Sociolinguistics*, M. Hellinger & U. Ammon (eds), 345–361. Berlin: Mouton de Gruyter.
- House, Juliane. 2003. Misunderstanding in intercultural university encounters. In *Misunderstanding in Social Life. Discourse Approaches to Problematic Talk*, J. House, G. Kasper & S. Ross (eds), 22–56. Harlow: Pearson.
- House, Juliane. 2004. English as a lingua franca and its influence on texts in other European languages. In *Mediazione Linguistica E Interferenza*, G. Garzone & A. Cardinaletti (eds), 21–48. Milano: FrancoAngeli.
- House, Juliane. 2006. Communicative styles in English and German. *European Journal of English Studies* 10(3): 249–267.
- Hyland, Ken. 2001. Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes* 20(3): 207–226.
- Hyland, Ken. 2002. Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics* 34: 1091–1112.
- Hyland, Ken. 2005. Stance and engagement: A model of interaction in academic discourse. *Discourse Studies* 7(2): 173–192.
- Koller, Werner. 2000. Übersetzungen ins Deutsche und ihre Bedeutung für die deutsche Sprachgeschichte. In *Sprachgeschichte: Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*, W. Besch, A. Betten, O. Reichmann & S. Sonderegger (eds), 112–129. Berlin: Mouton de Gruyter.
- Kotthoff, Helga. 1989. *Pro und Kontra in der Fremdsprache*. Frankfurt: Lang.
- Kretzenbacher, Heinz. 1994. 'Just give us the facts!' The connection between the narrative taboo, the ego taboo and the metaphor taboo in scientific style. *Lingua e Stile* 29(1): 115–130.

- Kretzenbacher, Heinz. 1998. Fachsprache als Wissenschaftssprache. In *Fachsprachen*, 1. Halbband, L. Hoffmann, H. Kalverkämper & H. E. Wiegand (eds), 133–142. Berlin: Mouton de Gruyter.
- Kuo, Chih-Hua. 1999. The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes* 18(2): 121–138.
- Mauranen, Anna. 1993. Contrastive ESP rhetoric: Metatext in Finnish – English economic texts. *English for Specific Purposes* 12: 3–22.
- Möller, Georg. 1958. *Guter Stil im Alltag. Eine neuartige Satzbauschule*. Leipzig: Verlag Enzyklopädie.
- von Polenz, Peter. 1981. Über die Jargonisierung von Wissenschaftssprache und wider die Deagentivierung. In *Wissenschaftssprache. Beiträge zur Methodologie, theoretischen Fundierung und Deskription*, T. Bungarten (ed.), 85–110. München: W. Fink.
- Seibicke, Wilfried. 1969. *Wie schreibt man gutes deutsch? Eine Stilfibel*. Mannheim: Duden Verlag.
- Schiffrin, Deborah. 1987. *Discourse Markers*. Amsterdam: John Benjamins.
- Tannen, Deborah. 1982. The myth of orality and literacy. In *Linguistics and Literacy*, W. Frawley, (ed.), 37–50. New York NY: Plenum Press.
- Vassileva, Irina. 1998. Who am I /who are we in academic writing?: A contrastive analysis of authorial presence in English, German, French, Russian and Bulgarian. *International Journal of Applied Linguistics* 8(2): 163–190.
- Weinrich, Harald. 2005. *Textgrammatik der deutschen Sprache*. Hildesheim: Georg Olms.
- Wierzbicka, Anna. 1991. *Cross-Cultural Pragmatics. The Semantics of Human Interaction*. Berlin: Mouton de Gruyter.

# Empirical studies of translations as a mode of language contact

## “Explicitness” of lexicogrammatical encoding as a relevant dimension<sup>\*</sup>

Erich Steiner

University of Saarbrücken

A brief initial review of the notions of *language contact* and of *multilinguality* is followed by a summary assessment of what a specifically multifunctional and feature-based perspective on them implies in terms of methodology and of domain. This perspective is here instantiated in a corpus architecture, querying tools, and research hypotheses. It is then discussed and evaluated against a first round of empirical results on explicitness and explication in English and German texts, obtained on the basis of relatively low-level linguistic features. Finally the issue will be raised of how the frequently employed notions of *directionality of change* and of *frequency of usage* may have a bearing on a modelling of language change which operates on properties of encoding, such as explicitness, density and directness of discourses.

**Keywords:** translation, language contact, explicitness, empirical methodologies

### 1. Introduction

Linguistic studies of *language contact* and of *multilinguality* are among the main motivating research activities for studies of language in general. An assessment

---

<sup>\*</sup> The research reported on here is sponsored by the German Research Foundation (DFG) as project no. STE 840/5-1. For current information cf. <http://fr46.uni-saarland.de/croco/>. The author would like to express his gratitude to the Sonderforschungsbereich “Mehrsprachigkeit” at the University of Hamburg (and two of its anonymous referees) for providing a discussion forum for issues related to the work reported on here, as well as to Silvia Hansen-Schirra, Stella Neumann, and Mihaela Vela for continuing collaboration, and to Kerstin Kunz and to Elke Teich for discussion and feedback.



of what a specifically multifunctional and feature-based perspective on those phenomena implies in terms of methodology and of domain should be a welcome addition to a field which usually has a broadly functional motivation, but a predominance of formal and structure-oriented methodologies in some respects. The multifunctional perspective outlined here is characterized by a due consideration of systems alongside structures, of the instance alongside the system, of more abstract (and at the same time, more empirical) types of contrast than have often been in the centre of theorizing, and by the assumption of a metafunctional modularization of linguistic structure. We would like to illustrate how it can feed into and enrich existing more structure-based modelling of language contact phenomena.

In a second and methodologically more specific cycle of argumentation, the perspective outlined above will be given some more detail, instantiated in a multi-layer electronic corpus architecture, a combination of querying tools for information extraction, and in a methodology for the empirical investigation of translations between English and German in particular. This perspective and methodology will then be discussed and evaluated against a first round of empirical results, obtained on the basis of relatively low-level linguistic features, such as lexical density, type-token statistics, and word-class distributions, in terms of which registers within and between languages, as well as originals and translations can be seen to differ. “Low-level” here means “of relatively low abstraction”. More high-level features, such as those operating over constituent structures or grammatical functions require more annotation (ideally up to tree-bank level) and advanced querying, both of which we are working on currently. The question will furthermore be addressed as to what extent these statistical findings can be interpreted as indicators of explicitness and explicitation in texts (for definitions of these two terms cf. the beginning of Section 3 of this paper). Our aim on a general level is that of making another attempt at closing, or at least narrowing, the gap between (top-down) modelling of language contact phenomena and (bottom-up) collection of empirical data.

Finally the issue will be raised of how the frequently employed notions of *directionality of change* and of *frequency of usage* may have a bearing on a modelling of language change which operates on properties of encoding in a multi-functional and feature-based view on language. In particular, we shall engage in some initial thoughts about how the likely increase in multilinguality, and translation as a special case of it, may have an influence on rate and direction of language change through properties such as different degrees of explicitness, density and directness of discourses.

## 2. Clarification of terms

*Language contact* is the situation in which languages, or rather, instantiations of language systems through their speakers, influence each other synchronically in shared socio-semiotic contexts (classical accounts include Weinreich 1953; Thomason & Kaufman 1988; Oesterreicher 2001). This is complementary to the historical axis, along which genetically related languages are in contact through time. Language contact applies to varieties within languages, as it does to different standard languages. Major topics of research are (cf. Thomason & Kaufman 1988: 65ff.):

- the interplay between synchronic contact and genetic inheritance
- linguistic vs. socio-cultural constraints on interference
- analytic frameworks for contact-induced language change (linguistic levels of change; borrowing vs. interference through shift; predictive power of the frameworks, external vs. internal explanations)
- language maintenance
- normal vs. exceptional transmission (creolization, pidgins)

In an attempt to generalize on the strength and on linguistic levels of language contact, a *borrowing scale* is postulated, ranging from lexical borrowing only through slight structural borrowing through moderate structural borrowing to heavy structural borrowing. Most studies to date have focussed on lexical items and/ or grammatical structures, rather than on features or properties of the linguistic systems and instances (discourses, texts) involved, although both perspectives have often been acknowledged as relevant (cf. also Heine, this volume, in his Figure 1 on contact-induced linguistic transfer).

*Multilinguality* is usually predicated either of individuals, or of linguistic communities as socio-cultural formations, or else of discourses/ texts (for a representative and comprehensive survey cf. Auer & Wei 2007). In the first sense, studies of multilinguality are often carried out as studies of language development/ acquisition of several languages in one speaker (bilingualism, trilingualism etc.). In the second sense, they are targeted at linguistic communities and are methodologically situated in sociolinguistics. A terminological distinction which reflects this division is that between *bilingualism* as referring to the individual, and *diglossia* as referring to communities. In the third sense, there are a few strands of research into *multilingual text production* (cf. Matthiessen 2001; Steiner & Yallop 2001; Teich 2001; Steiner 2004a, b, 2005a, b, c), cross-cultural pragmatics (House 1997, 2002) and information structure across languages

(Hasselgard et al. 2002; Fabricius-Hansen & Ramm forthc.), in which *multilinguality* is treated as a property of discourses which are assumed to have interesting and typical properties compared to monolingual discourses (cf. several contributions in Franceschini (ed.) 2005, in particular v. Stutterheim & Carrol 2005). If we say that discourses are multilingual, then we imply that they show special discourse properties of *directness vs. indirectness*, *orientation towards self vs. other*, *orientation towards content vs. interaction*, *explicitness vs. implicitness*, *routine-orientedness vs. ad-hoc formulation* (as e.g. in House 1997:84), ultimately to be realized as lexicogrammatical properties such as interference, borrowing, code-/language switching, special metafunctional orientations in terms of ideational, interpersonal, or textual biases, directness, density, explicitness, and others. These discourses thus instantiate specific contact varieties, or registers. In our own research, we regard translations as an important venue of influence in language contact (cf. Frawley 1984; Baker 1996 for translations as a special text type or even code). But this venue of influence is additional to, and different from, more traditional venues of contact through borrowing or interference. It is less obvious, the resulting varieties are superficially close to native ones, and it applies intra-lingually, across registers, as much as it does inter-lingually.

Investigations of multilinguality are meaningful on all of the levels mentioned above, provided the empirical claims that are being made by the ascription of the property to individuals, communities, or discourses are clear. Furthermore, in the case of discourses, it must be clear whether empirical claims are made about properties on the level of text/ discourse, or else on the level of lexicogrammar – or about both of them. A multifunctional and feature-based perspective will usually encompass the discourse-oriented perspective, at least as an important component, and certainly as a prominent object of study. Multilinguality of discourses can be assumed to be a property which is both a result of, and an environment for, language contact and change.

In a first attempt to characterize our own research efforts relative to the substantial tradition of research briefly characterized above, it will be obvious that they owe most to Systemic Functional Linguistics (cf. Halliday & Hasan 1976; Halliday 1978; Halliday & Martin 1993; Halliday & Matthiessen 1999, 2004). We have additionally drawn on other comparative perspectives with some functional leanings (cf. Hopper & Thompson 1982; Hawkins 1986; Thomason & Kaufman 1988; Biber 1995; Simon-Vandenberg & Steiner 2005; Traugott & Dasher 2005), and on insights from certain strands in translation studies, contrastive linguistics and cross-cultural pragmatics (Doherty 1996, 2002, 2006; Fabricius-Hansen 1996; House 1997/1977, 2002). In terms of methodology, a perspective of the kind advocated here will give due consideration

- to systems alongside structures,
- to the instance alongside the system,
- to more abstract types of contrast, for example in terms of *explicitness*, than have often been in the centre of theorizing, and
- to the metafunctional modularization of language.

Corpus-based work in our own group on original and translated texts in English and German (cf. Hansen 2003; Neumann 2003; Teich 2003) shows how an instance-based orientation of work on multilingual discourses can yield new insights and methodologies in addition to the more traditional system-based investigations. It is also within this instance-based perspective that properties of discourses come into view which are below the threshold, or outside the realm of, borrowing of lexical or structural patterns across languages. A typical case are “good” translations, which often show no lexical or structural trace of language contact, but which may have a characteristically different “feeling” to them, which is the result of different frequencies and proportionalities of native patterns, rather than the result of borrowing or interference on the lexical or structural levels. What these pieces of research do, methodologically, is to combine a Biber-type corpus orientation (cf. Biber 1995 and elsewhere) with multi-layer corpus architectures and annotations, elaborated querying techniques and modelling of multilinguality against the background of more structured linguistic theories, especially of functional orientations.

In my own work, I have attempted to derive from lexicogrammatical patterns some more abstract (and at the same time, more empirical) properties than have often been in the centre of theorizing. One of these, *explicitness*, is in focus here. Languages, through their instantiations in texts and discourses, influence each other in contact situations if there is some relevant sense of a contrast. Traditionally, these contrasts have often been sought in the non-existence of lexical items and their immediate grammatical environment in a receptor language. Beyond the lexical level, borrowing scales such as the one postulated by Thomason & Kaufman (1988: 65ff.), are an attempt at systematizing processes of borrowing into receptor languages, or interference from source languages, in terms of grammatical structure. There is an underlying assumption of gaps in the receptor language, or otherwise a strong influence of the source language through *shifting* speakers. Again the expectation is one of some relevant contrast inviting the borrowing or interference, which is an expectation shared in our work. However, we posit additional levels of observation and modelling: in the first place, the relevant contrasts may be in terms of higher level text- and discourse structures, and only through these in lexis and grammar. In the second place, the contrasts may

manifest themselves initially and for a substantial period in terms of changing frequencies of existing configurations of lexicogrammatical features, rather than in the borrowing or interference of individual lexicogrammatical structures. Pressure towards language change builds up, as it were, through changed frequencies of existing constructions long before it manifests itself in new structures on any one of the linguistic levels. This does not mean that any of the more traditional studies of language contact and change are unimportant or obsolete, but rather that perceived differences in ways of structuring discourses often have to do with changes in relative frequencies. These are then perceived as making a text more or less *explicit*, *direct*, or *dense* than some received norm in some language or variety within a language. These differences are often hardly above the threshold of perception, and thus constitute much more of a cline of perceived properties of texts/ discourse, than coarse binary distinctions such as *natives vs. non-native* command of a language or variety would suggest. And in this sense, translations can be expected to constitute a prime example of contact varieties (cf. Baumgarten & Özçetin “Language variation through language contact in translation”; this volume).

Finally, several linguistic frameworks have postulated a modularization of linguistic structures along different dimensions, usually adopting some diversification into, roughly, referential/ ideational/ propositional vs. interactional/interpersonal vs. textual/ organizational meanings. The latter dimension is an enabling function yielding structure in terms of Theme vs. Rheme, Topic vs. Comment and Given vs. New information. A model giving prime architectural place to these distinctions is Systemic Functional Linguistics (SFL). Within such a model, avenues of language contact will be modularized by metafunction, and will be conceptualized to operate on properties (features), rather than on structures primarily. The more structure-oriented tradition in grammaticalization studies has focused on explorations of morphosyntactic change, building on Lehmann’s (1995) classic study on processes and parameters of grammaticalization. This type of grammaticalization research mainly focuses on the change of free syntactic units into highly constrained morphemes with a grammatical function. A significant step in the more system-based and multifunctional direction has, outside of SFL, been taken in some more recent work by Traugott & Dasher (2005), who focus on semantic-pragmatic change in grammaticalization. They hypothesize semantic change to proceed along the following cline: *propositional towards textual towards expressive*. A cline such as the one postulated here is, of course, strongly reminiscent of the metafunctional modularization in SFL of dimensions of grammatical structure into ideational, textual and interpersonal, an architecture which has been exploited in recent typological work on a range of languages (cf. Caffarel et al. 2004; Steiner & Teich 2004; and Matthiessen 2004 in particular).

In the following chapter, a methodology will be described for empirical investigations of translations, which lend themselves particularly well to empirical research in a multifunctional and feature-based perspective as outlined above.

### 3. A methodology for the empirical investigation of translations

In this chapter, the perspective outlined above will be given some more detail. It is instantiated in a multi-layer electronic corpus architecture (cf. Neumann & Hansen-Schirra 2005; Hansen-Schirra, Neumann & Vela 2006), in a combination of querying tools for information extraction, and in a methodology for the empirical investigation of translations between English and German in particular. This perspective and methodology will then be discussed and evaluated against a first round of empirical results, obtained on the basis of relatively low-level linguistic features, such as lexical density, type-token statistics, word-class distributions, in terms of which registers within and between languages, as well as originals and translations can be seen to differ. The question will furthermore be addressed as to what extent these statistical findings can be interpreted as indicators of explicitness and explication in texts.

#### 3.1 Corpus architecture, types of contrast, hypothesis formation, querying techniques<sup>1</sup>

Figure 1 below graphically represents the structure of our corpus. Our arrangement of sub-corpora permits the following relevant types of contrast to be investigated:

- **Contrast C1** (*reference corpora, cross-register*) between the English Reference Corpus (ER) and the German Reference Corpus (GR), each consisting of 34 000 words (17 registers of 2000 words each). Contrasts under C1 yield a cross-register profile for original texts in the languages English and German.
- **Contrast C2** (*register controlled*) between the registers of *Essay, Fiction, Instruction, Popular Science, Letters to Shareholders, Speeches, Tourism, Web* for each of English originals (EO), German originals (GO), English translations (ETrans), and German translations (GTrans), each of the sub-corpora having 31.250 words per register, each register sample comprising at least 10 texts,

---

1. We are currently developing some statistics for interpreting data and results, in particular for the comparison of groups, measuring relationships such as (normal) distribution and similarity, and also ultimately clustering techniques (cf. among others Oakes 1998).

- 250 000 words altogether for each of EO, GO, ETrans, GTrans. The translations are all translations of the corresponding samples of matching originals. Within this contrast, we can separately investigate **Contrast C2.1** (*within one register, between languages, differentiated into 8 sub-contrasts by register EO vs. GO*), and **Contrast C2.2** (*between registers, within each of the languages English and German, yielding 8 contrasts within each of the corpora EO and GO*).
- **Contrast C3** (*translations vs. originals within each of the two languages*): EO vs. ETrans and GO vs. GTrans, yielding either one contrast per language globally, or, if intersected by register, 8 contrasts between originals and translations for each of the languages.
  - **Contrast C4** (*originals and their translations across languages, i.e. EO vs. GTrans and GO vs. ETrans*); this contrast is the only one between originals and their translations. We differentiate this into **C4.1 Sub-corpora (and texts) as wholes (without alignment)** and **C4.2 Aligned corpora, i.e. explication by translation units**, and we investigate translations between English and German in both directions.

Let us now, after specifying the corpus architecture and the types of contrast which it allows to investigate, proceed to a discussion of hypothesis formation and querying techniques to test these hypotheses. Staying with contrast C.4.2. for the moment, we will exemplify some hypotheses and associated queries on

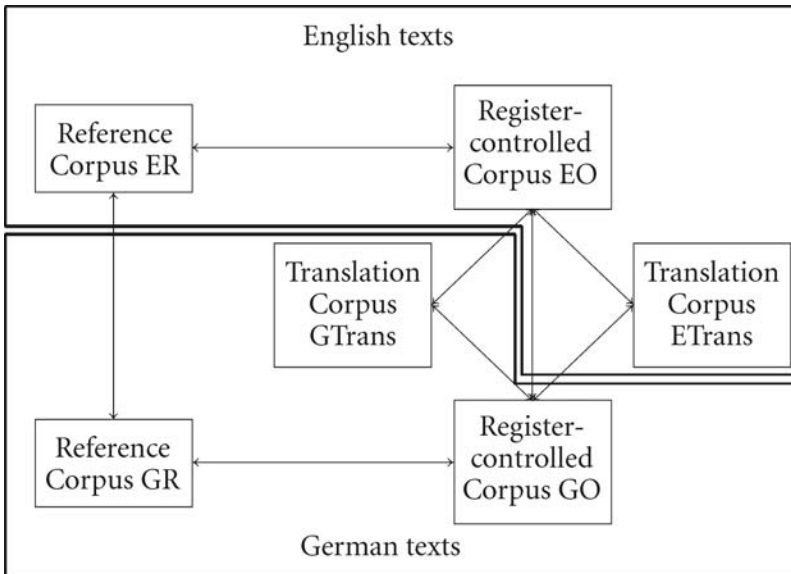


Figure 1. Structure of the CroCo-Corpus.



the basis of the annotation and alignment for the cohesion markers described by Halliday & Hasan (1976) for English, and some equivalents for German (cf. Hansen-Schirra, Neumann & Steiner 2007 for more detail).

1. to 7. below are examples of hypotheses about cohesion to be tested on the data. For either a given pair of non-aligned text segments globally, or else for a given aligned source – target fragment of two texts in a translation relationship (C.4.2), we expect differences along the following parameters:

1. the proportion of explicit to implicit referents;
2. the proportion of phoric to fully lexical (auto-semantic) phrases;
3. the number of newly introduced discourse referents per discourse segment;
4. the amount of cohesive ellipsis and substitution;
5. the strength of lexical chains (patterns of lexical cohesion) as measured by various ratios between content and function words, and as measured by type-token relationships;
6. the strength (internal connectivity) of lexical chains as measured by average number of items per lexical chains;
7. the ratio between explicit and implicit encoding of conjunctive relations.<sup>2</sup>

An indicator for the first hypothesis mentioned above could be the proportion of explicit (pronominal) referents vs. implicit ones in comparing German relative clauses with their non-finite English correspondences. The relevant evidence is reflected in the annotation and alignment at word level. Relative pronouns receive the part-of-speech tag *prels* and if they occur in both languages, they are linked to each other. In cases of a relative pronoun in a German translation without translational correspondence in the English original text, the former is not aligned at all – it receives a so-called empty link. Figure 2 shows the XML representation of the tokenized corpus, the part-of-speech tagging and the alignment on the word level. In the token index file each token is assigned an index number. The part-of-speech annotation of tokens refers back to this index file via *xlinks* specifying the index number of the respective token. The word alignment then refers to the index numbers of both source and target token in turn. In cases of empty links, they receive the value *undefined*.

For the investigation of explicit pronominal referents in German relative clauses vs. implicitly encoded English referents, all German tokens with the

---

2. Observe that in comparing any text fragments which are not in a unit-of-translation-relationship (all of C1–C4.1.), we are testing for the global property of (relative) explicitness. However, whenever we are comparing a specific aligned and instantiated source-target (translation) unit, we are testing “explicitation” (or its opposite, implicitation), as defined in Section 3 of this paper.



English original: ... a palmist, inferring the future out of his own lined flesh		
German translation: ... ein Handleser, der seine Zukunft aus den eigenen Linien ableitete		
( a palmist who his future out-of the own lines inferred )		
<b>token index file</b>	<b>part-of-speech annotation</b>	<b>word alignment</b>
<token id="t64" strg="ein"/>	<token pos="art" xlink:href="#t64"/>	<token> <align xlink:href="#t55"/>
<token id="t65" strg="Handleser"/>	<token pos="nn" xlink:href="#t65"/>	<align xlink:href="#t66"/> </token>
<token id="t66" strg=","/>	<token pos="yc" xlink:href="#t66"/>	<token> <align xlink:href="#t56"/>
<token id="t67" strg="der"/>	<token pos="prels" xlink:href="#t67"/>	<align xlink:href="#t74"/> </token>
<token id="t68" strg="seine"/>	<token pos="pposat" xlink:href="#t68"/>	<token> <align xlink:href="#undefined"/>
<token id="t69" strg="Zukunft"/>	<token pos="nn" xlink:href="#t69"/>	<align xlink:href="#t67"/> </token>

Figure 2. XML corpus annotation and alignment on word level including empty links.

```

for $k in $doc//tokens/token
let $fileName := $doc//translations/translation[ @n='1']/@trans.loc
let $fileNameNew := replace($fileName,"tok","tag" )
where ($k/align[1][ @xlink:href != "#undefined"] and $k/align[2]
[ @xlink:href = "#undefined" ] and doc($fileNameNew)//token
[ @xlink:href eq $k/align[1]/@xlink:href][ @pos eq "prels" ]

```

Figure 3. XQuery for relative pronouns with empty links.

part-of-speech tag *prels* (for relative pronoun) have to be extracted which are not aligned on the word level. The respective XQuery is shown in Figure 3.

The output of this query are sentences like the ones displayed in Figure 2 earlier. This example (taken from the fiction sub-corpus) is interpreted as explicitation since participant role (and thus the reactivation of the referent), tense and mood are explicitly realized in the finite relative clause of the German translation, whereas they are implicit in the English original.

A similar query could be posed for conjunctive relations. Here, all German tokens with the part-of-speech tag *kous* (for conjunction) are to be extracted which are not aligned on word level (since the conjunctive relation is encoded implicitly, for instance through a participle clause). The results for this query displayed in Figure 4 are taken from the sub-corpus of shareholder information. Here, all examples show explicitation in the German translations, since the implicit conjunctive relation encoded in the English participles (marked in bold face) are translated explicitly with German conjunctions.

Another interesting phenomenon in terms of our hypotheses 1–7 above is ellipsis in translations, since it is a direct indicator for implicitation. The alignment of our corpus on different layers enables us to find ellipsis through empty alignment links on all alignment levels. On the basis of these empty links, we can furthermore investigate in which syntactic functions ellipses occur, whether they

```

<result no="13"><ori_en>Baker Hughes Business Support Services has assumed accounting, payroll, benefits and IT
support duties for many of the company's U.S. operations, eliminating duplicate efforts by division personnel. </ori_en>
<trans_ge>Baker Hughes Business Support Services hat die Buchführung, Gehalts- und Sozialleistungen sowie IT-
Aufgaben für viele Niederlassungen des Unternehmens in den Vereinigten Staaten übernommen, wodurch doppelte Arbeit
durch das Personal in den Tochterunternehmen vermieden werden konnte. </trans_ge></result>
<result no="14"><ori_en>In this environment, Baker Hughes revenue declined 22% to $4.5 billion for 1999, compared to
$5.8 billion in 1998. </ori_en>
<trans_ge>Vor diesem Hintergrund sanken die Umsatzerlöse von Baker Hughes im Jahre 1999 um 22% auf 4,5 Mrd.
Dollar, während sie 1998 noch 5,8 Mrd. Dollar betragen hatten. </trans_ge></result>

```

Figure 4. Results for conjunctions with empty links.

prefer finite or non-finite clauses or how they are dealt with in different translation directions.

### 3.2 Some “low-level” results on explicitness of encoding<sup>3</sup>

In this section, an initial discussion of a first round of low-level results will be presented. The limitation is two-fold in that we shall have to restrict ourselves to automatically assigned low-level features, and we shall not yet be able to discuss aligned corpora, thus staying with contrasts C1.–C4.1. in this section. The first restriction is due to the additional time needed to hand-annotate substantial parts of the corpus (for clauses, syntactic functions, phrases), while the second is due to the fact that alignment other than alignment by sentence, can only be a product of these higher-level annotations. However, even our as yet restricted data allow us a first impression of what types of hypotheses can be developed and what types of results can be expected. Our overall numerical results for the dependent variables of *lexical density* (LD), *type-token ratios* (TTR) and *for part-of-speech* (POS) proportionalities are given in the Appendix. Please observe that we are explaining the calculation of our dependent variables in footnote 6 in the Appendix. Only a few interesting first results will be interpreted in this paper, at this stage still without statistical computations for *significance*, which will be undertaken somewhat later.

Starting with a comparison of the *reference corpora* (C1), our first dependent variable is lexical density as percentage of lexical words of all words in a (sub-) corpus (cf. *Lexical Density* in the corresponding statistics in the Appendix: E: 57.49 vs. G: 56.98). The reference corpora thus deviate slightly from *total language*

---

3. Error margins as checked by us are all below 3% for lemmatization, and below 4% for POS tagging, with the exception of ETrans, which have an error margin of almost 7%, but only due to a high number of one case of error in two of the sub-corpora. These are currently being isolated and reduced.

in our register-controlled corpora (E: 55.44 vs. G: 55.62), and they deviate somewhat more from those for the originals only (EO vs. GO), and for some individual registers. They are also somewhat higher than the figures reported for English in Biber et al. 1999 (62ff., especially 65). But more significantly, they point into opposite directions between the two corpora. The difference in lexical density (LD) in our reference corpora will later be taken as a base against which LD-differences in C2–C4 are interpreted.

The difference in type-token relationships (cf. Table ‘*Type Token Ratio*’ in Appendix) between our reference corpora is E: 15.64 against G: 21.71. This looks like a surprisingly high difference, particularly as it cannot be attributed to a higher lexical density of our German reference corpus, which one might have expected given the morphologically more analytical character of English compared to German. This difference in TTR, though, is confirmed by figures from our register-controlled corpora. Note that TTR-comparisons are only valid between corpora of equal size (cf. Biber et al. 1999: 53), and of course only on lemmatized corpora, as in our case. The two reference corpora have 34 000 words each, register-controlled sub-corpora all have around 31 000, while each of EO, GO, ETrans and GTrans have around 250 000.

The differences in overall word class distribution in terms of *Part of Speech* in our reference corpora are to be found in the top and bottom lines respectively of our Table ‘*Parts of Speech per Sub-corpus in %*’ in the Appendix. The proportions for some selected *verbal: nominal* oppositions are:

- |                                  |                                       |
|----------------------------------|---------------------------------------|
| – <i>verb: noun</i>              | (E: 15.72: 24.60 vs. G: 13.04: 22.93) |
| – <i>adverb: adjective</i>       | (E: 4.63: 6.24 vs. G: 5.02: 9.20)     |
| – <i>conjunction: adposition</i> | (E: 5.12: 10.55 vs. G: 4.47: 8.08)    |
| – <i>pronouns</i>                | (E: 5.46 vs. G: 8.45)                 |

There seems to be a degree of dominance of nominal word classes in German (N+Adjective+Adposition) vs. Verbal Classes (V+Adverb+Conjunction) (40.21 vs. 22.53) compared to English (Nominal: 41.39, Verbal. 25.47), especially if we take into account the fact that for the pronouns, German once more has 8.45, against English having 5.46 only.<sup>4</sup>

There is no straightforward interpretation as to overall explicitness here, because the stronger verbal orientation of English may easily be counterbalanced by

---

4. In our nominal: verbal oppositions, our grouping of *conjunctions* with *verbal* and *adpositions* as *nominal* has been done largely based on their preferred complementation patterns: conjunctions project more clearly verbal constituents than adpositions, which project more nominal ones. Obviously, we rely on graded notions of “nominality/ verbality” here, as is usually the case in syntax nowadays.

the stronger morphological marking in German. However, if the stronger nominal orientation of German can later be shown to be due to a higher number of arguments per predicate, this would be one type of explicitness. *Explicitation* does not enter into the picture here, as the two reference corpora are not in a translation relation.

Returning to lexical density, higher measures may be taken to be evidence for higher experiential explicitness, though in no way higher explicitness of logical relations (*and, or, if* etc.) or text structuring devices (*theme, focus* etc.). It is not clearly related to any interpersonal meanings (*clause mood, modality, affect*).

As for type-token statistics, higher values for the German reference corpus may indicate strength of lexical (rather than pronominal) cohesion, possibly explicitly lexical specification of domains.<sup>5</sup> Any direct comparisons between the findings for German and English here will not be interpreted in themselves, but will be taken as a background profile against which later findings on C2–C4 will be interpreted.

Moving on now to **Contrast C2** (*register controlled*), we shall begin with C2.1., that is contrasts within one register of original texts between languages, differentiated into 8 sub-contrasts by register. All the relevant data are in the Table ‘*Lexical Density in %*’ in our Appendix. The relevant columns there are the two columns under “Original” and the entries under “Reference Corpora”. The first phenomenon to be noted here is that the figures for *Total\_Reg* are almost identical for English and German (56.03 (E) and 56.06 (G)), unlike in the case of the reference corpora (57.49 vs. 56.98). The *Total percentages* are thus lower by about 1 to 1.5% in the corpora of originals than in the reference corpora. Both findings are likely to be a reflection of the different composition of the two sub-corpora.

As for the direct comparison by register, there are differences larger than 1% (Fiction, PopSci, Share, Speech, Tourism) for some registers of the originals, with the direction of differences pointing in both directions, depending on register. In other words, lexical density seems to vary depending on the language, because of our findings for the reference corpora, but more strongly depending on the register as shown here.

*Type-token relationship.* The relevant figures are in Table ‘*Type Token Ratio in %*’ in the Appendix, the columns for ‘*Originals*’ (E) and ‘*Originals*’ (G) in particular, again compared to the values under ‘*Reference Corpora*’. The values for

---

5. Comp. very strongly Biber et al. (1999: 43) “... a high type/ token ratio... serves to increase the semantic precision and informational density of a written text, and thus it can be required to perform a challenging ideational task, at the same time, it is a reflection of considerable opportunity for careful production, as in many writing situations, as opposed to real-time production constraints characterizing most speaking situations”.

German are much higher than for English for all registers, where it should be emphasized once more that our findings are based on lemmatized corpora, so morphology does not account for the differences.

If we then move over to the *overall word class distribution in terms of PoS (EO vs. GO)* (cf. Table 'Parts of Speech per Sub-corpus in %' in the Appendix), we find that for the combined 8 registers of originals, the nominal word classes seem to score higher in German than in English. English is characterized by the following part-of-speech percentages: N (27.21) + Pron (4.73) + Adj (7.16) + Adpos (10.62) = 49.72 (Nominal words). German scores higher overall on nominal words: N (24.51) + Pron (9.32) + Adj (9.41) + Adpos (8.34) = 51.58 (Nominal Words).

The reverse is true for the verbal classes, where English has: V (14.90) + Adv (3.93) + Conj (5.44) = 24.27 (Verbal Words). German scores overall V (12.28) + Adv (3.99) + Conj (5.37) = 21.64 (Verbal Words).

As for an overall interpretation, whereas in terms of lexical density German registers and English registers vary as to which is the more explicit, and whereas in terms of type token ratios, it seems to be the German registers throughout which are more explicit, English registers seem to be higher throughout in terms of verbal orientation than the German ones. Again the question remains, what precisely this implies for explicitness, but this will be further investigated.

The next contrast to be discussed here is Contrast C2.2., that is between registers, within each of the languages English and German, yielding 8 contrasts within each of the languages.

Starting with *Lexical Density* (for complete figures cf. Appendix, Table on 'Lexical Density'), the spread between registers in English originals (EO) is 59.14 – 47.73, that is 11.41 percentage points. For German (GO) the spread is 60.74 – 51.05 = 9.69. The spread of variation among English registers is larger in terms of lexical density than it is for German. Interestingly, the extreme registers for both languages are *Fiction* (low) and *Tourism* (high). We can also see that the spread in lexical density *within* languages is much higher than for any register *across* the two languages, although the precise interpretation of these numerical findings must await further data to be intersected with.

Moving on to 'Type-Token-Ratio (TTR)', the spread for English originals is 12.66 – 7.00 = 5.66, for German originals 24.17 – 14.29 = 9.88. The spread of variation among German registers is thus numerically larger in terms of TTR than it is for English. This time, though, the extreme registers for the two languages are *Instruction* (low) and *Fiction* (high) for English, but *Instruction* (low) and *Tourism* (high) for German, with additional differences in ranking between the remaining registers.

Let us next proceed to *overall word class distribution in terms of PoS* (within EO and GO, but between registers). We compute a simple proportion between a

generalized “nominal super-class” (N+Pro+Adj+Adp) divided by a generalized “verbal super-class” (V+Adv+Conj), on the basis of PoS statistics in the two sub-corpora of originals (cf. Appendix “Parts of Speech in English (resp. German) Sub-corpora in %”). For individual registers, the proportionalities are as follows, ranked from “high” to “low”:

English:

Tourism (2.749) > Share (2.334) > Web (2.167) > Essay (2.060) > Instr. (1.997) > Speech (1.991) > PopSci (1.755) > Fiction (1.564)

German:

Tourism (3.059) > Web (2.681) > Share (2.675) > Essay (2.339) > Instr. (2.252) > Speech (2.093) > PopSci (1.978) > Fiction (1.798)

It is very interesting that in terms of general nominal orientation, our registers are ranked almost identically in the English and German corpora, which is one language-internal corroboration of our intuitive register sampling, initially based on contextual and cultural categories. It should also be noted that while German comes out as more nominal in our general sense, this tendency does not hold if we simply compare the register specific frequencies of nouns (without pronouns), where English frequently scores higher than German.

Next we shall briefly turn to **Contrast C3**, between translations and originals within each of the two languages (EO vs. ETrans and GO vs. GTrans). Starting once more with lexical density, we find that in English, lexical density is higher for originals than for translations (56.03 : 54.85). This appears due to interference if we take the difference in the reference corpora as a basis. The LD-figures for *total registers* or *total language*, however, marginally point the other way, so that the interpretation as interference needs to be qualified. Increased LD might be interpreted as increased experiential explicitness in originals (because of explicit lexical specification), but possibly also as increased logical explicitness in the translations (increased use of conjunctions and prepositions). Interestingly, LD-comparison goes in different directions depending on register (cf. LD-statistics in *Appendix*), so apart from the process of translation, there seem to be register-specific factors involved. In German, the overall tendency also shows translations as less dense than originals, but interestingly, the registers in which the LD is higher for translations than for originals are not the same as in English. What seems to be clear is that interference from the source language cannot play a role here.

If next we move on to *Type-Token Relationship (TTR)* compared between originals and translations within each of the two languages, we find that for English, TTRs go both ways, although the global figure is higher for originals than for translations (10.33 vs. 10.13 as averages over TTRs; 5.80 vs. 5.60 for counts on

aggregate accumulated running words, i.e. on the whole corpus as one file). Direct numerical comparisons with ER are not valid, due to different corpus size and compositions. The same holds true for German, with the difference being even bigger, and with only one register (*Share*) showing a higher TTR for translations over originals. In view of the baseline figures for the reference corpora, and for our own corpora, this could be largely explained as an effect of interference. One of the interesting cases is the *Share* register in German, where the increased TTR in German translations very clearly cannot be due to interference at all.

Proceeding then to *overall word class distribution in terms of PoS (EO vs. ETrans in "Parts of Speech in the English Sub-corpora in %")*, we see that for the combined 8 registers of English (*Total*), the nominal word classes seem to score higher in originals than in translations. English originals have a percentage of N (27.21) + Pron (4.73) + Adj (7.16) + Adpos (10.62) = 49.72 (Nominal words), whereas for translations, this percentage is N (26.14) + Pron (4.54) + Adj (7.35) + Adpos (11.30) = 49.33 (Nominal Words).

Interestingly, the verbal classes also seem to score higher in English originals (V (14.90) + Adv (3.93) + Conj (5.44) = 24.27) than in English translations (V (14.34) + Adv (4.82) + Conj (4.80) = 23.96). In English translations it is adjectives, adverbs, adpositions and articles which become more frequent, relative to the overall nominal and verbal word classes. In terms of explicitness, this would indicate increased entity and event *modification*, as well as increased logical explicitness and nominal determination for translations over originals within English.

For the combined 8 registers of German originals, the nominal classes score higher in originals than in translations: For nominals, the figure is (N (24.51) + Pron (9.32) + Adj (9.41) + Adpos (8.34) = 51.58), for German translations, we find (N (23.84) + Pron (8.67) + Adj (9.69) + Adpos (8.39) = 50.59). The verbal classes, though, show a slight increase in German translations over originals, which means that German translations have an increased verbal quality relative to register parallel originals. German originals show (V (12.28) + Adv (3.99) + Conj (5.37) = 21.64), whereas German translations have the slightly higher proportion of (V (12.08) + Adv (4.89) + Conj (4.85) = 21.82) for the combined verbal classes.

In summary, in both languages, translations are less dense, lexically less rich (TTR), and partly more verbal than originals. They show increased *entity*- and *event* modification, as well as increased logical explicitness and nominal determination.

Moving finally over to **Contrast C4** between originals and their translations (English translations ETrans of German originals GO and German translations GTrans of English originals EO), we are for the first time in the current context comparing source and target texts. Starting with contrast **C4.1.** between



*Sub-corpora (and texts) as wholes (without alignment)*, we shall discuss English register controlled corpora and their German translations (EO, GTrans):

*Lexical Density* is higher in English originals than in their German translations (56.03 vs. 55.19 in *Total Register*, cf. Table 'Lexical Density' in the Appendix.). This seems to be in line with the tendency in our reference corpora, although even against that background, the size of the difference is surprising. It is not in line at all with the relationship between EO and GO, because this would not lead us to expect a difference (EO 56.03 vs. GO 56.06). We therefore suspect an influence of the translation process. Very significantly, in *Fiction* and *PopSci*, the lexical density of the German translations is even higher than that of the English originals, counter to the overall tendency, which indicates strong register-specific influences. German originals also have higher LD than the English translations (56.06 vs. 54.85), with only *Instruction* and *Share* going against that trend.

*Type Token Relationship (TTR)* is higher in German translations than in EO, but generally (10.33 vs. 16.71, cf. Table 'Type Token Ratio' in the Appendix) not by as much as between the two corpora of originals. Translation as a process seems to have an effect there. The difference between German originals and English translations is – predictably – even wider generally (18.34 vs. 10.13), though not to the same extent in all cases.

Turning our attention to *overall word class distribution in terms of PoS (EO vs. GTrans)* for the combined 8 registers of English originals and their German translations, the nominal word classes seem to score lower for English – though only due to the high percentage of pronouns and adjectives in the German translations. English originals have 49.72 % nominal words (N (27.21) + Pron (4.73) + Adj (7.16) + Adpos (10.62)), whereas German translations have 50.59 % nominal words (N (23.84) + Pron (8.67) + Adj (9.69) + Adpos (8.39)).

The verbal classes, on the other hand, seem to score higher in English originals than in German translations: 24.27 % (V (14.90) + Adv (3.93) + Conj (5.44)) for English originals, vs. 21.82 % for German translations (V (12.08) + Adv (4.89) + Conj (4.85)). The proportionalities N/V are 49.72: 24.27 = 2.048 in EO, vs. 50.59: 21.82 = 2.318 in GTrans.

Comparing these proportionalities with those between originals in the languages we obtain: 2.048 for EO (as above), but 51.58: 21.64 = 2.385 for GO. This means that the German translations have moved somewhat towards their English originals in comparison to German originals.

Moving over now to the corresponding opposite pair, that is to say *overall word class distribution in terms of PoS in German originals and their English translations*, we find the following: GO have 51.58: 21.64 = 2.385 Nominal vs. Verbal



Classes, whereas ETrans have  $49.33: 23.96 = 2.058$  Nominal vs. Verbal Classes. This means that the English translations have moved somewhat towards their German originals in comparison to English originals. Altogether, then, the difference in “nominal orientation” is bigger between originals than between the translations: GO: 2.385 – EO: 2.048 GTrans: 2.381 – ETrans: 2.058, that is to say the difference between originals is 0.337, whereas between translations it is only 0.323.

For contrast C.4.2. *On aligned corpora, i.e. explicitation by translation units* figures are not yet available, due to the incomplete alignment of our corpora at this stage. We are currently publishing some first figures for explicitation in the realization of cohesion (noun vs. pronoun, ellipsis, substitution, lexical cohesion) in translation units (cf. Hansen-Schirra, Neumann and Steiner 2007), but a fuller exploitation of aligned corpora must await further progress in our corpus alignment, which is under way.

#### 4. Explicitness of encoding, directionality of change, frequency of use

Finally the issue will be raised of how the notions of *explicitness*, and ultimately also *density* (cf. Bickel 2003; Noonan ms. and this volume)) and *directness*, alongside the frequently employed notions of *directionality of change* (Traugott & Dasher 2005) and of *frequency of usage* (Bybee & Hopper 2001) may have a bearing on models of language contact and language change operating on properties of encoding in a multi-functional and feature-based view on language.

Our point of departure will be a working definition of the notions of *explicitness* and *explicitation* (cf. Hansen-Schirra, Neumann & Steiner 2007):

*Explicitness* on the lexicogrammatical level is conceptually related to *density* and *directness*. These three are properties of (lexico-)grammatical constructions (cf. Steiner 2004b, 2005a, b, c). The opposite of *explicit* in this usage is *lexicogrammatically not realized, but still part of the construction* (unrealized participant roles, unrealized features in non-finite constructions, grammatical ellipsis, projection of units of meaning onto different grammatical categories, grammatical metaphor, transcategorization, etc.).

On the textual level, *explicitness* is related to properties such as *simplified, normalized, levelled-out, sanitized, (in-)direct, participant-oriented, content oriented* and similar ones (cf. Baker 1996; House 1997, 2002 for relevant work), that is to say properties predicated on entire texts, rather than on domains of grammaticalization, such as the clause.

The *explicitness* of higher level units such as texts/discourses is not simply the sum total of *explicitness* features of clauses. It is an *emergent property* on a higher

level in the sense that the properties on the text level are perceived as a result of the interaction of clause level features, such as *explicitness*, *directness*, *density*, with textual features due to cohesion, markers of genre, and register. All of the latter will, in turn, be realized as lexical and/or grammatical patterns, but their function is not accounted for by lexicogrammar. *Explicitness* on this level can furthermore be a result of global textual patterns (such as type-token ratio, lexical density, etc.), which are *epiphenomena* of lexicogrammatical patterns, but not lexicogrammatical themselves. *Explicitness* – a property of lexicogrammatical or cohesive structures and configurations in one text – is measured through operationalizations of the type we have indicated in Section 2.1 for cohesion, and in earlier publications for lexicogrammar (Steiner 2005a: 56ff.).

*Explicitation*, on the other hand, is a process or a relationship between intralingual variants and/or translationally related texts. The texts resulting from *explicitation* are more *explicit* lexicogrammatically and cohesively than their counterparts. *Explicitation* is thus defined on instantiated, indexed and aligned pieces of discourse/text, translations in particular, which share all or some of their meaning.

**Definition:** We assume *explicitation* if in a translation (or language-internally in a pair of register-related texts) meanings (not only ideational, but including interpersonal and textual) are realized in the more explicit variant which are not realized in the less explicit variant, but which are in some theoretically-motivated sense implicit in the latter. The resulting text is more *explicit* than its counterpart.

Note that this definition is meant to exclude the indefinitely many possibilities through which meaning can simply be added to some text/discourse, without being in any motivated sense *implicit* in the source variant.

*Directionality of change* (Traugott & Dasher 2005) and *frequency of usage* (Bybee & Hopper 2001) may now have implications for a modelling of language contact (and change) in terms of explicitness and related properties:

*Directionality of change* is a notion which can be predicated on different types of structure. In earlier versions, we encounter hypothetical developments from morphologically synthetic to analytic language types, between types of basic word order, or between types of marking relations such as head-marking, dependent marking, mixed-marking etc. Some influential work has postulated cycles of development, driven by the dialectical needs of language users towards increased expressiveness on the one hand, and maximal economy on the other (e.g. Hagege 1993: 147ff.). In more recent times, directionality has sometimes been linked to multi-functional, or multi-dimensional models of language, as for example in the work of Traugott & Dasher, who refer to Halliday's multifunctional hypothesis (Traugott & Dasher 2005: 94f.). Halliday and colleagues have, indeed, in several places raised the issue of language change (e.g. Halliday & Matthiessen

1999: 227ff., 507ff.; Matthiessen 2004: 655ff.), frequently in connection with ideas from general systems theory, without so far having had occasion to spell out all the implications.

Traugott & Dasher (2005: 19ff.) trace a line of theorizing which assumes interactions between *subjectivity*, *intersubjectivity* and *objectivity* in language (use). In earlier versions, Traugott (2005: 94, but originally in 1982) had postulated an unidirectional development of semantic change along the lines of *propositional* > (*textual*) > *expressive*, which later on she differentiated into sub-types (2005: 281). Very interestingly, *subjectivity*, *intersubjectivity* and *objectivity* seem to be properties of grammatical constructions, much in the same way as we conceive of *explicitness*, *density* and *directness*. Where we see our role relative to this interesting line of research is in a comparison of the kinds of abstraction we are making, in their relationship to the multifunctional hypothesis, and, importantly, in our attempts at developing empirical research methodologies based on electronic corpora. Our aim also is to trace the contribution of situations of multilinguality and translation to language contact and change. Finally, we would like to investigate particular registers as sites of contact and change (cf. Traugott & Dasher 2005: 283f., also their remarks on historical pragmatics and historical discourse analysis 99ff; importantly House e.g. 2002). What needs to be clarified is the precise locus of the phenomena we are talking about: grammar, semantics, discourse, or the mapping between them (Traugott & Dasher 2005: 282f.). One attempt at this clarification is made in our remarks about explicitness and explicitation at the beginning of this section (for more detail cf. Steiner 2005c; but also Doherty 2006: 49ff.).

Staying with *explicitness* for a moment, we would speculate that it is a property of constructions and configurations both on the textual and on the lexicogrammatical levels. We would furthermore like to suggest that translations and other forms of multilingual discourse show degrees of explicitness differing from the explicitness of encoding in registerially related non-translational or otherwise monolingual discourses. And we would also assume that it is partly these differences through which pressures towards expressiveness or economy exert their force, thus becoming driving forces of change. Such change, though, would not simply presuppose the existence of relevant differences in explicitness and other properties of that type, but also certain critical levels of frequency before they become effective.

*Frequency* of use can be found in several of the studies in Halliday (2005) (for example pp. 93ff.), where it is argued that linguistic sub-systems can be more or less stable as a consequence of proportional frequencies between relevant types of construction, for example the relationship between positive and negative clauses in the environment of primary tense in a big corpus. These frequencies, and the resulting *markedness*, may be among the driving forces of change (cf. also

Johanson this volume and his notion of “frequential copying” (Section 8)). In Bybee & Hopper (2001: 2f.), frequency is accorded a key role in the *emergence* of structure in discourse (cf. in particular MacWhinney 2001: 449ff., 464f.). It affects the strength of a pattern, works differently on types, tokens and collocations, has effects on pattern productivity, may preserve old structures, may positively work towards fusions, contraction and affixation, may increase accessibility to (sound) change, or accessibility to semantic bleaching and other functional changes. Particularly within an approach operating with *properties* of constructions alongside the constructions themselves, frequency of a construction may itself affect properties (such as explicitness). And finally, the frequency of more or less explicit discourses may be a driving force in change. All of these processes can be assumed to be influenced by the degree of multilinguality of discourses, and this is what we would like to explore in more detail in future work.

## 5. Summary and conclusions

We started out with a brief review of notions of language contact, proceeding then to an instantiation-based notion of *multilinguality*. The latter we predicated on discourses, rather than on either individual speakers or socio-cultural formations. We suggested that while an anchoring of the notion of *multilinguality* in socio-cultural formations and in individuals remains a valid field of investigation, studies of multilinguality had much to gain from a focus on properties of (instantiated) discourses.

As a next step, we argued that a multifunctional and feature-based perspective would add much-needed depth to considerations of

- systems alongside structures, and thus of the linguistic potential,
- the instance alongside the system, and thus to empirical research methodologies,
- more abstract types of contrast than have often been in the centre of theorizing, allowing us to reflect on properties, rather than structures,
- the metafunctional modularization of language, thus creating an interface to functional/ communicatively oriented models of language and culture.

We proceeded to describe a methodology for the empirical investigation of translations, outlining a corpus architecture, our hypothesis formation and our essential querying techniques. Based on these, we discussed some first results on low-level properties of our texts and their contribution to explicitness and explicitation. It was shown how texts in different languages, registers, and between originals and

translation, vary along parameters having a bearing on *explicitness* of discourses. More, and more high-level, data will be obtained within the next year or so.

In a final summary, we think that *explicitness of encoding*, like *subjectivity*, *intersubjectivity*, *objectivity*, but also like *density* and *directness*, may be of different degrees, and that there may be a directionality in language contact situations such that variants (including languages) may influence each other, in principle in either direction, although there is evidence for directionality for at least some of them. We also think that frequency of use will at least through its quantitative impact have a bearing on the strength of contact. For example, relative frequency of translations between given languages and given registers may play a significant role here. In short, languages may well influence each other through direct mechanisms of borrowing and shift, on lexical, on structural and on discourse levels. However, in addition to, and possibly underlying these processes, they may influence each other through the subjectivity, intersubjectivity and objectivity of their discourses in contact, as well as through the relative explicitness of encoding along several functional dimensions.

## References

- Auer, Peter & Wei, Li (eds). 2007. *Handbook of Multilingualism and Multilingual Communication* [Handbooks of Applied Linguistics HAL 5]. Berlin: Mouton De Gruyter.
- Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In *Terminology, LSP and Translation Studies in Language Engineering*, H. Somers (ed.), 175–186. Amsterdam: John Benjamins.
- Biber, Douglas. 1995. *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: CUP.
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Bickel, Balthasar. 2003. Referential density in discourse and syntactic typology. *Language* 79(4): 708–736.
- Bybee, Joan & Hopper, Paul J. (eds). 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.
- Caffarel, Alice, Martin Jim R. & Matthiessen, Christian M. I. M. (eds). 2004. *Language Typology. A Functional Perspective*. Amsterdam: John Benjamins.
- Doherty, Monika (ed.). 1996. *Information Structure as a Problem of Translation*. [Special Issue of *Linguistics*. 34].
- Doherty, Monika. 2002. *Language Processing in Discourse. A Key to Felicitous Translation*. London: Routledge.
- Doherty, Monika. 2006. *Structural Propensities*. Amsterdam: John Benjamins.
- Fabricius-Hansen, Cathrine. 1996. Informational density: A problem for translation theory. In *Information Structure as a Problem of Translation*. [Special Issue of *Linguistics*. 34], M. Doherty (ed.), 521–565.

- Fabricius-Hansen, Cathrine & Ramm, Wiebke (eds). Forthc. *Subordination vs. Co-ordination in Sentence and Text. A Cross-Linguistic Perspective*. (Studies in Language Comparison series 98) Amsterdam: John Benjamins.
- Franceschini, Rita (ed.). 2005. *In einer anderen Sprache*. [Themenheft der Zeitschrift für Literaturwissenschaft und Linguistik Jahrgang 35, Heft 139]. Stuttgart: Metzler Verlag.
- Frawley, William. 1984. Prolegomena to a theory of translation. In *Translation. Literary, Linguistic & Philosophical Perspectives*, W. Frawley (ed.), 159–175. Newark NJ: University of Delaware Press.
- Hagège, Claude. 1993. *The Language Builder*. Amsterdam: John Benjamins.
- Halliday, Michael A. K. 1978. *Language as Social Semiotic. The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Halliday, Michael A. K. 2005. *Computational and Quantitative Studies*. Vol. 6 of *Collected Works of M. A. K. Halliday*. London: Continuum.
- Halliday, Michael A. K. & Hasan, Ruqaiya. 1976. *Cohesion in English*. London: Longman.
- Halliday, Michael A. K. & Martin, Jim R. 1993. *Writing Science: Literacy and Discursive Power*. London: The Falmer Press.
- Halliday, Michael A. K. & Matthiessen, Christian M. I. M. 1999. *Construing Experience through Meaning. A Language-Based Approach to Cognition*. London: Cassell.
- Halliday, Michael A. K. & Matthiessen, Christian M. I. M. 2004. *An Introduction to Functional Grammar*. London: Arnold (earlier versions by Halliday in 1985/1994).
- Hansen, Silvia. 2003. *The Nature of Translated Text. An Interdisciplinary Methodology for the Investigation of the Specific Properties of Translations* [Saarbrücken Dissertations in Computational Linguistics and Language Technology 13]. Saarbrücken: University of Saarbrücken.
- Hansen-Schirra, Silvia, Neumann, Stella & Vela, Mihaela. 2006. Multi-dimensional annotation and alignment in an English-German translation corpus. In *Proceedings of the 5th Workshop on NLP and XML (NLPXML-2006) at EACL 2006*, Trento, 4. April 2006, 35–42.
- Hansen-Schirra, Silvia, Neumann, Stella & Steiner, Erich. 2007. Cohesive explicitness and explicitation in an English-German translation corpus. *Languages in Contrast* 7(2): 241–265.
- Hasselgard, Hilde, Johansson, Stig, Behrens, Bergljot & Fabricius-Hansen, Cathrine (eds). 2002. *Information Structure in a Cross-Linguistic Perspective*. Amsterdam: Rodopi.
- Hawkins, John. 1986. *A Comparative Typology of English and German*. London: Croom Helm.
- Hopper, Paul J. & Thompson, Sandra A. (eds). 1982. *Studies in Transitivity* [Syntax and Semantics 15]. New York NY: Academic Press.
- House, Juliane. 1997/1977. *A Model for Translation Quality Assessment. A Model Revisited*. Tübingen: Gunter Narr.
- House, Juliane. 2002. Maintenance and convergence in translation – some methods for corpus-based investigations. In *Information Structure in a Cross-Linguistic Perspective*, H. Hasselgard, S. Johansson, B. Behrens & C. Fabricius-Hansen (eds), 199–212. Amsterdam: Rodopi.
- Lehmann, Christian. 1995. *Thoughts on Grammaticalization*. München: Lincom (originally 1982).
- MacWhinney, Brian. 2001. Emergentist approaches to language. In *Frequency and the Emergence of Linguistic Structure*, J. Bybee & P. Hopper (eds), 449–470. Amsterdam: John Benjamins.



- Matthiessen, Christian M. I. M. 2001. The environments of translation. In *Exploring Translation and Multilingual Text Production: Beyond Content*, E. Steiner & C. Yallop (eds), 41–126. Berlin: Mouton de Gruyter.
- Matthiessen, Christian. M. I. M. 2004. Descriptive motifs and generalizations. In *Language Typology. A Functional Perspective*, A. Caffarel, J. R. Martin & C. M. I. M. Matthiessen (eds), 537–674. Amsterdam: John Benjamins.
- Neumann, Stella. 2003. *Die Beschreibung von Textsorten und ihre Nutzung beim Übersetzen. Eine systemisch-funktionale Korpusanalyse englischer und deutscher Reiseführer*. Frankfurt: Peter Lang.
- Neumann, Stella & Hansen-Schirra, Silvia. 2005. The CroCo Project. Cross-linguistic corpora for the investigation of explicitation in translations. In *Proceedings from the Corpus Linguistics Conference Series* 1(1): 1–11. <http://www.corpus.bham.ac.uk/PCLC/cl-134-pap.pdf>
- Noonan, Michael. ms. *A Crosslinguistic Investigation of Referential Density*. <http://www.uwm.edu/~noonan>.
- Oakes, Michael P. 1998. *Statistics for Corpus Linguistics*. Edinburgh: EUP.
- Oesterreicher, Wulf. 2001. Historizität – Sprachvariation, Sprachverschiedenheit, Sprachwandel. In *Language Typology and Language Universals* [Handbücher zur Sprach- und Kommunikationswissenschaft 20(1) & 20(2)], M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds), Vol. 2, 1554–1596. Berlin: Mouton de Gruyter.
- Simon-Vandenberg, Annemarie & Steiner, Erich (eds). 2005. *Functional Approaches to Discourse*. Vol. 27(6). Special Issue of *Language Sciences*.
- Steiner, Erich. 2004a. *Translated Texts: Properties, Variants, Evaluations*. Frankfurt: Peter Lang.
- Steiner, Erich. 2004b. Ideational grammatical metaphor: Exploring some implications for the overall model. *Languages in Contrast* 4(1): 139–166.
- Steiner, Erich. 2005a. Some properties of texts in terms of ‘information distribution across languages’. *Languages in Contrast* 5(1): 49–72.
- Steiner, Erich. 2005b. Some properties of lexicogrammatical encoding and their implications for situations of language contact and multilinguality. In *In einer anderen Sprache* [Themenheft der Zeitschrift für Literaturwissenschaft und Linguistik 35. 139], R. Franceschini (ed.), 54–75. Stuttgart: Metzler Verlag.
- Steiner, Erich. 2005c. Explicitation, its lexicogrammatical realization, and its determining (independent) variables – towards an empirical and corpus-based methodology. *SPRIK-reports* no. 36. Oslo. [http://www.hf.uio.no/forskningprosjekter/sprik/docs/pdf/Report\\_36\\_ESteiner.pdf](http://www.hf.uio.no/forskningprosjekter/sprik/docs/pdf/Report_36_ESteiner.pdf)
- Steiner, Erich & Teich, Elke. 2004. Metafunctional profile for the grammar of German. In *Language Typology. A Functional Perspective*. A. Caffarel, J. R. Martin & C. M. I. M. Matthiessen (eds), 139–184. Amsterdam: John Benjamins.
- Steiner, Erich & Yallop, Colin (eds). 2001. *Exploring Translation and Multilingual Text Production: Beyond Content*. Berlin: Mouton de Gruyter.
- von Stutterheim, Christiane & Carrol, Mary. 2005. Subjektwahl und Topikkontinuität im Deutschen und Englischen. In *In einer anderen Sprache* [Themenheft der Zeitschrift für Literaturwissenschaft und Linguistik 35. 139], R. Franceschini (ed.), 7–27. Stuttgart: Metzler Verlag.
- Teich, Elke. 2001. Towards a model for the description of cross-linguistic divergence and commonality in translation. In *Exploring Translation and Multilingual Text Production: Beyond Content*, E. Steiner & C. Yallop (eds), 191–228. Berlin: Mouton de Gruyter.

- Teich, Elke. 2003. *Cross-linguistic Variation in System and Text – A Methodology for the Investigation of Translations and Comparable Texts*. Berlin: Mouton de Gruyter (published version of Habilitationsschrift of 2001).
- Thomason, Sara G. & Kaufman, Terrence. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley CA: University of California Press.
- Traugott, Elizabeth. 1982. From propositional to textual and expressive meanings; some semantic-pragmatic aspects of grammaticalization. In *Perspectives on Historical Linguistics*, W. Lehmann & Y. Malkiel (eds), 245–271. Amsterdam: John Benjamins.
- Traugott, Elisabeth & Dasher, Richard. 2005. *Regularity in Semantic Change*. Cambridge: CUP.
- Weinreich, Uriel. 1953. *Languages in Contact*. The Hague: Mouton.



Appendix: Some low-level statistics on corpus data<sup>6</sup>

	Lexical Density in %					
	English		German		Reference Corpora	
	Original	Translation	Original	Translation	English	German
Essay	56.76	55.99	57.15	54.39		
Fiction	47.73	48.40	51.05	50.20		
Instruction	58.77	59.56	58.36	55.04		
PopSci	55.55	51.29	54.18	56.98		
Share	58.02	56.21	56.16	56.13		
Speech	54.02	51.19	52.79	52.80		
Tourism	59.14	59.24	60.74	59.08		
Web	58.21	56.88	58.03	56.93		
Av. TOTAL_REG	56.03	54.85	56.06	55.19	57.49	56.98
Av. TOTAL_LANG	55.44		55.62			

	Type Token Ratio in %					
	English		German		Reference Corpora	
	Original	Translation	Original	Translation	English	German
Essay	9.19	9.37	18.61	14.63		
Fiction	12.66	11.85	19.14	18.33		
Instruction	7.00	6.82	14.29	14.05		
PopSci	10.98	10.38	17.55	17.16		
Share	9.39	9.25	16.01	16.52		
Speech	9.83	8.56	15.87	14.73		
Tourism	12.58	13.33	24.17	19.93		
Web	10.99	11.45	21.07	18.38		
Av. TOTAL_REG	10.33	10.13	18.34	16.71	15.64	21.71
Av. TOTAL_LANG	10.23		17.53			
TOTAL	5.80	5.60	11.25	9.92		

6. The basis for the calculation of our dependent variables is as follows: *Lexical Density* is calculated as percentage of content words per sub-corpus, *Type Token Ratio* as the percentage of different lexical lemmas (types) per sub-corpus (based on lemmatized texts), and the percentages given in the *Parts-of-Speech* calculations are to be read as percentages of a given word class per sub-corpus. Importantly, for *TTR* it makes a huge difference whether we compare averages over sub-corpora, or else *TTRs* over the complete running words of a total corpus, which is why we have separately given both counts in our Table *Type Token Ratios* in the Appendix.

Parts of Speech in the English Sub-corpora in %														
		Noun	Pronoun	Verb	Adjective	Adverb	Adpos	Article	Conj	Particle	Number	Punct	Others	
English	TRANS	Essay	26.82	2.58	13.89	8.75	5.10	11.97	12.12	4.79	2.01	2.08	9.81	0.06
		Fiction	19.94	9.94	15.44	4.73	6.51	10.01	10.51	4.83	2.73	1.11	14.11	0.14
		Instruction	29.82	2.05	16.55	5.42	4.04	9.53	11.93	4.24	1.78	2.76	11.58	0.30
		PopSci	22.14	4.74	15.25	7.31	5.54	11.49	12.53	5.66	2.09	1.52	11.31	0.41
		Share	27.55	5.24	14.05	7.90	4.12	12.32	10.48	4.35	2.26	2.46	9.19	0.09
		Speech	23.22	5.37	15.84	8.40	5.53	11.19	11.85	5.17	2.74	1.23	9.44	0.03
		Tourism	31.75	2.19	10.60	7.61	3.65	11.94	11.71	4.48	1.42	2.46	12.02	0.18
		Web	27.87	4.22	13.08	8.71	4.03	11.92	11.12	4.85	1.77	2.00	10.25	0.16
		TOTAL	26.14	4.54	14.34	7.35	4.82	11.30	11.53	4.80	2.10	1.95	10.96	0.17
	ORI	Essay	27.81	3.61	15.36	8.29	3.66	10.95	10.53	5.55	2.92	1.57	9.71	0.04
		Fiction	19.11	10.06	16.66	5.30	5.81	9.65	10.15	5.73	2.40	0.99	13.97	0.15
		Instruction	29.67	3.61	15.89	5.20	3.18	9.15	11.35	4.78	2.75	2.01	12.18	0.22
		Popsci	25.57	3.14	16.17	7.76	4.98	11.31	11.53	5.53	2.20	1.48	10.13	0.19
		Share	29.14	5.34	13.83	6.97	3.15	11.01	9.02	5.49	2.18	2.95	10.73	0.19
		Speech	25.58	5.34	15.92	8.60	3.55	10.94	10.22	5.87	3.03	1.23	9.62	0.09
		Tourism	31.65	2.45	10.88	8.38	3.63	11.50	10.81	5.12	1.62	1.79	12.13	0.04
Web		29.11	4.34	14.47	6.74	3.46	10.48	9.57	5.45	2.35	2.15	11.81	0.08	
TOTAL		27.21	4.73	14.90	7.16	3.93	10.62	10.40	5.44	2.43	1.77	11.29	0.12	
TOTAL_EN		26.67	4.64	14.62	7.26	4.37	10.96	10.96	5.12	2.27	1.86	11.12	0.15	
	REF_EN	24.60	5.46	15.72	6.24	4.63	10.55	10.32	5.12	2.64	2.41	11.97	0.34	

Parts of Speech in the German Sub-corpora in %														
		Noun	Pronoun	Verb	Adjective	Adverb	Adpos	Article	Conj	Particle	Number	Punct.	Others	
German	TRANS	Essay	25.49	8.13	12.14	11.20	3.10	8.38	11.83	5.40	1.82	0.87	11.16	0.47
		Fiction	17.16	14.64	15.12	7.52	5.67	6.91	7.99	6.18	2.44	0.48	15.60	0.30
		Instruction	25.27	8.63	12.79	7.45	2.72	8.00	10.94	5.28	2.24	0.83	14.82	1.02
		PopSci	22.85	8.81	12.78	10.65	6.00	8.55	10.03	4.71	1.79	1.07	12.43	0.33
		Share	25.08	9.15	11.40	9.47	4.32	9.39	9.35	5.28	1.56	3.10	10.90	0.99
		Speech	24.80	9.63	12.81	9.63	3.78	8.17	11.14	5.62	1.99	0.66	11.41	0.36
		Tourism	29.13	6.78	8.92	10.28	3.40	8.95	10.35	5.01	0.95	1.28	13.19	1.77
		Web	26.26	8.75	12.29	9.07	2.90	8.35	8.54	5.45	1.92	1.46	13.60	1.40
		TOTAL	23.84	8.67	12.08	9.69	4.89	8.39	10.29	4.85	1.75	1.47	13.34	0.72
	ORI	Essay	24.35	7.05	11.69	10.88	5.12	8.38	11.23	4.84	1.77	1.60	12.49	0.61
Fiction		17.57	13.58	14.43	7.37	6.25	7.23	8.75	4.76	2.64	0.52	16.46	0.43	
Instruction		26.22	5.94	13.41	7.69	3.43	8.22	11.01	4.50	1.71	2.08	14.83	0.96	
PopSci		20.28	10.03	12.69	9.82	6.50	7.61	10.78	4.94	1.90	0.99	13.89	0.57	
Share		25.30	9.26	11.64	10.69	4.30	9.27	10.30	4.44	1.52	2.19	10.45	0.64	
Speech		21.86	10.81	13.42	9.88	5.51	7.83	10.12	5.13	1.98	0.75	12.23	0.49	
Tourism		28.24	4.86	8.40	10.73	4.17	9.95	10.08	5.01	1.06	2.19	14.03	1.29	
Web		26.90	7.83	10.96	10.46	3.88	8.60	10.07	5.22	1.45	1.44	12.37	0.82	
TOTAL		24.51	9.32	12.28	9.41	3.99	8.34	10.02	5.37	1.84	1.22	12.89	0.83	
		TOTAL_DE	24.17	8.99	12.18	9.55	4.44	8.36	10.16	5.11	1.80	1.34	13.12	0.78
	REF_DE	22.93	8.45	13.04	9.20	5.02	8.08	9.88	4.47	1.90	2.27	14.12	0.63	

Parts of Speech per Sub-corpus in %													
	Noun	Pronoun	Verb	Adjective	Adverb	Adpos	Article	Conj	Particle	Number	Punct	Others	
English	Reference	24.60	5.46	15.72	6.24	4.63	10.55	10.32	5.12	2.64	2.41	11.97	0.34
	Original	27.21	4.73	14.90	7.16	3.93	10.62	10.40	5.44	2.43	1.77	11.29	0.12
	Translation	26.14	4.54	14.34	7.35	4.82	11.30	11.53	4.80	2.10	1.95	10.96	0.17
German	Original	24.51	9.32	12.28	9.41	3.99	8.34	10.02	5.37	1.84	1.22	12.89	0.83
	Translation	23.84	8.67	12.08	9.69	4.89	8.39	10.29	4.85	1.75	1.47	13.34	0.72
	Reference	22.93	8.45	13.04	9.20	5.02	8.08	9.88	4.47	1.90	2.27	14.12	0.63



# Author index

## A

Abaev, Vladimir I. 119  
 Abbi, Anvita 109, 120, 121  
 Adam, Jean-Michel 245  
 Adams, Marianne 148  
 Adams, Marilyn 278, 279, 281, 282  
 Ädel, Annelie 299  
 Aijmer, Karin 310  
 Aikhenvald, Alexandra Y. 3, 8, 33, 68, 72, 76  
 Åkerlund, Walter 171  
 Akinçi, Mehmet Ali 240, 242, 250  
 Aksu-Koç, Ayhan A. 242  
 Alarcos Llorach, Emilio 192  
 Alexiadou, Artemis 148, 149, 150, 154, 156  
 Alo, M. A. 263  
 Anagnostopoulou, Elena 148, 149, 150, 154, 156  
 Antonsen, Elmer 167  
 Ard, Josh 5  
 Auger, Julie 157

## B

Backus, Ad 240  
 Badia Margarit, Antoni M. 192  
 Baker, Mona 320, 334  
 Bakker, Peter 21, 22, 23  
 Bamberg, Michael 242, 249, 250  
 Barbosa, Pilar 148, 149, 150, 152, 155, 156  
 Baroni, Marco 311  
 Baskaran, Loga 263  
 Baumgarten, Nicole 8, 293, 302, 322  
 Beal, Joan 265  
 Behnstedt, Peter 141, 147, 155, 156  
 Belletti, Adriana 140  
 Benzing, Johannes 119  
 Berg, Thomas 123  
 van Bergen, Linda 220

Bergsland, Knut 24, 26, 29, 30  
 Berman, Ruth A. 242, 246, 250  
 Bernardini, Silvia 311  
 Bernhardt, Stephen A. 297, 298  
 Besch, Elmar 309  
 den Besten, Hans 114  
 Bhatt, Rakesh M. 270  
 Biber, Douglas 297, 320, 321, 328, 329  
 Bickel, Balthasar 97, 334  
 Bisang, Walter 233  
 Blanc, Michel 240  
 Bliss, Alan J. 219, 221, 222, 223  
 Bolander, Maria 284  
 Bollée, Annegret 109  
 Bolten, Jürgen 302  
 Bonnesen, Matthias 135, 155  
 Bonnotte, Isabelle. 242, 244, 245, 251  
 Botha, Rudolf P. 114  
 Böttger, Claudia 293, 302, 306  
 Braunnüller, Kurt 165, 167, 173, 176  
 Brendemoen, Bernt 73  
 Breu, Walter 44  
 Brigaudiot, Mireille 245, 247, 248  
 Brines, Vivenc 277, 278  
 Brinzeu, Félicien 117  
 Bronckart, Jean-Paul 246, 251  
 Broselow, Ellen 189  
 Bruening, B. 119, 123, 124  
 Bührig, Kristin 302, 310  
 Bungarten, Theo 297  
 Burkhart, Dagmar 114  
 Butterworth, Guy 277, 280, 281, 282  
 Bybee, Joan 232, 334, 335, 337  
 Byrnes, Heidi 296  
 Bzdęga, Andrzej Z. 119

## C

Caffarel, Alice 322

Calder, George 112  
 Campbell, Lyle 52, 53, 54  
 Cancino, Herlinda 277, 280  
 Capidan, Theodor 117  
 Cardinaletti, Anna 149, 156  
 Carrol, Mary 320  
 Cassidy, Frederic G. 264  
 Cazden, Courtney B. 282, 283  
 Chafe, Wallace 297  
 Chambers, J. K. 265  
 Chikovani, Guram 48, 55  
 Chirita, Diana 174  
 Chomsky, Noam 153  
 Christoffersen, Marit 172  
 Claudi, Ulrike 71  
 Clements, J. Clancy 128  
 Clyne, Michael 74, 77, 296  
 Comeau, Liane 240  
 Comrie, Bernard 9, 10, 76  
 Corne, Chris 68  
 Coulmas, Florian 310  
 Coveney, Aidan 140, 147, 155  
 Cristofaro, Sonia 174, 175  
 Croft, William 5, 167, 233  
 Csató, Éva Á. 63, 73  
 Curnow, Timothy Jowan 8, 100  
 Cysouw, Michael 25

## D

Damrad-Frye, Robin 249  
 Dasher, Richard 320, 322, 334, 335, 336  
 DeCarrico, Jeanette 284  
 DeCesaris, Janet Ann 192  
 Denison, David 213, 220  
 Déprez, Viviane 156  
 Deuchar, Margaret 240, 258  
 Dewaele, Jean-Marc 147  
 Dezső, Pais 119, 120  
 Dixon, Robert M. W. 3, 8, 33, 68  
 Doherty, Monika 296, 320, 336  
 Dolz, Joaquim 242, 245, 246, 251, 254

- Downing, L. J. 125  
 Drescher, Martina 297  
 Dressler, Wolfgang U. 123  
 Drijkoningen, Franck 151  
 Druetta, Ruggero 157  
 Dryer, Matthew S. 34  
 Dupuis, Fernande 156
- E**  
 Eckman, Fred R. 188, 189  
 Ellis, Rod 277, 278  
 Elsig, Martin 7, 144, 155, 156, 157  
 Elšík, Viktor 19  
 Elsness, Johann 220  
 Escure, Geneviève 263
- F**  
 Faarlund, Jan Terje 167, 168  
 Fabricius, Anne H. 121  
 Fabricius-Hansen, Cathrine 296, 320  
 Fähnrich, Heinz 114  
 Fanego, Teresa 213, 228  
 Fayol, Michel 242, 244  
 Felix, Sascha 277  
 Fenton, James 227  
 Field, Frederic W. 15  
 Filppula, Markku 219, 220, 221, 223, 224, 262, 264, 265, 286  
 Fitzpatrick, David 219, 231  
 Flores Farfán, José Antonio 42  
 Fortescue, Michael 49, 50, 51  
 Foulet, Lucien 140, 141  
 Fox, Cynthia A. 156, 157  
 Franceschini, Rita 320  
 Frawley, William 320  
 Friedemann, Marc-Ariel 154, 156  
 Friedman, Victor A. 35, 48, 49
- G**  
 Gagnepain, Jean 215, 216, 217, 219  
 Gass, Susan 5  
 Gazdar, Kaevan 302  
 Genesee, Fred 240  
 Georg, Stefan 87, 91  
 Gil, David 108  
 Givón, Talmy 43  
 Glover, Jessie R. 86  
 Glover, Warren 86
- Godel, Robert 114  
 Goffman, Erving 299  
 Goldberg, Adele 215, 233  
 Golovko, Evgenij 24, 25, 26  
 Grannes, Alf 117, 119, 124, 125  
 Greene, David 217  
 Grunow-Hårsta, Karen 89
- H**  
 Haase, Martin 53, 72  
 Hagège, Claude 335  
 Haig, Geoffrey 103  
 Halliday, Michael A. K. 302, 303, 308, 320, 325, 335, 336  
 Hamers, Josiane 240  
 Hansen-Schirra, Silvia 317, 323, 325, 334  
 Hård, John Evert 173  
 Harris, Alice 52, 53, 54, 223  
 Harris, John 223  
 Harwood, Nigel 297, 298  
 Hasan, Ruqaiya 320, 325  
 Haspelmath, Martin 82, 111, 119  
 Hasselgard, Hilde 320  
 Hatch, Evelyn 277, 280, 281, 282  
 Haugen, Einar 101, 166  
 Hawkins, John 296, 311, 320  
 Heine, Bernd 3, 8, 9, 10, 21, 36, 37, 46, 54, 68, 70, 71, 72, 75, 174, 179, 232, 319  
 Henry, Alison 264  
 Herrick, Dylan 189, 208, 209  
 Heston, W. L. 121  
 Hetzer, Armin 114  
 Hickey, Raymond 221, 223  
 Hickmann, Maya 250  
 Hilbert, Michaela 5, 7, 154  
 Hildebrandt, Kristine 86, 89, 99  
 Holmer, Arthur 169  
 Honda, Isao 87  
 Hopper, Paul J. 320, 334, 335, 337  
 Horcajada Diezma, Bautista 113, 114  
 Hoshi, Michiyo 86  
 Höskuldur Þráinsson 168  
 House, Juliane 293, 296, 302, 306, 310, 313, 319, 320, 334, 336  
 de Houwer, Annick 240  
 Huang, Joseph 277, 280, 281  
 Hulk, Aafke C.J. 151, 156, 193, 240
- Hünнемeyer, Friederike 71  
 Hyland, Ken 297, 298
- I**  
 Idiazabal, Itziar 7, 244  
 Inkelas, Sharon 108
- J**  
 Jakobson, Roman 188, 192  
 Jankuhn, Herbert 167  
 Jiha, Michel 116, 119  
 Johanson, Lars 3, 7, 8, 10, 34, 40, 42, 49, 56, 62, 66, 73, 74, 75, 76, 77, 116, 337  
 Jón Helgason 172  
 Jones, Mari C. 55, 56  
 Jörgensen, Nils 177  
 Juvonen, Päivi 257
- K**  
 Kahr, Joan Casper 34  
 Kaifer, Anton 244  
 Kaiser, Georg A. 148, 156, 157  
 Kaufman, Terrence 3, 8, 34, 35, 36, 266, 267, 268, 284, 285  
 Kayne, Richard 140, 154, 156  
 Kehoe, Margaret 192  
 Keller, Wolfgang 220  
 Kelly, Patricia 224  
 Kern, Barbara 242, 250  
 Kiesler, Reinhard 174  
 Kiral, Filiz 39  
 Kirchhoff, Klaus Rainer 302  
 Kloss, Heinz 163, 166  
 Koch, Peter 165, 166  
 Koller, Werner 293, 294  
 Koptjevskaja-Tamm, Maria 108  
 Kortmann, Bernd 262, 264, 265  
 Kotthoff, Helga 296  
 Kouwenberg, Silvia 109  
 Krause, Wolfgang 167  
 Kretzenbacher, Heinz 297  
 Kuo, Chih-Hua 297, 298, 304  
 Kuteva, Tania 3, 8, 33, 36, 37, 68, 70, 71, 72, 174, 179  
 Kytö, Merja 213, 220
- L**  
 Lafitte, Pierre 119  
 Lanza, Elizabeth 240  
 Larsson, Carl 171, 172

Lasch, Agathe 173  
 Lehmann, Christian 322  
 Leisiö, Larisa 43  
 Leroy, Maurice 114  
 Leslau, Wolf. 35  
 Lewis, Geoffrey L. 116  
 Li, Charles N. 34  
 Lindblad, Gustaf 172  
 Lindström, Jan 111  
 Lleó, Conxita 5, 6, 107, 188, 189,  
 192, 193  
 Lönnroth, Lars 177  
 López del Castillo, Lluís 189,  
 207  
**M**  
 Mac Eoin, Gearoid 214  
 MacWhinney, Brian 337  
 Maddieson, Ian 191, 192  
 Mahboob, Ahmar 263  
 Mahootian, Shahrzad 119  
 Mair, Christian 265  
 Marchello-Nizia, Christiane 141  
 Martin, Jim R. 320  
 Masica, Colin 91  
 Matras, Yaron 19, 20, 21, 22, 24,  
 34, 41, 77  
 Matthiessen, Christian M. I. M.  
 302, 319, 320, 322, 335, 336  
 Maurant, Anna 299  
 Mayer, Mercer 242  
 Mayerthaler, Willi 127  
 Mazaudon, Martine 87, 88  
 McCafferty, Kevin 219, 221, 222,  
 223, 224  
 Meisel, Jürgen M. 7, 135, 154, 155,  
 156, 158, 240  
 Menz, Astrid 72, 75  
 Mesthrie, Rajend 263  
 Michailovsky, Boyd 90  
 Miller, Jim 265  
 Miller, Kerby 219  
 Miller, Roy Andrew 34  
 Minassian, Martiros 117, 118, 119  
 Mittendorf, Ingo 220  
 Möller, Georg 309  
 Montgomery, Michael 227  
 Moravcsik, Edith 101, 107  
 Mörtz, Karlheinz 114  
 Mukattash, Lewis 280, 281,  
 282, 284

Müller, Hans-Georg 116, 117, 118,  
 119, 123, 125  
 Müller, Natascha 193, 240  
 Myers-Scotton, Carol 72  
**N**  
 Nadasdi, Terry 157  
 Nagy, Naomi 157  
 Nattinger, James R. 284  
 Nettle, Daniel 34, 172  
 Neumann, Stella 317, 321, 323,  
 325, 334  
 Nichols, Johanna 99, 100  
 Noonan, Máire 151, 154  
 Noonan, Michael 5, 6, 10, 82,  
 85, 86, 87, 92, 93, 94, 95, 96, 97,  
 100, 102, 334  
 Norde, Muriel 164  
**O**  
 Oakes, Michael P. 323  
 Öberg, Jan 173  
 Ó Corráin, Ailbhe 215, 216, 217,  
 218, 221, 222  
 Oesterreicher, Wulf 165, 166,  
 319  
 O'Neill, Francis 227  
 Ó'Rahilly, Thomas F. 218  
 Ó Siadhail, Mícheál 214, 215,  
 219  
 van Oostendorp, Marc 192  
 Özçetin, Demet 8, 322  
**P**  
 Pagliuca, William 232  
 Panhuis, Dirk G. J. 174  
 Paradis, Johanne 188  
 Paul, Ludwig 117, 119  
 Payrató, Lluís 189, 207  
 Perkins, Revere 232  
 Peters, Ann 284  
 Pettersson, Gertrud 171  
 Pfaff, Carol W. 240  
 Picard, Marc 140, 154  
 Pietsch, Lukas 5, 9, 10, 33, 107,  
 117, 119, 224  
 Pinkster, Harm 174  
 Pla Fulquet, Joan 189, 208, 209  
 Plähn, Jürgen 115  
 Pohl, Jacques 141, 147, 155  
 von Polenz, Peter 297

Pollock, Jean-Yves 140, 156  
 Pons, Eva 186  
 Poplack, Shana 144, 155, 156, 157  
 Poppe, Erich 220  
 Poudel, Kedar Prasad 95, 96  
 Propp, Vladimir 245

**Q**  
 Quasthoff, Uta M. 242, 249  
 Quay, Suzanne 240

**R**  
 Rafel i Fontanals, Joaquim 192  
 Raible, Wolfgang 174  
 Rainer, Franz 110  
 Rakow, Martin 189  
 Ramisch, Heinrich 112  
 Ramm, Wiebke 320  
 Rand, David 143  
 Ratcliffe, Robert R. 51  
 Ravem, Roar 277, 280, 281, 282  
 Recasens, Daniel 189, 208, 209  
 Reh, Mechthild 54  
 Riehl, Claudia Maria 40  
 Rinke, Esther 135, 155, 158  
 Rizzi, Luigi 136, 142, 148, 151, 153  
 Roberge, Paul 114  
 Roberge, Yves 156  
 Roberts, Ian 136, 138, 140, 148,  
 150, 151, 152, 154  
 Rohlf, Gerhard 113  
 Roland, Françoise 250  
 Romaine, Suzanne 240  
 Ronan, Patricia 216  
 Rosat, Marie-Claude 242  
 Rösler, Irmtraut 173  
 Ross, Malcolm D. 3, 8, 9, 34,  
 35, 45, 46, 47, 48, 77, 100, 101,  
 102, 103  
 Rubino, Carl 109, 110, 111, 112,  
 121  
 Rudanko, Juhani 220

**S**  
 Sakel, Jeanette 41  
 Sánchez-Prieto Borja, Pedro  
 113, 114  
 Sankoff, David 143, 157  
 Sankoff, Gillian 157  
 Sansò, Andrea 109, 110, 121  
 Schiffrin, Deborah 299



- Schneuwly, Bernard 246  
Sebastián, Eugenia 250  
Sebba, Mark 264  
Seibicke, Wilfried 309  
Sekerina, Irina A. 24, 25, 26, 28, 29, 30  
Selinker, Larry 269  
Sgroi, Salvatore C. 113  
Shapira, Rina G. 278, 280, 281  
Siemund, Peter 33, 135, 266, 287  
de Silva Jayasuriya, S. 128  
Simon-Vandenberg, Annemarie 320  
Skoda, Francoise 121  
Slobin, Dan I. 242, 246, 250  
Söll, Ludwig 141, 155  
Sportiche, Dominique 154  
Sridhar, Kamal K. 285  
Sridhar, Shikaripur N. 285  
Starke, Michal 149, 156  
Steiner, Erich 8, 10, 319, 320, 322, 325, 334, 335, 336  
Stephany, Ursula 250  
Stimm, Helmut 72  
Stolz, Thomas 5, 10, 107, 108, 109, 110, 113, 121, 127  
Stowell, Tim 155  
von Stutterheim, Christiane 320  
Sussex, Ronald. 38, 44  
Szmrecsanyi, Benedikt 262
- T  
Tannen, Deborah 297
- Teich, Elke 317, 319, 321, 322  
Ternes, Elmar 52, 53  
Terry, Robert M. 141, 155  
Thibault, Pierrette 157  
Thomas, Peter W. 112  
Thomason, Sarah Grey 3, 4, 5, 8, 34, 35, 36, 38, 39, 100, 101, 127, 172, 266, 267, 268, 284, 285, 319, 320, 321  
Thompson, Sandra A. 320  
Þórhallur Eypórsón 167  
Thun, Nils 111, 119  
Tietze, Andreas 117, 120  
Tiphine, Ursula 277, 278, 281  
Tophinke, Doris 173  
Tosco, Mauro 37  
Trask, Robert L. 72  
Traugott, Elizabeth 320, 322, 334, 335, 336  
Triggs, Jefferey 220  
Tuson, Ampar 189, 207
- U  
Uí Bheirn, Ú. 218
- V  
Valois, Daniel 156  
Vance, Barbara 136, 148, 150  
Vassileva, Irina 297, 298, 300  
Vela, Mihaela 317, 323,  
Vennemann, Theo 164  
Veny, Joan 189, 208  
Vezzosi, Letizia 220
- Viaplana, Joaquim 192  
Viereck, Karin 112  
Viereck, Wolfgang 112  
Vila, Francesc X. 186  
Visser, Frederic T. 225
- W  
Wagner, Heinrich 220  
Wagner, Max L. 113  
Wälchli, Bernhard 108  
de Weck, Geneviève 242  
Wehr, Barbara 53  
Weinert, Regina 284  
Weinreich, Uriel 3, 8, 36, 65, 73, 115, 215, 233, 319  
Weinrich, Harald 309  
Wenning, E. Alf 172  
Wessén, Elias 166, 169, 170, 179  
Wierzbicka, Anna 110, 310  
Wiese, Richard 111  
de Wind, Maarten 140, 151, 156  
Winford, Donald 3, 8, 34, 267, 269, 285  
Wolfram, Walt 263  
Wollin, Lars 177
- Y  
Yallop, Colin 319
- Z  
Zeevaert, Ludger 8, 34, 170  
Zoll, Sheryl 108  
Zubizarreta, Maria Luisa 139

# Language index

## A

Abkhaz 117, 118, 119, 123–125  
 Aboriginal English 264  
 African-American Vernacular English (AAVE) 263, 264, 267  
 Afrikaans 114, 115  
 Akkadian 34  
 Albanian 35, 111, 113, 114, 117, 118, 119  
 Aleut 24–27, 29–31, *see also* Bering Island Aleut, Copper Island Aleut  
 Algonquian 21–23  
 American English 115, 267, 270  
 Amharic 35  
 Appalachian English 264  
 Arabic 6, 9, 41, 48, 51, 55, 64, 113, 114, 117, 118, 119, 277, 280, *see also* Bukhara Arabic, Egyptian Arabic, Lebanese Arabic, Neo-Arabic varieties, Old Arabic, Qashqa-Darya dialect of Arabic, Siculo-Arabic  
 Armenian 111, 114, 118, 119, 123, *see also* Modern Eastern Armenian, Old Armenian  
 Aromunian 117  
 Australian English 267  
 Austronesian 34, 120  
 Azerbaijani 6, 39  
 Azeri 111, 117, 118, 119

## B

Bashkir 117, 118  
 Basque 6, 7, 70, 71, 72, 79, 110, 111, 117, 119, 120, 239–245, 247–257  
 Bavarian 112  
 Belfast English 264  
 Belize English 263  
 Bengali 120

Bering Island Aleut 26, 27  
 Black South African English 264  
 Bombay Hindi 35  
 Breton 6, 51–54  
 British Creole 263, 264  
 British English 219, 226, 230, 267, 270, 271, 273, 299  
 Bukhara Arabic 51  
 Bulgarian 114, 117, 118, 119, 124, 126, 300  
 Byzantine Greek 113

## C

Cameroon English 264, 267  
 Catalan 5, 6, 185–196, 198–201, 204–210, *see also* Central Catalan  
 Celtic 51–52, 54, 112, 214, 215, 220, 286  
 Cèmuhi 68  
 Central Catalan 186, 194, 196  
 Central Himalayish 83, 95  
 Chantyal 81, 82–85, 86, 87–98, 99–103  
 Chinese 6, 34, 277  
 Classical French 135, 136, 142, 151, 157  
 Classical Latin 174  
 Classical Nahuatl 42  
 Classical Tibetan 86  
 Colloquial American English 264  
 Colloquial French 6, 7, 53, 54, 136, 142, 155–157, 159  
 Copper Island Aleut 6, 9, 15, 16, 24–31  
 Cree 21–24, 31  
 Croatian 44, 73, *see also* Molisean  
 Cymbrian 112, 113  
 Czech 70

## D

Dai 120  
 Danish 6, 50, 64, 164, 165  
 Dari 48, 55  
 Domari 6, 9, 41, 43  
 Dravidian 34, 120, 121  
 Drubéa 68  
 Dutch 6, 41, 71, 72, 74, 114

## E

Early Modern Icelandic 172  
 Early Modern Irish 217  
 Early Old Swedish 171, 178  
 Early Romani 19, 20  
 East African English 264  
 Egyptian Arabic 51  
 English 5, 6, 7, 8, 9, 16, 17, 22, 23, 31, 39, 44, 49, 50–51, 55, 57, 63, 64, 71, 72, 74, 97, 100, 111, 112, 113, 115, 116, 119, 174, 213–215, 219–224, 225, 228, 231–234, 250, 261–265, 266–269, 277–280, 283, 284–286, 293, 295–297, 299–308, 310–313, 317, 318, 321, 323–326, 328–334, *see also* Aboriginal English, African-American Vernacular English, American English, Appalachian English, Australian English, Belfast English, Belize English, Black South African English, British English, Cameroon English, Colloquial American English, East African English, Hiberno-English, Indian English, Indian South African English, Irish English, Malaysian English, Middle English, Modern English, New Zealand English, Newfoundland English, Nigerian English, Nigerian

- Pidgin English, Old English,  
Pakistani English, Scottish  
English, Singapore English,  
South African English,  
Standard English, Tyneside  
English, Ulster English, Welsh  
English
- Eskimo 49–51, 57
- F
- Faroese 164
- Finland Swedish 43
- Finnish 6, 9, 43, 110, 111, 299
- Flemish 40, 41
- French 6, 7, 21–24, 31, 32, 40, 52,  
54, 55, 67, 68, 71, 72, 110, 111,  
135–137, 140, 141, 142, 144, 147,  
148, 150–159, 186, 241, 242,  
244, 245, 250, 253, 277, 300,  
*see also* Classical French,  
Colloquial French, Middle  
French, Old French, Québec  
French, Renaissance French,  
Standard French
- G
- Gagauz 72, 75
- Gascon 53, 71
- Georgian 6, 111, 114, 117, 118, 119,  
123, *see also* Old Georgian
- German 6, 8, 9, 11, 18, 19, 31, 40,  
41, 64, 72, 73, 111, 112, 113, 119,  
127, 135, 148, 171, 173, 179, 192,  
250, 293–297, 299, 300–313,  
317, 318, 321, 323–326, 328–334,  
*see also* High German,  
Low German, Middle Low  
German, Modern High  
German, Old High German
- Ghale 81, 83
- Greek 6, 16, 17, 19–21, 31, 73,  
110, 111, 113, 114, 118, 119, 250,  
*see also* Byzantine Greek,  
Modern Greek
- Guernésiais 55, 56
- Gullah 264
- Gurung 83, 86, 87, 89, 90, 94, 95
- H
- Hiberno-English 213–215, 217,  
219, 221, 223, 224, 225, 228, 230,  
233, 234
- High German 173
- Hindi 6, 107, 120, 128, *see also*  
Bombay Hindi
- Hispanized Xalitla Nahuatl 42
- Hungarian 34, 109, 111, 117, 118,  
119, 120
- I
- Iberian Spanish 114, 187
- Icelandic 164, 165, 172, *see also*  
Early Modern Icelandic
- Indian English 5, 7, 261–263,  
264, 267, 269, 270, 271, 273,  
276, 284, 285, 286
- Indian Ocean Creoles 21
- Indian South African English  
264
- Indic languages 90
- Indo-Aryan 19, 34, 41, 120, 121
- Indo-European 34, 35, 48, 55, 81,  
82, 97, 107, 109, 110, 277
- Inuktitut 6, 9, 50
- Irish English 100, 213, 219, 220,  
223, 224, 225, 226, 227, 230, 235,  
262, 264, 267, 268, 286, *see also*  
Hiberno-English
- Irish 5, 6, 9, 52, 213, 214, 215–225,  
227, 231, 233, 234, 262, 264, 286,  
*see also* Early Modern Irish,  
Old Irish
- Italian 6, 40, 44, 110, 111, 113, 186
- Italo-Romance 113
- J
- Jamaican Creole 263, 264, 267
- Japanese 34
- K
- Kadiwéu 6, 9, 38, 39
- Kalmyk 117, 118, 119
- Karachay 117, 118, 119
- Karaim 63, 64, 72
- Kazakh 117, 118
- Kham-Magar 81
- Khoekhoe 114
- Korlai Portuguese 128
- Kurdish 6, 117, 118, 119, 121
- Kusunda 81, 82
- Kyyrölä Russian 43, 44
- L
- Lango 100
- Late Old Swedish 169, 171, 176,  
177, 178
- Latin 6, 8, 16–19, 25, 31, 64, 71,  
163, 164–166, 172–174, 177–180,  
*see also* Classical Latin,  
Medieval Latin
- Lebanese Arabic 116, 118
- Lezgian 111, 117, 118, 119
- Low German 6, 164, 166,  
172–174, 179
- M
- Ma'a 35
- Macedonian 35, 49
- Magar 81, 83, 89, 90
- Malay 6, 114
- Malaysian English 263, 264, 267
- Manange 82, 83, 86, 87, 89,  
94–99
- Marathi 35, 128
- Medieval Latin 165, 173
- Medieval Spanish 113, 114
- Michif 6, 9, 15, 16, 21–24, 25, 31
- Middle English 220, 232
- Middle French 142, 158
- Middle Low German 64, 163,  
165, 166, 172, 173, 179
- Modern Eastern Armenian 118
- Modern English 174, 220, 233
- Modern Greek 110, 113, 117
- Modern High German 174
- Modern Persian 121
- Modern Swedish 163, 166, 168,  
169, 171, 174, 177
- Molisean 6, 44
- Mon-Khmer 120
- Montana Salish 6, 39
- N
- Nahuatl 6, 42, *see also* Classical  
Nahuatl, Hispanized Xalitla  
Nahuatl
- Nar-Phu 82, 83, 85–87, 89–99,  
102
- Neo-Arabic varieties 114

Nepali 6, 81–98, 99–103, *see also*  
Standard Nepali

Newari 83

Newfoundland English 264

New Zealand English 267

Nigerian English 263, 267

Nigerian Pidgin English 263,  
267

Norwegian 164, 165, 280

## O

Occitan 186

Old Arabic 117

Old Armenian 114

Old East Nordic 172

Old English 23, 174

Old French 136, 140, 148–150,  
154, 155, 158

Old Georgian 114

Old High German 174

Old Iranian 121

Old Irish 216, 219, 232

Old Norse 163, 167, 169

Old Swedish 8, 163, 164–166,  
169–180

Old Turkic 118

Old West Nordic 172

Ossetic 117, 118, 119, 121

## P

Pakistani English 263

Persian 39, 64, 117, 118, 119, *see*  
*also* Modern Persian

Pipil 71

Polish 6, 9, 17–19, 21, 44, *see*  
*also* Polish dialect of Silesia

Polish dialect of Silesia 73

Polska Roma 21

Portuguese 6, 9, 38, 39, 72, 76,  
114, 128, 278, *see also* Korlai  
Portuguese

Punjabi 120, 278

## Q

Qashqa-Darya dialect of Arabic  
48, 55

Québec French 156, 157

## R

Renaissance French 151, 158

Romance 40, 53, 54, 70, 72, 110,  
113, 135, 149, 150, 151, 155, 158

Romani 6, 9, 15, 16, 19–21, 31, 34,  
*see also* Early Romani

Rumanian 21, 49, 114, 117, 118, 119

Russian 6, 9, 24–31, 43, 44, 64,  
73, 115, 116, 117, 118, 119, 300, *see*  
*also* Kyyrölä Russian

## S

Salish 39, *see also* Montana  
Salish

Sardinian 111, 113, 186

Scottish English 264, 265, 286

Scottish Gaelic 52, 216

Seke 82, 83, 87, 94, 95

Serbian 21, 73, 114

Serbian Gurbet 21

Siculo-Arabic 113

Singapore English 5, 7, 154,  
261, 262, 264, 267, 269, 270,  
273–276, 283, 284–286

Sino-Tibetan 92, 97

Slavic 34, 38, 39, 44, 49, 64, 72,  
110

Slovenian 70

Sorbian 70

South African English 264, 267

Spanish 5, 6, 7, 42, 71, 72, 113, 114,  
185, 186, 187, 189, 190–194, 195,  
201, 204, 205–210, 239, 240–  
245, 247, 248, 249, 250–256,  
277, 278, 282, *see also* Iberian  
Spanish, Medieval Spanish

Sranan 72

Standard English 219, 220, 221,  
225, 230, 231, 232–233, 261, 263,  
270, 275

Standard French 6, 7, 136, 140,  
142, 155, 157, 158

Standard Nepali 91, 102

Standard Turkish 35, 48, 49

Sumerian 34, 121

Sursilvan 72

Swedish 6, 63, 71, 72, 111, 163,  
164–166, 173, 177, 179, 180,  
*see also* Early Old Swedish,

Finland Swedish, Late Old  
Swedish, Modern Swedish,  
Old Swedish

## T

Taiwanese 280

Tajik 48, 51, 55

Takia 6, 34, 35, 45–48

Tamang 83, 87, 90, 94, 95, 96, 102

Tamangic languages 6, 81, 82,  
86–98, 100–103

Tamil 120

Tayo 67

Telugu 120

Thakali 83, 84, 85, 87, 89, 90,  
91, 94

Tibetan 34, 81, 82, 83, 85, 86, 87,  
94, 95, 96, *see also* Classical  
Tibetan

Tibeto-Burman 81, 82, 85, 97,  
102, 120

Turkish 6, 35, 39, 48, 49, 73,  
107, 111, 114, 116–127, 242, *see*  
*also* Standard Turkish, West  
Rumelian Turkish

Turkmen 48, 55

Tyneside English 265

## U

Ulster English 227

Ulster Scots 227

Uzbek 48, 51, 55

## W

Waskia 6, 34, 35, 45–48

Welsh 52, 112, 216, 264

Welsh English 264, 286

West Greenlandic 49, 50

West Rumelian Turkish 35,  
48, 49

Wutun 34

## Y

Yakut 117

Yiddish 115, 116

## Z

Zhangzhung 83



# Subject index

## A

academic discourse 297–298,  
300, 304  
acculturation 166  
adstrate languages 37, 116  
AF order 169  
allomorphy 17, 19, 188, 192, 193  
allophony 91, 188, 192, 193  
anaphoric pronouns 74  
anchoring 242, 247, 250, 251,  
254, 256, 337  
angloversals 261, 265  
A-position 152  
A'-position 150, 151  
areal linguistics 108  
aspect 66, 93, 99, 102, 213, 215,  
216, 220, 243, 244, 251, 252,  
301  
Ausbau 163, 166, 174, 177, 179,  
180

## B

back vowels 190  
borrowability 15, 122, 127  
borrowing 5, 7, 9, 10, 15, 16,  
19, 21, 31, 34, 36, 37, 57, 62,  
65, 66, 81, 86, 87, 89, 92, 94,  
95, 98, 100, 101, 102, 115, 122,  
127, 222, 319–322, 338, *see*  
*also* structural borrowing,  
syntactic borrowing  
borrowing scale 100, 127, 319,  
321  
business communication 8, 293,  
295, 296, 300, 304, 307, 313  
business texts 304, 305, 313

## C

calque 113, 223  
calquing 36, *see also*  
grammatical calquing  
cardinal vowels 190

carry-over insertion 63, 64  
casual style 142  
clause linkage types 174  
clitic left dislocation 137  
code 61–77, 87, 201, 240, 320  
code copying 61–66  
code switching 62, 87  
combinational copying 64, 66,  
74, 76  
communicative action 246, 247  
communicative conventions  
295–297, 300, 307  
communicative event 297, 299  
communicative norms 313  
communicative preferences  
294, 295, 296, 302  
communicative styles 293, 296,  
299  
complex inversion 137, 138, 140,  
142–147, 151–155, 157, 158  
complexity 4, 6, 27, 62, 99,  
122–126, 188, 189, 192, 193,  
207, 208, 210, 247  
congruent realizations 308  
constituent interrogatives 264,  
270, 273–276, 280–282  
Construction Grammar 215, 233  
contact-induced change 3, 4, 7,  
9, 44, 48, 68, 81, 82, 98, 100,  
172, 215, 266, 267  
contact-induced innovation 35,  
179, 180  
contact linguistics 7, 108, 262,  
266  
contact variety 262, 268, 269,  
285, 320, 322  
contextual markedness 50, *see*  
*also* markedness  
conventionalization 61, 66, 76  
copying 7, 8, 49, 57, 61–71, 73,  
76, 77, 123, 127, *see also* code  
copying, combinational

copying, frequential copying,  
global copying, insertional  
copying, selective copying,  
target of copying  
creoles 21, 37, 67, 72, 128, 263,  
264, 267, 269  
cultural elite 177

## D

decatégorialization 33, 45, 46  
desemanticization 45  
developmental sequence 282  
dialectal variation 224  
dialect speakers 43, 44  
diphthongs 190  
directionality of change 318,  
334, 335  
discourse 8, 9, 23, 33, 38, 41,  
49, 96, 97, 102, 140, 141, 155,  
250, 251, 253, 255, 256, 279,  
295, 297–301, 304–310, 313,  
318–322, 325, 334–338  
discourse functions 298, 301,  
304–306, 313  
dominant language 40, 42  
dominant tense 242, 253, 254  
*do*-periphrasis 271–275, 278,  
279, 281, 282  
dynamic agreement 153

## E

echo-word formation 107, 109,  
115, 118, 121, 126, 128, 129  
embedded clause 261, 270, 276,  
283, 286  
embedded interrogatives  
261–264, 270, 275–277, 282,  
283, 286  
empirical methodologies 317  
engagement 298  
ergative syntax 96, 102

*est-ce que* 137, 139, 141, 143, 144,  
147, 148, 151–153, 155–158  
explicitation 317, 318, 323, 324,  
325, 326, 329, 334–337  
explicitness 296, 317, 318, 320,  
321, 323, 325, 327, 328, 329, 330,  
331, 332, 334–338  
external factors 163, 185, 188,  
193, 194, 207, 208, 210

## F

FA order 169  
feature-based perspective 317,  
318, 320, 323, 337  
fixed chunk 272, 275, 276  
focus construction 33, 52, 53  
formal equivalence 178, 234  
formally equivalent translation  
180  
formal style 140, 141, 156, 157  
formulaic language 261, 284  
fossilisation 269, 273, 285  
frame-changing developments  
76  
free inversion 137, 140, 148, 149,  
152, 154, 155, 158  
free variation 26, 91  
frequency 7, 8, 9, 38, 43, 44, 46,  
54, 55, 63, 64, 74, 75, 92, 107,  
108, 111–114, 120, 128, 129, 144,  
145, 147, 177, 188, 191, 202, 214,  
224, 225, 227, 229, 233, 249,  
252, 271, 273, 279, 281, 284,  
294, 297, 299, 301, 302, 303,  
306, 313, 317, 318, 334–338  
frequency of use 9, 46, 54, 55,  
74, 334, 336, 338  
fractional copying 8, 61, 62,  
64, 66, 74, 75, 337  
front vowels 90  
full vowels 192

## G

genre 97, 111, 113, 177, 225, 242,  
246, 251, 253, 256, 293, 297,  
298, 300, 304, 307, 313, 335  
genre variation 242  
global copying 61, 64  
grammatical calquing 36, 66, 77

grammaticalization 8, 37, 42,  
43, 45, 46, 48, 51, 53, 54, 55,  
57, 61, 68–71, 75, 76, 140,  
154, 155, 156, 158, 174, 178,  
322, 334, *see also* target of  
grammaticalization  
grammatical metaphor 308,  
309, 334  
grammatical replication 36, 37,  
43, 47, 54, 57, 66  
group language shift 268, *see*  
*also* group shift  
group shift 268, 285

## H

high register 87, 88  
high vowels 187  
historical syntax 135  
horizontal variation 208  
hypotactic structure 178, 180

## I

ICE corpora 262, 270  
identity 63, 67, 109, 298, 303  
immersion programmes 7, 239,  
240, 241, 242, 246, 256  
imperfective 159, 242–245, 251,  
253, 256  
indirect morphosyntactic  
diffusion 66  
inflectional class 17–19  
inflectional morphology 9, 15,  
16, 18, 19, 21, 22, 25, 26, 31,  
101, 164  
insertional copying 63, 65  
interaction 3, 33, 37, 62, 64, 70,  
176, 188, 193, 210, 247, 285,  
293, 297, 298, 320, 335, 336  
internal factors 163, 172, 188,  
189, 193, 207, 208, 210  
internal variation 277  
interrogative constructions 261,  
263  
interrogative inversion 7, 261,  
262, 269, 281  
interrogative pronouns 72,  
176, 270  
interrogatives 72, 136, 148,  
150, 154, 155, 157, 158, 262,  
273, 274, 277, 278, *see also*

constituent interrogatives,  
embedded interrogatives,  
interrogative constructions,  
interrogative inversion,  
interrogative pronouns,  
interrogative system, main  
clause interrogatives, polar  
interrogatives

interrogative system 7, 135, 136,  
140, 141, 148–158  
intersubjective positioning 298  
intonation questions 137, 139,  
141–147, 155, 156  
inversion 137, 141, 147, 148,  
153, 155, 156, 157, 158, 159,  
261–265, 270–283, 286, *see*  
*also* complex inversion,  
free inversion, interrogative  
inversion, pronominal  
inversion, simple inversion,  
stylistic inversion, subject-  
verb inversion

## L

L2 acquisition 7, 268, 269, 277,  
280, 284, 285  
L2 variety 263, 265, 267–269,  
277, 285  
language acquisition 4, 6, 7, 154,  
240, 261, 262, 266, 268, 269,  
277, 283, 284, 285, 286, 287  
language change 4–10, 68, 135,  
158, 159, 163, 164, 176, 179, 185,  
266, 268, 294, 317, 318, 319,  
322, 334, 335  
language contact 4–10, 15, 33,  
34, 36–40, 42, 43, 46, 49, 51,  
54, 57, 64, 68, 73, 74, 75, 81, 98,  
103, 107, 108, 109, 112, 113, 114,  
116, 118, 119, 125, 127, 128, 129,  
158, 163, 171, 172, 174, 176, 177,  
180, 187, 210, 214, 239, 256,  
266, 268, 287, 293, 295, 301,  
317–322, 334–338  
language shift 4, 16, 100, 173,  
263, 266, 267, 268, 269, 284  
language variation 294  
left dislocation 137, 138  
lingua franca 165  
linguistic variation 294, 295, 301

literary language 165, 167  
 loan semantics 66  
 loanshift 66  
 loan syntax 66  
 loan translation 66  
 low register 87–89

## M

main clause 76, 164, 167–169,  
 174, 179, 180, 261, 263, 264,  
 270, 271–273, 275–277, 282,  
 283  
 main clause interrogatives 261,  
 263, 264, 270, 276, 282, 283  
 markedness 4, 6, 188, 189,  
 191, 193, 207, 208, 210, 301,  
 336, *see also* contextual  
 markedness  
 marked vowels 6, 192, 194  
 metaphor 232, *see also*  
 grammatical metaphor  
 metaphorical realization 307  
 metatypy 45, 77, 102  
 mid vowels 185, 189–191, 195,  
 203, 205, 207, 209  
 mixed language 6, 15, 16, 19, 21,  
 24, 31, 267  
 modal verb 231, 278, 281, 305  
 model language 7, 8, 36, 38, 40,  
 43–47, 49, 51, 55, 56, 57  
 morphological typology 21  
 morphophonological variation  
 190  
 morphosyntax 6, 99, 100  
 multifunctional perspective 318  
 multilinguality 317–321, 336, 337

## N

[N]-feature 153  
 narrative phases 248, 251  
 narrative skills 242  
 narrative structure 242, 245,  
 247, 248  
 narrowing 38–40, 55, 318  
 New Englishes 262, 263, 286  
 nondialect speaker 43  
 non-null subject language 149  
 non-standard variety 117, 228,  
 262, 263  
 noun cohesion 334

null subject language 149–151,  
 155, 158  
 null subjects 149

## O

overgeneralization 282

## P

perfective 242–245, 251, 252,  
 253, 256  
 personal pronouns 22, 25, 30,  
 31, 295, 297–300, 302, 304,  
 310, 311  
 phonological change 135, 194,  
 210  
 pidgins 128, 263, 267, 269, 319  
 polar interrogatives 263, 264,  
 270–282  
 pragmatic unmarking 38, 42,  
 179  
 process type 302, 303, 306, 313  
 productivity 23, 42, 214, 233, 337  
 pronominal inversion 137, 138,  
 140, 142–147, 151, 153–155,  
 157, 158  
 pronominal markers 25  
 pronouns, 8, 25, 29, 30, 66, 73,  
 95, 140, 144, 152, 157, 158, 171,  
 293, 295, 297, 298, 300, 301,  
 302, 304, 313, 328, 331, 333, *see also* anaphoric pronouns,  
 interrogative pronouns,  
 personal pronouns, relative  
 pronouns

## Q

questions 117, 157, 198, 208, 264,  
 270, 277, 280, 281, 283, *see also* intonation questions,  
 wh-questions, yes/no-  
 questions

## R

reduplication 5, 10, 73, 107–116,  
 118, 121, *see also* partial  
 reduplication  
 referential density 97, 99  
 register 62, 87, 88, 89, 98, 136,  
 208, 225, 229, 293, 295, 297,  
 298, 301, 313, 318, 320, 323,  
 324, 328–338

relative pronouns 94, 170, 179,  
 325, 326  
 replica language 8, 36, 38, 40,  
 45, 46, 57, 68  
 replication 5, 7, 8, 10, 36, 37,  
 39, 51, 57, 62, 179, 213, 233,  
 234, *see also* grammatical  
 replication, syntactic  
 replication  
 restructuring 37, 63, 77, 267

## S

segmental variation 119, 120  
 selection 34, 40, 64, 311  
 selective copying 64, 65, 76  
 shift variety 267, 284  
 simple inversion 137, 140, 148,  
 150–153, 155, 158  
 single vowels 190  
 sociolinguistic variables 193  
 speaker-deictic element 297,  
 300, 301  
 speech style 97–99, 198  
 spoken variety 147, 177  
 sprachbund 37  
 stance 298  
 standard variety 113, 155, 159  
 stressed vowels 190, 192  
 structural borrowing 36, 66, 319  
 style 97, 98, 99, 296, 300,  
*see also* casual style,  
 communicative styles, formal  
 style, speech style, written  
 style  
 stylistic inversion 137, 138, 140,  
 144, 147, 148, 150, 154–156, 159  
 stylistic variation 297  
 subject-verb inversion 135, 136,  
 147, 157, 158, 159, 278  
 subjunctions 176, 178  
 subordinate clause 96, 163, 164,  
 167–180  
 subordination 76, 96, 102, 103,  
 174, 175, 179, 180  
 subordination marking 175, 179  
 substance 7, 36, 57, 63, 64, 65,  
 66, 70  
 substrate language 21, 37, 121,  
 262, 286  
 substratum interference 266



- successive bilingualism 240, 257  
 superstrate language 37, 262  
 syllable structure 122, 123, 124, 125  
 synchronic variation 269, 284  
 syntactic borrowing 101, 102  
 syntactic change 52  
 syntactic integration 175, 176  
 syntactic productivity 213  
 syntactic replication 213  
 syntax 5, 21, 23, 35, 36, 41, 45, 47, 51, 52, 65, 85, 91, 93, 135, 163, 164, 173, 177, 216, 240, 261, 328, *see also* ergative syntax, historical syntax, loan syntax, morphosyntax  
 Systemic Functional Linguistics 320, 322
- T**  
 take-over insertion 63  
 target of copying 64, 67, 68, 69, 72–74  
 target of grammaticalization 68, 69  
 target vowels 199, 202, 209, 210, 212  
 tense 27–30, 66, 92, 93, 99, 142, 213, 215, 216, 220, 230, 239, 242, 243, 244, 247, 250, 251–256, 278, 279, 301, 326, 336  
 text type 8, 163, 164, 177, 180, 298, 300, 320  
 thematic structure 245  
*-ti/-tu* 137, 138, 140, 141, 153, 154, 155, 157, 158  
 topicalization 9, 42, 51  
 translation 18, 39, 66, 77, 110, 114, 172, 177, 178, 208, 219, 245, 293–295, 300–313, 318, 320–327, 329, 331–336, 338, 340, 342, 345  
 translationese 311  
 translation relation 301, 325, 329  
 triglossic 165
- U**  
 unstressed vowels 190
- V**  
 V2-hypothesis 149  
 V2-language 148, 149  
 V2-rule 148  
 variability in production 201  
 variation 20, 56, 74, 148, 150, 170, 173, 206, 262, 266, 269, 294, 295, 313, 330, *see also* dialectal variation, free variation, genre-variation, horizontal variation, internal variation, language variation, linguistic variation, morphophonological variation, segmental variation, stylistic variation, synchronic variation, vertical variation  
 variety 26, 28, 44, 63, 66, 112, 140, 149, 165, 174, 177, 186, 194–196, 205, 207, 210, 263–267, 273, 284, 322, *see also* contact variety, L2 variety, non-standard variety, shift variety, spoken variety, standard variety, written variety  
 verbal routine 296, 310, 311, 312  
 verb cohesion 243  
 verb density 242, 247, 248, 249, 256  
 verb-final order 55, 173  
 verb-late 164, 170–175, 177, 179  
 verb-late patterns 170, 171  
 verb-late word order 164, 170–175, 177, 179  
 verb movement 135, 150, 151, 154, 155, 157, 158  
 verb types 271, 272, 273, 274, 283, 284, 298, 301  
 vernacular universals 261, 265, 284  
 vertical variation 208
- vowels 19, 20, 28, 89, 91, 121, 124, 125, 185, 187–196, 199–210, *see also* back vowels, cardinal vowels, diphthongs, front vowels, full vowels, high vowels, marked vowels, mid vowels, single vowels, stressed vowels, target vowels, unstressed vowels
- W**  
 wedge construction 171  
 wh-criterion 135, 136, 148, 152, 153, 158  
 wh-fronting 137, 139, 141, 144, 147, 153, 155, 156  
 wh-in situ 137, 139, 141, 147, 155, 156  
 wh-operator 149, 151–153, 156–158  
 wh-questions 135, 138, 141, 144, 147, 148, 157, 158, 159, 270, 283  
 word order 9, 21, 34–36, 38–44, 46, 47, 49–52, 54, 55, 56, 57, 74, 75, 91, 99, 101, 135, 136, 141, 154, 157, 160, 163, 164, 167–177, 179, 180, 181, 182, 264, 269, 272, 276, 277, 283, 335, *see also* verb-late word order  
 word order change 33, 35, 36, 38, 40, 42, 46, 47, 49, 56, 75, 136  
 written language 39, 165, 173, 177, 217, 308  
 written style 311  
 written variety 155, 177, 180
- Y**  
 yes/no-questions 136, 138, 140, 142–145, 147, 148, 157

In the series *Hamburg Studies on Multilingualism* the following titles have been published thus far or are scheduled for publication:

- 7 **SIEMUND, Peter and Noemi KINTANA (eds.):** Language Contact and Contact Languages. 2008. x, 358 pp.
- 6 **THIJE, Jan D. ten and Ludger ZEEVAERT (eds.):** Receptive Multilingualism. Linguistic analyses, language policies and didactic concepts. 2007. x, 328 pp.
- 5 **REHBEIN, Jochen, Christiane HOHENSTEIN and Lukas PIETSCH (eds.):** Connectivity in Grammar and Discourse. 2007. viii, 465 pp.
- 4 **LLEÓ, Conxita (ed.):** Interfaces in Multilingualism. Acquisition and representation. 2006. xiv, 284 pp.
- 3 **HOUSE, Juliane and Jochen REHBEIN (eds.):** Multilingual Communication. 2004. viii, 359 pp.
- 2 **BRAUNMÜLLER, Kurt and Gisella FERRARESI (eds.):** Aspects of Multilingualism in European Language History. 2003. viii, 291 pp.
- 1 **MÜLLER, Natascha (ed.):** (In)vulnerable Domains in Multilingualism. 2003. xiv, 374 pp.