

CLIP 1



**Korpuslinguistik und interdisziplinäre
Perspektiven auf Sprache**
**Corpus Linguistics and
Interdisciplinary Perspectives on Language**

Bd./Vol. 1

Herausgeber / Editorial Board:
Holger Keibel, Marc Kupietz, Christian Mair

Gutachter / Advisory Board:
Heike Behrens, Mark Davies, Martin Hilpert,
Reinhard Köhler, Ramesh Krishnamurthy, Ralph Ludwig,
Michaela Mahlberg, Tony McEnery, Anton Näf,
Michael Stubbs, Elke Teich, Heike Zinsmeister

Marek Konopka/Jacqueline Kubczak
Christian Mair/František Štícha
Ulrich H. Waßner (Hgg.)

Grammatik und Korpora 2009

Dritte Internationale Konferenz

Grammar & Corpora 2009

Third International Conference

Mannheim, 22.-24.09.2009

narr |
VERLAG

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet über <<http://dnb.d-nb.de>> abrufbar.

© 2011 Narr Francke Attempto Verlag GmbH + Co. KG
Dischingerweg 5 · D-72070 Tübingen

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt.
Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne
Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für
Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und
Verarbeitung in elektronischen Systemen.
Gedruckt auf chlorfrei gebleichtem und säurefreiem Werkdruckpapier.

Internet: www.narr.de
E-Mail: info@narr.de

Layout: Tröster, Mannheim
Printed in Germany

ISSN 2191-9577
ISBN 978-3-8233-6648-5

Inhalt / Contents

Vorwort / Preface	9
-------------------------	---

I. Plenarvorträge / Keynote Speeches

Bruno Strecker: Korpusgrammatik zwischen reiner Statistik und „intelligenter“ Grammatikografie	23
Douglas Biber / Bethany Gray: Is conversation more grammatically complex than academic writing?	47
Mirjam Fried: Grammatical analysis and corpus evidence	63
Françoise Gadet: What can be learned about the grammar of French from corpora of French spoken outside France	87

II. Korpusgestützte Grammatikforschung / Corpus-based Grammar Research

Christa Dürscheid / Stephan Elspaß / Arne Ziegler: Grammatische Variabilität im Gebrauchsstandard: das Projekt „Variantengrammatik des Standarddeutschen“	123
Reinhard Fiehler: Korpusbasierte Analyse von Univerbierungsprozessen	141
Hagen Hirschmann: Eine für Korpora relevante Subklassifikation adverbialer Wortarten	157
Franziska Münzberg: Korpusrecherche in der Dudenredaktion: Ein Werkstattbericht	181
Per Bærentzen: Einige neue Regularitäten im Gebrauch der Pronominalformen <i>deren</i> und <i>derer</i>	199
Geert Stuyckens: Zum Wesen der Subjektlücken in Verbzweitkoordination auf der Grundlage eines deutsch > niederländischen Übersetzungskorpus	213

Elma Kerz: The role of low-level schemas in English academic writing. A usage-based constructionist approach.....	229
Markéta Malá: Copular clauses in English and in Czech – a comparative corpus-based approach	253
Svetlana Gorokhova: The role of frequency effects in the selection of inflected word forms: A corpus study of Russian speech errors.....	267
Francesca Strik Lievers: Constructing Judgments. The Interaction between Adjectives and Clausal Complements in Italian	287
Lisa Brunetti / Stefan Bott / Joan Costa / Enric Vallduví: A multilingual annotated corpus for the study of Information Structure.....	305

III. Methodologie korpuslinguistischer Grammatikforschung / Methodologies of corpus-linguistic Grammar Research

Holger Keibel / Cyril Belica / Marc Kupietz / Rainer Perkuhn: Approaching grammar: Detecting, conceptualizing and generalizing paradigmatic variation	329
Oliver Mason: Reconciling Phraseology and Grammar	357
Milena Hebal-Jeziarska / Neil Bermel: Frequency and oppositions in corpus-based research into morphological variation	373
Stella Neumann: Contrasting frequency variation of grammatical features.....	389
Thomas Herbst / Susen Faulhaber: Optionen der Valenzbeschreibung. Ein Valenzmodell für das Englische.....	411
Amir Zeldes: On the Productivity and Variability of the Slots in German Comparative Correlative Constructions.....	429
Cyril Belica / Marc Kupietz / Andreas Witt / Harald Lungen: The Morphosyntactic Annotation of DeReKo: Interpretation, Opportunities, and Pitfalls	451

IV. Einblicke in die aktuelle Forschung / Insights into current studies

František Štícha: Der kommunikative und der systembezogene Status grammatischer Phänomene mit niedriger Häufigkeit	473
Said Sahel: Monoflexion als Erklärung für Variation in der Nominalphrasenflexion des Deutschen	485
Georg Albert: Innovative Sprachverwendungen: Verbreitung und Kontext.....	495
Eva Breindl / Maik Walter: Kausalverknüpfungen im Deutschen. Eine korpusbasierte Studie zum Zusammenspiel von Konnektorbedeutung, Kontexteigenschaften und Diskursrelationen.....	503
Manfred Stede / Uwe Küssner: Kausale Konnektoren in der Automatischen Textanalyse.....	513
Julia Richling: Diachrone Analyse eines Newsgroup / Webforum-Korpus	521
Tomas By: The Prolog version of the Tiger Dependency Bank.....	531
Silke Scheible / Richard Jason Whitt / Martin Durrell / Paul Bennett: Investigating diachronic grammatical variation in Early Modern German. Evidence from the <i>GerManC</i> corpus.....	539
Christopher Cox: Quantitative perspectives on syntactic variation: Investigating verbal complementation in a corpus of Mennonite Plautdietsch	549
Silvia Hansen-Schirra: Empirical profiling of LSP grammar	557
Olga O. Boriskina: Noun Cryptotype Analysis as an Approach to Corpus-driven Modelling of N+V Collocations	567
Siaw-Fong Chung / Yu-Wen Tseng: Learning Prepositions: A Corpus-based Study in Taiwan EFL Contexts.....	575
Svetlana Savchuk: The <i>Russian National Corpus</i> as a Tool for Research on Grammatical Variability.....	585
Ruska Ivanovska-Naskova: Italian-Macedonian parallel corpus	599

CHRISTOPHER COX

Quantitative perspectives on syntactic variation: Investigating verbal complementation in a corpus of Mennonite Plautdietsch¹

Abstract

This study presents a quantitative analysis of syntactic variation in the ordering of verbal constituents in a series of bipartite verbal complementation constructions in Canadian Mennonite Plautdietsch. The alternation in verb-final contexts between v_1-v_2 and v_2-v_1 complement orders attested in corpus data is rendered statistically through the application of generalized linear mixed effects modelling. Such quantitative, corpus-based methods, this study concludes, might serve not only in evaluating the simultaneous effects of multiple hypothesized predictors of variation, as is undertaken with the Plautdietsch corpus data, but also in permitting a greater number of linguistic data and potentially competing predictors to bear upon the task of analysis than would otherwise be possible.

1. Introduction

A common problem in linguistic description lies in analyzing grammatical variation – in seeking to derive perspicuous and yet empirically adequate accounts of variable phenomena in language. While the threads of variation might be perceived to run throughout the fabric of grammar, the theoretical and practical issues such variability poses are felt with perhaps particular acuteness in contemporary syntactic research, where analyses commonly adopt introspection and categorical judgements as their primary bases of argumentation. As Bresnan et al. (2007) contend, however, introspective methods applied without systematic consideration of empirically-attested variation in both attestation and judgement may risk unintentionally underestimating the range and complexity of licit variation, and thus ultimately provide only partial

¹ Thanks are due to R. Harald Baayen, Grzegorz Kondrak, John Newman, Sally Rice, and members of the Department of Linguistics at the University of Alberta for their helpful comments on this research. The support of the Trudeau Foundation, the Killam Trusts, and the Social Sciences and Humanities Research Council of Canada (SSHRC) is gratefully acknowledged here.

accounts of the phenomena of interest. On this view, an active, analytical engagement with empirically-attested variation is essential to empirically-adequate grammatical description.

Corpora may play a natural role in this analytical process, inasmuch as they provide contextualized examples of grammatical variation which corpus-based techniques might then assist in understanding. The present paper pursues one such quantitative, corpus-based account of grammatical variation, concentrating upon verb placement alternations in West Germanic verbal complementation constructions (VCC). The remainder of this paper is organized as follows: after introductory consideration of the typological status of verbal complementation in the Continental West Germanic languages, specific attention is given to variation in VCC attested in Mennonite Plautdietsch (ISO 639-3: pdt). A quantitative analysis is then conducted upon instances of naturally-occurring variation in such constructions drawn from a corpus of Canadian Mennonite Plautdietsch. Both the results and the methodology used to obtain them are briefly discussed in the conclusion, with special reference to the larger problems which such variable constructions pose for syntactic analysis generally.

2. Verbal complementation

Verbal complementation (VC) refers here to the introduction of verbal elements v_{i+1} as a syntactic argument of another verb v_i . From a typological perspective, VC represents a common strategy for the analytical construction of complex predicates, both globally (cf. Dixon 2006) and in the Continental West Germanic languages specifically (cf. Wurmbrand 2004, 2005; Zwart 2005). In the latter group of languages, verbal complementation commonly finds use in the expression of aspectual, epistemic, and deontic information relating to the predicate under construction, as is the case with lexical verbs introduced as the complements of auxiliary (in perfective constructions, expressing aspectual distinctions) and modal (in epistemic and deontic constructions) verbs.

Despite cognate patterns of morphological marking in and the considerable frequency of use of verbal complementation constructions across the Continental West Germanic languages, acceptable complement orders differ notably between individual constructions and speech communities. As examples (1a-c) demonstrate, even within relatively simple, bipartite modal-infinitive constructions in verb-final contexts, striking variation exists in the reported grammaticality of different verbal complement orders:

Modal-infinitive construction, verb-final context

- (1a) ...*weil er das Buch *muss₁ kaufen₂ / kaufen₂ muss₁* [*1-2 / ✓ 2-1]
 ‘because he had₁ to buy₂ the book.’ (Standard German: Wurmbrand 2004: 74)
- (b) ...*da Valère dienen boek wilt₁ kuopen₂ / *kuopen₂ wilt₁* [✓ 1-2 / *2-1]
 ‘that Valère wants₁ to buy₂ this book.’ (West Flemish: Wurmbrand 2005: 331)
- (c) ...*dat Jan het boek kan₁ lezen₂ / lezen₂ kan₁* [✓ 1-2 / ✓ 2-1]
 ‘that Jan can₁ read₂ the book.’ (Standard Dutch: Wurmbrand 2005: 324)

Such variation in the relative acceptability of even these common constructions is relatively unexpected, given the close genetic and geographic relationship between most of these languages, and raises questions as to the ultimate motivations of this variability, which neither consistently identifies the same logically-possible linear orders as being grammatical or ungrammatical cross-linguistically (1a-b), nor presents a categorical delimitation between such alternants in all cases (1c).

3. Analyzing verbal complementation in Mennonite Plautdietsch

Such non-categorical variation in verbal complement orders is also attested in verb-final contexts in Mennonite Plautdietsch (MP), one such Continental West Germanic language (cf. Epp 1993, Kaufmann 2005). The corpus sentences presented here as (2a) and (2b) mirror the variability noted in (1c), with both v_1 - v_2 and v_2 - v_1 orders being attested:

Modal-infinitive construction, verb-final context

- (2a) *Mi heat sikj daut soo aus wann doa waut opp'em Hoff **mucht₁ senne₂**.*
 ‘It sounds to me as if there might₁ be₂ something on the yard.’
 (RE1972.S0000881)
- (b) *Aus mien Hollända daut een Stoot toojehorcht haud, fruag he mi meteenst waut fer,ne Sproak soont **senne₂ mucht₁**.*
 ‘Once my Dutchman had listened to that for a while, he suddenly asked me what (kind of a) language that might₁ be₂.’ (RE1972.S0000929)

Formal statistical modelling of this ordering alternation is adopted in this study to assess the relevance of a number of factors proposed in the literature on VCC to bear upon the acceptability of particular alternants in Continental West Germanic languages. A sample of bipartite verbal complementation constructions occurring in verb-final contexts, such as (2), was drawn from a small, orthographically normalized and POS-tagged corpus of Mennonite Plautdietsch (approx. 124 000 tokens), representing the recent works of two prominent Canadian Mennonite authors, Reuben Epp (RE) and Jacob M. Fehr (JMF).

Queries of this corpus via POS tags were able to retrieve 4 711 unique instances of VC. Subsequent inspection and manual annotation of each such construction revealed 1 382 instances of VC involving a bipartite complementation structure in a verb-final context. These data were then coded for four distinct classes of predictors:

- 1) *Textual and authorial predictors*: author identity (i.e., RE or JMF), text identity, text genre (i.e., verse or prose), text translation status (i.e., composed in Plautdietsch or translated into Plautdietsch from another language)
- 2) *Morphosyntactic predictors*: morphological regularity of participant verbs, presence or absence of separable verbal prefixes, tense and subject agreement features of the finite verb, clause passivity, clause type (e.g., causal, temporal, etc.), complementizer identity, constructional context (e.g., modal-infinitive, auxiliary-participle), presence or absence of coordination involving the complement verb
- 3) *Lexical and semantic predictors*: verb lemma, verb class (i.e., lexical, modal, auxiliary), verb-pair mutual information scores
- 4) *Processing-related predictors*: verb frequency (of tokens and lemmata), Kullback-Leibler divergence from the ordering patterns of other verbs appearing in the same construction (serving as a measure of the degree to which a given order is unexpected in a particular constructional context), syntactic weight (via clause length in orthographic words), structural parallelism (via sentence distance from the immediately preceding instance of VC in the same text)

The alternation between v_1-v_2 and v_2-v_1 order was rendered statistically through generalized linear mixed-effects modelling (cf. Baayen 2008), treating the observed verb order as a binary dependent variable and all of the above factors as potential predictors. As well, adjustments for idiosyncratic differences be-

tween individual verbs (e.g., in their frequency and distribution across constructional contexts) were made automatically within the model by including both v_1 and v_2 as random effects.

4. Results and discussion

Summarized results of this statistical modelling are given in Table 1, presenting those fixed effects from the list given above which emerged as statistically significant predictors of verbal order.

Fixed Effect	Estimate	Std. Error	z-Value	p	
(Intercept)	5.12994	0.83442	6.148	7.85e-10	***
Author=RE	0.70619	1.58228	0.446	0.655372	
Syntagm=Modal-Inf	-0.10564	0.41579	-0.254	0.799446	
TextGenre=Verse	-1.31984	0.39968	-3.302	0.000959	***
V1ProportionInV1V2Order	-8.71853	1.05551	-8.260	< 2e-16	***
poly(ClauseLength, 2)1	-16.63070	3.60994	-4.607	4.09e-06	***
poly(ClauseLength, 2)2	10.83093	2.89583	3.740	0.000184	***
Coordination=CoordinatedV2	-0.98559	0.45916	-2.146	0.031835	*
V1ProportionalUseByRE	21.99279	6.56676	3.349	0.000811	***
V2Prefix=Separable	-1.58729	0.35311	-4.495	6.95e-06	***
V1Tense=Present	0.72030	0.26833	2.679	0.007376	**
StructuralParallel=NoPrecedent	-0.58674	0.26330	-2.228	0.025855	*
Author=RE : Syntagm=Modal-Inf	2.63046	0.66376	3.963	7.40e-05	***
Author=RE : Clause=Temporal	4.04043	1.71708	2.353	0.018619	*
Author=RE : TextGenre=Verse	-3.41530	0.72884	-4.686	2.79e-06	***

Table 1: Summarized fixed effects structure of the generalized linear mixed-effects model of bipartite VCC in verb-final contexts ($n = 1382$). Negative estimates favour v_1 - v_2 order

Several observations might be made on the basis of this model. Significant differences appear to exist between the two authors represented in the corpus sample as regards their ordering preferences in particular syntagms (v_2 - v_1 order is favoured by RE in modal-infinitive constructions), clause types (RE prefers v_2 - v_1 order in temporal clauses), and text genres (verse commonly features v_1 - v_2 order with both authors, although to a greater extent with RE).

These strong author-specific effects notwithstanding, it is still possible to observe more general effects of coordination (sentences with coordinated v_2 favour v_1 - v_2 order), syntactic weight (longer clauses overwhelmingly favour v_1 - v_2 order), complement verb prefixation (v_1 - v_2 order is more likely to be observed if v_2 features a separable prefix), and, perhaps most unexpectedly, finite verb tense (present tense favours v_1 - v_2 order, albeit weakly). Taken together, these results suggest that the represented speakers of MP are sensitive to specific constructional, morphological, lexical, and processing factors in their use of this alternation – a hypothesis which might be tested by extending the present model to further speakers, speech communities, and proposed predictors.

Indeed, while some caution must necessarily be exercised in interpreting results based upon such a limited linguistic sample, the quantitative, corpus-based approach taken in the analysis of these constructions might itself be advanced as one viable means of assessing the empirical adequacy of hypothesized motivations for VC ordering variation, insofar as these proposed predictors are capable of being consistently formalized and applied to the coding of corpus data. In the case of the present MP corpus data, such methods were effective not only in discerning even comparatively subtle effects (such as that of finite verb tense) upon verbal constituent ordering, suggesting directions for further research, but were able to do so given a relatively limited amount of data and a comparatively wide range of possible predictors. The extension of quantitative, corpus-based analysis to further speakers and speech communities might be hoped to render possible the simultaneous consideration of more extensive collections of linguistic data and more detailed sets of predictors than could receive comparable treatment by hands and eyes alone – and thus ultimately contribute to the perspicuity and empirical adequacy of the resulting analyses.

References

- Baayen, R. Harald (2008): *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bresnan, Joan / Cueni, Anna / Nikitina, Tatiana / Baayen, R. Harald (2007): Predicting the dative alternation. In: Boume, Gerlof / Krämer, Irene / Zwarts, Joost (eds.): *Cognitive foundations of interpretation*. Amsterdam: Royal Netherlands Academy of Science, 69-94.
- Dixon, Robert M.W. (2006): Complement clause types and complementation strategies in typological perspective. In: Dixon, Robert M.W. / Aikhenvald, Aleksandra Y. (eds.): *Complementation*. Oxford: Oxford University Press, 1-48.
- Epp, Reuben (1993): *The story of Low German & Plautdietsch. Tracing a language across the globe*. Hillsboro, Kansas: Reader's Press.
- Kaufmann, Göz (2005): Der eigensinnige Informant. Ärgernis bei der Datenerhebung oder Chance zum analytischen Mehrwert. In: Lenz, Friedrich / Schierholz, Stefan J. (eds.): *Corpuslinguistik in Lexik und Grammatik*. (= *Stauffenburg Linguistik* 37). Tübingen: Stauffenberg, 61-95.
- Wurmbrand, Susi (2004): West Germanic verb clusters: The empirical domain. In: É. Kiss, Katalin / van Riemsdijk, Henk (eds.): *Verb clusters. A Study of Hungarian, German, and Dutch*. Amsterdam: Benjamins, 43-85.
- Wurmbrand, Susi (2005): Verb clusters, verb raising, and restructuring. In: Everaert, Martin / van Riemsdijk, Henk (eds.): *The Blackwell Companion to Syntax*. Malden, MA: Blackwell, 227-341.
- Zwart, Jan-Wouter (2005): Continental West Germanic languages. In: Cinque, Guglielmo / Kayne, Richard S. (eds.): *The Oxford handbook of comparative syntax*. New York / Oxford: Oxford University Press, 903-946.