



PROJECT MUSE®

Corpus-based approaches to sentence structures (review)

Matti Rissanen

Language, Volume 85, Number 1, March 2009, pp. 237-240 (Review)

Published by Linguistic Society of America

DOI: <https://doi.org/10.1353/lan.0.0099>



➔ *For additional information about this article*

<https://muse.jhu.edu/article/262996>

- ROSS, JOHN R. 1970. On declarative sentences. *Readings in English transformational grammar*, ed. by Roderick A. Jacobs and Peter S. Rosenbaum, 222–72. Waltham, MA: Ginn.
- SAFIR, KEN. 2003. Anaphora, movement and coconstrual. *Grammar in focus, vol. 2: Festschrift for Christer Platzack*, ed. by Lars-Olaf Delsing, Cecilia Falk, Gunlog Josefsson, and Halldor Sigurdsson. Lund: Department of Scandinavian Languages, Lund University.
- SAFIR, KEN. 2004. *The syntax of (in)dependence*. Cambridge, MA: MIT Press.
- SAFIR, KEN. 2005. Person, context and perspective. *Rivista di Linguistica—Italian Journal of Linguistics* 16.107–54.
- THRÁINSSON, HÖSKULDUR. 1991. Long-distance reflexives and the typology of NPs. In Koster & Reuland, 49–76.

Utrecht University
 Utrecht Institute of Linguistics OTS
 Janskerkhof 13
 3512 BL Utrecht
 The Netherlands
 [eric.reuland@let.uu.nl]

Corpus-based approaches to sentence structures. Ed. by TOSHIHIRO TAKAGAKI, SUSUMU ZAIMA, YOICHIRO TSURUGA, FRANCISCO MORENO-FERNÁNDEZ, and YUJI KAWAGUCHI. (Usage-based linguistic informatics 2.) Amsterdam: John Benjamins, 2005. Pp. vi, 315. ISBN 9789027233141. \$143 (Hb).

Reviewed by MATTI RISSANEN, *University of Helsinki*

This volume has been produced by the Usage-Based Linguistic Informatics (UBLI) project at Tokyo University of Foreign Studies. The project is a part of the 21st Century Center of Excellence Program, funded by Japan's Ministry of Education, Sports, Culture, Science and Technology. The UBLI program aims to produce 'an overall integration of Theoretical and Applied Linguistics', which 'will be realized on the basis of Computer Sciences' (3). The research concentrates on linguistic usage, which is illustrated with large amounts of textual data. Thus, it is only natural that most of the articles of the volume are based on evidence derived from corpora or other text sources.

The research carried out by UBLI and reported on in the present volume is in many respects related to the area of evidence-based linguistic study that sees language mainly as a means of communication within a speech community, first and foremost between the speaker/writer (SP/W) and audience/reader (AD/R), with due attention paid to discourse and pragmatics (see e.g. Traugott & Dasher 2002). In recent decades this approach, which pays special attention to linguistic variation, has rapidly gained ground, particularly since the introduction of computerized corpora. The study of variation takes into account both extralinguistic factors—mainly sociolinguistic, genre-based, or regional—and language-internal trends of variation, such as metaphoric and grammaticalization. Theoretical issues are, of course, duly included in the analysis and in the conclusions drawn. Many of the articles in the present volume offer interesting starting points for more elaborate and extensive studies of variation. We can also hope that in the compilation of new corpora the extralinguistic factors causing variation will be kept in mind.

The most remarkable feature of the volume, which consists of three introductory chapters and fourteen articles, is the variety of the languages discussed, both non-Indo-European and Indo-European. They include Japanese, Korean, Chinese, Malay, Tagalog, Nuuchahnulth, Turkish, Arabic, Russian, German, French, English, and Spanish. Nuuchahnulth deserves a special mention: it is a Wakashan language spoken by only a handful of people in British Columbia. In a way, it is a pity that no Finno-Ugric languages are included in any of the studies. Finnish, for instance, which contains both original non-Indo-European and contact-based Indo-European syntactic features, might have offered an interesting point of comparison for some of the constructions discussed.

Most of the articles in the collection deal with questions related to verbal syntax. TAKAYUKI MIYAKE discusses constructions formed with the causative verb *shi* in Mandarin Chinese, comparing them with other types of causative construction and calling particular attention to the subject noun phrase of the construction. The study is based on corpora of modern Beijing Chinese representing colloquial, literary, and newspaper language.

Valence and voice marking are discussed in four articles. TSUNEKAZU MORIGUCHI concentrates on the interaction of valence and topicalization, with special attention to subject, focus, and the passive voice. His discussion of the various types of topicalization in Japanese, Philippine-Formosan aboriginal languages, and various Western European languages is most insightful; the typological differences among the languages become obvious. Finnish, which does not tolerate the agentive passive but readily uses word order for topicalization, might provide one more worthwhile source for comparison. ROBERT R. RATCLIFFE discusses valence from the point of view of the derived verb system in Arabic. He argues that the system has a productive core, which expresses valence and the number and role of verbal arguments. He points out that corresponding systems of morphological valence marking can be found in most languages. HIDEHIKO NAKAZAWA analyzes the ways of expressing the passive voice in Russian, paying special attention to the use of animate subjects with reflexive verbs formed with *-sja* to indicate perfective or imperfective passives. He examines the acceptability of a number of constructions on the basis of a questionnaire, which allows him to quantify the results. TOSHIHIRO TAKAGAKI compares two Spanish passive constructions: the periphrastic passive formed with the auxiliary *ser* and the reflexive passive formed with *se*. His research, based on the KLM Corpus (newspapers, novels, and conversations), shows that the *ser* passive is less frequent than the *se* passive, and that in some contexts the two are interchangeable. Aspectual factors play a role in the choice of the type of passive.

KAZUYUKI URATA's paper deals with the forms of the predicate verb in English adverbial clauses introduced by *lest*. Using American and British English corpora (Freiburg-Brown for American English and Freiburg-LOB and the British National Corpus for British English, supplemented by material from the *Time Almanac* and *The Times*), he is able to show that in American English the present subjunctive is clearly favored in *lest* clauses, while in British English there is more variation between the subjunctive, the indicative, and *should* + infinitive constructions. Of all the articles in this volume, this survey uses evidence derived from corpora most extensively.

The verb also plays an important role in the two studies that concentrate on word-order analysis. YOICHIRO TSURUGA examines the alternative positions of subject and object NPs with the French verb *planter*: the three acceptable types N_0 -V- N_1 -de- N_2 (*Luc plante le jardin de roses* 'Luc plants roses in the garden'), N_0 -V- N_1 -avec- N_2 (*Luc plante le jardin avec des roses*), and N_0 -V- N_2 -Loc- N_1 (*Luc plante des roses dans le jardin*), and the ungrammatical N_2 -V- N_1 (**Des roses plantent le jardin* 'Roses plant the garden'). He argues that there are only thirteen verbs in French that follow this pattern of acceptability. The article makes use of corpus evidence and the author presents relevant comments on the usefulness and importance of corpus-based study in the concluding chapter of the article. KIYOKO SOHMIYA compares the word-order patterns in English and Japanese. She points out that English favors transitive constructions and their variants, while Japanese 'shows a generous display of passive constructions in contexts where English would not allow it' (233). English focuses on the time axis: 'the unmarked, canonical SVO word order and its derived versions express cause and effect in a visual iconic way', while in Japanese word order is relatively free, and 'the whole sentence describes a state of affairs in an event rather than cause-effect dynamics' (249). Observations of this kind would seem most relevant to the description of basic typological differences between languages. It seems that the indication of possessive relations in a locally-colored way, as, for instance, in Russian *u menja jest'* or Finnish *minulla on* (literally I.ADESS + is 'on me is'), compared with the (Germanic) type *I have*, reflects the same kind of basic typological difference in perceptual meaning; see also the Latin *mihi est* and *habeo* (e.g. Heine 1997).

The morphology of the verb, mainly with reference to Korean and Japanese, plays an important role in HIDEKI NOMA's article, in which morphology is linked with other aspects of structural

analysis, specifically morphosyntax and supra-morphosyntax, which takes into account text and discourse. The author emphasizes the importance of defining and analyzing words ‘in their living state’, within a language hierarchy model, which extends from ‘linguistic field’ through the levels of text/discourse, utterance, sentence, clause, and word combination to the level of individual words.

Word-class definition and the ways of linking linguistic items are discussed in several articles. ISAMU SHOHO and HIROSHI UZAWA concentrate on Malay constructions in which a clause is connected with an adjective resembling an adverb of manner, either by an explicit complementizer or by a zero link. The authors ask the question of whether there is ‘any justification for setting up the independent word class of adverbs distinctive from adjectives’ (129). They come to the conclusion that Malay adjectives should be divided into three groups: those that function exclusively as adverbs, those that function both as adjectives and as adverbs with a varying degree of adverbness, and those that function exclusively as adjectives. YUJI KAWAGUCHI analyzes two Turkish clause-linkage suffixes, *-DIK-* and *-mE-*, making use of the two-million-word multi-genre METU Turkish Corpus. Following Givón’s categorization (2001:40–41), the author divides the complement-taking verbs into three semantic classes—manipulation verbs, modality verbs, and perception-utterance-cognition verbs—and shows how the semantics of the main clause verb (VERB2) is relevant to the choice of linking suffix: *-mE-* prevails with manipulation and modality verbs and *-DIK-* is more common with perception-utterance-cognition verbs.

Nuuchahnulth possessive constructions are analyzed by TOSHIHIDE NAKAYAMA. In this language, there are two possible ways of indicating possession in certain types of construction: the possessive suffix can be attached either to the nominal argument or to the predicate. The author examines the nature of this alternation and the factors that affect the choice of one or the other strategy. The alternative used depends on whether the possessor or the possessed is the direct argument of the predicate. The choice of argument structure interacts ‘with discourse salience that is shaped by factors including referentiality, agentivity, involvement/affectedness, topicality, and definiteness’ (31). The study is based on some thirty texts in the Ahousaht and Tseshaht dialects.

HIDETO ITO’s article is the only one with a clearly historical focus. The author compares the Late Medieval Korean translation of *Mengshan’s sayings* with the original Early Baihua text, and discusses the ways in which the grammatical markers, mainly verbal affixes, were rendered in the translation. Mistranslations imply that the translators did not quite understand the meaning of the aspect markers typical of Baihua.

One of the articles, by FRANCISCO MORENO-FERNÁNDEZ, discusses in detail the methodology of corpus compilation. The subject of the article is a corpus connected with the ‘Project for the Sociolinguistic Study of Spanish from Spain and America’. The corpus will consist of samples of spoken texts by speakers representing different genders, age groups, and levels of education in a number of Spanish-speaking countries (at the time of the writing of the article Argentina, Columbia, Guatemala, Mexico, Puerto Rico, Spain, and Venezuela).

More attention could perhaps have been paid to the proofreading and layout of the volume. To mention a few details, the position of notes varies from article to article; they are placed either at the bottom of each page or as endnotes. The number of misprints is also fairly high; to refer to Otto Jespersen as ‘JESPERSEN, D’, and to Ferdinand de Saussure as ‘SAUSSURE, F.’ is slightly embarrassing.

Minor formal shortcomings do not, however, diminish the most favorable impression given by the volume. It describes and analyzes a large number of syntactic and morphological features in a large number of languages, with due reference to the semantics of the constructions and pertinent observations on the structural and typological similarities and differences between various languages. Theoretical considerations are successfully combined with textual analysis. The volume stands as evidence of the high level of research carried out within the UBLI project. It is to be hoped that the studies in this volume, and the work of UBLI in general, will further enhance the compilation of corpora even in languages for which none exist so far.

REFERENCES

- GIVÓN, TALMY. 2001. *Syntax: An introduction*, vol. 2. Amsterdam: John Benjamins.
- HEINE, BERND. 1997. *Possession: Cognitive sources, forces, and grammaticalization*. Cambridge: Cambridge University Press.
- TRAUGOTT, ELIZABETH CLOSS, and RICHARD B. DASHER. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.

Department of English
P.O. Box 24 (Unioninkatu 40)
00014 University of Helsinki
Finland
[matti.rissanen@helsinki.fi]