

IP 01, University of Leipzig
“Differential agreement vs. differential case”

Team:

Main researchers:

Balthasar Bickel (Principal Investigator)

Eva van Lier (postdoc)

Robert Schikowski (PhD student)

Cooperative researchers:

Alena Witzlack-Marakevich (postdoc)

Sabine Stoll (senior researcher)

Tyko Dirksmeyer (PhD student)

Taras Zakharko (PhD student)

a) Aims and objectives

The goal of this project is to elucidate the nature of referential scale effects on verb agreement in comparison to case marking. Both marking systems can be affected by referential scales. The most obvious parallel is between differential (i.e. optional) object agreement and differential case marking on objects, which both seem to follow similar principles: higher ranking arguments favor overt agreement and overt case, lower-ranking arguments disfavor overt agreement and overt case (cf., e.g. Swahili and Spanish). Less well-known is case marking that is sensitive to a direct vs. inverse scenario distinction. This is attested for example in Yurok or Sahaptin and parallels verb-coded direct vs. inverse distinction, but this is attested. But there are also systematic differences between referential scale effects on case vs. on agreement: (i) split alignment in agreement does not seem to follow a universal referential scale (Bickel in press) while for case, at least a handful of families reflect some versions of the scale (Bickel & Witzlack-Makarevich 2008); (ii) while differential case marking may be driven by Zipfian principles (favoring zero exponence on the most frequent NPs, i.e. high-ranking agents and low-ranking patients; cf. Jäger 2007), no such principle apparently underlies agreement, which is instead favored by high-ranking arguments regardless of their role (Siewierska 2004, Bickel in press). (iii) differential agreement can affect both non-agentive arguments of ditransitives simultaneously, while differential case marking is often blocked by ditransitives (as in Nepali) or extends only to themes (as in Spanish). Differential agreement affecting both non-agentive arguments has been demonstrated, for example, in Chintang, a Sino-Tibetan language of the Kiranti group in Nepal (Bickel et al. in press, and below):

- (1) a. huṣa-ŋa Joge citthi hakt-o-ko.
 DEMs-ERG J.[-NOM] letter[-NOM] [3sA-]send-3sP-NPST
 ‘He sends the letter to Joge.’
- b. hungo kam citthi hak-no.
 DEMs[-NOM] 3sPOSS-friend[-NOM] letter[-NOM] [3sS-]send-NPST
 ‘He sends letters to friends.’ (in general)

In (1a), the verb is inflected transitively, with *-o* signaling a third person singular primary object (‘P’). In (1b), the verb lacks object agreement, and this entails that both arguments must be interpreted as being of unknown cardinality, which typically means nonspecific reference. As a result, neither *kam* ‘friend’ nor *chitthi* ‘letter’ can be preceded by demonstratives or marked by number affixes. The effect is obtained with all ditransitive verbs, regardless of whether they assign nominatives to all objects, as in (1), or nominatives and instrumentals (cf. ‘cover someone with something’), or nominatives and locatives (cf. ‘send someone somewhere’).

In order to understand the extent to which differential agreement and differential case marking are typologically similar, to what extent they differ, and what principled reasons there could be for similarities and differences, we choose a two-pronged approach: (a) a case study comparing differential agreement and case marking in the corpora of two languages, and (b) an analysis of the typological distribution of referential scale effects on agreement vs. on case. Only a case study involving corpora makes it possible to reach any degree of precision in understanding the nature of the relevant referential factors; and only a typological study makes it possible to distinguish ‘natural’ and therefore universal trends in this from idiosyncracies. The case study is planned as a Ph.D. dissertation project; the typological study is planned in close cooperation with Siewierska’s IP and involves one post-doc position shared with that IP.

Language choice

For the case study, we choose the language illustrated above: Chintang. The choice is based on the fact that Chintang has differential object agreement but no differential case marking on objects, and that it is in close contact with a language that has differential case marking on objects but no differential object agreement. This language is the Indo-European language Nepali, the *lingua franca* of the Chintang-speaking area and much else beyond in the Central Himalayas. Chintang has been in contact with Nepali for at least about 150 years, and for the past few generations, most Chintang speakers have been bilingual in Nepali. Currently, all speakers are bilingual (in fact trilingual as they also speak a closely-related neighboring language, Bantawa).

The advantage of choosing a pair of languages that are in intensive contact is that (a) convergence in other parts of grammar (word order, clause linkage, discourse strategies, particle usage etc.) allows reducing confounding factors in typological comparison, and (b) we can also investigate to what extent differential agreement and case can mutually influence each other or whether there are any principled differences between the two systems that may block such influences. This second issue ties into a broader research issue that has been of long-standing interest in Himalayan linguistics: to what extent did Sino-Tibetan and Indo-European influence each other and in what directions?

Research questions

1. Syntax of differential agreement in Chintang. At present, we understand only the basics of differential object agreement in Chintang, and one key issue concerns the syntax of the structure: in what precise ways is the structure similar to, or different from, antipassivization on the one hand and incorporation on the other hand? How does it compare to other detransitivization phenomena in other languages (referred to variously as ‘noun stripping’, ‘binominatives’ etc. in the literature). This research can initially draw on earlier research in a neighboring language (Puma, Bickel et al. 2007) but will then also require close cooperation with other members of the CRP and consultation of the typological and theoretical literature.

2. Semantic and pragmatics of differential agreement in Chintang. Our impression is that the key semantic difference between forms with vs. without object agreement in Chintang is whether or not the forms entail a specific cardinality of the set of possible referents. This is inspired by research on Puma, but it needs thorough investigation in Chintang. What has not been studied at all so far, and what will be a key issue in the proposed project is the question of what discourse conditions and which referential factors — animacy, agency potential, topicality, focus etc. — affect the likelihood of object agreement. This question also requires research on the way discourse particles modulate information structure (various markers of topicality and ‘emphasis’).

3. Comparison of Chintang and Nepali. To what extent are differential agreement in Chintang and differential case marking in Nepali similar? Where exactly do they differ in terms of (a) semantic entailments and (b) discourse factors that favor one or the other form? Is there any evidence for convergence due to language contact?

4. Typological comparison. What kinds of referential factors are involved in differential agreement and differential case marking elsewhere in the world? While it is impossible to determine such issues at the same depth as in a case study, we can at least compare what is known about the effects of categorical factors (like [\pm speech-act participant], or [\pm person], or [\pm animate]) in a sample of languages worldwide.

Corpora needed for research topics 1-3

In Chintang, a project sponsored by the DoBeS program of the Volkswagen Foundation (www.uni-leipzig.de/~ff/cpdp) has developed a corpus that currently (September 2008) consists of about 185,000 words.¹ About half of this is spontaneous multi-party conversational data and about 20% is narrative data; the rest consists of various genres including ritual speech. Of the entire corpus, about 80% is currently glossed and each morpheme is tagged for the language it belongs to (enabling research on code-switching). All data are in addition fully translated into English and Nepali. This data will be available for research in the current project, but since it was gathered by researchers in another project, all use is subject to the Fair Use Code of Conduct of that project. This Code of Conduct, which is available at the web site of the Chintang Documentation Project (<http://www.uni-leipzig.de/~ff/cpdp>), regulates how contributions to the corpus should be acknowledged and under what conditions corpus-based publications require invitation of other contributors to be co-authors.

For Nepali, we have access to a 260,000-word corpus of spontaneous speech in various genres (http://www.bhashasanchar.org/ncorpus_spoken.php) that was developed at the Department

¹ This is complemented by a longitudinal child language corpus containing about additional 80K.

of Linguistics at Tribhuvan University, Kathmandu, with which the Department of Linguistics at U. Leipzig has a formal cooperation agreement, and with which there has been intensive and successful cooperation for many years.

In addition, we have a corpus of 10 Pear Stories (Chafe 1980) in Nepali (collected and transcribed by Bickel; cf. Bickel 2003) and this will be complemented by the same number of Pear Stories in Chintang in the present project, to be recorded under the same experimental conditions as the Nepali Pear Stories. The Pear Story collection will allow parallel corpus research and systematic control of factors influencing differential agreement and differential case marking.

All corpora will need to be tagged for possible factors — animacy, focus, topicality, generic vs specific reference etc — that are suspected to be relevant for the choice of agreement (Chintang) and case (Nepali) forms. The coding set will be developed by the PhD student in collaboration with Bickel and other partners in the CRP, and we will then decide to what extent the actual coding will be done by trained native speakers and to what extent it will be done by the PhD student. (There are currently eight well-trained native speaker transcribers and coders in the Chintang Documentation Project.). For the practical coding work, the Chintang Documentation Project has developed a Java tool that assists researchers in the consistent coding of corpora. (The tool is currently in pilot use for coding transitivity and will soon be made available publicly.)

All additional annotations and all additional Chintang data that will be gathered during this project, including elicited data, will be deposited in the same archive as the currently existing corpus, viz. in the DoBeS archive (www.mpi.nl/DOBES). Future use of the data will then be subject to the same regulations as the ones holding for the current corpus. This includes the Code of Conduct mentioned before and a regulation of access rights (also available at the Chintang Documentation Project web site). Basically, access to data is granted for scientific purposes on request (and with sufficient anonymization of speakers), but any access is monitored by an Advisory Board of the Chintang community.

All enrichment of the Nepali data and the Pear Stories will be archived together with the existing Nepali corpus archive at Tribhuvan University in Kathmandu.

Typological database needed for research topic 4

We have a database with detailed information on referential splits in case marking systems of over 330 languages as part of the AUTOTYP (www.uni-leipzig.de/~autotyp) network (for first results, see Bickel & Witzlack-Makarevich 2008). The database is structured in such a way that it can be straightforwardly filled by parallel data on agreement. In fact, part of this will be accomplished within an ongoing DFG-sponsored project on grammatical relations (BI 799/3-1 “Typological variation in the processing of grammatical relations”), but since that project does not specifically focus on referential conditions, additional manpower is needed. In order to integrate this research with the database development in Siewierska’s IP we apply for one one-year post-doc position to be filled by the same person who will be hired for two years in Siewierska’s IP. Given the state of our database, the equivalent of one year work-time will suffice for entering agreement information and performing analyses together with Bickel and Siewierska. In order to facilitate data acquisition both within and outside the CRP, we will design questionnaires in the first few months of the project.

The database will be archived as part of the AUTOTYP network, but it will also be published as an electronic appendix to the first publication drawing on the database.

b) Methods

The corpus project will involve additional recording of Pear Stories in Chintang under controlled conditions, as specified in Bickel (2003). In addition, we plan classical elicitation work covering research questions 1-3 on Chintang and Nepali. All such fieldwork will be done in agreement with the ethics guidelines of the Department of Linguistics at U. Leipzig (<http://www.uni-leipzig.de/~ff/ethicguidelines.html>).

For the analysis of the corpora we will employ multiple regression modeling, following Baayen (2008). For the statistical analysis of the typological database we will use the same methods, but adapted to the specific challenges of typological data (non-random sampling, impossibility to stratify on language families because of isolates, etc; see Bickel 2007). In addition, we will explore the adaptation of advanced explorative methods (various clustering methods, high-dimensional regression etc.) for discovering patterns in referential scales across families, areas and, above all, across the difference between case and agreement (starting with Bickel & Witzlack-Makarevich’s (2008) Multi-Dimensional Scaling analysis of scales in case marking).

c) Workplan

The project aims at starting in June 2009. In the first six months, the doctoral student will survey the theoretical and typological literature on differential object marking. During the same time, he or she will

work on the glossing of Chintang texts so as to become familiar with the language. (We have made excellent experience in the past with student assistants quickly learning how to gloss texts using the semi-automated routines offered by the program 'Toolbox' and the by-now fairly rich dictionary that we have.) In August 2009 the student will attend an intensive course in Nepali that is regularly offered at the University of Heidelberg and familiarize himself/herself with the Nepali corpus. Based on the work on Chintang and Nepali, the student will then work with Bickel and other CRP members in developing a set of codes for further tagging of the corpora. At this point, we will decide which part of the coding is better done by native speaker assistants and which by the Ph.D. student. The first half of 2010 will be devoted to fieldwork by the student. This will include recording of Pear stories and intensive elicitation work on both languages. The second half of 2010 will be spent in Leipzig, working up the fieldwork results, further developing the corpus tagging and conducting corpus searches based on this. In early 2011, a second field trip of about two months is planned to answer new questions arising from the analyses, but the major part of 2011 and the first half of 2012 will be used for writing up the dissertation and articles based on it and on archiving the data and annotations.

The student will be supervised by Bickel throughout the lifetime of the project and Bickel will travel to Nepal in 2010 in order to advise the student on the ongoing fieldwork and, if necessary, to help train native speakers in corpus coding.

The post-doc will be hired two thirds in Siewierska's IP and one third in the present IP. The workload will be allocated accordingly, and work on the databases will proceed simultaneously during the first two years of the project. The last year is reserved for performing analyses and writing up results in collaboration with Siewierska, Bickel and other interested members of the CRP.

References

- Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics*. Cambridge: Cambridge University Press.
- Bickel, B. (in press). On the scope of the referential hierarchy in the typology of grammatical relations. In G. G. Corbett & M. Noonan (Eds.), *Case and grammatical relations: papers in honor of Bernard Comrie*. Amsterdam: Benjamins.
- Bickel, B. (2003). Referential density in discourse and syntactic typology. *Language* 79, 708 – 736.
- Bickel, B. (2007). A general method for the statistical evaluation of typological distributions. Ms. University of Leipzig, <http://www.uni-leipzig.de/~bickel/papers>.
- Bickel, B., Gaenszle, M., Rai, A., Rai, P. D., Rai, S. K., Rai, V. S., Sharma (Gautam), N. P. (2007). Two ways of suspending object agreement in Puma: between incorporation, antipassivization, and optional agreement. *Himalayan Linguistics* 7, 1 - 18.
- Bickel, B., Rai, M., Paudyal, N., Banjade, G., Bhatta, T.N., Gaenszle, M., Lieven, E., Rai, I.P., Rai, N.K., Stoll, S. (in press). Ditransitives and three-argument verbs in Chintang and Belhare (Southeastern Kiranti). In A. Malchukov, M. Haspelmath, & B. Comrie (Eds.), *Studies in ditransitive constructions*. Berlin: Mouton de Gruyter.
- Bickel, B. & Witzlack-Makarevich, A. (2008). Referential scales and case alignment: reviewing the typological evidence. In A. Malchukov & M. Richards (Eds.), *Scales*. Leipzig: Institut für Linguistik.
- Chafe, W. (1980). *The Pear Stories: cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- Jäger, G. (2007). Evolutionary Game Theory and typology: a case study. *Language* 83, 74 - 109.
- Siewierska, A. (2004). *Person*. Cambridge: Cambridge University Press.