Deriving the weight of syntactic constraints from experience

Markus Bader
University of Konstanz

While the gradedness of syntactic constraints has been widely recognized, most theories aimed at modeling graded acceptability in syntax explicitly deny the possibility that constraint weights can be learned from experience (e.g., *Linear Optimality Theory*, Keller, 2006; the *Decathlon Model*, Featherston, 2005). This contrasts with much work in phonology (e.g., *Stochastic OT*, Boersma & Hayes, 2001; *Harmonic Grammar*, Coetzee & Pater, 2008). One of the reasons for this is that research comparing graded acceptability ratings with corpus-derived frequencies in the domain of syntax (e.g., Featherston, 2005; Kempen & Harbusch, 2008) has uncovered certain mismatches between perceived grammaticality and corpus frequency. One mismatch that has been found is that syntactic structures that all occur with zero or near-zero frequency can nevertheless receive significantly different grammaticality ratings. Mismatches of this kind have been used to argue that constraint weights cannot be learned from experience.

This talk presents a case study on verb cluster formation in German showing that such mismatches between acceptability and language use only hold when acceptability is related to language use on the level of global sentence probabilities, as in Stochastic OT, but not when this relationship is considered on the atomic level of individual constraints, as in Harmonic Grammar. The relevant experimental data on graded acceptability are partially taken from the literature (Bader & Schmid, 2009) and partially new. To obtain the relevant frequency data, we analyzed the DeWac Corpus of internet texts (Baroni, Bernardini, Ferraresi & Zanchetta, 2009).

Consider first three-verb clusters with a modal verb in the perfect tense, as illustrated in (1).

(1) a. . . . dass Maria ein Buch **hat lesen müssen**. $\sqrt{}$Aux=1, $\sqrt{}$V<Mod
          that M.   a   book has read   must
          '. . . that Maria had to read a book.'
     b. *. . . dass Maria ein Buch **lesen müssen hat**.   *Aux=1, $\sqrt{}$V<Mod
     c. *. . . dass Maria ein Buch **müssen lesen hat**.   *Aux=1, *V<Mod

In (1-a), the order of lexical verb (*lesen* 'read'), modal verb (*müssen* 'must') and finite auxiliary (*hat* 'has') follows the rules of prescriptive grammar. First, the lexical verb precedes the modal verb, in accordance with the default rule for verbs in German (selected verb before selecting verb). Second, the finite auxiliary occupies the initial position of the verb cluster. This is required by a special linearization rule for modal verbs in the perfect tense. The two sentences in (1-b) and (1-c) deviate from the grammatical order to different degrees. In (1-b), the auxiliary is in final position instead of the required initial position but the main verb still precedes the modal verb. In (1-c), the auxiliary is again in final position and in addition the modal verbs precedes the main verb.

Experiments show that sentences with the grammatical order are highly acceptable whereas sentences with deviating verb orders are rejected most of the time. Importantly, sentences like (1-c) are judged as even worse than sentences like (1-b). In terms of corpus frequencies, both ill-formed orders occur with zero or near-zero frequency. Despite initial appearance, this is not necessarily at odds with the hypothesis that language use determines acceptability. After all, (1-c) violates two constraints on verb clusters whereas (1-b) violates only a single constraint. The seeming discrepancy between judgment and frequency data can therefore be reconciled by making two assumptions. First, constraint violations act in a cumulative way on the acceptability of sentences (see Sorace & Keller, 2005). Second, constraint weights are learned from experience.

This talk will first present an analysis of three-verb clusters in the framework of Harmonic Grammar as presented in Boersma & Pater (2008) and Coetzee & Pater (2008). Simulations using the Praat program show that experimental acceptability jugdments can be predicted from learned constraint weights. As a further test of the resulting model, the acceptability of sentences with four-verb clusters was investigated using the method of magnitude estimation (Bard, Robertson & Sorace, 1996). The results of this experiment will be presented in conjunction with additional corpus data on four-verb clusters. Again, simulations show that acceptability can be predicted from

learned constraint weights. The presented results will finally be discussed with regard to the general relationship between grammar and language use.

References

Bader, M. & Schmid, T. (2009). Verb clusters in Colloquial German. *The Journal of Comparative Germanic Linguistics* 12, 175–228.

Bard, E. G., Robertson, D. & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language* 72, 32–68.

Baroni, M., Bernardini, S., Ferraresi, A. & Zanchetta, E. (2009). The WaCky Wide Web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation Journal* 23, 209–226.

Boersma, P. & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32, 45–86.

Boersma, P. & Pater, J. (2008). Convergence properties of a gradual learning algorithm for Harmonic Grammar. Unpublished manuscript, University of Massachusetts at Amherst (available at http://roa.rutgers.edu/).

Coetzee, A. W. & Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26, 289–337.

Featherston, S. (2005). The decathlon model of empirical syntax. In M. Reis & S. Kepser (Eds.), *Linguistic evidence*, (pp. 187–208). Berlin: de Gruyter.

Keller, F. (2006). Linear optimality theory as a model of gradience in grammar. In G. Fanselow, C. Féry, R. Vogel & M. Schlesewsky (Eds.), *Gradience in grammar: Generative perspectives*, (pp. 270–287). New York: Oxford University Press.

Kempen, G. & Harbusch, K. (2008). Comparing linguistic judgments and corpus frequencies as windows on grammatical competence: A study of argument linearization in German clauses. In A. Steube (Ed.), *The discourse potential of underspecified structures*, (pp. 179–192). Berlin: de Gruyter.

Sorace, A. & Keller, F. (2005). Gradience in linguistic data. *Lingua* 115, 1497–1524.