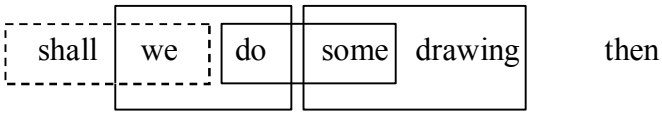
Agreement groups coverage of Hungarian mother-child language

Abstract

Agreement groups constitute a distributional framework for analysing linguistic datasets. The analysis is based on groups of minimally contrasting utterances. Such groups can also be considered as representing agreement relations (as understood in the generalised sense of Drienkó 2012b). Agreement groups provide a means for processing novel utterances on the basis of utterances already encountered. Analysing 2-5 word long English mother-child utterances, Drienkó (2012a) found that at any stage of linguistic development some percentage of novel utterances were compatible with the agreement groups extracted from the body of utterances encountered that far. The results were slightly improved when a "category guessing" mechanism was added. The author also claimed that the formation of groups may support categorisation, and the actual emergence of grammatical agreement. Similar results are to be reported in Drienkó (2013a) for Hungarian and Spanish. Drienkó (2013b) is meant to be a first step towards extending agreement groups analysis for sequences larger than the utterances in the groups: it is investigated how agreement groups can account for fragments of longer utterances, i.e. to what extent longer utterances can be "covered" by agreement groups. We extracted a collection of agreement groups from the CHILDES files corresponding to the first 32 Anne sessions of the Manchester corpus (Theakston et al. 2001), and applied our method to the utterances of the next 30-minute session. We found the average "coverage" of an utterance to be about 78%.

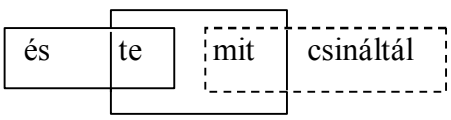
Example 1: Coverage of sentence *'shall we do some drawing then'*.
Boxes indicate fragments compatible with some agreement group. Continuous lines are for fragments which would be novel utterances. The discontinuous box means that *'shall we'* is not novel, it is an explicit utterance of the corpus.



Coverage (5 words of 6): 5/6=83%

The present work aims to replicate Drienkó's (2013b) experiment, this time with Hungarian data. Our CHILDES dataset comes from the Miki files of the Réger corpus (Réger 1986, Babarczy 2006). We extracted agreement groups from the first 30 sessions and tested coverage on the last (31st) file.
We find the average coverage to be about 42%. If coverage by only novel fragments is considered, the result is 29%.

Example 2: Coverage of sentence *'és te mit csináltál'*



Coverage 1: 4/4=100%
Coverage 2 (only novel): 3/4=75%

Our results illustrate that, similarly to the case of English, the linguistic processing of Hungarian may depend on both "prefabricated" elements stored in memory and novel phrases analysable distributionally, i.e. via agreement groups.

The difference between our results for Hungarian, and our earlier results for English may partly be due to differences in corpus parameters (corpus size, length of sessions, transcribing conventions, etc.). On the other hand, the lower level of coverage for Hungarian might be ascribed to morphological differences: several forms of a base word were regarded as different word types, which must have had a more significant impact on the distributional structure of the Hungarian dataset.

References

Babarczy, A. (2006) The development of negation in Hungarian child language. *Lingua* 116, pp. 377-392.

Drienkó, L. (2012a) Agreement groups analysis of mother-child discourse. Talk presented at the *4th UK Cognitive Linguistics Conference, King's College London*, 10-12th July 2012.

Drienkó, L. (2012b) *A linguistic agreement mapping-system model: agreement relations for linguistic processing*. LAP-Lambert Academic Publishing

Drienkó, L. (2013a) Distributional cues for language acquisition: a cross-linguistic agreement groups analysis. Poster presentation for the *11th International Symposium of Psycholinguistics, Tenerife, Spain* 20-23 March, 2013

Drienkó, L. (2013b) Agreement groups coverage of mother-child language. Talk to be presented at the Child Language Seminar conference, Manchester, 24-25 June, 2013

MacWhinney, B. (2000) The CHILDES Project: Tools for analyzing talk. 3rd Edition. Vol. 2: The Database. Mahwah NJ: Lawrence Erlbaum Associates.

Réger, Z. (1986) The functions of imitation in child language. *Applied Psycholinguistics* 7. 323–352.

Theakston AL, Lieven EV, Pine JM, Rowland CF.(2001) The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *J Child Lang.* 28(1) 127-52.