Subject-Verb Agreement with Conjoined NP's in the Ancient Greek Dependency Treebank

Francesco Mambrini[1] and Marco Passarotti[2]

Universität zu Köln, Germany[1]

fmambrin@uni-koeln.de

Università Cattolica del Sacro Cuore, Milan, Italy[2]

marco.passarotti@unicatt.it

EALC, University of Oslo, June 14-17 2012

Partial and Resolved Agreement: a corpus-based approach

Different languages show different means to deal with situations in which a verb or a predicate (the "target") is required to agree with a conjoined subject phrase (the "controller"). In some languages the agreement can be triggered by only one of the multiple conjuncts ("partial agreement", PA), while in others the agreement is controlled by the "resolved" features of the coordinated argument as a whole ("resolved agreement", RA). Hence, in the case of agreement in number between target and controller, if the conjunction of two subjects results in singular agreement, PA is implicated; if it results in dual or plural agreement, the agreement reflects resolution (RA).

The agreement of coordinated subjects with the verb is a well known issue in current syntactic research. However, no systematic treatment of this topic can be found in ancient Greek grammars. The lack of resources for quantitative approaches to such general phenomena as verb-subject agreement has been perhaps one of the causes for this absence.

The Ancient Greek Dependency Treebank (AGDT) of Perseus Project can now provide the dataset to start the research on resolved and partial agreement. Currently, the AGDT includes the whole text of "Iliad" and "Odissey", along with the complete opus of Hesiod, Aeschylus and several tragedies of Sophocles. All the texts included in the corpus are syntactically annotated according to a dependency-based representation style (Bamman et al. 2009). A dependency-based representation of a sentence is a tree-graph in which each word of the sentence corresponds to one node of a tree and the nodes of the tree are connected in head-dependent fashion.

In this paper, we aim at presenting a first survey of the phenomenon of resolved and partial agreement with conjoined subject phrases in Ancient Greek. We concentrate on morphological number agreement only, as complex verbal forms (i.e. those including a past participle and an auxiliary verb) are quite uncommon in our data. We will focus on the sample of 729 cases where a group of coordinated subjects is governed directly by a verb, whose different behavior in agreement can be easily studied.

Agreement with coordinated NP's: a first survey

In opposition to what happens in languages such as English (where non-resolved agreement is confined to or-coordinates and a very limited number of specific constructions), PA is a very pervasive phenomenon in the AGDT. Cases of PA amount to 55.65% of our sample. As expected, the ratio of non-resolved agreements with or-coordinates is considerably higher than with and-coordinates in ancient Greek too (83.87% and 53.34% respectively). But for the and-coordinates, it must be observed that PA covers 68.69% of the truly ambiguous cases, i.e. those where the coordinated subjects are not all plural and where both constructions are in fact possible.

Four parameters can be singled out in order to test their influence in the choice between PA and RA:

- ✠ distance (i.e. number of the tree nodes between the verb and the nearest subject in the conjoined NP);
- ✠ "breadth" of the NPs (defined as the number of nodes occurring between the nearest and the furthest subjects in the NP);
- ✠ position of the controller in the word-order: preverbal, postverbal, and mixed (i.e. coordinate subjects are placed both on the left and the right of the verb);
- ✠ animacy (Zaenen et al. 2004).

1

The average distance between the verb and the nearest subject of a conjoined NP is 2.24 nodes; the average breadth of the coordinated phrase that plays the role of subject is 4.08. Our data show that, in the cases of PA, the average distance is lower (1.83 nodes vs 2.55 in case of RA), while the whole conjoined NP displays a greater breadth (4.20 nodes in PA vs 3.65 RA). It will be also shown that sub-groups of data with singular-only or mixed-number subjects display the same trend even more distinctly: conjoined NPs that trigger PA feature a lower distance between subjects and a higher breadth.

The order of the constituents plays also a role in the data. RA is more frequent with preverbal subjects (49.07% RA vs 50.93% PA) than with postverbal subjects (41.38% RA vs 58.62% PA). In case of singular subjects only, a more marked trend is reflected: postverbal subjects trigger PA in 78.46% of cases; PA drops at 59.73% with preverbal subjects. PA dominates also in the case of discontinuous phrases, especially when one of the conjoined subjects is isolated in preverbal position, with the other constituent(s) of the NP occurring after the verb (67.17% PA vs 32.83% RA). The percentage raises at 88.71% of PA in case of NPs formed by singular subjects only.

Animacy seems to be relevant factor as well. We crossed our data with the animacy lexicon for the New Testament in Greek built for the annotation of the PROIEL corpus (Haug and Jøhndal 2008); the missing words, which were attested in our sample but not in PROIEL animacy lexicon, were annotated according to the same tagset. We were thus able to verify that PA dominates (93.33%) in the case of two or more non-concrete coordinated subjects (121 cases); conversely, human coordinated subjects license singular agreement only in 46.52% of the cases.

Future work and open questions

Current debate to partial and resolved agreement is especially centred on the contrast between clausal and phrasal interpretation of PA (see e.g. Aoun et al. 1994, Munn 1999). Advocates of clausal interpretation have pointed out a number of lexical items (such as the verb "meet" or the adverb "together" in English) that license plurality and are therefore incompatible with PA in a number of languages. The existence of comparable words and their behaviour in relation to agreement in ancient Greek is still to be explored.

The question of how to assess the clausal or phrasal nature of the attested cases of partial agreement can only be introduced at present and is open to methodological discussion.

References

Aoun, J., E. Benmamoun, and D. Sportiche (1994). Agreement, word order, and conjunction in some varieties of arabic. *Linguistic Inquiry 25* (2).

Bamman, D., F. Mambrini, and G. Crane (2009). An Ownership Model of Annotation: The Ancient Greek Dependency Treebank. In *Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories (TLT8), Milan*, pp. 5-15.

Haug, D. T. T. and M. L. Jøhndal (2008). Creating a parallel Treebank of the old Indo-European Bible Translations. In Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008), pp. 27-34.

Munn, A. (1999). First conjunct agreement: Against a clausal analysis. *Linguistic Inquiry 30* (4), 643-668.

Zaenen, A., J. Carletta, G. Garretson, J. Bresnan, A. Koontz-Garboden, T. Nikitina, M. C. O'Connor, and T. Wasow (2004). Animacy encoding in english: why and how. In *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, pp. 118-125.