

Data from Student Treebanking as a Pedagogical Resource

The direct advantages of using dependency treebanking in the Latin and Greek classrooms are becoming more familiar. Among these is the encouragement which treebanking gives the students to become mindful of every syntactic relationship. Treebanking also constitutes a crucial element of a rapidly developing field of research to which undergraduates can contribute in a meaningful way. It is against this background that this presentation will shift its focus to the more indirect advantages of classroom treebanking, in particular, the ways in which the data generated can be made to serve educational goals.

I have been treebanking in my Latin classes for more than five years using the resources of the Alpheios Project, the Arethusa Annotation Framework, the Perseids Project, and the Perseus Digital Library. Thus, I have available a great deal of hard data about student performance. These can reach a fine degree of granularity, since Arethusa generates an error matrix for each word of each sentence for each student. Student work is evaluated under four rubrics, *lemma*, *postag*, *relation* and *head*. The first two most directly reflect the student's control of vocabulary and word forms (*postag* refers to part of speech and associated morphology), *head* designates the item upon which the focus word depends, while *relation* is the grammatical label of a given dependency (*e.g.*, indirect object, or absolute clause). A typical semester's treebanking homework might include 45 assignments of three sentences per assignment, each sentence averaging c. 10 words. Thus, the system produces c. 5400 data points per student. For a single class of 25 students, a computational study of a single semester's treebanking results would be based on a set of c.135,000 items.

I have begun the process of analyzing these data and results will be presented in the CAMWS session. The first line of investigation is to identify significant correlation between

language features and student error rates. For example, what items of vocabulary are associated with greater than expected error in the *head* or *relation* data categories, or what syntactic structures give the students trouble regardless of the vocabulary items involved? Results can be surprising: many students have shown confusion about the dependency structure of the simple indirect object, although they are able to translate it without difficulty. We may also examine more complex relationships: what combination of features produces an error rate higher than we would predict from the errors associated with the individual components? The most interesting results will be presented to the panel with emphasis on the possible pedagogical use of the information.

The second section of this paper turns to the inherent difficulties in making student treebanking data comparable outside the walls of a single classroom. A salient problem is the result of one of Arethusa's most appreciated advantages, its great flexibility. Among other options, Arethusa allows users to build their own tag sets for labelling syntactic constructions. This function is vital for classroom treebanking, since the teacher can adopt the terminology of the chosen textbook, deciding, for example, to treat objective and subjective genitives as separate structures or combining them under a single label. Unfortunately, the resultant data sets cannot be compared directly, since syntactic structures are diversely described.

The concept of Dependency Distance may provide a remedy for this situation. A construction of linguists working in dependency grammar, DD is the distance in words from a given word to the word upon which it depends grammatically. Thus, for a sentence of at least two words, the minimum possible DD is 1 and the maximum one less than the number of words in the sentence. A sentence, too, has a DD, which is the average DD for its constituent words.

Average DD has been proposed as a good measure of syntactic complexity. As illustration, here are sentences of given DDs from Vergil (based on the treebanks published by Perseus):

iuvenum manus emicat ardens litus in Hesperium (*Aeneid* 6.5-6; DD: 1.5)

*gelidus Teucris per dura cucurrit ossa tremor, funditque preces rex pectore ab
imo:* (*Aeneid* 6.54-55; DD 2.5)

*ille meum comitatus iter maria omnia mecum atque omnis pelagique minas caelique
ferebat, invalidus, viris ultra sortemque senectae.* (*Aeneid* 6.122-123; DD 3.5)

Using the data drawn from my classroom trees, I will explore the effectiveness of DD as a proxy for uniformity of syntactic tag sets. The DD for all assigned sentences will be compared to its error rate, particularly in the *relation* and *head* categories. Positive correlation will support the proposition that student treebank data allows for meaningful comparison and aggregation, in spite of possible differences in the tag set used (or the author annotated). Such a result would permit collation between different classrooms and institutions, providing the basis of a deeper investigation into important details of student competence in Greek and Latin.