

Constructional Potentiality: Priscianic grammar as a disambiguation  
technique in the automatic recognition of Latin syntax

In most languages word order plays the major role in determining which words form a single phrase or constitute. A tree structure can be abstracted automatically from a sentence by linear determination of the major syntactic constituents. However, in certain highly-inflected languages, of which Latin is perhaps the most extreme example, constituents may be broken up in the word order of the sentence. Inflectional endings form the signals for placing each constituent in its constituent. Theoretically the higher degree of syntactic explicitness provided by inflectional morphemes should make these languages easier to analyse mechanically. Yet it proves that there is a high degree of ambiguity in the morphemes or morphophonemic alternations used to mark inflection and that native speakers rely heavily on their ability to identify both the semantic meaning and the inflectional habits of a given word.

Recognizing this state of affairs those institutes involved with automatic parsing of Latin, the Laboratoire pour l'Analyse Statistique des Langues Anciennes, Liège, and the Centro Automazione Analisi Linguistica, formerly in Gallarate but now in Pisa, have constructed rather large parsing dictionaries which identify the word-class and inflectional habits of lexical lemmata. These parsers do not, to my knowledge, produce tree structures of analysed sentences or give any deep-structure representation.

An alternative approach to the automatic syntactic analyses of Latin is offered by the constructional syntax of Latin developed by Priscian from Greek grammatical theory. Each word is a member of at least one 2-member construction. There are two types of constructional relationship

convenientia (agreement), (a) verb agrees with subject in number, (b) adjective agrees with noun; regimen (government), (a) verb governs object, (b) preposition governs noun. Several word-classes do not participate in constructions, notably connectives, some ordinal numerals, modifiers and adverbs, but these can be located in the syntax of the sentence, usually by word order, occasionally by the reduction of all possible constituencies to one. Priscian's list of constructions must be extended in some directions such as to include the genitive (dependent on the nearest noun) and the ablative.

The deep structure of the Latin sentence can be shown to consist of a left side containing, as in Chinese, a cluster of presentences, labels and topics, and a right side, which proceeds logically towards the speaker's comment on the main topic. The items of the deep structure are formed into a syntax tree structure (similar to the Russian  $\Delta$ -structure) by virtue of their membership in constructions legal in Latin syntax.

The first word of the sentence, if inflected, has a constructional potentiality assigned to each possible inflectional segment matched in a look-up of inflectional morphemes. Each inflectional morpheme has a code of constructional potentiality assigning to it all possible syntactic roles. The probability factor is not the same for this potentiality in the first position as for this potentiality further to the right in the sentence. The parsing program assigns construction <sup>n</sup> modes to each possible constructional pair until it arrives at a series of surface-level <sup>n</sup> modes which make a well-formed  $\Delta$ -structure.

By adding to the dictionary the constructional potentiality of all lexical items which are basic verbs one can assign values and semantic features to the words belonging to the multi-member constructions of transitive and bi-transitive verbs.