

Counting nouns is not always the right question

In many well-known languages, “counting” occurrences of nouns and verbs in a discourse is a relatively simple task. In those languages (e.g. English), three properties converge on the same categorial distinction: the morphological distinction between nouns and verbs, the syntactic distinction between noun phrases and verb phrases, and (simplifying) reference vs. predication. In this paper, based on a corpus study of Oneida, a Northern Iroquoian language, we show that “counting” is not so easy and that some languages challenge the very premise of the workshop in that the structure of a language can lead to quite distinct counts depending on which property is deemed relevant (morphological, syntactic nouniness/verbiness, reference vs. predication).

Oneida words belong to one of four distinct word classes: Verbs, Particles, Nouns, and Kinship terms (which have both verbal and nominal properties). Verb roots outnumber noun roots (roughly three to one). Interestingly, verbs are the only truly open class. New categories of entities are morphological verbs that are lexicalized as referring expressions (e.g. *kawistótha?* 'it makes it cold, refrigerator') and, occasionally, borrowings from English, often with Oneida inflections (e.g. *akotpony* 'her pony').

Identifying distinct noun or verb *types*, according to morphological criteria, is straightforward. But, determining how many noun phrase *tokens* or referring expressions occur is harder. This is because (i) referring expressions can come from any one of the four morphological classes and (ii) the syntactic difference between what one could call noun phrases and verb phrases in Oneida follows from whether the phrase is a referring expression or not, rather than from the morphological category of the root. To elaborate:

First, the dissociation between being a morphological noun and being the head of a referring expression has a drastic effect on how many “nouns” there are in Oneida discourse. Our analysis is based on a corpus of naturally-produced Oneida discourse (70 transcribed discourses, 14 speakers, 50,000+ clauses) and it shows that (i) morphological nouns are infrequent (about 5% as reported previously in Iroquoian literature), but (ii) the number of referring expressions is much higher (between 18-48% of clauses if one includes temporal/locatives expressions).

Second, the absence of a distinct *noun* phrase syntax means that counting noun phrases cannot be a way out of the quandary raised above. Because the syntax of Oneida is not based on the selection of dependents by heads (syntax is direct in Koenig and Michelson’s (2012) terminology) and involves mostly adjunctive and appositive structures, distributional differences that may seem to differentiate apparent noun phrases and verb phrases reduce to differences in semantic/pragmatic types (referring expressions vs. predication) so that counting noun phrases is actually counting referring expressions.

In conclusion, our paper shows that a close look at languages that differ drastically from Indo-European languages challenges the assumption that the right cross-linguistic question is asking how many nouns or verbs occur in discourse. Asking how many referring expressions there are seems a more fruitful, cross-linguistically valid question.