How many word-classes are there after all?

István Kenesei
Research Institute for Linguistics, HAS, Budapest

Morphology is concerned primarily with the internal properties of words, i.e., their structures, constituents, inflection, derivation, etc., rather than with such external properties such as how and why they belong to what word-class, yet these questions tend to resurface regularly.

Analysis of word structure must make reference to word-class, when, for instance, determining which affix can be attached to which stem, or what types of items can undergo compounding in a given language. And although word-class based derivational processes have recently been called into question (cf., e.g., Plag's 2004 critique of Aronoff 1976), it is taken for granted that words (lexemes, lexical items/units, listemes, etc.) are labelled for their categories and derived forms carry word-class labels.

Clearly, an answer along the familiar lines of simply identifying word-class labels with the terms *noun, verb, article*, etc. will not do, since we want to know precisely *what* it is for a word to be a noun, verb, article, etc. In the classical literature the answer was simple: give a notional definition, provide a few representative examples, and the reader will draw the conclusion. Compare this passage: "Les objets de nos pensées, sont ou les choses, comme *la terre, le Soleil, l'eau, le bois,* ce qu'on appelle ordinairement *substance.* Ou la maniere des choses; comme d'estre *rouge,* d'estre *dur,* […] &c. ce qu'on appelle *accident.* […] Car ceux qui signifient les substances, ont esté appellez *noms subsantifs*; & ceux qui signifient les accidens, en marquant le sujet auquel ces accidens conuiennent, *noms adjectifs.*" (Lancelot & Arnauld 1660/1967). The order of presentation may have varied, but this approach has been the dominant one, cf. Lyons (1977), which demonstrates that part-of-speech definitions are based on morphological, syntactic and semantic criteria, invoking for the latter ontological distinctions of first-order entities (persons, animals, things), second-order entities (events, processes, states-of-affairs), and third-order entities (propositions).

In spite of common beliefs, even the founder of the American structuralist tradition has this to say: "The noun is a word-class; like all other form-classes, it is to be defined in terms of grammatical features […] When it has been defined, it shows a class-meaning which can be roughly stated as follows 'object of such and such a species'; examples are *boy, stone, water, kindness.*" (Bloomfield 1933). And although Hockett (1958) based his definition on purely formal criteria, he lumped together minimal forms (words) with maximal forms (phrases), cf.: "[The pattern of interchangeability] defines a form-class which includes *she, he, it, John, Mary, the man at the corner, my friend Bill,* and so on endlessly, but which by no means inlcudes all forms, since we can name many which are excluded: *her, him, them, me, yes, no, ripe, find her, go with us tomorrow.*"

The question we address is, however, not how to determine the word-class of an item in a stretch of talk, utterance, etc., but what type of information word-class 'labels' encode, and how many such labels we have to do with.

To begin with, the inclusion of 'class-meaning' or any other semantic is obviously redundant since the meaning of the word, together with all its ontological consequences, must be part of its characterisation. But it is also obvious that a semantic characterisation will not determine the word-class as is argued on the basis of countless cases like the well-known Russian example based on the adjective *belij* 'white' and the verb *belet'* 'show white, occur as white', which can be mostly used interchangibly as far as semantics, but not syntax, is concerned.

It is precisely this point that word-class information can be pinned down: however much an adjective may resemble a verb in its meaning, if they differ in their word-classes, they will differ in their behaviour in syntax and morphology, that is in what syntactic and/or

morphological environments they can occur. An adjective can, among other things, be used in an attributive position, which is prenominal in some languages and postnominal in others; it can have comparative and superlative affixation (in some languages), etc. A verb takes up finite inflection, it takes part in syntactic processes, such as agreement, it may require a complement, etc. It follows then that information concerning the word-class of an item is information about what morphological and syntactic environments it can occur in, in other words, it is a set of instructions as to what derivational and inflectional affixes it can take, and what syntactic environment it determines.

As far as the latter aspect is concerned, in current grammatical theory a lexical item is characterised by its 'edge features', i.e., the properties that determine the structure they are (to be) constituents of. In other words, words project structure. In that sense, whether lexical or syntactic operations are concerned, lexical items are characterised by information that serves to identify with what type of items or constituents they can be merged with (cf. Chomsky 2005). Ultimately, word-classes are seen as encoding this type of information, specifying it in finer detail as major categories are broken down into subclasses, such as transitive, ditranstive, etc., verbs, or attributive and/or predicative, etc., adjectives, and so forth.

To pursue this line further, it transpires that traditional classifications of 'minor' word-classes, such as interjections, conjunctions, various types of adverbials, let alone the notorious 'particles' (e.g., *only, too, even, just* and their equivalents in other languages), are based on convenience rather than theoretical considerations; they arise out of a need to classify them for purposes of lexicography or grammar writing and are often based on common semantic or pragmatic properties (hence the term 'particles', because they have little, if anything, in common). Other dubious word-classes are set up on the distinction whether their members form a closed class, cf. the case of pronouns, exemplified in modern textbooks as Radford (2009).

As is well-known, grammatical words, or in current terminology, functional categories, are distinguished from notional or thematic categories by having clear-cut complementation structures but no thematic grids. Moreover, they all form closed classes, i.e., no productive word formation process (nor borrowing) can target them. Consequently, whether it is sensible to set up a category of 'articles' of four items (*a/an, the, some,* and zero) is highly questionable, since their 'edge features' will differ as to what NP is merged with them and are only held together by belonging to a common semantic/pragmatic domain. Again, this category might be due to the interest of grammarians and lexicographers in grouping all words under some heading. I will argue that word-class information is relevant only in case of open class items, and that the edge features of closed class items must be learned, listed, etc., one by one by the language learner or in her grammar.

References
Aronoff, Mark. 1976. *Word formation in generative grammar.* The MIT Press, Cambridge.
Bloomfield, L. 1933. *Language.* Allen and Unwin, London (1935/1967).
Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36, 1-22.
Hockett, Charles F. 1957. *A course in modern linguistics.* Macmillan, New York.
Lancelot, Claude, & Antoine Arnauld. 1660/1967. *Grammaire générale et raisonnée.* The Scolar Press, Menston.
Lyons, John. 1977. *Semantics.* Cambridge U.P., Cambridge.
Plag, Ingo. 2004. "Syntactic category information and the semantics of derivational morphological rules," *Folia Linguistica* 38: 3-4, 193-225.
Radford, Andrew. 2009. *Analysing English sentences,* Cambridge U.P., Cambridge.