

On defining parts of speech with Generative Grammar and NSM¹

Avery D. Andrews

1. Introduction

There is a long-standing controversy as to whether it makes sense to try to define parts of speech cross-linguistically, and if so, how to do it. The generativist position, advocated by Chomsky (1970), Baker (2003), Chung (2012a, 2012b), Panagiotidis (2015) and many others, has almost always assumed that it does, focusing on noun (N), verb (V) and adjective (A), applying a combination of syntactic and semantic ideas and concepts. On the other hand, functionally oriented typologists such as Dryer (1997), Croft (2001) and Haspelmath (2010, 2012) seem to have largely come to the conclusion that it does not; while Wierzbicka (2010) argues that it can be done, using the grammatical treatment of certain Natural Semantic Metalanguage (NSM) primes as universal exemplars. Here I will propose a different NSM-based approach, intended to be complementary to Wierzbicka's, which uses NSM to define 'semantic cores', whose members are treated equivalently by certain kinds of grammatical phenomena, restricted in order to avoid the problem of 'methodological opportunism' discussed by Croft (2001). The goal is definitions that can be applied

1 I would like to acknowledge the anonymous reviewers and Dominic Dessaix for helpful questions and suggestions.

universally, and also in an empirically meaningful manner, so that it is possible, in principle, for them to yield the result that the defined part of speech does not exist in a given language, whether or not this happens in fact. We will find definitions for noun and verb that apply universally and always seem to find something, whereas this is not the case for adjectives. A major difference from Wierzbicka (2010) is that the approach can be applied without working out the prime set of the language being investigated, with its attendant complexities such as identifying allolexes, since the relevant NSM specifications will take the form of NSM formulas which can be stated in the NSM of any language, including that of the investigator. The present approach can therefore be applied at a much earlier stage of work on a language.

The approach can be seen as a development of that of Lyons (1968), Schachter and Shopen (2007) and Wierzbicka (1986), aspiring to meet the criticisms by the latter of the former two, although I won't be able to go through this in detail. Since parts of speech are a complex problem, in section 2, I go through a number of preliminary issues for defining them, and then examine nouns in section 3. Then I will consider adjectives in section 4, verbs in 5 and some general conclusions in 6.

2. Preliminaries

Generative syntacticians generally believe that 'parts of speech', however ultimately defined, involve labels or features of some kind that occur in syntactic structures, which have a degree of autonomy from both sound and meaning. There are various ways of accomplishing this, but for convenience I will adopt the approach used in Lexical Functional Grammar as well as original Transformational Grammar (TG) and various other approaches, where they are labels on the nodes of trees, either all levels (TG), or only overt c-structure (LFG).

But, given the labels, there is a further issue of labels vs classes of lexical items. It is possible, for example, that there are universal labels such as N or V for syntactic positions, determining aspects of the surrounding syntactic environment (e.g. 'nominal' vs 'verbal' projections), but no restrictions on what open class lexical items can occupy them. That is, there might be just one open class of 'contentives', capable of functioning as nouns, verbs, etc. depending on their position in the structure. Such a position has in fact been argued for certain languages, including Austronesian and

Salishan languages, and also Nootka and Mundari. However, as discussed by Panagiotidis (2015: 29–37), these arguments are not convincing; see Chung (2012a, 2012b) for discussion of nouns vs verbs in Chamorro (Austronesian), arguing that in spite of the existence of many stems that can function as either nouns or verbs, these are distinct parts of speech due to unpredictable meaning shifts between the uses of putative ‘noun’ or ‘verb’ stems in the different positions. See also Evans and Osada (2005) for similar points on Mundari from a non-generative point of view, and issue 41(2) of *Studies in Language* for a number of further relevant studies (brought to my attention by an anonymous reviewer). Therefore, the current preponderance of research seems to indicate that all languages have distinct classes of lexical items associated with what are normally analysed as nominal and verbal positions, even though in some languages there are many words that can appear in both.

Assuming that we regard parts of speech as classes of items rather than kinds of syntactic positions, there are at least two approaches to trying to give them a characterisation in universal terms, one in terms of structure, the other in terms of meaning. Purely structural approaches do not appear to have been worked out to any extent, while the traditional definitions took a meaning-based approach, although with shortcomings discussed by Wierzbicka (2010) and, previously, many others. Baker (2003) takes a mixed approach, arguing that verbs are characterised structurally by the property of projecting a specifier, while for nouns he draws on a proposal by Geach (1962) and Gupta (1980) that the meanings of nouns, but not members of other parts of speech, involve ‘criteria of identity’. Panagiotidis (2015) on the other hand proposes to differentiate nouns from verbs on the basis of their semantics, following Acquaviva (2009),² but in a much more abstract way than Wierzbicka. I won’t claim that the approaches of Baker and Panagiotidis are wrong, only that they are somewhat abstract and theoretical; the present proposal aspires to more immediate applicability by people who don’t have much time to delve into complex theoretical and philosophical issues, as well as providing fewer potential loopholes for saving proposals that are in difficulty. Wierzbicka’s on the other hand requires working out the prime set of a language, making it difficult to apply at the early stages of work on a language.

2 References are to the version available through LingBuzz (ling.auf.net/lingbuzz).

Parts of speech are sometimes defined in terms of the possibilities for substitution, along roughly the following lines:³

- (1) A word class is a set of words that can be substituted for each other without affecting grammaticality

This definition depends on the concept of ‘word’, whose nature is another long-term controversy in linguistics; see Goddard (2011) for discussion with respect to NSM. I will not attempt to say anything substantial here about this controversy, but will merely note that WORDS is an NSM prime (however this works out for formulating technical definitions), and, for all languages, some kind of reasonably satisfactory notion of ‘word’ seems to be available (even if it is coextensive with ‘formative’, or is somewhat arbitrary in certain respects), and that the way in which controversial issues of word-hood are resolved doesn’t seem to have much substantive effect on issues of what parts of speech are taken to exist in a language, and none on what open ones are to be accepted.

But (1) encounters a problem with inflection, in that different inflected forms of the same word are frequently not inter-substitutable in all environments, and in addition, a further problem, first discussed at length by Gross (1979), is that on a rigorous application of the criterion of (1), words fall into very large numbers of very small classes, and the traditional parts of speech would only be the highest divisions in the category system, and would not be very clearly defined. Mass and collective nouns, for example, have different substitutional properties than count or plural ones, and split into further subcategories on various grounds, as discussed by Goddard (2017: 256); would they be one part of speech as traditionally assumed, or two? In the opposite direction, adjectives and adverbs have certain things in common, so perhaps they should be combined into one part of speech, and, indeed, sometimes are.

The original definition of part of speech was, however, given in terms of inflection, and we do not wish to abandon that. What I suggest is the following: in each language, each part of speech has a system of inflectional categories, such that any form that belongs to that part of speech must

3 Cf. the definition of ‘syntactic category’ in Culicover (1976: 13): ‘A syntactic category is a group of words or sequences of words in a given language that can replace one another in a sentence of the language whatsoever without affecting grammaticality’. This generalises the definition in (1) to cover phrases as well as single words.

belong to at least one of the categories. For example, in English, a given noun might lack a singular or plural form, but must appear in one form or the other.

But in languages with little or no inflection, there needs to be more, which I suggest is based on word order. In English, for example, most adjectives come before the head noun in the noun phrase unless they have a prepositional complement, in which case they come after. What we do *not* want is the possibility or necessity of basing parts of speech distinctions on compatibility with certain kinds of complements or determiners, or phenomena of selection, such as the choice of unitisers with different kinds of mass nouns (Goddard 2009, 2017):

(2) I need several items of cutlery/?furniture/*water/*meat/*fork

I therefore propose a relatively ‘theoretical’ definition of the concept of part of speech, compatible with LFG works and other frameworks that give a relatively direct account of overt structure:

- (3) A part of speech is a class of lexemes defined by:
- (a) how it is involved in the phrase structure rules or equivalent, that is, with the constraints that are responsible for the overt structure of relative ordering of words in utterances, and
 - (b) the requirement that it has a system of inflectional categories such that the form of any word that is functioning as a member of that part of speech belongs to at least one of the categories (specified in all dimensions if there are multiple ones, such as gender, number and case, in German or Russian).

Observe that (3), a definition of the concept of part of speech, is a ‘comparative’ concept as described by Haspelmath (2010), based on formal ideas, while the definitions of specific parts of speech to come have a semantic aspect, but will still be comparative concepts in Haspelmath’s sense. By limiting the criteria, we aspire to avoid, or at least reduce, methodological opportunism (Croft 2001).

With these preliminaries in hand, we can proceed to the problem of defining the part of speech ‘noun’.

3. Nouns

As discussed above, our strategy, essentially derived from Lyons and Schachter, will be to define a semantic core, such that everything or almost everything that falls into it will be a noun. A problem we wish to avoid, however, is to give the appearance of assuming that nouns will already be present (although that appears to be the case). Therefore we formulate the definition in two parts, an existential statement that is empirical, and might be falsified (but hopefully is not), and the definition itself, based on the empirical assertion. We state our assertion as follows, mostly now in NSM, but using ‘[T]’ to mark first use of technical terms which we have not and will not here attempt to explicate. The need for the ‘open[T]’ qualification will be discussed below:

(4) There is an open[T] part of speech[T] like this:

A word[T] is in this part of speech if one of these things are true:

This word says what kind of thing something is

This word says what kind of person someone is

This word says what kind of place some place is

This word says what kind of living thing some living thing is

Having made our empirical claim, we can then use it to define a part of speech:

(5) People can say what this part of speech is with the word ‘noun’

Although I’ve expressed (5) in something very close to NSM, the role of NSM is less essential here.

The empirical claim of (4) is that words that fall into the semantic core provided by (4) obey the requirement of belonging to one of the inflectional classes associated with some single part of speech, and will also not behave differently from each other under the principles determining relative order. That is, a word meaning ‘dog’ will not precede a definite article while one meaning ‘tree’ follows it. On the other hand, we can observe that this core is split in extremely random ways by grammatical gender classifications, which, however, do not appear to ever affect relative order, and so therefore do not distinguish parts of speech by our definition. A potential problem with this definition is exactly what a ‘kind’ is taken to be. Is a big dog a kind of dog, for example (Dominie Dessaix,

pers. comm.)? I propose not, on the basis that if somebody says ‘what kind of animal is that’, and somebody else answers ‘a (bloody) big one’, they are clearly indicating that they don’t know the answer to the question. But the issue of ‘core’ vs ‘extended’ uses of the primes, is nevertheless important.

A subtlety of this definition is that it does not preclude the existence of languages with only one part of speech, as discussed in the previous section. For this, we need a principle that requires the existence of at least one additional part of speech. The most obvious proposal for this is ‘verb’, for which we propose a definition in section 5. We do on the other hand have an effective safeguard against the problem of indefinite subdivisions of word classes; there can be subclasses of noun, such as proper names or mass nouns (not covered by (4)), but these will not be parts of speech according to the definition, since mass nouns do not obey different principles of ordering than those that fall under (4), in English or other languages known to me.

The proposal is not inconsistent with that of Wierzbicka (2010), but has somewhat different properties. She proposes that the primes *THING(S)* and *PEOPLE* provide a grammatical prototype for nouns, which implies that there are at least two nouns, but not, actually, that there are any more, although one might be able to derive some expectations on the basis of further assumptions. A somewhat peculiar point is that (4–5) do not actually imply that the exponents of *THING(S)* and *PEOPLE* are nouns, since they do not designate kinds of things and people, but rather the entire supercategories. The present proposal furthermore differs from Wierzbicka’s in that it does not require that the investigator find the exponents of these primes and ascertain their syntactic properties. Both approaches appear to do something useful, and there is more to be done to fit them together.

But an issue arises with the semantic core: what should really be in it? (4) provides for four kinds, but there are others that might plausibly be included. One is proper names, which often have some special properties, but appear to be nouns nevertheless according to our criteria. Another (pointed out by Martin Haspelmath, pers. comm.) is mass nouns. Both of these raise some puzzles, but they are not critical, because although it would be good to make the semantic core as large as it can be, (4) already asserts the existence of one big enough to provide the basis of a part of speech definition.

Another problem that needs to be addressed is provided by classifiers, reviewed extensively by Aikhenvald (2000). These occur in a number of different types, sometimes together in a single language, occurring in combinations with nouns and other items, especially numerals and other quantifiers, in various environments. The classifier used with a nominal expression is normally determined by its meaning, sometimes also saying what kind of thing something is (along with other possibilities), in accordance with (4). But they are clearly different from nouns, in the languages that have them, appearing in different positions. For example, the 'noun classifiers' *naj* and *no7* in Kanjobolan (Mayan, Craig 1992: 284)⁴ apply to people and animals, respectively, occurring before the noun:

- (6) *xil* [*naj* *xuwan*] [*no7* *lab'a*]
 saw CL:MAN John CL:ANIMAL snake
 'The man saw the snake.'

For this chapter, the important fact about classifiers is that there are a finite number of them, although the number can be reasonably large, up to several hundred (Aikhenvald 2000: 84). Classifiers are therefore a closed class in languages that have them, and where they are independent words.

For this reason we have included in (4) the specification that the class be open. A potential problem is presented by so-called 'repeaters', which are nouns that appear to be used as classifiers, appearing in classifier positions, often instead of the normal classifiers (Aikhenvald 2000: 361). But repeaters can be treated as nouns, since any noun can be a repeater. Classifiers are very intricate, and there may be more problems for our definition lurking within them; for example, Truqueuse with its 'virtually open' system of relational classifiers (which classify possessed nouns according to their properties and mode of possession; Aikhenvald 2000: 187); is this really open, or not? However, I don't think that this problem needs to be solved immediately.

4 Cited in Aikhenvald (2000: 82).

4. Adjectives

Adjectives are the most questionable of the three categories we are considering. Even ‘traditional’ adjectives (as recognised by the grammatical traditions of the languages in question) are sometimes difficult to distinguish from either nouns, verbs or both. In some languages, they arguably not distinct at all. And our conclusion will be that they are not a universal category, at least in the sense that nouns and verbs are. Familiar European languages clearly have an open ‘adjectival’ part of speech distinct from noun (arguably, in some cases, according to our criteria, two of them, as we shall see). In English, for example, they inflect differently from both nouns and verbs (showing neither tense nor number, but having limited comparative), and appear in a special position in the noun phrase (NP), between determiners and nouns. But ‘adjective’ fails as a universal category under our criteria, for a number of reasons, including lack of a reliable semantic core, failure in some languages to differentiate clearly from other categories such as noun and verb, and failure to have a consistent treatment in terms of principles of ordering.

The absence of a semantic core was in effect observed in the classic paper Dixon (1977), who discussed languages in which the apparent adjective class had as few as eight members (Igbo) or even seven (Mulluk-Mulluk) or, possibly, none at all (Samoan). Although the properties expressed by the closed adjective classes are similar (predominantly size and dimension, newness, colour, and quality), the only ones that actually appear on all the lists of supposed adjectives are ‘big/large’ and ‘small/little’. The semantic core would have to be ‘size’, but this is not enough to be useful, because a semantic core needs to contain a reasonable number of things in order for unsplitability to be an interesting property.

Then there are the various languages that have been claimed to have no adjective class at all. Warlpiri has, for example, been proposed to be a language where putative adjectives are not distinct from nouns (with demonstratives also included in this category, which is sometimes called ‘nominals’). One basis for this claim is that there appear to be no inflectional differences, and while there are differences in ordering tendencies between words translated into English as nouns vs adjectives, they are apparently not very strong. Some additional motivation is provided by the claimed interpretations of example (7), due ultimately to Ken Hale (Simpson 1991: 265):

- (7) *kurdu-ngku wita-ngku ka Wajilipi-nyi*
 child-ERG small-ERG PRES chase-NONPAST

‘The small child is chasing it.’

‘The childish small thing is chasing it.’

In the first, more standard-looking, gloss, ‘child’ plays the role of noun (in Simpson’s LFG analysis, supplying the PRED-value for the NP), ‘small’ that of the modifier (within the ADJUNCT attribute), while the second gloss has the roles reversed. But this is a rather subtle intuition due to Ken Hale (1973), and perhaps it is time to re-examine the situation. Consider for example the following two variants of (8):

- (8) a. *Kurdungku ka wajilipinyi*
 b. *Witangku ka wajilipinyi*

Can (8a) really mean ‘the childish thing’ is chasing it, as well as ‘the child is chasing it’? And can (8b) simply mean ‘the small thing is chasing it’, or ‘something small is chasing it’, or does it rather require a contextually determined understanding of what kind of thing is being described as small (child? dog? drone?), as would be the case for an English sentence such as ‘a/the small one is chasing it’? At this point, we don’t know, but a closer look seems warranted.

Stative verbs are the other category from which adjectives sometimes seem to not be distinct. As examples of this, Dixon cites Yurok and Samoan, neither refuted, and the latter supported by subsequent work (Mosel and Hovdhaugen 1992: 73–74). Baker (2003) discusses these and additional cases in some detail, concluding that the languages do have adjectives after all, but this conclusion can be challenged on the basis that he doesn’t provide any criteria distinguishing the putative adjectives from stative verbs. We might consider postulating that languages can lack stative verbs, using only adjectives for this purpose, but as far as I know, whenever there is a clear distinction between adjectives and verbs, there are also some stative verbs alongside the adjectives, often with no very clear difference in meaning. For example, in Greek, we have the adjective *kurazménos* ‘tired’, which we can combine with the copula *íme* ‘I am’ to say *íme kurazménos* ‘I am tired’, as well as the verb *kurázome* ‘I am tired’. If the verb category always spills over to include some stative verbs when it is easily distinguishable from ‘adjectives’, then it is probably theoretically better to say that we have stative verbs but not adjectives when such a distinction does not appear to exist.

A case that might be of this kind that has attracted substantial discussion is the Austronesian language Chamorro (Chung 2012a, 2012b). Chung argues that Chamorro has standard noun, verb and adjective categories, disagreeing with Topping (1973), who proposes that there are only two major categories, noun and verb, with ‘adjectives’ being (stative) verbs. Chung discusses a modification construction that putative adjectives can appear in but not verbs, but does not provide any worked out contrast between ‘adjectives’ on the one hand, and stative verbs on the other (although morphologically transitive but stative *tungu* ‘know’ appears as a verb (Chung 2012a: 14)). Topping (1973: 231–32) also notes that intransitive ‘verbs’ but not ‘adjectives’ can take manner adverbs, while the latter but not the former can take intensifiers. But the predicates shown as taking the manner adverbs are all non-stative, and those taking intensifiers stative; therefore, a distinction between stative verbs and adjectives is not established.

On the other hand, the nature of the modification construction that Chung discusses suggests that there is a structural role of ‘modifier’, distinct from relative clauses, and not available to non-stative verbs. This suggests that there is something real behind the label ‘adjective’, even if it does not meet the requirements to be a label for a universal part of speech in the present sense.

The opposite problem for a universal adjective category is too many arguably distinct word classes that are candidates for the title of ‘adjective’. The classic case is Japanese, with its ‘noun-like’ and ‘verb-like’ adjectives, but there is by now a fairly robust consensus that the differences between these are essentially a matter of morphological realisation of the same inflectional categories; furthermore, the positional criteria appear to be the same. Therefore, the two types can be regarded as the same part of speech. For further discussion, see Yamakido (2005).

But in Romance languages especially, and to a limited extent in English and Greek, the positional criteria don’t fully work. As is well known, in Romance languages, adjectives usually follow the head N, but there are some that can only precede, others that only follow, and many that can appear in either position, often with differences in meaning, including restrictive (postnominal) vs non-restrictive (prenominal), and various kinds of more idiosyncratic shifts, as discussed for example by Demonte (2008):

- (9) *El hombre pobre*
 the man poor
 ‘The poor [not rich] man.’ (Demonte 2008: 80)

- (10) *El pobre hombre*
 the poor man
 ‘The poor man.’ (who evokes pity (cf. Demonte 2008: 76))

Nationality adjectives, on the other hand, appear only to follow (*empresarios americanos*/**americanos empresarios* ‘American businessmen’), while modal adjectives can only precede (*el presunto asesino*/**el asesino presunto* ‘the alleged murderer’).

On the basis of positional criteria, the implication is that these two kinds of ‘adjectives’ are different parts of speech. This conclusion is strengthened by the fact that they differ (as in English) in the kinds of complements they can take. Postnominal adjectives can take prepositional phrase (PP) complements, while prenominal ones cannot:

- (11) *las condiciones buenas para nosotros prolongarán nuestra vida*
 the conditions good for us will prolong our life
 ‘Conditions good for us will prolong our lives.’ (this example found online has been lost, but searching for ‘condiciones buenas para’ will turn up more like it)

- (12) **las buenas para nosotros condiciones*

This cannot be due to some kind of universal restriction on adjective phrase structure inside NP, because in Greek, prenominal adjectives can take PP complements (and, according to my observations, do, at a rate of about 1 per 5,000 words of newspaper text⁵), and even sometimes in spoken TV news. This is possible even for modal adjectives, which, as in English, cannot be used as predicate adjectives:

5 Based on a bit more than 50,000 words’ worth of articles from the online edition of *To Βήμα* (www.tovima.gr/), selected on the basis of being the leftmost three on the banner that used to appear at the top of the webpage, or fewer if the articles were long, on multiple days chosen on no systematic basis.

- (13) *i ipotitheméni apo ton Iosíf apistía tis Mariás*
 the suspected by the Joseph infidelity of(GEN) Mary(GEN)
 ‘Mary’s suspected/alleged infidelity according to Joseph.’ (Available at: www.sostis.gr/blog/item/490-h-gennhsh-sta-evaggelia.)

English and Greek lack a distinction between postnominal and prenominal adjectives, but they have some comparable phenomena, concerning adjectives which occur prenominally but not as predicate adjectives, as well as, in Greek, in the ‘polydefinite’ construction (Alexiadou 2001: 233). On the other hand, the traditional adjectives in Greek and Spanish have the same kind of agreement morphology, marking the nominal inflectional categories.

All in all, ‘adjectives’ make a rather poor showing as a universal category, at least by comparison with nouns:

- (14) a. No semantic core.
 b. Sometimes fall into multiple types by morphological criteria emphasising form.
 c. Sometimes fall into multiple types under position-based criteria.
 d. Sometimes putative adjectives are not clearly distinct from verbs or nouns

A property that does seem to unify different kinds of adjectives to some degree is degree modification, sometimes by specific words. In Spanish, for example, *muy* ‘very’ expresses extreme positive degree for both prenominal and postnominal adjectives, but not verbs, which use *mucho*. But adverbs also take *muy*, so if we want to use this as a criterion, we should be prepared to accept adverbs and adjectives as a single part of speech. This is in fact advocated by Baker (2003), but is not universally accepted, and arguably not a very useful position for people interested in the primary description of languages. Furthermore, there are many traditional adjectives that reject degree modification.

The overall conclusion is that ‘adjective’ is probably not best regarded as a universally present or even definable part of speech, at least by the general approach used here. On the other hand, it is plausible that there is a universal ‘modification’ construction, so that if there is a part of speech to which all of its fillers belong, we could call that ‘adjective’. But the distinctions between different kinds of adjectives in Spanish and Greek

above indicate that there might be more than one kind of ‘modification’ construction, with different kinds of fillers. This conclusion is supported by a great deal of additional work on adjectival constructions, going back to Siegel (1976), recently surveyed by Pfaff (2015), in the course of an extremely detailed treatment of adjectival modification in Icelandic.

5. Verbs

Perhaps surprisingly, verbs do not submit to the kind of treatment that works for nouns. Suppose we try something like (15), where ‘...’ represents some kinds of things that happen that are only expressible by verbs:

- (15) There is a part of speech into which a word falls if it says that something like this happens: ...

Unfortunately, there doesn’t seem to be any such part of speech. The problem is primarily presented by what are often called ‘complex verb constructions’, which consist of two words, one, traditionally called a so-called ‘light verb’, that usually give some kind of generic information about the event, but sometimes very little, and another word, sometimes called a ‘coverb’, often an apparent noun, adjective, which provides more detailed information about the event. Constructions similar to these occur in most languages. In English, for example, we have a choice between (16a) and (16b):

- (16) a. I went to the liquor store
 b. I had/made/did a trip/run/dash to the liquor store

In (16a) we have a conventional verb that indeed says what kind of thing happens, but in (16b), the same information is conveyed by one of several possible nouns, with at least three possible choices of verb, none of which provides significant information about the nature of the event, which comes from the nouns. In some languages making extensive use of such constructions, the traditional verb often makes a more substantial contribution to saying what kind of thing happened, but the noun usually says more.

In English, this kind of construction can be regarded as a minor pattern, but in Hindi (Mohanani 1994), Urdu (Butt 1995), Japanese (Matsumoto 1996) and many other languages, they are more prominent, while in certain

other languages such as Kalam (Papua New Guinea, Pawley 2006) and Jaminjung (Australia, Schultze-Berndt 2002) they are the major technique for describing events. In fact, in both of these languages, the ‘verb’ category is closed, containing 125 members in Kalam, 25 in Jaminjung. A further difference is that in English, the ‘coverb’ is clearly a noun heading an NP, while in the latter two languages, its status as a part of speech is unclear, and it is very clear that it is not head of an NP. This is because of the extremely limited possibilities for modifying or otherwise extending the coverb, which are limited to certain morphological operations that can be regarded as derivational. (17) from Jaminjung illustrates a simple coverb, (18) a reduplicated one, reduplication of the coverb being ‘very frequent—almost obligatory—with coverbs representing repetitive events’ (Schultze-Berndt 2002: 77):

- (17) *yalumbarra* [V *marrug ga-jga-ny*] *yarrajgu, warnda-bina*
 King.Brown hidden 3SG-GO-PST afraid grass-ALL
 ‘The King Brown snake went into hiding—being afraid—into the grass.’ (Schultze-Berndt 2002: 120)

- (18) *mang-mang* *ganu-ngg-m*
 RDP-move.knees.outward 3SG:3SG-SAY/DO-PRS
 ‘She is moving her knees in and out, dancing.’ (Schultze-Berndt 2002: 77)

The normal order is coverb first and then light verb, although the opposite is possible, similarly with verbs and preverbs in Warlpiri.

Wierzbicka (2010: 294–95) suggests defining verbs via the prototypes provided by the exponents of DO and HAPPEN, and while this seems viable, it does not guarantee the existence of more than two verbs, and I think that it is useful to try to define an open class making a stronger claim. To enable this, I propose extending the notion of parts of speech to include what can be called ‘complex words’, which I will define as words that have a certain amount of internal structure, but arguably sit under a single lexical node, which we can call ‘V’ for combinations meeting a workable definition of ‘verb’. To cover Jaminjung, this internal structure must be sufficient to allow clitics to intervene, but not major categories (Schultze-Berndt 2002: 120). Such ‘semipermeable V’ constructions are also found with Warlpiri preverb–verb combinations (Nash 1986: 51–52) and serial verbs in Tariana (Aikhenvald 2000), languages where the (traditional) verb

category is open. I suggest they would also be plausible for languages such as Romance ones and Standard Modern Greek, where object clitics attach to the verb, but show signs of syntactic autonomy (Preminger (2019) and previous work cited there).

With this adjustment, which is intriguingly not necessary for nouns, we can get a universal open category of verbs, corresponding reasonably well to what is found in linguistic descriptions, with a two-step definition, assuming a prior definition of noun:

- (19) a. There is a single open, possibly complex, part of speech, different from ‘noun’, into which words saying what kind of thing happened can fall.
- b. A word saying what kind of thing happened is in this part of speech if it is not in ‘noun’. We can call it ‘verb’.

This seems complex, but I think that the complexity is necessary, because it is clear that on the one hand nouns can provide sufficiently rich descriptions of events in order to count as ‘saying what kind of thing happened’, whereas other widely recognised parts of speech (adjectives, adverbs, prepositions) do not (setting aside some rather complex issues involving ‘mixed categories’, as discussed recently by Panagiotidis (2015)). The aspect of this where NSM is most critically relevant is ‘what kind of thing happened’, which includes things that people do.

Not so clear is the status of the parts of complex verbs, often ‘coverbs’. These are very diverse, and sometimes regarded as a heterogeneous collection of nouns, adjectives and adverbs, and the question of what part of speech they belong to is generally regarded as difficult (e.g. Pawley 2006: 10). Schultze-Berndt (2002: 71–76) argues that they should be taken as a distinct part of speech, distinct from adverbs and nominals. Another possibility is that they might be regarded as essentially morphology rather than syntax, and have no category label at all. Adapting a somewhat similar proposal by Panagiotidis (2015: 18) within the framework of Distributed Morphology for Greek, we might propose this as the structure for ‘dive’ in Jaminjung (Schultze-Berndt 2002: 475):

- (20) [*V buwu* [*V irdbaj*]]
 Enter.water FALL
 ‘dive’

We want the second component as well as the whole thing to be a V because that is what the cross-reference and TAM (tense–aspect–mood) marking prefixes apply to, but, according to this idea, the first component has no category label at all. Regardless of the issue of what, if any, label coverbs have, it is clear that they do not project higher phrasal structure. So if they have a syntactic category label, it is a non-projecting one in the sense of Toivonen (2001).

A further complexity arises in certain languages such as the Germanic ones, whereby additional words such as idiomatically interpreted prepositions and particles may combine non-compositionally with the verb, such as in the numerous idiomatic and one compositional meaning of ‘put down’, or the combination ‘put up with’. Unlike the components of complex verbs, these ‘supplementary items’ are from closed rather than open classes, and therefore can provisionally be set aside. There is also a question of what percentage of the vocabulary they constitute, both in the lexicon and in actual use.

The main empirical claim made by (19) is that if a word describes what kind of thing happens, it has two choices of part of speech, one being ‘noun’, the other being the additional one that we call ‘(possibly complex) verb’. Comparably to with nouns, we just do not seem to find words meeting the semantic criteria of (19) divided between three rather than just two parts of speech. For states, however, things are different: in many languages states can be expressed by a third part of speech, traditionally called ‘adjective’ (or maybe even by two or more additional parts of speech, depending on how we treat ‘nominal’ vs ‘verbal’ ‘adjectives’ in Japanese). Another interesting fact is that there doesn’t seem to be anything corresponding to nominal gender for verbs: one could imagine a language in which verbs were divided into semi-arbitrary classes calling for different forms of adverbs, but this does not seem to happen.

This proposal is not empirically contradictory to that of Wierzbicka (2010); indeed, it seems plausible that they could be integrated along the lines that the non-complex verbs might always include exponents of DO and HAPPEN. However it can be hard to find these. In Jaminjung, the verb *yu(nggu)* ‘do/say’ seems plausible as an exponent for DO (Schultze-Berndt 2002: 355), but I have not yet been able to identify any exponent for HAPPEN. This of course does not mean that there isn’t any, but indicates that even extensive and careful descriptive semantic work can fail to identify the exponent(s) of a prime if that task is not specifically one of the research goals.

6. Conclusion

The main conclusions are as follows:

- (21) a. There are at least two universally definable and universal parts of speech, which can be called noun (or 'nominal', when there is no contrast with adjectives or demonstratives), and verb.
- b. These have semantic cores that can be stated using NSM, which do not split in ways indicative of a part of speech difference (evidence for different phrase structure rules or a different inflectional system).
- c. There is an asymmetry in that words in the semantic core of 'verb' can be nouns or verbs, but those in the semantic core of 'noun' can only be nouns.
- d. There is a further asymmetry in that we need a notion of 'complex word' to deal with languages with (traditionally so-called) coverb + verb constructions, but no such thing is needed for nouns.
- e. The universality of adjectives is not established, and indeed, seems dubious, although there is plausibly universal 'modification' construction restricted to states.

There are many areas wanting further investigation. Some of the ones closest to the topics I have discussed are:

- (22) a. Can the definition of 'noun' be simplified to a single specification without lists that are essentially disjunctions (such as 'things, living things, or ...'), or widened to include mass nouns and various kinds of collections, or, ideally, both?
- b. Is there after all a distinction between 'adjectives' and 'stative verbs' in languages where this is not obvious from factors such as morphology and the use of copulas?
- c. Given that 'adjectives' are dubious, what about 'adverbs'?
- d. Is there a universal asymmetry between nouns and verbs with respect to the closed class supplements?

Although the existence of at least two universal parts of speech might be taken as evidence that there is an autonomously syntactic Universal Grammar that includes a list of categories, including required and perhaps possible ones, it also seems plausible that the apparently universal parts of speech are aspects of the interface between grammar and meaning, and are essentially cognitive and semantic rather than syntactic.⁶ The difference between kinds of things (living things, etc.), on the one hand, and kinds of things that can happen, on the other, is after all very important in life (for example, the former can be returned to, and often retain traces of what has happened to them, while the latter cannot be returned to, and so cannot retain traces of their history), so it would not be a surprise if this difference injected itself rather forcefully into the structure of language.

References

- Acquaviva, Paolo (2009). The roots of nominality, the nominality of roots. *LingBuzz*. Available at: ling.auf.net/lingbuzz/000824.
- Aikhenvald, Alexandra Y. (2000). *Classifiers: A Typology of Noun Classification*. Oxford: Oxford University Press.
- Alexiadou, Artemi (2001). Adjective syntax and noun raising: Word order asymmetries in the DP as the result of adjective distribution. *Studia Linguistica* 55: 217–248.
- Baker, Mark C. (2003). *Lexical Categories: Verbs, Nouns, and Adjectives*. Cambridge: Cambridge University Press. doi.org/10.1017/cbo9780511615047.
- Butt, Miriam (1995). *The Structure of Complex Predicates in Urdu*. Stanford, California: CSLI Publications.
- Chomsky, Noam (1970). Remarks on nominalizations. In R. Jacobs and P. Rosenbaum (eds), *Readings in English Transformational Grammar*. Waltham, Massachusetts: Ginn, 184–221.
- Chung, Sandra (2012a). Are lexical categories universal? The view from Chamorro. *Theoretical Linguistics* 38: 1–56. doi.org/10.1515/tl-2012-0001.

⁶ Note, however, that it is perfectly possible to have a universal grammar (UG) that says nothing at all about specific categories, requiring these to be stated in the actual grammars provided by the theory of UG. The Xerox Linguistic Environment (XLE) is an example of this.

- Chung, Sandra (2012b). Reply to the commentaries. *Theoretical Linguistics* 38: 137–43.
- Craig, Collette Grinewald (1992). *Classifiers in a Functional Perspective*. In R.E. Ascher (ed.), *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press. 565–69.
- Croft, William (2001). *Radical Construction Grammar*. Oxford: Oxford University Press.
- Culicover, Peter (1976). *Syntax*. New York: Academic Press.
- Demonte, Violeta (2008). Meaning–form correlations and adjective positions in Spanish. In Louise McNally and Christopher Kennedy (eds), *Adjectives and Adverbs*. Oxford: Oxford University Press, 71–100.
- Dixon, R.M.W. (1977). Where have all the adjectives gone? *Studies in Language* 1: 19–80. doi.org/10.1075/sl.1.1.04dix
- Dryer, Matthew S. (1997). Are grammatical relations universal? In Joan Bybee, John Haiman and Sandra Thompson (eds), *Essays on Language Function and Language Type*. Amsterdam: John Benjamins, 115–43.
- Evans, Nicholas and Toshiki Osada (2005). Mundari: The myth of a language without word classes. *Linguistic Typology* 9: 351–90. doi.org/10.1515/lity.2005.9.3.351.
- Geach, Peter (1962). *Reference and Generality*. Ithaca, New York: Cornell University Press.
- Goddard, Cliff (2009). A piece of cheese, a grain of sand: The semantics of mass nouns and unitizers. In Francis Jeffrey Pelletier (ed.), *Kinds, Things and Stuff*. New York and Oxford: Oxford University Press, 132–65. doi.org/10.1093/acprof:oso/9780195382891.003.0008.
- Goddard, Cliff (2011). The lexical semantics of ‘language’ (with special reference to ‘words’). *Language Sciences* 33: 40–57. doi.org/10.1016/j.langsci.2010.03.003.
- Goddard, Cliff (2017). Furniture, Vegetables, Weapons: Functional collective superordinates in the English lexicon. In Zhengdao Ye (ed.), *The Semantics of Nouns*. Oxford: Oxford University Press, 246–81. doi.org/10.1093/oso/9780198736721.003.0010.
- Gross, Maurice (1979). On the failure of generative grammar. *Language* 55: 859–85.
- Gupta, Anil (1980). *The Logic of Common Nouns*. New Haven, Connecticut: Yale University Press.

- Hale, Kenneth L. (1973). Person marking in Warlpiri. In Stephen Anderson and Paul Kiparsky (eds), *A Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston, 308–44.
- Haspelmath, Martin (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86: 663–87. doi.org/10.1353/lan.2010.0021.
- Haspelmath, Martin (2012). Escaping ethnocentrism in the study of word-class universals. *Theoretical Linguistics* 38: 91–102. doi.org/10.1515/tl-2012-0004.
- Lyons, John (1968). *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- Matsumoto, Yo (1996). *Complex Predicates in Japanese*. Palo Alto, California: CSLI Publications.
- Mohanan, Tara (1994). *Argument Structure in Hindi*. Stanford, California: Center for the Study of Language and Information.
- Mosel, Ulrike and Even Hovdhaugen (1992). *Samoan Reference Grammar*. Oslo: Scandinavian University Press.
- Nash, David G. (1986). *Topics in Warlpiri Grammar*. New York: Garland Press.
- Panagiotidis, Phoivos (2015). *Categorical Features*. Oxford: Oxford University Press.
- Pawley, Andrew (2006). Where have all the verbs gone: Remarks on the organisation of languages with small, closed verb classes. Paper presented at *Eleventh Biennial Symposium on Intertheoretical Approaches to Complex Verb Constructions*. Houston, Texas: Rice University.
- Pfaff, Alexander Peter (2015). Adjectival and Genitival Modification in Definite Noun Phrases in Icelandic. Unpublished PhD dissertation, University of Tromsø.
- Preminger, Omer (2019). What the PCC tells us about ‘abstract’ agreement, head movement, and locality. *Glossa* 13: 1–42. doi.org/10.5334/gjgl.315.
- Schachter, Paul and Timothy Shopen (2007). Parts-of-speech systems. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*. Cambridge: Cambridge University Press, 1–60.
- Schultze-Berndt, Eva (2000). *Simple and Complex Verbs in Jaminjung: A Study of Event Categorization in an Australian Language*. MPI Series in Psycholinguistics. Neimegen: MPI.

- Siegel, Muffy (1976). Capturing the Russian adjective. In Barbara Partee (ed.), *Montague Grammar*. New York: Academic Press, 293–309. doi.org/10.1016/b978-0-12-545850-4.50016-9.
- Simpson, Jane H. (1991). *Warlpiri Morpho-Syntax*. Boston: Kluwer Academic.
- Toivonen, Ida (2001). The Syntax of Non-Projecting Words. Unpublished PhD dissertation, Stanford University.
- Topping, Donald M. (1973). *Chamorro Reference Grammar*. Honolulu: University of Hawai'i Press.
- Wierzbicka, Anna (1986). What's in a noun? (Or: How do nouns differ in meaning from adjectives?). *Studies in Language* 10 (2): 353–89. doi.org/10.1075/sl.10.2.05wie.
- Wierzbicka, Anna (2010). Lexical prototypes as a universal basis for cross-linguistic identification of parts of speech. In Petra M. Vogel and Bernard Comrie (eds), *Approaches to the Typology of Parts of Speech*. Berlin/New York: Mouton de Gruyter, 285–317. doi.org/10.1515/9783110806120.285.
- Yamakido, Hiroko (2005). The Nature of Adjectival Inflection in Japanese. Unpublished PhD thesis, Stony Brook University.

This text is taken from *Meaning, Life and Culture: In conversation with Anna Wierzbicka*, edited by Helen Bromhead and Zhengdao Ye, published 2020 by ANU Press, The Australian National University, Canberra, Australia.

doi.org/10.22459/MLC.2020.18