1   Parts-of-speech systems

Paul Schachter and Timothy Shopen[†]

0   Introduction

Parts of speech is the traditional term for the major classes of words that are grammatically distinguished in a language. While all languages make parts-of-speech distinctions, there are rather striking differences between languages with regard to both the kind and the number of such distinctions that they make. A field worker investigating an unfamiliar language may therefore find it useful to know what generalizations can be made about parts-of-speech systems. What, for example, can be said about the ways in which, and the limits within which, parts-of-speech inventories may differ from one another? Which parts-of-speech distinctions are universal and which language-specific? What are the ways in which languages that lack a particular part of speech express the semantic equivalent? And what relations are there between the parts-of-speech system of a language and the language's other typological characteristics? It is the aim of this chapter to provide some answers to such questions.

By way of orientation, the present section sets forth some general assumptions that underlie the presentation in the rest of the chapter. First, then, it is assumed here that the primary criteria for parts-of-speech classification are grammatical, not semantic. As has been amply demonstrated in the linguistic literature (see, for example, Fries (1952)), the familiar notional parts-of-speech definitions, such as 'a noun is the name of a person, place, or thing', fail to provide an adequate basis for parts-of-speech classification, since there are many cases in which their applicability or inapplicability is unclear. Grammatical criteria, on the other hand, are not open to this objection.

The grammatical properties of a word that are here taken to be relevant to its parts-of-speech classification include the word's distribution, its range of

1

syntactic functions, and the morphological or syntactic categories for which it is specifiable. Consider, in this connection, the three words of the sentence:

(1)   Boys like girls

The words *boys* and *like* can be shown to differ in their distributions (**Like boys girls* is ungrammatical), in their functional range (*boys* can function as a subject but *like* cannot) and in their categorizations *(boys* is categorized for number but not for tense, while *like* is categorized for both). Thus these two words are assigned to distinct parts-of-speech classes. On the other hand, the words *boys* and *girls*, having highly similar distributions (cf. *Girls like boys*), functional ranges, and categorizations, are assigned to the same parts-of-speech class. There are, to be sure, cases that are less clearcut than these – cases, for example, involving *partial* similarities of distribution, functional range, or categorization, which may require dividing a parts-of-speech class into subclasses. (For some further discussion, see section 1.) But, by and large, the grammatical properties in question constitute a serviceable basis for parts-of-speech classification.

While it is assumed here that the assignment of words to parts-of-speech classes is based on properties that are grammatical rather than semantic, and often language-particular rather than universal, it is also assumed that the *name* that is chosen for a particular parts-of-speech class in a language may appropriately reflect universal semantic considerations. Thus, although the familiar notional definition of nouns mentioned above does not always provide an adequate basis for deciding whether or not a given word is a noun, once the words of a language have been assigned to parts-of-speech classes on grammatical grounds and it is found that one of these classes includes the preponderance of words that are the names of persons, places, and things, then it is perfectly reasonable to call this class the class of nouns, and to compare the class so named with the similarly named classes of other languages. (On this point, see Lyons (1968:317–19).) Thus the words *boys* and *girls* are assigned to the same parts-of-speech class, and the word *like* to a different class, on language-particular grammatical grounds, but it is on universal semantic grounds that the class to which *boys* and *girls* are assigned is called the class of nouns, while that to which *like* is assigned is called the class of verbs.

An interesting recent proposal concerning universal semantic grounds for the identification of parts of speech is to be found in Wierzbicka (2000): namely, the use of universal exemplars, basic words that are presumably found in all languages, such as the equivalents of *person* and *thing* for nouns, *do* and *happen* for verbs. In any language, Wierzbicka suggests, the parts of speech that have been established on grammatical grounds that contain translations of these words can be said to be nouns and verbs respectively. She goes on to develop

this approach for a full range of parts of speech, arguing for the approach's
superiority to other 'prototype approaches', which do not, she says, contain
exemplars that are found in all languages. Notable exemplars are found in
Dixon (1995), as well as Lyons (1977:vol. II), Croft (1984), Givón (1984a) and
Hopper and Thompson (1984).

Another assumption reflected in this chapter is that all languages make a
distinction between *open* and *closed* parts-of-speech classes. Following Robins
(1964:230), we can describe open classes as those 'whose membership is in
principle unlimited, varying from time to time and between one speaker and
another' and closed classes as those that 'contain a fixed and usually small
number of member words, which are [essentially] the same for all the speakers
of the language, or the dialect'. Thus open classes are classes such as nouns
and verbs, and closed classes are classes such as pronouns and conjunctions.

That all languages contain open classes is beyond doubt, despite occasional
apocryphal reports to the contrary: i.e., reports of languages whose vocabu-
laries consist of only a few hundred words. A more serious question can be
raised about the universal status of closed classes. It is certainly true that closed
classes play a rather minor role in some languages, and it has in fact some-
times been claimed that there are languages in which they play no role at all.
The languages in question are invariably so-called *synthetic* languages: that is,
languages that favour morphologically complex words. (Synthetic languages
are commonly contrasted with analytic languages, in which words consisting
of a single morpheme are the norm. If a scale were established, ranging from
highly synthetic languages, such as Eskimo, to highly analytic ones, such as
Vietnamese, modern English would be somewhat closer to the analytic than
to the synthetic end of this scale.) The relation between a language's position
on the synthetic–analytic scale and the role of closed classes in that language
is discussed more fully in section 2. That section also considers, and rejects,
the claim that there are known instances of languages with no closed classes
at all.

The distinction between open and closed parts-of-speech classes provides the
basic organizing principle of the remainder of this chapter, with open classes
being dealt with in section 1 and closed classes in section 2.

1   Open classes

The open parts-of-speech classes that may occur in a language are the classes
of *nouns*, *verbs*, *adjectives*, and *adverbs*. Typically, each of these classes may
be divided into a number of subclasses on the basis of certain distinctive gram-
matical properties. For example, the class of nouns in English may be divided
into such subclasses as common and proper (on the basis of whether or not the

nouns occur with articles like *the*: *the girl* vs **the Mary*), count and mass (on the basis of whether or not they occur in the plural: *chairs* vs **furnitures*), etc. And the class of English verbs may be divided into such subclasses as transitive and intransitive (on the basis of occurrence with objects: *enjoy it* vs **smile it*), active and stative (on the basis of occurrence in the progressive: *is studying* vs **is knowing*), etc. Such subclasses are not ordinarily identified as distinct parts of speech, since there are in fact properties common to the members of the different subclasses, and since the label *parts of speech* is, as noted earlier, traditionally reserved for 'major classes'. In any case, the discussion of open parts-of-speech classes in this chapter does not include a systematic account of the subclassification of these classes, but instead offers only a few observations concerning subclasses that are particularly widespread, or that seem particularly interesting from a typological viewpoint.

It must be acknowledged, however, that there is not always a clear basis for deciding whether two distinguishable open classes of words that occur in a language should be identified as different parts of speech or as subclasses of a single part of speech. The reason for this is that the open parts-of-speech classes must be distinguished from one another on the basis of a *cluster* of properties, none of which by itself can be claimed to be a necessary and sufficient condition for assignment to a particular class. And the fact is that languages vary considerably in the extent to which the properties associated with different open word classes form discrete clusters. Typically there is some overlap, some sharing of properties, as well as some differentiation. In English, for example, although nouns, verbs, and adjectives are clearly distinguished from one another in various ways, there are still certain properties that they share. Thus nouns and adjectives, as well as verbs, may be subclassified as active vs stative on the basis of occurrence in the progressive (compare *John is being a boor / boisterous* and **John is being my brother / tall*). And in certain other languages, as will become clear in the following sections, nouns and verbs, or nouns and adjectives, or verbs and adjectives, may have very much more in common than they do in English. What this means is that there may in some cases be considerable arbitrariness in the identification of two open word classes as distinct parts of speech rather than subclasses of a single part of speech. Thus some rather celebrated questions – for example, whether or not all languages make a distinction between nouns and verbs – may ultimately turn out to be more a matter of terminology than of substance (cf. section 1.2).

In the following presentation of the open parts-of-speech classes, nouns, verbs, adjectives, and adverbs are discussed in turn. In each case, the characteristic grammatical and notional properties of the class are enumerated, with relevant examples. Certain subclasses are also noted, and, where appropriate, there is a discussion of the question of the universality of a particular parts-of-speech distinction (see section 1.2), or of the ways in which languages that

lack a particular distinction express the semantic equivalent (see sections 1.3
and 1.4).

1.1   Nouns

The distinction between *nouns* and *verbs* is one of the few apparently universal
parts-of-speech distinctions. While the universality of even this distinction has
sometimes been questioned, it now seems that the alleged counter-examples
have been based on incomplete data, and that there are no languages that cannot
be said to show a noun–verb distinction when all relevant facts are taken into
account. We shall look further into the matter of languages which allegedly fail
to distinguish nouns and verbs at the end of section 1.2, after the characteristic
properties of these two parts of speech have been described.

For convenience we can adapt the traditional definition of nouns, assigning
the label *noun* to the class of words in which occur the names of most persons,
places, and things. As was explained in the introductory section, this type of
notional correlation is not the basis for determining membership in a class, but
merely the basis for assigning a name to a class established on other grounds. It
is therefore not a matter of concern if the class of nouns includes, as it typically
does, words that are not the names of persons, places, or things, or if some such
names are found in some other class.

It may be useful, however, to try to go beyond such traditional definitions to a
deeper understanding of the semantics. Let us briefly consider in this connection
some proposals made by Ronald W. Langacker (1987) and Anna Wierzbicka
(1986). Langacker, working exclusively with English data, argues for certain
universal semantic properties of nouns and verbs. Nouns, he proposes, do not
foreground relations, but instead designate 'a region in some domain'. Verbs,
on the other hand, do foreground relations. (For more on Langacker's views on
verbs, see section 1.2, below.) Consider, for example, the difference in meaning
between the following sentences:

(2)   The principal is speaking in the next room

(3)   The principal's speech is in the next room

The first sentence, using the verb *speak*, evokes an image of an audience and
the principal communicating with them. The second sentence, using the noun
speech, on the other hand, may simply serve to locate a physical entity; there is
not necessarily any audience or any communication. Of course someone could
read the speech (and the principal himself or herself could in fact read or recite
it aloud to an audience), in which case communication would take place, but
it could also be left unread with no communication. Thus the communication
relation is not foregrounded.

Moreover, the speech in the context of the second sentence is likely to be understood as being in writing, which makes it easier to think of it as a 'region in some domain'. People can, of course, also say

(4)   The principal is giving a speech in the next room

in which case the noun necessarily represents something spoken and there is communication with an audience. In Langacker's scheme, however, this could be said to come from the force of the verbal expression with the verb *give*.

Langacker also makes an interesting suggestion about how the semantics of count and mass nouns can explain their syntactic differences. (Count nouns are nouns whose typical referents are countable and which may therefore be pluralized; mass, or noncount, nouns are nouns whose typical referents are not countable and which may therefore not normally be pluralized.) He suggests that count nouns can take plural inflection because their referents are 'bounded' in space. 'Bounded' means that, whether the count noun refers to a single entity (*dog*, *tree*) or to a set of entities (*crowd*, *herd*), the referent is conceived of as being defined in space. Mass nouns (*milk*, *sincerity*), on the other hand, refer to things that are conceived of as not 'bounded' but instead as having an indeterminate extent in space.

Like Langacker, Wierzbicka contrasts the semantic properties of nouns with those of another part of speech, but in her case this other part of speech is adjectives. Using examples from a variety of languages, she seeks to show how the semantics of nouns and adjectives can account for differences in how they are used. Nouns, she proposes, tend to refer to groupings of the permanent and/or conspicuous characteristics of entities. This is in contrast to adjectives, which tend to refer to a single temporary and/or less conspicuous characteristic. For example, to say *She is a cripple*, using the noun *cripple*, categorizes the person permanently, saying something about what kind of person she is. To say *She is sick*, with the adjective *sick*, on the other hand, says nothing about what kind of person she is, but instead refers to a single characteristic that the person has for the moment.

Because of these semantic differences, Wierzbicka argues, nouns are used for reference and categorization more easily than adjectives, while adjectives are used attributively more easily than nouns. Thus in a phrase such as *a sick woman*, the noun *woman* provides a broad categorization of the referent while the adjective *sick* serves to refine the categorization. By contrast, *a cripple woman*, with the two nouns providing a double categorization, is awkward.

An interesting example of the awkwardness of the attributive use of nouns is the following headline from the *Canberra Times* of 20 October 1999: *Diplomat murder accused granted bail*, which involves a recursive use of noun attributes. *Diplomat murder* means 'murder of a diplomat', *diplomat murder accused*

means 'person accused of murder of a diplomat', so the entire headline means 'bail has been granted to a person accused of the murder of a diplomat'.

The most common function for nouns is as arguments or heads of arguments – for example, as (heads of) subjects or objects, as in the case of the italicized words in:

(5) The little *boy* was eating *candy*

Nouns may also function as predicates, however, either with an accompanying copula, such as English *be* (6) or Hausa *ne* (7), or without any copula, as in Tagalog (8) or Russian (9):

(6) They are *teachers*

(7) Su *malamai* ne
they teachers COP
'They are teachers'

(8) Mga *guro* sila
PL teacher they
'They are teachers'

(9) Oni *učitelja*
they teachers
'They are teachers'

Typical categories for which nouns may be specified, either morphologically or syntactically, are case, number, class or gender, and definiteness. Case marking indicates grammatical functions (such as subject, direct object, and indirect object; cf. Andrews in chapter 3, and Dryer in chapter 4, of this volume for illuminating discussions of these functions), as in the following examples from Latin (10) (in which case is marked morphologically, by suffixation) and Japanese (11) (in which case is marked syntactically, by postpositions).

(10) Femin-a mal-um puell-ae dedit
woman-NOM apple-ACC girl-DAT gave
'The woman gave an apple to the girl'

(11) Onna ga shojo ni ringo o ataeta
woman SUBJ girl DAT apple OBJ gave
'The woman gave an apple to the girl'

Number marking distinguishes singular from plural, and, more rarely, dual, as in English *house/houses*; Eskimo *iglu* 'house' */ iglut* 'houses' */ igluk* 'two houses'; or Tagalog *bahay* 'house' */ mga bahay* 'houses'. Class or gender marking partitions the set of nouns into subsets, each of which has its own distinctive marking and/or necessitates a distinctive marking on certain other words which

show agreement with nouns. Typically, the classification is in part semantically based and in part semantically arbitrary. Examples include the gender systems of Indo-European languages (e.g. German *der Mann* (the-masculine man) 'the man', *die Frau* (the-feminine woman) 'the woman', *das Mädchen* (the-neuter girl) 'the girl'), the class systems of Bantu languages (such as Swahili, in which most nouns that refer to human beings are in class I, which takes the prefix *m-*, e.g. *mtu* 'person', *mtoto* 'child', *mgeni* 'stranger', but in which some of the other classes have little semantic coherence), and the noun-classifier systems of such languages as Thai (cf. section 2.2, below). Some examples of definiteness distinctions are *a man* vs *the man*, Norwegian *en mann* 'a man' vs *mannen* 'the man', and Hebrew *ish* 'a man' vs *ha-ish* 'the man'.

In most languages some grammatical distinction is made between *common nouns*, which are used to refer to any member of a class of persons, etc. (e.g. *girl*, *city*, *novel*), and *proper nouns*, which are used to refer to specific persons, etc. (e.g. *Mary*, *Boston*, *Ivanhoe*). The precise character of the grammatical distinction, however, as well as its precise semantic correlates, may show considerable variation from language to language. For example, while common nouns in English differ from (most) proper nouns by occurring with articles, in Tagalog (which has no articles) common and proper nouns take different case markers and topic markers, as the following examples illustrate:

(12)   Malapit *sa*   babae   *ang* bata
        near   OBLIQ woman TOP child
        'The child is near the woman'

(13)   Malapit *kay*   Maria *si*   Juan
        near   OBLIQ Maria TOP Juan
        'Juan is near Maria'

Moreover, the Tagalog classes that are distinguished on this basis are not semantically coextensive with the English classes of proper and common nouns. The Tagalog nouns that take the markers of (13) are restricted to those that refer to specific *persons*; nouns that refer to specific *places*, etc., take the other set of markers, although their English equivalents are clearly proper, rather than common, nouns:

(14)   Malapit *sa*   Maynila *ang* Pasay City
        near   OBL   Manila   TOP Pasay City
        'Pasay City is near Manila'

Apart from making a distinction between common and proper nouns, languages may make various other kinds of subclass distinctions within the set of nouns: for example, the distinction between count and mass nouns and the gender distinctions mentioned above. As was explained at the start of this

section, however, such subclass distinctions generally go beyond the scope of
this chapter. (See Corbett in vol. III, chapter 4, on gender and noun class.)

1.2   Verbs

Verb is the name given to the parts-of-speech class in which occur most of the
words that express actions, processes, and the like. As in the case of nouns,
Langacker (1987) has proposed a deeper, more general account of the seman-
tics, proposing, as noted above, that verbs, unlike nouns, foreground relations.
In Langacker's scheme, however, the foregrounding of relations is not a unique
property of verbs, since there are certain other parts of speech that can also
foreground relations. What is distinctive about verbs, he suggests, is the fore-
grounding of *temporal* relations (relations that are anchored in time) or of
relations concerned with process. Atemporal relations, on the other hand, are
foregrounded by adpositions, adjectives, and adverbs (as well as by infinitives
and participles, which Langacker does not classify as verbs).

The characteristic function of verbs is as predicates, as in:

(15)   The people *danced*
          The student *solved* the problem

In some languages, however, verbs can also occur as arguments as in the fol-
lowing example from Tagalog:

(16)   Pinanood ko   ang mga *sumasayaw*
          watch   I   TOP PL   were. dancing
          'I watched the ones who were dancing'
          cf. *Sumasayaw*   ang mga tao
              were. dancing   TOP PL   person
              'The people were dancing'

The use of a verb as an argument is to be distinguished from the probably more
common use of a verbal *noun* as an argument, as in Akan:

(17)   Mehwɛɛ   *asaw*   no
          I. watched dancing the
          'I watched the dancing'

The verbal noun is a noun which is morphologically related to a verb, but which
does not itself occur as a verbal predicate. For example, the verbal noun *asaw*
of (17) is related to the verb *saw* 'dance' but could never itself be used as a
predicate.

The categories for which verbs may be specified include tense, aspect, mood,
voice, and polarity. (As in the case of nouns, the categorization may be mani-
fested either morphologically or syntactically. Only morphological illustrations

will be given in this section, however. For some syntactic illustrations, see the
presentation of auxiliaries in section 2.3. See also, for further information on
tense, mood, and aspect, vol. III, chapter 5, by Timberlake, and, for a detailed
treatment of mood, section 4 of that chapter.) Tense marking indicates time
relative to the time of the utterance: for example Haya *akaija* 'he came (earlier
than a few days ago)', *alaizile* 'he came (within the past few days)', *yaija* 'he
came (earlier today)', *alaija* 'he will come (in the near future)', *aliija* 'he will
come (in the distant future)'. Aspect marking indicates whether the action of the
verb is regarded as complete or incomplete, durative or momentaneous, etc.: for
example, Classical Greek *bebouleûsthai* 'to have already decided', *bouleúesthai*
'to be deciding', *bouleúseasthai* 'to decide' (unspecified for completeness or
durativeness). Mood marking involves distinctions such as indicative (actual)
vs subjunctive (possible) or declarative vs interrogative: for example, French
(qu')il viendra '(that) he will come' vs *(qu')il vienne* '(that) he may come';
Menomini *pi·w* 'he is coming, he came' vs *pi·ʔ* 'is he coming?, did he come?'.
Voice marking has to do with the role of the subject in the action expressed
by the verb, the most common voice distinction being active vs passive, as in
Latin *videt* 'he sees', *videtur* 'he is seen'. And polarity marking distinguishes
affirmative from negative, as in Akan *tu* 'pulls', *ntu* 'doesn't pull'. (In addition
to being marked for inherently verbal categorizations, verbs in some languages
are marked to indicate certain categorizations (person, number, class) of their
subjects and, less frequently, their objects: for example, Latin *video* 'I see',
videmus 'we see'; Swahili *wa-ta-ni-uliza* (they-future-I-ask) 'they will ask me',
ni-ta-wa-uliza (I-future-they-ask) 'I will ask them'.)

In all languages it is possible to subclassify verbs as transitive or intransitive
on the basis of whether or not they occur with objects. In some languages the
transitive–intransitive distinction entails certain other grammatical distinctions.
For example, in Bambara the past tense is expressed by an auxiliary (*ye*) with
transitive verbs but by a suffix (*-la*) with intransitive verbs:

(18)   U   *ye*   a   san
         they PAST   it buy
         'They bought it'

(19)   U   boli-*la*
         they walk-PAST
         'They walked'

Many languages also have a subclass of copulative verbs, like English *be*, that
occur with predicate nominals or adjectives. In other languages, however, there
is either no copula at all (as in Tagalog – cf. example (8)) or the copula is
not a verb (as in Hausa – cf. example (7)). (For further discussion of non-verb
copulas, see section 2.5.)