

Cambridge University Press

978-0-521-58158-5 - Language Typology and Syntactic Description: Grammatical Categories and the Lexicon, Second Edition - Volume III

Edited by Timothy Shopen

Excerpt

[More information](#)

1 Typological distinctions in word-formation

Alexandra Y. Aikhenvald

0 Introduction

This chapter deals with patterns of word-formation, their classification and parameters of cross-linguistic variation. Grammatical words (section 1) in most languages have an internal structure; the typological parameters which account for their cross-linguistic variation are discussed in section 2. Word-formation processes correlate with syntax in different ways depending on language type. One such word-formation process – known as ‘the most nearly syntactic of all’ (Mithun (1984)) – is noun incorporation, discussed in section 3.

The structure of words in a language can be more or less iconically motivated (see section 4). Word-formation, traditionally, falls into compounding and derivation. A compound consists of morphemes which could be free (see section 5), while derivation involves the use of different classes of bound morphemes and of morphological processes to form words (see section 6). Word-formation processes vary in terms of their productivity – see section 7. Word-formation processes are prone to distinct patterns of grammaticalization and lexicalization – see section 8. A brief summary is given in section 9, and in section 10 I provide suggestions for field workers describing word-formation in previously undocumented or poorly documented languages.

1 The word

Word-formation accounts for the structured organization of the lexicon. The lexicon is usually conceived of as a list of the form–meaning correspondences conventionalized by speakers, but which are largely arbitrary. However, this list may be structurally organized. The principal function of word-formation is the enrichment of the lexicon by forming new words; for instance, *red* and *reddish* in English are regular derivations based on *red*.

What is a word? ‘Word’ has, for a long time, been recognized as a universal unit by scholars of varied persuasions. The concept of the word is, however, at least twofold. Many languages make a distinction between *phonological* and *grammatical* word (though the majority of grammars do not pay enough

Cambridge University Press

978-0-521-58158-5 - Language Typology and Syntactic Description: Grammatical Categories and the Lexicon, Second Edition - Volume III

Edited by Timothy Shopen

Excerpt

[More information](#)

attention to this distinction: see Dixon (1977, 1988); Foley (1991); S. R. Anderson (1985a)).

A *phonological* word can be defined as a prosodic unit not smaller than a syllable. Cross-linguistic criteria used to distinguish the phonological word include: (i) stress and other prosodic characteristics; (ii) phonotactics, and phonological rules which apply either word-internally or across word boundaries. See further discussion in Dixon and Aikhenvald (2002).

A *grammatical* word consists of a number of grammatical elements which (i) always occur together, rather than scattered through the clause (the criterion of cohesiveness); (ii) occur in fixed order; and (iii) have a conventionalized coherence and meaning (Dixon and Aikhenvald (2002); see also Dixon (1977:88, 1988:21–31); Matthews (1991)). Criterion (iii) relates to both the number of morphemes per word and the expression of grammatical categories which are obligatory for a grammatical word to be well-formed in a given language. In most non-isolating languages (see section 2), a grammatical word must include at least one inflectional morpheme. For instance, in Yidiny it can have only one (Dixon (1977)). In North Arawak languages of South America a grammatical word must contain at least one root morpheme and not more than one prefix. The presence of inflectional morphemes is not obligatory in grammatical words in Kaingang (Gê), which shows a general tendency toward isolating typology (Wiesemann (1972)).

Grammatical and phonological words often, but not always, coincide (e.g. Lehiste (1964); Dixon and Aikhenvald (2002)). Thus, many languages have clitics which constitute grammatical words on their own but must be attached to another grammatical word within one phonological word and thus cannot form a phonological word on their own, e.g. *-n't* as in English *mustn't*.

Further distinctions within the concept of word include word as an orthographic unit (a useful tool for counting the number of words while composing a telegram; however, it is applicable only to languages with an institutionalized writing system) and word as a lexical unit – that is, a unit which can be treated as one entry in a dictionary (see Mugdan (1994:2551)). Lexical units, whose form–meaning association is hardly predictable on the basis of the meaning of their components, are not limited to a list of words only. Often, a combination of words – a phrase, or even a sentence – can be *idiomatic*, or non-compositional. In English, expressions like *she spilt the beans* or *willy-nilly* ought to be included in lexical listings, based on the arbitrariness of lexical information.

In this chapter, we will limit ourselves only to words as *grammatical* units, concentrating on discovering the principles of the internal structure of words and their cross-linguistic variability, rather than on the arbitrariness of the form–meaning correlations. For this reason idiomatic combinations of words will not be discussed any further. Throughout the chapter, when we say ‘word’, we are

Cambridge University Press

978-0-521-58158-5 - Language Typology and Syntactic Description: Grammatical Categories and the Lexicon, Second Edition - Volume III

Edited by Timothy Shopen

Excerpt

[More information](#)

referring to ‘grammatical word’ (see Dixon and Aikhenvald (2002), for further discussion).

2 Morphological typology and word-formation

The traditional parameters used for morphological typology of languages starting from the nineteenth century were largely based on the differences in their internal word structure. These parameters are of two kinds. The first one is based on the transparency of morphological boundaries between the morphemes within a grammatical word, and the second one relates to the degree of internal complexity of words (see E. Sapir (1921)).

2.1 Transparency of word-internal boundaries

Based on this parameter, three types of language are recognized: *isolating*, *agglutinating*, and *fusional*.

An *isolating language* typically has a one-to-one correspondence between a morpheme and a word; that is, in such a language every morpheme is an independent word. An example of an almost perfectly isolating language is Vietnamese, as illustrated in (1) (Thompson (1987:207)).

- (1) Chì ây quên
 s/he ANAPHORIC forget
 ‘She (or he) forgets’, or ‘She (or he) has forgotten’, or
 ‘She (or he) will forget’

Every word in this sentence is invariable. There is no morphological variation for tense, or for grammatical function. Where English grammar would require a reference to time in the verb in every sentence, in speaking Vietnamese one is not required to have this. The time reference is understood from the context; so (1) could also be translated as ‘She (or he) has forgotten’ or as ‘She (or he) will forget’. If time reference is important, a time word or an aspect marker – also a separate word – can be inserted. In (2), an ‘anterior’ aspect marker is used in the same sentence as (1) to indicate that the action of ‘forgetting’ started before the time of the utterance.

- (2) Chì ây đã quên
 s/he ANAPHORIC ANTERIOR forget
 ‘She (or he) forgot’ or ‘She (or he) has forgotten’

It is in general true that every word in Vietnamese consists of just one morpheme; however, the existence of productive compounding and its lexicalization results in the creation of words of more complicated structure, e.g.

Cambridge University Press

978-0-521-58158-5 - Language Typology and Syntactic Description: Grammatical Categories and the Lexicon, Second Edition - Volume III

Edited by Timothy Shopen

Excerpt

[More information](#)

4 Alexandra Y. Aikhenvald

hôm nay (day now) ‘today’, *hôm kia* (day that) ‘day before yesterday’, *hôm kia* (day that; more remote than *kia*) ‘two days before yesterday’.

In an *agglutinating language*, a word may consist of several morphemes but the boundaries between them are clearcut. There is typically a one-to-one correspondence between a morpheme and its meaning, and a morpheme has an invariant shape which makes it easy to identify. Hungarian and Turkish are classic examples. A noun is easily segmentable into a lexical stem, a number affix and a case affix. An extract from the Hungarian noun declension paradigm for *ember* ‘man’ is illustrated below.

	Singular	Plural
Nominative	<i>ember</i>	<i>ember-ek</i>
Accusative	<i>ember-et</i>	<i>ember-ek-et</i>
Dative	<i>ember-nek</i>	<i>ember-ek-nek</i>
Locative	<i>ember-ben</i>	<i>ember-ek-ben</i>

In *fusional* – sometimes misleadingly called (in)flexional – *languages* there is no clear boundary between morphemes, and thus semantically distinct features are usually merged in a single bound form or in closely united bound forms. Extracts from Russian nominal paradigms for *dom* ‘house’ and *koška* ‘cat’ below illustrate this point.

	Declension 1		Declension 2	
	Singular	Plural	Singular	Plural
Nominative	<i>dom</i>	<i>dom-a</i>	<i>košk-a</i>	<i>košk-i</i>
Accusative	<i>dom</i>	<i>dom-a</i>	<i>košk-u</i>	<i>košek</i>
Dative	<i>dom-u</i>	<i>dom-am</i>	<i>košk-e</i>	<i>košk-am</i>
Instrumental	<i>dom-om</i>	<i>dom-ami</i>	<i>košk-oj</i>	<i>košk-ami</i>

An affix like *-ami* cannot be segmented into a marker for number and a marker for case; and in a word like *košek* (‘cats’ accusative plural) the stem itself is fused with case and number. Along similar lines, in Latin the final *-a* of *femina* ‘woman’ expresses the meanings: nominative case, singular number and feminine gender (as well as first declension).

The term *(in)flexional*, sometimes used in place of *fusional*, is misleading: we will see in section 11 that both *fusional* and *agglutinating languages*, as opposed to *isolating languages*, can have *inflectional morphology*.

Fusion and agglutination are best treated as quantitative notions. Even the ‘classic’ *agglutinating languages* such as Turkish or Hungarian may be problematic with respect to the treatment of boundaries and the existence of variants of morphemes (allomorphs). These languages are known for vowel harmony across morphemic boundaries, e.g. Hungarian *ember-ek-ben* (man-PL-LOC) ‘in men’, but *asztal-ok-ban* (table-PL-LOC) ‘in tables’. In addition, Hungarian has a

Cambridge University Press

978-0-521-58158-5 - Language Typology and Syntactic Description: Grammatical Categories and the Lexicon, Second Edition - Volume III

Edited by Timothy Shopen

Excerpt

[More information](#)

certain amount of stem alternation in the formation of plurals (e.g. *szó* ‘word’, pl. *száv-a-k*) (see Hagège (1990) on the tendency of an agglutinating morphology to develop into a fusional, or partly fusional, type). Various phonological processes apply across morpheme boundaries, and, as a consequence, the morpheme boundaries may become blurred, which yields the creation of fusional morphology (see section 6).¹

2.2 *Internal complexity of grammatical words*

The second typological parameter has to do with the number of morphemes per word. This typological dimension is largely complementary to that described in section 2.1.

Analytic languages tend to have a one-to-one correspondence between a word and a morpheme; they have few if any bound morphemes. Vietnamese (1–2 above) or Mandarin Chinese are good examples of analytic languages.

In contrast, in *synthetic* languages a word consists of several morphemes, and there are numerous bound morphemes. Hungarian or Russian are representative of synthetic languages.

Polysynthetic languages (also sometimes called ‘incorporating’: see section 3, on the reasons for distinguishing these terms) are characterized by extreme internal complexity of grammatical words. Here, the bound morphemes often express semantic content reserved for lexemes in languages of other types. Polysynthesis basically refers to the possibility of combining large numbers of morphemes (lexical and grammatical) within one word, as in the following example from West Greenlandic (Fortescue (1994:2602)):

- (3) anigu-ga-ssa-a-junna-a-ngajal-luinnar-simassa-galuar-put
avoid-PASS-PART-FUT-be-no.longer-almost-really-must-however-
3PL.INDIC
‘They must really almost have become unavoidable but . . .’

Interest in polysynthesis has grown considerably since the 1990s, due to an increasing amount of new data from different parts of the world (Foley (1986, 1991); De Reuse (1994); Fortescue (1994); among others). The following traits tend to cluster in polysynthetic languages, although none of them is defining by itself (Fortescue (1994:2601)):

- (1) noun stem incorporation within the verbal complex, and incorporation of adjectival stems within nouns (see section 3);

¹ E. Sapir (1921) suggested a fourth type: *symbolic* languages. These languages utilize internal changes, such as ablaut, vowel and consonant changes, and changes in stress and tone, as a means of marking grammatical contrasts. This type has never been as widely used in typological classification of languages as the others, mainly because these internal changes are also widely used in fusional languages, and it is hard to draw a boundary.

Cambridge University Press

978-0-521-58158-5 - Language Typology and Syntactic Description: Grammatical Categories and the Lexicon, Second Edition - Volume III

Edited by Timothy Shopen

Excerpt

[More information](#)6 *Alexandra Y. Aikhenvald*

- (ii) a large inventory of bound morphemes, together with a limited set of independent stems;
- (iii) derivational processes productive in the formation of individual sentences, the verbal word being a minimal sentence;
- (iv) pronominal cross-referencing of subjects, objects, and sometimes also of other arguments (obliques, or datives) on the verb, and of possessors on nominal forms;
- (v) integration of locational, instrumental and other adverbial elements (manner, etc.) into the verb complex as bound morphemes;
- (vi) many possible affixal 'slots', just a few of them obligatory, within a verbal word.

Concomitant properties of polysynthetic languages include relatively free pragmatic constituent order, possibilities of variable morpheme ordering and head-marking.

Many, but not all, polysynthetic languages have noun incorporation (section 3). Most can have a wide range of recursively occurring affix types (verbalizers, nominalizers, adverbial type 'postverbs') with an extremely large overall stock of affixes (e.g. 400–500 in West Greenlandic, and 200 in Kwakwala). Yet other languages are typified by a large number of affixes attached to different slots only within a verbal complex ('field-affixing': Fortescue (1994:2602)). They can be suffixing (Yupik, or West Greenlandic), or suffixing and prefixing (e.g. Nadëb, from the Makú family; Guahibo languages from Colombia; or North Australian languages).

The combination of these properties is also attested. A combination of incorporation and 'field'-affixing can be illustrated with the structure of the verb complex in Traditional and Modern Tiwi (Osborne (1974); Lee (1987: 152–3)) – see table 1.1. (Modern Tiwi, spoken by the younger generation, has been simplified within a contact situation: Lee (1987:155–6).)

Example (4) shows a chain of prefixes in Traditional Tiwi. All these prefixes are said to be obligatorily used.

- (4) warta a-watu-wuji-ngi-mangi-rr-akupuraji yiripuwarta
 bush 3SG.MASC-morning-CONT-CV-water-CV-fall high.tide
 'The high tide is falling [literally 'water-falling'] [exposing the]
 land (bush)' (Jennifer Lee, p.c.)

Historically, polysynthetic morphology often arises from the combination and subsequent grammaticalization of independent roots. Thus, Fortescue (1992) suggests 'that contemporary Eskimo languages may have developed their complex morphophonemic patterns from a more agglutinative pre-Proto-Eskimo stage' (cf. also Foley (1997) for Yimas; Aikhenvald (2003) for Tariana).

Since polysynthetic structures are most often found in head-marking languages, Nichols (1986) suggested that there are no polysynthetic nouns.

Cambridge University Press

978-0-521-58158-5 - Language Typology and Syntactic Description: Grammatical Categories and the Lexicon, Second Edition - Volume III

Edited by Timothy Shopen

Excerpt

[More information](#)

Typological distinctions in word-formation

7

Table 1.1 *Morpheme slots in Tiwi verb (Lee (1987:152–5))*

Traditional Tiwi	Modern Tiwi
1. Subject	yes
2. Tense: past, non-past	yes
3. Locative: distant, directional, distant in time	yes
4. Mood 1: subjunctive, frustrative	yes
5. Mood 2: irrealis	yes
6. Temporal 1: 'in the morning'	no
7. Direct object or indirect object	no
8. Aspect 1: durative or non-past habitual, inceptive, common activity	yes
9. Stance: away from camp, or distant in time; walking along	no
10. Emphatic	yes
11. Connective	yes
12. Temporal 2: 'in the evening'	no
13. Concomitative	no
14. \pm 1 or 2 incorporated forms	no
15. Verbal root	yes
16. Voice: causative, completive, reflexive, reciprocal	yes
17. Aspect 2: movement; 'on the way'	yes
18. Aspect 3: repetitive, past habitual	yes
19. Locative	no

However, nouns in some Australian languages (Dench and Evans (1988)) and in some languages from South America (Aikhenvald (1999c)) have been shown to be inflectionally polysynthetic, since they have multiple marking of grammatical function known as 'double case' (see also Plank (1995)).

The distinction between analytic and synthetic languages is a continuum rather than a dichotomy, since languages display different degrees of synthesis. The degree of synthesis or analysis in a given language can be calculated, for instance, by dividing the number of morphemes in a sentence by the number of words. Some languages are considered more synthetic than others. Linguists often talk about 'mildly' polysynthetic languages. This is reflected in the approach of Greenberg (1954) who suggested the use of a quantitative index, M(orpheme) per W(ord) to calculate the degree of synthesis in a language. See Comrie (1981a:44–5) for further discussion of problems which arise there.

Languages which can be considered almost entirely analytic are the isolating languages of Southeast Asia – e.g. Mandarin Chinese, Classical Chinese and Vietnamese – and of West Africa – e.g. Igbo. The languages of Europe, Asia and North Africa are predominantly synthetic, while polysynthetic languages are concentrated in North and South America, in Siberia, in the north of Australia and in some parts of Papua New Guinea (Foley (1986)).

Cambridge University Press

978-0-521-58158-5 - Language Typology and Syntactic Description: Grammatical Categories and the Lexicon, Second Edition - Volume III

Edited by Timothy Shopen

Excerpt

[More information](#)8 *Alexandra Y. Aikhenvald*2.3 *Integrating the two parameters*

The degree of synthesis and the treatment of morphological boundaries are relatively independent typological parameters. For a description of a previously undocumented language, it is not enough to say that it is ‘analytic’, or that it is ‘isolating’. It is true that isolating languages tend to be analytic, but the reverse would be wrong: English, which has some fusional morphology, makes extensive use of analytic constructions.

Polysynthetic languages are often agglutinative in that the morpheme boundaries are clearcut, and there is little allomorphic variation. However, some polysynthetic languages do have elements of fusion. For instance, Greenlandic has a well-developed array of fused portmanteau inflections with a great morphophonemic complexity – see Fortescue (1992). The fusion of morphemes in a polysynthetic language is illustrated by (5), from Chiricahua Apache, an Athabascan language (Hoijer (1945:15)). Fused morphemes are underlined.

- (5) hà-n-ʔàh
 out.of-2SUBJ+IMPF-handle.a.round.object+IMPF
 ‘you take a round object (out of enclosed space)’

The degree of morpheme fusion and of synthesis have to be defined independently of one another. Figure 1.1 illustrates how the two can be plotted together. Examples of languages are given underneath the diagram.

techniques of joining morphemes

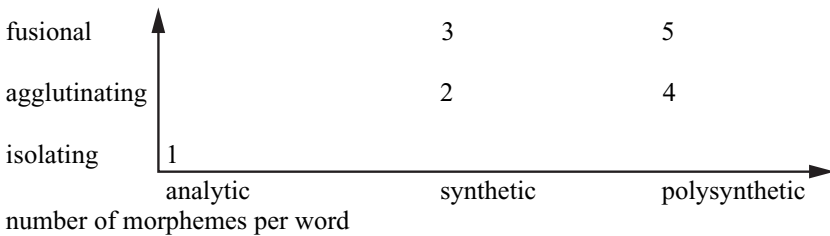


Figure 1.1 Interaction of two types of parameters in word-formation.

- (1) Vietnamese and Classical Chinese are typical examples of isolating analytic languages.
- (2) Hungarian is a typical agglutinating synthetic language.
- (3) Russian is a fusional synthetic language.
- (4) Yupik Eskimo is a polysynthetic agglutinating language.
- (5) Chiricahua Apache is a polysynthetic fusional language.

Cambridge University Press

978-0-521-58158-5 - Language Typology and Syntactic Description: Grammatical Categories and the Lexicon, Second Edition - Volume III

Edited by Timothy Shopen

Excerpt

[More information](#)2.4 *Word-formation and syntax in languages of different types*

The two sets of parameters illustrated in Figure 1.1 correlate with other properties. Isolating analytic languages tend not to have obligatory grammatical categories ordinarily shown in fusional or agglutinating languages, such as tense and case or agreement in gender or number (see examples (1–2) from Vietnamese).

As we will see in the following sections, compounding is widespread in isolating languages, while derivation is a property of languages of other types; this follows from the tendency to have a one-to-one correspondence between a morpheme and a word in isolating languages.

Analytic languages employ periphrastic constructions in syntax whereas synthetic languages tend to express similar meanings within an individual word by means of its affixes.

In Japanese, a synthetic language, passive – whereby the object of a transitive verb becomes the subject of an intransitivized verb and the original subject of the erstwhile transitive verb gets demoted – is expressed with an affix, as in (7). Example (6) is the underlying transitive clause.

- (6) Naomi-ga Seiji-o ut-ta
Naomi-SUBJ Seiji-O hit-PAST
'Naomi hit Seiji'
- (7) Seiji-ga Naomi-ni ut-are-ta
Seiji-SUBJ Naomi-by hit-PASS-PAST
'Seiji was hit by Naomi'

In contrast, an analytic language, such as Vietnamese, typically employs a periphrastic passive construction, as illustrated in (9), the passive of (8).

- (8) thầy pha tôi
teacher punish I
'The teacher punishes me'
- (9) tôi bị thầy pha
I suffer teacher punish
'I am punished by the teacher'

English, also a fairly analytic language, tends to employ periphrastic constructions which correspond to affixal constructions in more synthetic languages. Examples (10) and (11) illustrate an active and a passive sentence, respectively, in Latin; translations show their English counterparts.

Cambridge University Press

978-0-521-58158-5 - Language Typology and Syntactic Description: Grammatical Categories and the Lexicon, Second Edition - Volume III

Edited by Timothy Shopen

Excerpt

[More information](#)10 *Alexandra Y. Aikhenvald*

(10) Mulier hominem videt
 woman man+ACC.SG see+PRES+3SG
 'The woman sees the man'

(11) Homo ā muliere videtur
 man by woman+ABL.SG see+PASS+PRES+3SG
 'The man is seen by the woman'

Analytic isolating languages, such as Mandarin Chinese, tend to have no marking of grammatical relations other than constituent order (whereby 'the actor of a verb, if expressed, must precede the verb': LaPolla (1995:297)). Compare (12) and (13).

(12) wǒ men tīen tsin
 I PL play piano
 'We are playing the piano' (or 'we are playing the pianos', 'we are going to play the piano', etc.)

(13) ta da wǒ men
 s/he hit I PL
 'She or he is hitting us', 'she or he will hit us', etc.

Since the overt noun phrases are often omitted, the participants have to be inferred from the context. Thus, isolating languages are heavily context-dependent; it has been argued that in Chinese there has been no grammaticalization of the syntactic relations 'subject' and 'object' (see LaPolla (1995), for further discussion).

Numeral classifiers as independent words tend to occur in analytic isolating languages (Aikhenvald (2000)). A numeral classifier is illustrated in (14), from Hmong, a Hmong-Mien language from China (see Bisang (1993); Jaisser (1987:172)):

(14) Lawv muaj rau tus me nyuam
 they have six NUM.CL:LIVING.BEING child
 'They have six children'

When inanimate nouns appear with different classifiers, these highlight different aspects of their meaning. A well-known example from Burmese (Becker (1975:113)) illustrates this point. 'River' can be spoken of in at least eight contexts, shown in table 1.2. Numeral classifiers here are comparable to derivational affixes in more synthetic languages. The specific classifiers can thus add information about the referent, since they allow speakers to distinguish one sense of the referent from all the others. The 'repeater' classifier *myi?* – identical to the noun itself – in table 1.2 indicates that a river is looked upon just as a river, and