The indeterminacy of word segmentation and the nature of morphology and syntax[1]

Martin Haspelmath
Max Planck Institute for Evolutionary Anthropology, Leipzig

The general distinction between morphology and syntax is widely taken for granted, but it crucially depends on a cross-linguistically valid concept of '(morphosyntactic) word'. I show that there are no good criteria for defining such a concept. I examine ten criteria in some detail (potential pauses, free occurrence, mobility, uninterruptibility, non-selectivity, non-coordinatability, anaphoric islandhood, nonextractability, morphophonological idiosyncrasies, and deviations from bi-uniqueness), and I show that none of them is necessary and sufficient on its own, and no combination of them gives a definition of 'word' that accords with linguists' orthographic practice. 'Word' can be defined as a language-specific concept, but this is not relevant to the general question pursued here. 'Word' can be defined as a fuzzy concept, but this is theoretically meaningful only if the continuum between affixes and words, or words and phrases, shows some clustering, for which there is no systematic evidence at present. Thus, I conclude that we do not currently have a good basis for dividing the domain of morphosyntax into morphology and syntax, and that linguists should be very careful with general claims that make crucial reference to a cross-linguistic 'word' notion.

Keywords: word, clitic, affix, morphology, syntax, morphosyntax, lexical integrity

---

1. Words and the morphology–syntax division

Linguists generally assume that morphology and syntax are two different levels of grammatical organization. Descriptive grammars contain separate sections dealing with word structure and sentence structure, introductory courses and textbooks deal with either morphology or syntax, and linguists sometimes define themselves as morphologists or syntacticians and attend specialized conferences. More importantly, linguists often propose special general principles just for morphology or just for syntax, they discuss the nature of the interface between morphology and syntax, and in concrete cases they ask whether a particular phenomenon should fall under morphology or syntax.

Linguists thus usually presuppose that the (*morphosyntactic*) *word* is a fundamental and universal category of language structure, because morphology and syntax are both defined in terms of the word: "Morphology deals with the composition of words while syntax deals with the combination of words" (Dixon & Aikhenvald 2002: 6).[2] But what is a word? How do we as linguists recognize a word and tell it apart from an affix or from a syntactic phrase? How do language learners do this, if we assume that they organize their knowledge of language structure in terms of a mental morphology and a mental syntax?

In this article, I argue that we do not have a good answer to the question of how to define the notion of word in a clear and consistent way that accords with our intuitions and with conventional practice, despite decades of research that has tried to address the issue (cf. Krámský 1969, Juilland & Roceric 1972 for an earlier period). Linguists generally employ a range of different criteria, but these are not uniformly applicable across contexts and languages, and where they are applicable, they do not always converge. I conclude from this that we have no good basis for a general, cross-linguistically viable word concept, and hence no basis for a general bifurcation between morphology and syntax.

But if words cannot be readily identified, why do we work with them all the time? There are two possible explanations for this: On the one hand,

---

[2] One also finds the reverse, definitions of word in terms of morphology/syntax. For instance, Anderson (1992: 17) says that words can be characterized as "the irreducible terminal elements of syntactic structure [and] as the domain of principles regulating the appearance of morphological material". But such definitions would be helpful only if there were a definition of morphology and syntax that is independent of the word concept. I have never seen such a definition.

it could be that words are real after all, and we just have not found a way of identifying them consistently yet. But another possibility that should be taken seriously is that the idea of universal words is due to the bias towards written language and the strong influence of the habit of word separation by spaces in Western languages that has been with us for about a thousand years. In all languages using Greek-derived alphabets (Greek, Latin, Russian, etc.) and in many other languages influenced by them, blank spaces make orthographic words very salient units of written language. However, linguists have overcome their written-language bias also in other respects (e.g. by distinguishing carefully between sounds and letters since the 19th century), so it is time that we reassess the evidence for wordhood in an unbiased way. It is of course quite possible that we will eventually find evidence for something like a cross-linguistic word notion, but we will see that at the present stage of our knowledge, we do not have the evidence.

In Section 2 of this article, I review the kinds of criteria that have been applied in defining the word, and I conclude that only morphosyntactic criteria are relevant for the current concerns. Then Section 3, the heart of the article, discusses ten morphosyntactic word criteria and concludes that none of them singles out a class of items that comes close to what is traditionally regarded as morphosyntactic words. In Section 4, I discuss the widespread strategy of combining several criteria and show that different linguists use different criteria so that the results are not comparable. The difficulties with defining the word have long been recognized, and some linguists have retreated from the strongest position to the weaker claim that the word can only be defined language-specifically (Section 5), or as a fuzzy concept (Section 6). In Section 7, I spell out the consequences of my negative findings: Descriptive practice is not really affected, but comparative claims about words and morphology need to be reassessed. In particular, the notion of lexical integrity is not well supported and should not be appealed to in explaining grammatical phenomena (Section 8). Moreover, for comparative purposes we will eventually need concepts that are defined in such a way that they are crosslinguistically applicable (i.e. comparative concepts), such as *formative* and *construct* (Section 9). Instead of a subdivision of the grammar of sign combinations into morphology and syntax, we can just work with a unified domain of morphosyntax.[3]

---

[3] It should be noted that there are two rather different ways in which 'word' may be assumed as a cross-linguistic concept: On the one hand, 'word' may be taken as a cross-linguistic category that is in some sense pre-established (e.g. taken to be innately given as part of Chomskyan Universal Grammar), so that the word categories of two different languages

2. Four kinds of potential word-defining properties: semantic, orthographic, phonological, morphosyntactic

2.1. Speaker intuitions

Before we consider word-defining criteria or properties, let us briefly consider whether it is necessary to use our skills as linguists to identify words, or whether we can simply ask speakers what the words in their language are. This possibility is sometimes suggested in the literature. For example, Coseriu (1964: 141–142) simply asserts: "Nous estimons la notion de 'mot' comme intuitivement établie" ['We regard the notion of word as intuitively established'], and according to Aronoff & Fudeman (2005: 36), "speakers – literate and illiterate – have clear intuitions about what is and what isn't a word" (similarly Langacker 1972: 36, Bauer 1988: 45, Himmelmann 2006: 255). But it is unclear what the basis is for this optimism.[4] Speakers who are illiterate in any language (the best test case) are becoming increasingly rare, and field linguists who do not already expect to find words are not common either. Moreover, the speakers would have to be able to communicate their intuitions, but as Dixon & Aikhenvald (2002: 3) note, "the vast majority of languages spoken by small tribal groups . . . have a lexeme meaning '(proper) name', but none have the meaning 'word'." In European languages, too, we can see that the current words for 'word' (e.g. English *word*, French *mot*, Russian *slovo*, etc.) originally started out with much more general meanings (act of speaking) and seem to have acquired the

---

can be fully equated. On the other hand, 'word' may be taken as a comparative concept that is created just for the purposes of comparing languages, without any claim that individual languages have language-particular descriptive categories that are identical to the comparative concept or can be equated with categories in other languages (see Haspelmath (2010) for the distinction between comparative concepts, descriptive categories, and pre-established cross-linguistic categories). In this article, I am not distinguishing between these two cross-linguistic concepts, as my arguments apply equally to both conceptions of a general 'word' notion.

[4] It could be that it was influenced by Sapir's (1921: 33–34) observation that "the naïve Indian, quite unaccustomed to the concept of the written word, has nevertheless no serious difficulty in dictating a text to a linguistic student word by word". This is of course an anecdotal observation, but it became widely known through a very popular text. An anecdotal observation that goes against this is Evans et al.'s (2008: 97) finding that Dalabon speakers (not literate in their language) are happy to make metalinguistic comments about parts of polysynthetic words. As Dixon & Aikhenvald (2002: 23–24) note, the occurrence of pauses (and hence also breaks in dictation) seems to be linked to phonological words, rather than morphosyntactic words, so it is of little help in determining the latter.

narrower sense of Greek *léxis* only through formal schooling, in particular writing and grammar teaching (see Haebler 2002). It is of course still an open question what kinds of intuitions illiterate speakers have, but it is clear that literate speakers of languages with word-separating writing systems have no intuitions that are independent of the writing rules they have learned.[5] And where the writing conventions are the subject of debates (e.g. in the discussion of German spelling reform between 1996 and 2006, cf. Fuhrhop 2007), advocates of reforms have a very hard time convincing the general public that they even have a reasonable case to make.[6]

   Thus, when two linguists disagree about word segmentation,[7] resolving the disagreement by asking the speakers is not an option that is generally available, and we have to resort to word-defining properties.

2.2. Semantic non-compositionality

Pre-modern definitions of word often include a reference to meaning, as in the quotations in (1).

(1)   a.   Zedler 1749: "Wort: ein vernemlicher Laut, der etwas bedeutet."
            ['word: a perceptible sound that means something']
      b.   Sapir 1921: 34: "the smallest, completely satisfying bits of isolated meaning into which the sentence resolves itself."
      c.   Meillet 1921: 30[1982] "Un mot est défini par l'association d'un sens donné à un ensemble donné de sons susceptible d'un emploi grammatical donné."
            ['A word is defined by the association of a particular sense with a particular set of sounds that has a particular grammatical use.']

---

[5] One interesting possibility is to test speakers who are literate in another language than the test language. The only experiment of this sort that I know is Peterson's (2008: 34–39) study of six Kharia speakers who are literate in Hindi. In all 12 sentences that were presented to them, the speakers differed with respect to the number of words that they used for the Kharia sentences, sometimes quite markedly (e.g. between 1 and 4 words for Peterson's sentence (3), between 7 and 10 words for Peterson's sentence (8)).

[6] This contrasts strikingly with phonetic spellings, which are at least acknowledged as rational by linguistically unsophisticated reform critics.

[7] When several linguists or missionaries independently create writing systems for a previously unwritten language, there are also bound to be many disagreements, as is notoriously shown by the case of Bantu languages, which are sometimes spelled 'conjunctively' (with preverbal elements written as prefixes) and sometimes 'disjunctively' (cf. e.g. Guthrie 1948).

Evidently, these definitions really refer to something like the morph (a minimal sound–meaning pair), or any combination of morphs (such as a phrase). It is now very widely recognized that many complex words are semantically compositional in exactly the same way as phrases and clauses, and that conversely many phrases are idiomatic and thus not semantically compositional. Phrases like *spill the beans* or *fat cat* must be learned and stored as wholes and are lexical entries, but not morphosyntactic words.

Still, meaning sometimes shows up as a criterion for wordhood in modern works. Thus, Dixon & Aikhenvald (2002: 19) postulate the criterion that the word elements "have a conventionalised coherence and meaning". Other recent authors who mention non-compositional meaning as a criterion for wordhood are Zwicky & Pullum (1983: 505), Kanerva (1987: 510–512), Mel'čuk (1993: 210), and Harris (2000: 599). These authors may be guided by the (quite possibly correct) feeling that non-compositionality is more characteristic of complex words than of phrases, but meaning cannot be used as a criterion for distinguishing words from phrases.

2.3. Orthography

Many orthographies, especially (but not only) those based on the Greek, Latin and Cyrillic alphabets, use spaces between words.[8] However, there are also many orthographies that do not use spaces, e.g. Chinese, Japanese, and Sanskrit. In the European languages, too, word spacing is an innovation; until about a thousand years ago, *scriptio continua* (continuous writing) was the norm in Western writing (cf. Saenger 1997). There is no doubt that the modern orthographic use of spaces is to some extent guided by language structure, but not in such a way that conventional spelling could be used to decide contentious issues. As Jespersen (1924: 92) noted, "spelling is often perfectly arbitrary and dependent on fashion or, in some countries, on ministerial decrees not always well advised". In many languages, there are obvious inconsistencies in the spelling rules, e.g. the variable spelling of the German infinitive marker *zu* (spelled separately in simple verbs, e.g. *zu gehen* 'to go', but without space with particle verbs, e.g. *wegzugehen* 'to go away').

---

[8] Other word-separating symbols are found only sporadically, e.g. a colon-like symbol that is used in Amharic and other Ethiopic languages.

2.4. Phonology

Linguists often mention phonological criteria for delimiting words, e.g. obstruent devoicing in Russian, vowel harmony in Turkish, or stress in Polish. However, it has been clear for quite a while that phonological criteria and grammatical criteria do not always give identical results (e.g. Bloomfield 1933: 182–183, Hockett 1958: 58, Pike 1967: 399ff.). Trager & Smith (1951) distinguish between phonemic words and grammatical words, and since the 1980s (Dixon 1977, Nespor & Vogel 1986) the consensus in the field has been that "it is clear that the phonological word does not always coincide with the morphological word" (Bresnan & Mchombo 1995: 182). Thus, phonological criteria cannot decide contentious cases, and for this reason I will not discuss them further in this article, which focuses on morphosyntactic words.[9] Moreover, phonological word domains do not necessarily converge within the same language (e.g. Turkish has different domains for vowel harmony and stress, so that it could be said to have two incompatible kinds of phonological words, cf. Kabak & Vogel 2001). In fact, there does not even seem to be a tendency for phonological word domains to converge (Schiering et al. 2010). I will argue below that the situation with morphosyntactic words is probably quite similar.

2.5. Morphosyntax

Thus, we see that in problematic cases, we cannot resort to intuition (Section 2.1), to semantic criteria (Section 2.2), to orthographic criteria (Section 2.3), or to phonological criteria (Section 2.4). To identify morphosyntactic words, we need to apply morphosyntactic criteria, hoping that they will converge and decide unclear cases that seem intermediate between affixes and words, or between words and phrases. Many different types of criteria have been mentioned in the literature, and in the next section I will discuss the ten most important ones.

---

[9] I take the following terms to be synonymous: *grammatical word* (e.g. Trager & Smith 1951, Matthews 1974, and many others), *morphological word* (e.g. Wurzel 1984: 35–36, Di Sciullo & Williams 1987, Bresnan & Mchombo 1995), *syntactic word* (e.g. Kroeger 2005: 318), *word-form* (Mel'čuk 1993, French *mot-forme*), and *morphosyntactic word* (e.g. Spencer 1991: 45). I opt for the latter, as it is the most informative term.

3. Morphosyntactic word criteria

Each of the ten criteria of morphosyntactic wordhood (listed in 2) singles out certain types of expressions, and the question is whether one of the criteria (or a combination of them) is necessary and sufficient to define a kind of expression that we would intuitively call words. By this I mean words as identified by written spaces in the normal spelling used by speakers and/or by linguists. Of course, linguists need not confine themselves to segmentations of this sort, but if we found a systematic segmentation level that bore no close resemblance to segmentation by orthographic spaces in Western languages, we would not want to call it 'word level'.

(2)   *Ten criteria of morphosyntactic wordhood*
    1. Potential pauses
    2. Free occurrence
    3. External mobility and internal fixedness
    4. Uninterruptibility
    5. Non-selectivity
    6. Non-coordinatability
    7. Anaphoric islandhood
    8. Nonextractability
    9. Morphophonological idiosyncrasies
    10. Deviations from biuniqueness

In the following subsections, I will examine the criteria in turn and show that none of them is necessary and sufficient for wordhood, and that many are problematic in one way or another.

3.1. Potential pauses

A simple widespread idea is that words can be identified by potential pauses. Although spoken words are of course normally not surrounded by pauses, speakers have the possibility of pausing in the middle of a sentence, and according to one popular criterion, potential pauses indicate word boundaries (Hockett 1958: 166–167, Langacker 1972: 41).

However, pauses are usually considered to be relatively shallow performance phenomena having to do more with processing than with language structure. Although there are quite probably conventions for when to pause (just as there are conventions for overt hesitation markers), pauses are not part of grammatically well-formed sentences, and linguists cannot

easily ask speakers about their intuitions concerning pauses. This limits the practical usefulness of the potential pause criterion.

Moreover, a potential pause is neither a necessary nor a sufficient indication of a word boundary. Clitics are generally considered words (Haspelmath 2002: 149), but no pause is possible between a clitic and its host. Conversely, in languages where linguists have traditionally identified particularly long words, there also seems to be a tendency to allow pauses in the middle of words (cf. Dixon & Aikhenvald 2002: 11–12). Evans et al. (2008) give a detailed account of such pauses internal to the morphosyntactic word in Dalabon. An example from dictated speech is given in (3) (where '. . .' stands for a pause; Evans et al. 2008: 103).

(3)   Dalabon (Gunwinyguan; Arnhem Land, Australia)
      ka-h-. . .rak-. . .m-iyan
      S3sg.A>3sg.P-As-. . .wood-. . .get-FUT
      'He. . .will get. . .firewood.'

Dixon & Aikhenvald (2002: 24) note that the possibility of pausing may be more closely related to phonological wordhood than to morphosyntactic wordhood, and they mention the possibility of inserting expletives inside long words (e.g. *abso-bloody-lutely*) that may be considered to consist of two phonological words. There is no doubt that better corpora of spoken languages will give us more insight into these matters in the future, but at the moment the potential pause criterion is evidently not of much help for identifying morphosyntactic words.

3.2. Free occurrence

Another popular criterion is the possibility of occurring as a well-formed complete (but possibly elliptical) utterance, e.g. an answer to a question (e.g. *Where are you? Here; What do you need? Money*). Bloomfield (1933: 160) called utterance segments that can occur on their own *free forms*, and he famously defined the word as "a free form which does not consist entirely of (two or more) lesser free forms; in brief, a word is a *minimum free form*" (Bloomfield 1933: 178; also Hockett 1958: 168). But this definition does not single out forms that correspond to our intuition of grammatical words. On the one hand, it is too strict, because by this definition compounds such as *firewater* or *blood-red* would not be words, but phrases, because they have constituents that are themselves free forms. On the other hand,

it is much too loose, because many phrases such as *a flower*, *to Lagos*, or *put it away* would count as words, because the elements *a, to, put*, and even *put it* cannot occur on their own without something following them. Now one might possibly be prepared to accept that English *a-* is in fact an indefinite-article prefix, *to-* is an allative prefix, and *-it* is a pronominal object suffix, because there are of course many languages where such things are traditionally assumed to exist. But in English and quite a few other languages, not even transitive verbs are free forms, because they cannot occur on their own, and one would have to regard sequences such as *take the money* as single words. Thus, the criterion of independent occurrence does not have real practical value, and it is not surprising that it is hardly used in recent wordhood controversies. It seems that it is mentioned in discussions of wordhood primarily because of Bloomfield's (and Hockett's) authority.

However, the criterion of independent occurrence is helpful at least in one regard: If an element can occur independently (i.e. as a complete utterance all by itself), it clearly cannot be an affix but must be minimally a morphosyntactic word.[10]

3.3. External mobility and internal fixedness

Characterizations of grammatical words often make reference to the criterion of mobility or fixedness: Words can occur in different positions, whereas affixes occur in a fixed order (Boas 1911: 30, Reichling 1935, van Wyk 1968, and many others). Let us first consider mobility as a criterion for wordhood, before we look at fixedness as a criterion for affixhood.

The first thing to note is that in most languages, most words have a fixed position with respect to some other words, i.e. words are only relatively free in their ordering. A few languages have been described as exhibiting truly free word order (e.g. some Australian languages), but at least the notional noun phrase constituents almost always occur together in almost all languages, and many languages are like English in that they have fairly rigid order at all levels. In English and similar languages, it is primarily adverbial expressions that can occur in different positions. Nonsubject arguments can be preposed in a topicalization construction (*That book I haven't read*), but in this and many other cases it is phrases, not words, that are mobile.

---

[10] Bauer & Huddleston (2002) mention colloquial English *ish* (*Was it exciting? Ish*) as a freely occurring suffix, but most linguists would instead say that such uses show that this item has become a word.

Thus, the mobility criterion is quite restricted in its applicability.

   Another serious problem with the mobility criterion is that it presupposes that the different ordering is the only difference between two structures. This situation does sometimes occur, e.g. in the contrast between *Yesterday I saw her* and *I saw her yesterday*, which can reasonably be taken as evidence for the mobility and hence word status of *yesterday*. However, in many other cases, a better description might be in terms of two different constructions, one with an affix and another with a free form. In the case of Latin *cum*, for instance, it is normally assumed that it is a word when it precedes its complement (e.g. *cum grano salis* 'with a grain of salt'), but it is regarded as a suffix when it combines with personal pronouns (e.g. *mecum* 'with me', *vobiscum* 'with you'). This is probably because Latin adpositions are not normally mobile, but clearly, one could alternatively decide that 'with me' represents the same construction and a different order, hence a postposition (*me cum*). Another case where the traditional description sees two different constructions rather than mobility is the Finnish negation pattern, involving the negative marker *e-* followed by the subject person forms (cf. 4b). In affirmative clauses, the subject person forms follow the verb. This could be interpreted as mobility, but traditionally it is not seen as evidence for clitic (i.e. word) status of the person forms.

(4)   Finnish
    a.   *mene-t*   'you go'
        go-2SG
    b.   *e–t mene*   'you don't go'
        NEG-2SG go

Conversely, the auxiliary element in Basque (*da/du* in the 3rd person singular), which is obligatory in the finite forms of almost all verbs, is normally regarded as a separate word (see 5–6). In affirmative clauses, it has to follow the main verb immediately, and nothing can come between it and the verb (cf. 5b). However, in negated clauses, marked by the negative morpheme *ez*, the auxiliary follows this morpheme immediately and both together usually precede the main verb (cf. 6a–b). The negation-auxiliary complex need not even be adjacent to the verb (cf. 6c). This alternation is seen as mobility, and hence *ez* and *da* are generally regarded as separate words. But a description in terms of a different construction (perhaps with a negative verb *ez-*, analogous to Finnish *e-*) would also be possible, so that the auxiliary could be seen as a suffix.

(5)   Basque (Hualde & Ortiz de Urbina 2003: 518)
    a.   *Etxe-a   erori da.*
        house-ART fallen AUX.PRES.3SG
    b.   **Erori etxe-a da.*
        'The house fell down.'

(6)   a.   *Etxe-a   ez   da   erori.*
        house-ART NEG AUX.PRES.3SG fallen
    b.   **Etxe-a ez erori da.*
    c.   *Ez da etxe-a erori.*
        'The house did not fall down.'

A case where a root is potentially mobile comes from Evans et al. (2008), who report on an alternation in Dalabon that they call 'prepound extraction'. This concerns compound verbs such as *walk-ka-* [hide-take-], whose first member ('prepound') may alternatively occur outside the verbal complex, as illustrated in (7a–b).

(7)   Dalabon (Evans et al. 2008: 96)
    a.   *ka-lng-walk-ka-rr-inj*
        3SG.S-SEQ-hide-take-REFL-PST.PFV
        'then he hid himself'
    b.   *walk ka-lng-ka-rr-inj*
        hide 3SG.S-SEQ-take-REFL-PST.PFV
        'then he hid himself'

Evans et al. (2008: 97) comment that "clearly these possibilities do not create alternative ordering *within the word* – rather they offer two alternative constructions, in one of which the relevant material occurs within the verbal root, and in the other of which it doesn't." But if one sees (7a–b) as the same construction, the mobility of *walk* provides evidence for three separate grammatical words, *ka-lng, walk*, and *ka-rr-inj*. Why this latter description should be less good is unclear.

    Another example is the Lithuanian reflexive marker *si*, which appears in two different positions: at the end of the verb (after person-number suffixes) when the verb has no prefix, and between the prefix and the verb stem in prefixed verbs.

(8)   Lithuanian
    1SG *skutuo-si*   'I shave'   *ne-si-skutu* 'I don't shave'
    2SG *skutie-si*   'you shave'   *ne-si-skuti* 'you don't shave'
    3SG *skuta-si*   'he shaves'   *ne-si-skuta* 'he doesn't shave'

Traditionally, *-si* is regarded as a suffix in the first set of forms, and as a prefix in the second set of forms, but one could also emphasize the mobility of this element and consider it as a clitic (much as in Polish, where the cognate reflexive *się* is regarded as mobile and thus as a clitic).[11]

One could also adopt the different-construction view for alternations like English *They will do it* vs. *Will they do it?*, especially since the two constructions are not quite identical phonologically, as in Lithuanian (cf. the possibility of *They'll do it*, vs. **'ll they do it*?). Under this view, *they-will* and *will-they* could be regarded as complex words.

Thus, the application of the mobility criterion is far from straightforward, and it would have to be made much more precise before it can be used reliably.[12] Let us now look at fixed order as a criterion for affixhood. While words are expected to be mobile with respect to other words, word-internal elements are expected to occur in a fixed order (e.g. Bauer 1988: 52, Mugdan 1994: 2552, Dixon & Aikhenvald 2002: 19). Bauer (1988: 52) notes that in a Latin word like *reg-e:-ba-nt-ur* [rule-STEM-PST-3PL-PASS] 'they were being ruled', the order of the morphs cannot be rearranged, while the word forms of a sentence are "much more movable". But on the one hand, especially function words tend to be quite rigid in their ordering – for example, while the order of genitive and noun is often variable, the order of adposition and noun is much less variable (Dryer 2005a, 2005b).

On the other hand, the order of traditional affixes is not always fixed: Variable affix order has been discussed by Stevens (1971) (for Madurese), Fulmer (1990) (for Afar), Noyer (1994) (for Huave), Luutonen (1997) (for Mari), and Blevins (2001: 118–119) (for Nhanda). Most prominently, Bickel et al. (2007) present detailed arguments for free prefix ordering in

---

[11] Similarly, in the Icelandic alternation in (i) one usually assumes two different constructions, although it would also be possible to assume a word order alternation and thus word status rather than suffix status for the article in the (a) example. (The phonological alternation *hinn/inn* is fairly regular).

(i)   Icelandic (Thráinsson 2007: 19)
       a. *rauði herstur-inn* 'the red horse'
          red horse-ART
       b. *hinn rauði hestur* 'the red horse'
          the red horse

[12] Bernard Comrie (p.c.) points out that an analogous difficulty exists for the criterion of free occurrence: One might argue that *in* is a free word, not a prefix, because it can be used without complement in cases like *Lee is in*, but this presupposes that this is the same element as *in* in *Lee is in the house*. Whether or not they are identical or not is difficult to say.

Chintang. These are relatively little-studied languages, but if one regards the Romance object person markers as affixes (as is done by Bally 1913: 34, Tesnière 1932, Miller 1992a, Monachesi 1999, among many others), one has to face the fact that they occur as prefixes and as suffixes (as in 9), and that they can sometimes even move to a higher clause (as in 10, where 10b shows clitic climbing):

(9) Italian
  a. *me lo darà*
    'she will give it to me'
  b. *da-mme-lo*
    'give it to me (imperative)'

(10) Spanish
  a. *Quiero besar-te.*
    'I want to kiss you.'
  b. *Te quiero besar.*
    'I want to kiss you.'

3.4. Uninterruptibility

The third core criterion of grammatical word status is uninterruptibility: While phrasal combinations may be interrupted by other material, words cannot be so interrupted (e.g. Bloomfield 1933: 180, Langacker 1972: 48). As Mel'čuk (1993: 173–174) notes, the semantic relations must remain intact if one applies this test, because otherwise almost any string of morphs could be shown to be interruptible. Another requirement is that one must be sure that the interrupting element is not itself an affix. Weinrich (1963: 172) notes that even Latin *canta-t* [sing-3SG] 'sings' is interruptible, but only by other affixes (e.g. *canta-bi-t* [sing-FUT-3SG] 'will sing'). So the fact that French *il chante* [he sings] can be interrupted by a negation particle and an object pronoun (e.g. *il ne lui chante pas* [he not to.him sings not]) does not show that *il* is a separate word, because the intervening elements are themselves best analyzed as prefixes (Miller 1992a).

Mugdan (1994: 2552) is worried by this: "There is reason to suspect that the inserted material must be a word. . .and that the test presupposes the very unit it should help to identify". This would be fatal for uninterruptibility as a criterion, but fortunately, there is an alternative: One can say that interruptibility by free forms is a sufficient criterion for two-word status, because free forms cannot be affixes (though not all words are free

forms, as we saw in Section 3.2). Thus, English *he loved* is not a single word, because it can be interrupted by *never* (*he never loved*), and *never* is a free form (unlike *he* and *loved*).

So can we define the morphosyntactic word as a 'maximal uninterruptible string of morphs'? Unfortunately, this is not possible because the definition is too loose: There are many uninterruptible combinations that are not normally considered words. For example, in (11), the combinations linked by a plus sign are not interruptible by anything else:

(11)   a.   both+my parents
       b.   even+Kim understands it
       c.   very+good food

To rule out the A+B combinations in (11) as complex words, one would need to invoke another criterion, e.g. the criterion of non-selectivity (see Section 3.5): Since *both* and *even* combine with words of diverse classes, they would not count as affixes, despite being very tightly (and uninterruptibly) combined with their hosts in (11a–b) (in 11c, not even the non-selectivity criterion gives the desired result, because *very* is selective, combining only with adjectives; I know of no way of arguing against inflectional prefix status of *very*). Another problem is that some combinations which are usually considered as single words can actually be interrupted by free forms, in particular in incorporation patterns. For example, Pawnee allows both (12a) and (12b), and in the latter the element *rīks* 'arrow' interrupts the word in (12a). (These data are also discussed in Julien 2002: 35).

(12)   Pawnee (Caddoan; Boas 1911: 31)
       a.   *tā-tu-k^{u}t*
            'I have cut it for you'
       b.   *tā-tu-rīks-k^{u}t*
            'I have cut your arrow'

3.5. Non-selectivity (or promiscuity)

An important word criterion that is much less often mentioned than the first three criteria (isolability, mobility, uninterruptibility) is non-selectivity (which may also be called promiscuity): While an affix tends to be highly selective with respect to the kinds of hosts it can combine with, (function) words are often able to combine with a wide range of hosts. For example, the Lezgian elements *-na* 'perfective past (Aorist)' and *-ni* 'and,

also, even' contrast in this way: While -*na* combines only with verb stems of a particular kind (cf. 13a), -*ni* combines with nouns, adjectives, verb stems, infinitival clauses, adverbial clauses (among others) (cf. 13b).

(13)   Lezgian (Haspelmath 1993: 131, 142, 327–329)
    a. *awu-na*   [do-PST]   'did'

a.	*awu-na*	[do-PST]	'did'
	fe-na	[go-PST]	'went'
	qaču-na	[take-PST]	'took'
b.	*buba-ni*	[father-too]	'father, too'
	bürq'ü-ni biši-ni	[blind-too deaf-too]	'blind and deaf'
	güzlemiš-ni	[wait-too]	'even wait'
	kaler aca-z-ni	[cows milk-INF-too]	'also to milk cows'
	wun xkwedaldi winik-ni	[you return before-too]	'also before you returned'

Both -*na* and -*ni* are called suffixes in Haspelmath (1993), as both are written without a space in Lezgian spelling, but there is clearly a marked difference between them in terms of selectivity.

Many linguists would intuitively classify an element such as Lezgian -*ni* as a clitic (i.e. a kind of word) rather than a suffix, even though it is not isolable, not mobile, and host+*ni* combinations are not interruptible (because -*ni* attaches to the head and Lezgian is head-final, nothing can come between the head and -*ni*). This shows that selectivity is an important criterion in practice. As we saw in the previous section, non-selectivity is also the main criterion that can be adduced in favour of classifying English elements such as *both*, *even*, and *very* as words (cf. 11a–c above).

The problem with non-selectivity as a sufficient criterion for wordhood is that there are good examples of non-selective elements that are excluded from word status by other considerations. Such non-selective or promiscuous non-word elements have even been fairly prominent in the literature since Zwicky (1987) (who calls them *edge inflection*). Stump (2001: 126–130) calls them *promiscuous inflection* (see also Lapointe 1990, Halpern 1995, O'Connor 2002, Tseng 2003, Anderson et al. 2006, Samvelian 2007, Bermúdez-Otero & Payne forthcoming, among others). One example of such non-selective inflection is the locative case form in Òko, a Benue-Congo language of Nigeria. Consider the examples in (14).

(14)   Òko (Atoyebi 2010: 58)
    a. *ùgbègbèn* 'mirror'
    b. *úgbègbèn* 'in the mirror'

 c. *Òsìbìna u-ùbo* [God POSS-house] 'God's house'
 d. *Ósìbìna u-ùbo*   'in God's house'

We see that in Oko, locative case is marked by a high tone on the first syllable when the locative phrase begins with a vowel. Since all noun stems are vowel-initial in Oko, this is by far the most common situation. This locative high tone is thus promiscuous, but it cannot be regarded as a word because words are expected to be segmental entities. A similar case of suprasegmental marking at the edge of a phrase is cited by Poser (1985) and Anderson (1992: 212) from Tongan. Such cases do not seem to be uncommon in languages that make some use of tone for grammatical purposes.

 Another example of a promiscuous marker that cannot have word status is the English possessive -*'s*, as illustrated in (15).

(15) English (Zwicky 1987: 140)
  a. children's ideas
  b. kids' ideas
  c. [anyone who likes children]'s ideas
  d. [anyone who likes kids]' ideas

Examples (15c–d) show that English possessive -*'s* is non-selective and thus would qualify for word status, but as Zwicky (1987) notes, the haplological omission of the exponent in (15d) would then be unexpected, as haplology, like other morphophonological processes, is characteristic for morphological markers, not for syntactic combinations (see Section 3.8 below).

 But if we accept such markers as 'edge inflection', i.e. as parts of words rather than as words of their own (as Zwicky, Lapointe, Halpern, Stump and others do), then we can no longer use non-selectivity as a criterion for wordhood. For elements such as Lezgian -*ni*, we have no decisive evidence that they cannot be words, but we have no decisive evidence that they cannot be affixes either.

3.6. Non-coordinatability

The next three word criteria have become prominent only in the last few decades, after coordination, anaphora and extraction became important research topics in syntax. The general idea is that words, but not parts of words, take part in these processes.

 The criterion of non-coordinatability (cf. Miller 1992b) concerns both bases and affixes. First, it has been claimed that bases cannot be coordi-

nated, or in other words, that affixes cannot undergo coordination ellipsis (or in yet other words, that affixes cannot have wide scope over coordination). For example, Monachesi (1999: 24) notes that Italian pronominal 'proclitics' cannot be omitted in the second conjunct in (16) (it has to be repeated: *Lo comprerà e lo indosserà alla festa*), as is expected if one regards them as prefixes rather than words.

(16)   Italian
   **Lo comprerà e   indosserà alla   festa.*
   it   she'll.buy and she'll.wear at.the party
   'She will buy it and wear it at the party.'

Thus, if an element can undergo coordination ellipsis, it cannot be an affix according to this criterion.

   However, in some languages elements that are called affixes can undergo ellipsis. In Turkish, both derivational affixes and inflectional affixes can be ellipted (Erdal 2007: 178, 180; Kabak 2007):

(17)   a.   *kum ve   çakıl-cı*
            sand and gravel-PROFESSIONAL
            'supplier of sand and gravel'
        b.   *kedi ve   köpek-ler-im-e*
            ([*kedi ve köpek*]*-ler-im-e* or [*kedi ve köpek-ler*]*-im-e*)
            cat   and dog-PL-1SG-DAT
            'to my cat(s) and dogs'

This phenomenon is quite well known under the name of 'suspended affixation' (Lewis 1967: 35). One might suspect that the elements that can be omitted in coordination are in fact clitics rather than affixes, although they are within the vowel harmony domain. And indeed, not all elements within the vowel harmony domain behave alike in Turkish. Kabak (2007) and Erdal (2007) do not agree on what one should call clitics in Turkish, but they agree that the elements that can be ellipted must include some that are affixes (see also Broadwell 2008). Affix coordination is also found elsewhere (cf. Wälchli 2005: 57–64), though in the modern European languages it tends to be restricted to certain derivational affixes:

(18)   a.   English *pro-choice and -gun control* (Chaves 2008: 263)
        b.   German *trink- und ess-bar* 'drink(-able) and eatable'
        c.   Hungarian *ajtó- és ablak-talan* 'door(-less) and windowless'
            (Kenesei 2007: 270)

It is even easier to find examples with verb phrase or clause coordination. For example, Japanese allows the ellipsis of verbal tense suffixes in coordination constructions such as (19) (Fukushima 1999: 297), where the past-tense suffix -*ta* has scope over both verbs:[13]

(19) *Taroo-ga uta-i   (sosite) Hanako-ga   odot-ta.*
    Taro-NOM sing-CONJ and   Hanako-NOM dance-PST
    'Taro sang and Hanako danced.'

Thus, it is very difficult to maintain the view that words, but not word parts can be ellipted in coordination.

On the other hand, it has been claimed that word parts cannot be coordinated (or in other words, that bases cannot be ellipted), so for example Italian has *piang-eva e piang-erà* [cry-PST.3SG and cry-FUT.3SG] 'she cried and will cry', but not **piang-eva e -erà*. But it is well-known that coordinated prefixes and suffixes are not uncommon in English and related languages, as illustrated in (20).[14]

(20) a. English *pro- and anti-war*
      b. German *be- und entladen* 'load and unload'
      c. Catalan *inter- o intraestatal* 'inter- or intra-state' (Chaves 2008: 302)

Since Booij (1985), there has been a tradition of claiming that such cases of coordination of parts of words (including stems of compounds, e.g. Dutch *wespe- en bijesteken* 'wasp and bee stings') are prosodic deletions, not syntactic ellipses. However, more recent research (e.g. Smith 2000, Kenesei 2007, Chaves 2008) seems to show that the conditioning factors are quite complex and involve semantic and morphosyntactic factors as well.

Thus, coordinatability of elements cannot be taken as a sufficient criterion for phrasal as opposed to word status.[15]

---

[13] Constructions such as (19) are sometimes regarded as showing subordination rather than coordination; an analysis in terms of tense suffix coordination ellipsis presupposes a coordination analysis (following Fukushima; there is no space here to rehearse the arguments for this).

[14] In connection with these examples, a reviewer wonders whether the distinction between inflection (e.g. *piang-eva*) and derivation (e.g. *pro-war*) is irrelevant. The answer is yes, because I do not think that this distinction is generally well-supported (cf. Plank 1994). Moreover, I have not seen any claims that coordinatability of affixes (or any of the other criteria for wordhood) should be restricted to either derivational or inflectional affixes.

[15] Non-coordinatability cannot be taken as a sufficient criterion for affix status either, as coordinatability is influenced by multiple factors. For instance, the English object pronouns cannot easily have scope over two verbs (Chaves 2008: 277, attributing the examples to Hankamer 1973 and Bresnan 1974):

3.7. Anaphoric islandhood

Since Postal (1969), it has been widely assumed that words are anaphoric islands, i.e. expressions whose parts cannot be anaphorically related to other parts of the sentence, and this has repeatedly been cited as a criterion for identifying words (e.g. Di Sciullo & Williams 1987: 107, Mel'čuk 1993: 207, Bresnan & Mchombo 1995: 189, Aydemir 2004: 468). For example, Spencer (1991: 42) cites the unacceptability of (21) as evidence that *tea-pot* is a single word rather than a phrase.

(21)   *He took the tea$_1$-pot and poured it$_1$ into the cup.

But of course words are not the only anaphoric islands, so the explanation in terms of the word status of *tea-pot* cannot be taken for granted. In fact, a very simple semantic explanation is available: Anaphoric pronouns like *it* must refer to an established referent, and in the expression *tea-pot*, the element *tea* is not referential and hence there is no referent for the anaphoric pronoun *it* to refer to. Non-referential elements do not allow outbound anaphora (i.e. anaphora referring to an element of the complex expression) either when the complex expression is not a word, as seen in (22a–c) (English *weapons of mass destruction*, Italian *giacca a vento*, German *Fahrrad fahren*).[16]

(22)   a.   English
              *Weapons of mass destruction are controversial because its effects
              (i.e. the effects of mass destruction) are terrible.
        b.   Italian (Masini 2009: 26)
              *Ha   indossato la giacca a vento perché tirava   forte.
              he.has put.on   the jacket for wind because blowed strong
              'He put on the windbreaker because it (i.e. the wind) was blowing
              hard.'

    (i) *Alice composed and Tim performed it.
    (ii) *He tried to persuade but he couldn't convince them.

[16] In English linguistics, *N-of-N* expressions are not normally considered compounds, but in Romance linguistics, the analogous expressions of the type *N de N* (or *N a N* as in 22b) are often subsumed under compounds (e.g. Mathieu-Colas 1996). If such an analysis were adopted, the examples in (22a–b) would be fully parallel to (21). However, not all cases of non-referential nouns can be regarded as compound members. German *fahre Fahrrad* in (22c) cannot be a compound because of the criteria of interruptibility (e.g. *ich fahre oft Fahrrad* 'I often cycle') and mobility (e.g. *ich kann Fahrrad fahren* 'I can cycle').

c.   German
   Ich fahre Fahrrad, wenn es nicht kaputt ist.
   I   ride bicycle   when it not   broken is
   'I cycle when it (i.e. the bicycle) is not broken.'

On the other hand, when a word part is a name and therefore referential, then outbound anaphora is quite possible:

(23)   McCarthyites are now puzzled by him. (Sproat 1988: 299)

That outbound anaphora is primarily semantically and pragmatically determined is now widely accepted (cf. Ward et al. 1991). Moreover, in languages with highly productive noun incorporation, incorporated nouns can be referential (e.g. Sadock 1986: 22–26), and outbound anaphora is normal.

Inbound anaphora is a different matter: Anaphoric pronouns as part of compounds or derivational formations[17] are quite impossible in English (**McCarthy was happy that many him-ites were at the meeting*). However, Harris (2006) has demonstrated in detail that anaphoric pronouns can occur in such structures in Georgian, and are in fact quite common. Sentences such as (24) are quite normal in Georgian (Harris 2006: 119).

(24)   *Ševardnaʒe   icnobs?   K'i, imis-ian-i–a.*
       Shevardnadze know.3sg yes him-ite-nom-cop
       'Does she know Shevardnadze? Yes, she is a him-ite.'

Thus, neither outbound nor inbound anaphora can serve as a reliable criterion for word status across languages.

3.8. Nonextractability

It is sometimes said that word parts cannot be extracted, so that if something can be extracted, we know that it cannot be a word part (e.g. Bresnan & Mchombo 1995: 187). Aronoff & Fudeman (2005: 37) give the two examples of attempted topicalization and *wh*-movement in (25).

---

[17] Anaphoric pronouns as inflectional affixes are undeniably quite common, cf. Bresnan & Mchombo (1995: 190–192). Since the inflection–derivation distinction seems to be just as problematic as the word–phrase distinction, this is an additional problem for anaphoric islandhood as a word criterion.

(25) a. *Possible, it's im-.
      b. *Which school did you see the -bus?

Similarly, Bisetto & Scalise (1999) mention non-extractability as a criterion for distinguishing between compounds and phrases in Italian. While a noun phrase such as (26a) allows topicalization out of it (cf. 26b), a compound such as (26c) does not allow topicalization (cf. 26d).

(26) Italian (Bisetto & Scalise 1999: 38–39)
      a. *il trasporto dei passeggeri*
          'the transportation of the passengers'
      b. *Dei   passeggeri, è efficiente il   trasporto _.*
          of.the passengers is efficient the transport
          'Of the passengers, the transportation is efficient.'
      c. *nave ospedale*
          'hospital boat'
      d. **Ospedale, hanno   costruito una nave _.*
          hospital   they.have built   a   boat
          'Hospital, they have built a _ boat.'

But while it is probably true that word parts cannot be extracted, the question is whether their non-extractability has anything to do with their word part status. After all, not even words can normally be extracted. What is extracted (topicalized, *wh*-fronted) is entire referential phrases, not individual words. And languages usually put strong restrictions on the kinds of phrases that can be extracted. In English, for example, pronominal modifiers can never be extracted. For example, the phrase *lavishly decorated* in (27a) cannot be questioned or topicalized.

(27) a. I bought a lavishly decorated cake.
      b. *What kind (of) did you buy a _ cake?

Thus, the fact that **Which school did you see the -bus?* is not possible in English has nothing to do with the putative word status of *school-bus*, and analogous considerations apply to Italian *nave ospedale*.

3.9. Morphophonological idiosyncrasies

It is generally said that morphophonological idiosyncrasies are very common in combinations of stems and affixes, but should not occur in host-clitic combinations (e.g. Zwicky & Pullum 1983: 505). Thus, the existence of

morphophonological idiosyncrasies could be taken as a sufficient criterion for morphosyntactic wordhood. But while it is quite possible that this is a general tendency, we do find morphophonological idiosyncrasies in what by other criteria look like host-clitic combinations. These idiosyncrasies may affect the short element (clitic/affix) or the host. One case where the short element shows an alternation is the ergative marker in Pitjantjatjara (Pama-Nyungan), which is *-tu/-'u/-tju* after a consonant-final host and *-ngku* after a vowel-final host. It always follows the last word of the ergative NP (Bowe 1990):

(28) a. *titja-ngku*   'teacher (ergative)' (p. 22)
          teacher-ERG
      b. *tjitji puḻka-ngku*   'big child (ergative)' (p. 30)
          child big-ERG
      c. *tjitji ninti pukuḻ-ṯu*   'clever happy child (ergative)' (p. 30)
          child clever happy-ERG
      d. *wati nyanga pukuḻpa mulapa-ngku*
          man that   happy   very-ERG
          'that very clever man (ergative)' (p. 31)

The ergative marker is a clitic by the criterion of non-selectivity. The alternation is phonologically conditioned, but the conditioning environment does not explain the alternation, which is idiosyncratic (see Anderson et al. 2006 for more cases of idiosyncrasies in edge case marking).

Morphophonological alternations of clitic-like elements may also be lexically conditioned, e.g. in the Russian consonant-final prepositions *bez* 'without', *v* 'in', *iz* 'out of', *k* 'to', *nad* 'above', *ot* 'from', *pered* 'before', *pod* 'under', and *s* 'with; from'. These prepositions sometimes occur in an alternative form ending in *-o* (*bezo, vo, izo, ko*, etc.) when the following word begins with two consonants. However, the precise conditioning is quite complex, and to a large extent lexically conditioned. For example, the preposition *s* 'with' must take the form *so* when combining with *mnoj* 'me' (*so mnoj* 'with me'), but with most other words that begin with *mn-*, it takes the form *s* (e.g. *s mneniem* 'with the opinion'). This would argue for prefix status for these prepositions, a very nonstandard view (the prepositions may combine with nouns and all kinds of prenominal modifiers, thus they are nonselective).

The converse case, of a short element with apparent clitic status conditioning morphophonological alternations on the host, is attested in Polish.

Polish has 1st and 2nd person clitics that occur when the verb is in the past tense form. These may occur on the verb or on some constituent earlier in the clause, as illustrated in (29).

(29)   Polish (Aguado & Dogil 1989: 191)
    a.   *Kiedy Jank-owi   pomogł-em. . .*
         when Janek-DAT helped-1SG
    b.   *Kiedy Jank-owi-m   pomógł. . .*
         when Janek-DAT-1SG helped
    c.   *Kiedy-m Jank-owi pomógł. . .*
         when-1SG Janek-DAT helped
         'When I helped Janek.'

When the clitic occurs on the verb, as in (29a), the verb stem has a different vowel ([o] as opposed to [u] in 29b–c; [u] that alternates with [o] is spelled *ó* in Polish). This alternation is lexically conditioned in Polish, so it qualifies as idiosyncratic. This would argue that the first-person singular marker -(*e*)*m* is a suffix, but its mobility seems to show that it is not a suffix.[18] Another case where a mobile clitic-like element influences the morphophonology of the host was seen earlier in (8) (Lithuanian reflexive marker *si*).

3.10. Deviations from biuniqueness (one-meaning–one-form)

Since Matthews (1972), morphologists have often emphasized that morphological patterns do not necessarily show a one-to-one-correspondence ('biuniqueness') between formatives and meaning elements (see e.g. Anderson 1992, Aronoff 1994, Stump 2001, Spencer 2004). The four deviations from biuniqueness in (30) are among the most frequently mentioned.

(30)   a.   zero marking (= no form, one meaning)
        e.g. Russian *knig-a* 'book (NOM.SG)', *knig-Ø* 'books (GEN.PL)'
    b.   multiple exponence (= several forms, one meaning)
        e.g. Latin *rēg-s-ī* (spelled *rexi*) 'I directed' vs. *reg-ō* 'I direct', where
        the Perfect meaning is expressed three times (stem vowel length,
        suffix -*s*, and 1SG suffix -*ī*)

---

[18] One might argue that -*em* is a suffix in (29a), but -*m* is a clitic word in (29b–c), i.e. that these cases do not show the same construction and hence do not show mobility of the same entity. But since there is no good criterion for deciding what is the same construction and what is a different construction, the mobility criterion cannot be applied consistently (as we saw in Section 3.2).

   c.   cumulative exponence (= one form, several meanings)
       e.g. Modern Greek *katalavén-ete* 'is understood (3SG.PASS)', where
       -ete expresses both 3SG and passive
   d.   morphomic patterns (= one form, no meaning)
       meaningless stem markers, e.g. the Latin 'third stem' (Aronoff 1994),
       as in *scribo* 'I write', *scrip-t-um* 'written', *scrip-t-or* 'writer', *scrip-t-ura*
       'writing' (third stem *scrip-t-*)

Such deviations from biuniqueness have been used as an argument for a realizational approach to morphology, where formatives realize pre-existing features of words, rather than introducing them, and for an inferential approach, where the relations between roots and inflected word forms are expressed by rules, rather than by lexical entries (Stump 2001: ch. 1). The morpheme (or morph) of the earlier structuralist tradition plays little or no role in this approach, and the word occupies a central place. The writings on the inferential-realizational approach rarely discuss the syntax–morphology distinction, but implicitly they strongly convey the idea that morphology is special in this regard and quite different from syntax.[19]

However, deviations from biuniqueness of the kind mentioned in (30a–d) are not confined to morphology (Spencer 2001). The relationship between form and meaning can be quite complex in syntax as well, even though the simplest textbook examples may give a different impression. In this section, we will see that all four deviation types exist in syntax as well.

3.10.1. *Zero marking*

Meaningful zeroes are widespread not only in what is usually called morphology, but also in what is usually called syntax (e.g. Mel'čuk 1988). A simple example is the zero article in possessed noun phrases in English:

  (31)   the house / the small house / the house across the street / Ø Tomek's
        house

When a possessor NP precedes a noun, no definite article is present, but the meaning is still definite. *Tomek's house* means 'the house that belongs to Tomek', not 'a house that belongs to Tomek' (this would be expressed

---

[19] Faarlund (2009: 622–623) is quite explicit: "[clitics and inflectional affixes] are basically distinct, especially if we adopt an inferential model of morphology, whereby inflectional forms of the word are created by morphosyntactic feature specification, rather than just added as separate morphemes".

by *a house of Tomek's*). There is no overt expression here that would correspond to the definite meaning, but the definite meaning can be inferred with certainty, much as in the morphological example of Russian *knig-Ø* 'books (GEN.PL)'.

Another example of a meaningful zero comes from Bulgarian, cited by Spencer (2001), who lists quite a few deviations from biuniqueness in syntax. In (32a–b), we see two different tense-aspect-mood forms, one with and one without the copula *e* 'is'. There is clearly a meaning difference between the two forms, and the 'apparently' meaning must be attributed to the absence of the copula in (32b).

(32)   Bulgarian (Spencer 2001: 296)
     a.   Present Perfect Indicative:
         Tja e   otišla.
         she be.3SG.PRS leave.PTCP.F.SG
         'She has left.'
     b.   Aorist Renarrated:
         Tja otišla.
         she leave.PTCP.F.SG
         'She left (apparently).'

3.10.2. *Multiple exponence*

Multiple exponence, too, is more common in syntax than most morphologists seem to suspect. A simple example is the English Perfect, or the French Passé composé in (33):

(33)   a.   we have eaten (Spencer 2001: 281)
     b.   *nous avons mangé* (Martinet 1968: 296)
         'we ate'

Here the perfect or past-like meaning is jointly expressed by the auxiliary verb *have/avoir* and by the past participle form of the verb. Neither *have/avoir* alone nor the past participle alone can be said to carry the perfect/past meaning, because in other contexts these forms have different meanings. Only when they come together do we get the right meaning. Another example is the preposition *s* in Russian, which expresses the comitative 'with' sense only when it combines with a noun phrase in Instrumental case:

(34)   *s brat-om* 'with brother'

When the same preposition occurs with a Genitive noun phrase (*s brat-a*), the meaning is completely different ('from'), and when the Instrumental occurs with other prepositions (such as *pod* 'under', *nad* 'above'), again the meaning is totally different. In conservative Indo-European languages, this kind of syntactic multiple exponence is thus very common.

3.10.3. *Cumulative exponence*

Cumulative exponence, too, has generally been associated with inflectional morphology since Matthews (1972) coined the term. For syntax, a corresponding term *superlexeme* (Zwicky 1992: 364, Stump 2001: 13) has been coined, but the phenomenon is much less widely discussed and is generally passed over in silence by textbooks. A superlexeme is a single word that fills two syntactic positions, giving rise to a two-to-one relation between terminal nodes and words. A well-known example is French *du* 'of the (masculine)', which contrasts with the regular sequence *de la* with feminine singular nouns:

(35) a. *le château du   roi*
          the castle   of.the king
          'the king's castle'
     b. *le   château de la   reine*
          the castle   of the queen
          'the queen's castle'

  A less well-known example is Tagalog *kita*, which is simultaneously an agent pronoun and a patient pronoun (Schachter & Otanes 1972: 89):

(36) a. *Nakita ko siya.*   'I (*ko*) saw him (*siya*)'
     b. *Nakita ko kayo.*   'I (*ko*) saw you-all (*kayo*)'
     c. *Nakita ka niya.*   'He (*niya*) saw you (*ka*)'
     d. *Nakita kita.*   'I saw you'

As in French, where the regular combination **de le* never occurs, the regular combination of agent pronoun *ko* and patient pronoun *ka* is ungrammatical (**Nakita ko ka*, **Nakita ka ko*).

3.10.4. *Morphomic patterns*

Finally, corresponding to meaningless stem markers in morphology that simply have to be there (e.g. the Latin 'third stem', as in *scrip-t-*, which

cannot be assigned any consistent meaning), syntax has meaningless *periphrasis forms* (Haspelmath 2000: 662, Spencer 2001). A particularly striking example from Slovene is provided by Spencer (2001: 294):

(37)   a.   *Pohvali-la   sem.*
             praise-PTCP be.1SG.PRS
             'I praised.'
       b.   *Pohvali-la   bom*
             praise-PTCP be.1SG.FUT
             'I shall praise.'

In Slovene, both the Past tense and the Future tense are formed with the *l*-participle (the *-la* form in (37) is the feminine form), which has no meaning itself and is a pure periphrasis form. When combining with the Present tense form of the copula (as in 37a), a past-tense meaning results, and when combining with the Future tense form of the copula (as in 37b), a future-tense meaning results.

As Spencer (2001) stresses, this pattern is completely analogous to Aronoff's most prominent morphomic pattern in morphology. Aronoff also cites inflectional classes as 'morphomic' phenomena. Such classes of lexemes that behave differently also occur in syntactic periphrastic constructions. A well-known example is auxiliary selection in West Germanic, Italo-Romance and Gallo-Romance languages. Just as speakers have to remember more or less arbitrary lexical classes for plural formation, they have to remember more or less arbitrary lexical classes for Perfect tense formation (e.g. French *j'ai été* 'I have been, I was' vs. *je suis allé* 'I have gone, I went').

3.10.5. *Realizational morphology and constructional syntax*

As was mentioned earlier, deviations from biuniqueness that are commonly found in morphology have often been used as an argument for a 'realizational' or 'paradigm-based' approach to morphology, where the formatives are not directly associated with meanings (as in 'morpheme-based' approaches), but are introduced in a top-down fashion via rules that take a word's feature array as input. The feature array is determined by the structure of the paradigm.

At first glance, it appears as if this model can only work for the smaller word-size units of morphology, not for phrases and clauses. Especially the

idea that inflection is best described by 'exemplary paradigms' (Matthews 1972) seems impossible to carry over to syntax. However, Matthews was dealing with Latin, which has a fairly modest number of inflected forms per verb. For Latin-type languages, one may even assume that all inflected forms of all verbs are stored in memory, but there are of course many languages with richer morphology. According to Hankamer (1989), each Turkish verb has over a million different inflected forms, so that the human memory capacity would not be sufficient to memorize all forms. Thus, clearly, 'paradigm' must be understood in a more abstract sense – we cannot describe Turkish verbal inflection by listing a complete paradigm of a single verb. This more abstract sense of paradigm can of course be applied to syntax as well, just as notions developed in syntax (such as trees) can be applied in morphology (and Spencer 2001 makes concrete proposals for realization-based syntactic rules).

Clearly, the form–meaning relationship is often straightforward and compositional, but it is also often more complex. For the latter cases, morphologists have used paradigms and realization-based rules, and syntacticians have used constructional idioms. The similarity between realization-based morphology and construction-based syntax has recently been emphasized especially by Gurevich (2006) and Booij (2010). As far as I have been able to determine, the differences between them mostly derive from different traditions, not from any substantive differences. Thus, if the morphology–syntax dichotomy is abandoned, both can be merged into a realizational/constructional morphosyntax.

4. Combining the criteria: persuasion by test batteries

Since there is not one single criterion that identifies words, and attempts at coming up with a set of jointly necessary and sufficient conditions have not been successful either, in practice linguists have often adopted the strategy of persuasion via test batteries. In this strategy, a number of criteria are selected and applied, and in the published accounts usually all of them point in the same direction. The more criteria converge, the more persuasive the argument becomes, but the method is not rigorous, because the criteria can be selected opportunistically by the author.[20] Table 1 provides

---

[20] Some authors acknowledge the non-coincidence of the criteria, e.g. Börjars (1998: 44): "The behaviour of elements is often not totally consistent. This means that in order to

Table 1. Nine studies that examine wordhood using test batteries

	Zwicky & Pullum 1983	Kanerva 1987	Bresnan & Mchombo 1995	Ackema & LeSourd 1997	Monachesi 1999	Harris 2000	Milićević 2005	Lieber & Scalise 2006	Bickel et al. 2007
Free occurrence				+			+		
External mobility and internal fixedness	+			+	+	+			
Uninterruptibility				+					+
Non-selectivity	+	+			+	+	+		+
Non-coordinatability			+	+	+		+	+	+
Anaphoric islandhood				+				+	
Nonextractability				+				+	
Morphophonological idiosyncrasies	+	+				+	+	+	
Deviations from biuniqueness									+

an overview of the way in which different criteria have been applied by nine different authors (or author teams) in such test batteries.

There is no doubt that the application of such test batteries yields important insights into the structure of the languages, but it does not provide solid evidence for a unitary word notion that can be applied across languages and that gives us a reliable way of distinguishing between morphology and syntax.

5. The word as a language-specific concept

The difficulties with a universally applicable word concept have been known for quite a while, and many linguists have concluded that the word cannot be defined universally, but only in a language-specific way:

(38) a. Bazell (1958: 35): "Now there is perhaps no unit over which there
             is less agreement than the word. If there is any agreement at all,

---

arrive at the conclusion that an element is either a clitic or an affix, certain criteria must be assumed to be less crucial." But if different authors assume that different criteria are 'less crucial', no agreement can be reached.

  it is that the word has to be differently defined for each language analysed."

b. Martinet (1960: 112): "Il serait vain de chercher à définir précisément cette notion [i.e. la notion de mot] en linguistique générale. On peut tenter de le faire dans le cadre d'une langue donné." ['It would be futile to attempt to define this notion [i.e. the word notion] in a precise way in general linguistics. One can try to do it within a particular language.']

c. Lyons (1968: 206): "It follows from these facts that what we call 'words' in one language may be units of a different kind from the 'words' in another language."

d. Wurzel (1984: 35): "Was in einer Sprache ein morphologisches Wort ist, wird durch die einzelsprachlichen morphologischen Regularitäten bestimmt." ['The language-specific morphological regularities determine what is a word in a language.']

e. Spencer (2006: 129): "There may be clear criteria for wordhood in individual languages, but we have no clear-cut set of criteria that can be applied to the totality of the world's languages. . ."

On this view, there is thus no general word concept, and the term *word* potentially has as many different meanings as there are languages. The same should apply to the terms *morphology* and *syntax*, which cannot be defined in any other way than in terms of some word concept.

  On such a view, the claim that "all languages have words" (Radford et al. 1999: 145) would be interpretable only in the weaker sense that "all languages have a unit which falls between the minimal sign and the phrase" (Bauer 2000: 255).[21] But this claim is very weak indeed if it is not specified what criteria can be used for identifying such a unit. In fact, given the criteria in Section 3, it is clear that in most languages one could set up a fairly large number of such intermediate units if one wanted. One

---

[21] Incidentally, another view for which the word notion ends up being non-universal is the idea that the cross-linguistic word concept is not applicable to all languages. For example, Hockett (1944: 255) claims that "there are no words in Chinese", van Wyk (1968: 557) claims that "French is an example of a language with fairly low overall word autonomy, whereas the word autonomy of English, Dutch and German is comparatively high in general", and Mugdan (1994: 2552) says that "some languages evidently lack a word level altogether" (cf. also Milewski 1951: 248–249). These authors thus seem to assume the word as a cross-linguistically valid category, but as a category that not all languages need to instantiate. This would entail that only some languages have a syntax–morphology distinction, while other languages lack it. This idea is weaker than the widespread assumption criticized throughout this article (that all languages have such a distinction), but to be viable, this approach would still have to come up with a way of identifying words consistently across languages.

might posit units such as (complex) stems, words with level-I affixes, words-plus,[22] words with type-III incorporation,[23] clitic groups, tight compounds, loose compounds, tightly knit phrases, configurational phrases, non-configurational phrases, and many further units for which new terms would have to be made up. Some of these will be very well motivated, some less well. The basic problem remains the same: The units are defined in a language-specific way and cannot be equated across languages, and there is no reason to give special status to a unit called 'word'. What we *can* assert with great confidence is that all languages have different degrees of tightness of minimal sign combinations. There is no language that is both thoroughly isolating and thoroughly non-configurational: isolating languages tend to be highly configurational, and non-configurational languages tend to be highly synthetic. This generalization is highly significant, as it is not difficult to imagine a language in which all sign combinations are equally tight. The organization of human languages in terms of different degrees of tightness of sign combinations is quite possibly an innate design feature of language. But this does not mean that 'word', 'morphology' and 'syntax' are useful concepts for general linguistics.

6. The word as a fuzzy concept

Another way of addressing the problems with defining the concept of morphosyntactic word is to posit a fuzzy (or 'prototypical', or 'canonical') concept of word. Something like this has been done implicitly by many authors who have written on the issue. Already Sweet (1913) regarded the English articles *the* and *a* as 'half-words', and Wurzel (1984: 36) revives this under the heading of 'semi-word' (his examples are cases intermediate between traditional words and phrases, such as Icelandic *hund-s-ins* [dog-GEN-DEF-GEN] 'the dog's', with internal case agreement, or English *the Queen of England's mother*, with word-like edge-located inflection). A few quotations that explicitly admit fuzzy boundaries between words and phrases, or words (clitics) and affixes are given in (39).

(39)   a.   Bloomfield (1933: 181): "many forms lie on the border-line between bound forms and words, or between words and phrases; it is impossible to make a rigid distinction . . ."

---

[22] cf. Kageyama (2001).   [23] cf. Mithun (1984).

b. Lyons (1968: 204): "even if [the English article] is taken to be a word, it is not so 'fully' a word as other elements to which all the relevant criteria apply."

c. Mel'čuk (1993: 183): "puisque le concept [vague] d'autonomie . . . est sou-jacent au concept de mot-forme, nous pouvons en déduire que ce dernier est vague lui aussi" ['since the [vague] concept of autonomy underlies the concept of word-form, we can deduce that the latter is also vague']

d. Mugdan (1994): "the differences [between clitic and affix] are gradual and clear dividing lines are not easy to draw."

e. Aikhenvald (2002: 43): "These parameters provide us with a scalar definition of clitics: each prosodically deficient morpheme occupies a particular place within a multidimensional continuum, from a fully bound to a fully independent morpheme."

If 'word' is a fuzzy concept, it might still be universal, i.e. the conclusions drawn by the authors cited earlier in (38) might be premature. However, if 'word' is a fuzzy concept, the consequence is that the difference between words and phrases cannot be modelled by positing two separate components of grammar, morphology and syntax. One could continue using the notions of morphology and syntax, but as fuzzy concepts.

In order to show that a fuzzy concept of word is theoretically significant, one would have to demonstrate that grammatical units are not randomly distributed over the continuum between fully bound and fully independent units (cf. Aikhenvald's conception in (39e)), but that they cluster significantly. Figure 1 schematizes two different hypothetical situations that are both compatible with a fuzzy word concept and a continuum view of the morphology–syntax distinction.

In the first hypothetical situation (clustering distribution), there are three clearly discernible clusters: If the dimension along which the units differ (the boundness scale) can be quantified, the clustering can be demonstrated by statistical techniques. There are intermediate cases between the clusters of affixes, clitics, and independent words, but these are few and are just exceptions to the rule.

In the second hypothetical situation (random distribution), the units are randomly distributed over the boundness scale. One can arbitrarily decide to call certain segments of the scale 'affixes', 'clitics' and 'independent words' (cf. the vertical lines in Figure 1B), but since the units do not cluster, such a subdivision would have no theoretical significance (it might

A. Clustering distribution

· · · · · · · ·············· · ·· · ··· · · ·· ············· ·· · · · ·· · · ······ ············ ·· · · · · ·

          affixes   clitics   independent words

B. Random distribution

· ·· · · · · · ·· · ·· ··· · · · · ·· ·· · · ·· · · ·· · · · ·· · ·· · · · ·· · · · · ··

          affixes | clitics | independent words

Figure 1. Clustering vs. random distribution of grammatical units

have practical significance, of course, e.g. for orthographic conventions). One could just as easily subdivide the continuum into two, four, or five named segments.

Both of the hypothetical distributions in Figure 1 could be described as entailing a fuzzy word concept. The authors cited in (39) presumably assumed the clustering distribution rather than the random distribution, but we do not have the empirical evidence yet to show that a clustering actually obtains in real languages. Our intuitions that tend to favour the clustering distribution may well be coloured by the long tradition of looking at language structure in terms of spelling-defined words and clauses.

So this is an open question, and systematic empirical research on the basis of a diverse range of languages is needed before we can make a judgement. We should also be open to the possibility that other kinds of clusters, e.g. 'affixoid', 'clitic group', 'tight phrases', or 'stems' will turn out to be more significant than the word clustering. This would have the consequence that the primary division within morphosyntax would not be between morphology and syntax, but along other lines. The question is an empirical one, and should be approached in this way,[24] rather than a priori.

---

[24] Unfortunately, an empirical test is difficult, because the nature of the population to be sampled is not entirely clear, and it is not clear how the boundness scale should be quantified. The population from which we want to test a sample is the set of grammatical units in the world's languages, but this is of course an open-ended set. Single-morph affixes and clitics are relatively straightforward (because they constitute a finite, relatively small set), but we also need to compare these to larger units (word-level and phrase-level), and the elements of these larger units are not a closed class. The boundness scale could be quantified by applying some of the criteria of Section 3, but it is quite unclear how these criteria should be weighted, whether all of them are relevant, and what other criteria might be important. Different linguists will have very different views. Aikhenvald calls the continuum 'multidimensional' (cf. 39e), and finding clusters in a multidimensional space is even more difficult.

7. Consequences of the indeterminacy of word segmentation

That the word is difficult to define has long been recognized; quotations such as those in (40) abound in the literature.

(40) a. Jespersen (1924: 92): "What is a word? and what is one word? These are very difficult problems. . ."
      b. Langacker (1972: 37): "The word is a difficult notion to define."
      c. Matthews (1991: 208): "There have been many definitions of the word, and if any had been successful I would have given it long ago, instead of dodging the issue until now."

However, linguists have generally shied away from the seemingly unavoidable conclusion:

(41) Linguists have no good basis for identifying words across languages, and hence no good basis for a general distinction between syntax and morphology as parts of the language system.

One occasionally finds statements such as Garvin's "It is not so certain that we know how to isolate words, and hence how to separate morphology from syntax" (1954: 345), or Weber's "morphology and syntax are not distinct . . . word boundaries and morpheme boundaries should not be distinguished" (1983: 178),[25] but the vast majority of linguists have continued doing business as usual.[26]

    I maintain that at least three practical consequences follow from the recognition of the indeterminacy of word segmentation:

---

[25] Schwegler's (1990: 45) negative conclusion is even more explicit, though he does not mention the morphology–syntax division:

After generations of 19th- and 20th-century linguists had taken the 'word' largely for granted, structuralists set out to define what in popular as well as in scientific circles was regarded as the basic unit of speech . . . this lively discussion eventually led to the now generally accepted conclusion that (a) the 'word' cannot be defined by a single (or for that matter, multiple) common denominator, and (b) not all segments of speech are 'words' in the proper sense of the term.

However, I do not think that this conclusion was ever 'generally accepted' (certainly not in the 1980s).

[26] Jacobs (forthcoming) (whose paper came to my attention after this article had been completed) arrives at the same negative conclusion on the basis of German data, but at the end of his paper, he notes that for many purposes, doing business as usual is not problematic. I agree, but here I focus on those theoretical concerns for which the negative conclusions do have significant repercussions.

(42)   a.   Linguists who believe that words exist as a cross-linguistically identifiable category should try to find ways of identifying words consistently.
       b.   Linguists who are not committed should stop worrying about the morphology–syntax division.
       c.   Linguists of either category should be very careful with claims that presuppose the word concept.

In addition, of course, comparative linguists should try to determine empirically whether the various word criteria yield a clustering distribution, as discussed in Section 6, but this is not as urgent as (42a–c), because clusters might be found in all kinds of places in languages where no traditional notions lead us to suspect them. The really serious problem resulting from the indeterminacy of word segmentation is that linguists often presuppose the word concept and the morphology–syntax division, and even try to use it for explanatory purposes. Some of the contexts in which the word as a cross-linguistic category and/or the syntax–morphology division are presupposed are listed in (43).

(43)   a.   the Lexical Integrity Hypothesis (e.g. Bresnan & Mchombo 1995; see Section 8 below)
       b.   claims about the nature of creole languages (e.g. McWhorter 2001, 2005, referring to lack of inflectional morphology)
       c.   claims about variable complexity of language structure (e.g. Sampson 2009)
       d.   claims about an analyticity preference in Second Language Acquisition (e.g. Klein & Perdue 1997: 311)
       e.   claims about morphological change driving syntactic change (e.g. Lightfoot, ed. 2002)
       f.   claims about a general preference for suffixing over prefixing (e.g. Greenberg 1957, Bybee et al. 1990)
       g.   typological claims about affixal vs. periphrastic marking in typology, e.g. Dahl & Velupillai (2005, on inflectional future tense), Iggesen (2005, on number of inflectional cases), Dryer (2005c, on pronominal possessive affixes)

All of these claims are problematic as they can be tested only with a good definition of morphosyntactic wordhood. Some of them may well turn out to be correct in some version, once the nature of the claim has been clarified, and others may turn out to be quite illusory. For example, Sampson (2009: 3) cites the huge number of inflected verbal forms in Archi (about 1.5 million according to Kibrik 1998) as an evident example of 'complex-

ity', but he does not say that the vast majority of these inflected forms are completely regular and predictable by rule. Any isolating language of course has a vast number of possible constructs involving verbs and various functional elements (negation, tense, aspect, mood, subject and object pronouns, etc.), resulting in exactly the same complexity, differing primarily in spelling. Speakers of contemporary Western European languages traditionally find languages with rich morphology such as Latin difficult, and languages with even longer words such as Turkish or Archi look even more forbidding, but this impression should not be confused with the idea that Turkish or Archi are in some general sense more complex than Spanish or Swedish. Similarly, creole languages are often thought of as lacking inflectional morphology (McWhorter 2001, 2005), and it is clear that they generally lack the inflectional affixes of their lexifiers. But creole verbs are typically associated with preverbal tense-aspect, modality and polarity markers (cf. (44) from Seychelles Creole; Susanne Michaelis, p.c.) which might well be prefixes rather than separate words.

(44)   *Nou pa ti pe manz.*
       we NEG PST PROG eat
       'We were not eating.'

8. The notion of lexical integrity

In structuralist linguistics of the first half of the 20th century, the word concept was downplayed by many theoretically oriented linguists (cf. Schwegler 1990: 45, Albrecht 2002). In American structuralism, "the traditional word was not part of the final formal analysis" (Robins 2002: 143), and the European structuralists Charles Bally and André Martinet used the term 'word' only in inverted commas (Matthews 2002: 266). This view was inherited by early generative grammar, which also made no distinction between syntax and morphology. But since the 1970s, the idea of a syntax–morphology distinction became prevalent again in generative circles, under the heading of 'lexicalism' or 'lexical integrity'. This was not prompted by any progress in defining the word in universal terms; on the contrary, such attempts were largely given up.[27] But linguists now focused

---

[27] Characteristically, Langacker (1972) was the last prominent linguistics textbook that included a substantial discussion of the word concept (pp. 36–55). Later textbooks of

on some of the differences in behaviour between word parts and phrasal elements, and soon the idea that there is a principled distinction between syntactic and morphological regularities was formulated and met with widespread acceptance. The Lexical Integrity Hypothesis was variously formulated as in (45).

(45)   a. Lapointe (1980: 8; Generalized Lexicalist Hypothesis)
           No syntactic rule can refer to elements of morphological structure.
       b. Selkirk (1982: 70; Word Structure Autonomy Condition)
           No deletion or movement transformation may involve categories of both W[ord]-structure and S[entence]-structure.
       c. Anderson (1992: 84; Lexicalist Hypothesis)
           The syntax neither manipulates nor has access to the internal structure of words.

The problem with such formulations is that in the absence of a full characterization of syntax (or S-structure) and morphology (or word), they are partly circular (cf. Sproat 1988: 298–299, Julien 2002: 28). If we found that something that is written as one word has parts that can be manipulated in the sense of being elliptied in coordination, fronted by topicalization rules, extended by additional modifiers, etc., then we would probably conclude that the item is not in fact one word. The claims in (45) thus come dangerously close to saying that linguistic sign combinations that cannot be manipulated by syntactic rules cannot be manipulated by syntactic rules. In any event, the empirical import of such claims is not particularly clear.[28]

Another way of stating the notion of lexical integrity that avoids the danger of circularity is the claim that "the principles that regulate the internal structure of words are quite different from those that govern sentence structure" (Katamba 1993: 217), or that "words are built out of different structural elements and by different principles of composition than syntactic phrases" (Bresnan & Mchombo 1995: 181). But such claims are very weak: While it may be that words and phrases are 'different' in some

general linguistics or of syntax simply presuppose that words exist as a universally valid category (though morphology textbooks often include a discussion of the word concept).
[28] Moreover, there is a large amount of potential counterevidence to lexical integrity that has been discussed in the literature; cf. Lieber & Scalise (2006) and Booij (2009) for recent reviews from a sympathetic perspective, and Julien (2002: ch. 1) for a critical assessment from a minimalist perspective. Booij (2009: 97) states that "the word remains an essential unit for stating regularities" but does not try to argue for this position.

ways, they are also very similar in many ways, and there are also striking differences between different kinds of words, as well as differences between different kinds of phrases. That the distinction between words and phrases should have a special status is not warranted by the evidence that has been presented in favour of lexical integrity.

9. What does this mean in practice?

Talking about words is deeply entrenched in linguists' habits, and a frequently asked question is: How do we go about our everyday grammatical analysis (e.g. as field linguists, as computational linguists, as psycholinguists) if we cannot rely on theoretical morphologists and syntacticians to provide us with a well-motivated universally applicable distinction between morphology and syntax, between words and phrases?

There are two answers, a simpler one and a more complex, more advanced one. The simpler answer is that nothing needs to change in our way of talking about our entities, as long as we change our way of thinking about them. We can simply continue calling an expression like *book+s* a word, and expressions like *the+book* a phrase, as long as we do not thereby imply that words and phrases necessarily have different properties (in particular, fundamentally different properties). After all, we continue using the usual spelling in many cases, simply because it is convenient, without implying that the orthographic representation has any theoretical status. A word would then simply be any linguistic expression that has a space before it and after it and no space in the middle, either in the conventional spelling or (for unwritten languages or languages with no spacing in the spelling) in the linguist's transcription. This would solve most of the practical problems of descriptive linguistics, because in the great majority of concrete research contexts, the kinds of issues discussed in this article are not relevant anyway.

But in cross-linguistic and general contexts (such as those of (43)), the issues are relevant, and here we need a more complex and advanced answer. I propose that for the purposes of cross-linguistic comparison, we limit ourselves to more primitive concepts that are readily definable in cross-linguistic terms (i.e. to comparative concepts, cf. Haspelmath 2010), such as those in (46).

(46) a. formative: a minimal coherent set of phonological features that plays a role in the language system (= a minimal sign)
      b. morph: a formative that biuniquely expresses a meaning
      c. root: a morph with a concrete meaning
      d. construct: a set of formatives that together play a role in the language system
      e. bound construct: a construct that cannot occur on its own as a complete utterance
      f. free construct: a construct that may occur on its own as a complete utterance

Thus, all words and phrases are constructs that can be decomposed into their component constructs until the level of formatives is reached. All clitics and affixes are bound constructs, and all lexical items have citation forms that are free constructs. Many more of these cross-linguistically applicable comparative concepts can easily be created (separatable constructs, coordinatable constructs, etc.).

These definitions are not without problems either, of course, but they do not appeal to language-specific properties, and they avoid the danger of bias from our writing habits. Reformulating cross-linguistic generalizations in terms of truly comparative concepts such as those in (46) will not be an easy task, but in view of the difficulties with the traditional word concept, there is probably no alternative to it.

To conclude this section, let me illustrate the practical consequences by answering four questions that one reviewer asked and that some readers may wonder about:

(47) *Question*: If there are no words, how can children have a one-word stage in talking? (why not a one-stem stage?)
      Answer: The 'one-word stage' is probably really a 'one-root stage', but this probably differs from language to language, and even from child to child.

(48) *Question*: If there are no words, how can those languages which use an orthography in which word-boundaries are marked agree on where they belong?
      Answer: 'Words' as language-specific units are often unproblematic (Section 5), but the criteria employed in different languages are often very different. And the rules may be idiosyncratic, without clear relation to the language system (Section 2.3).

(49)   *Question*:   If there are no words, how come dictionaries are so useful?
        Answer: Dictionaries generally list the roots and unpredictable morph combinations of a language, i.e. words and many kinds of phraseological units. The idea that dictionaries are just lists of 'words' is too simplistic anyway.

(50)   *Question*:   If there are no words, what is a word-for-word translation?
        Answer: Morpheme-by-morpheme translations and phrase-by-phrase translations are also often practiced, and in many cases they are more useful than word-for-word translations. Words do not have special status in literal translations.

10. Conclusion

In the closing session of the 6th International Congress of Linguists, held in Paris in 1948, congress president Joseph Vendryes remarked that modern linguistics was in a crisis, and that linguists were not even in agreement on what a word is, one of the fundamental concepts of their object of studies (cf. Togeby 1949: 97). In this article, I do not offer a more optimistic conclusion on the issue of defining the word, but I do not think that this is an expression of a crisis of our field. On the contrary, as long as we are biased by writing habits and by the structures of a few languages that happen to be very well studied and widely known, our field is not fully mature. The very search for a definition of the concept 'word' seems to be guided by the unstated presupposition that something like the word must exist in languages, just as it exists in alphabetic writing. But a scientific approach to language structure should imply that we do not take any traditional concept for granted, and that we posit only those categories that we can argue for in describing language structure. Moreover, we should describe each language in its own terms, rather than assume aprioristically that a concept that has been found useful for one language must also be applicable for another language.

   I conclude from the arguments presented in this article that there is no definition of 'word' that can be applied to any language and that would yield consistent results that are in accord with our writing habits. This is not a problem for descriptive (i.e. language-specific) linguistics, because ad hoc notions can be easily created for each particular language, and often it will be possible to define 'word in language X' in such a way that the spelling is predicted by the definition.

However, we also want to compare languages with respect to issues such as analyticity and syntheticity, complexity, and types of morphological marking (e.g. cumulative vs. separatist encoding, stem alternation vs. pure concatenation). For such questions, we need comparative concepts that are universally applicable (cf. Section 9). This is an important task for future research.

The conclusion that we do not know what words are also means that we have no good basis for a morphology–syntax distinction. The part of (the study of) language structure that deals with sign combinations can be called morphosyntax, and for theoretical purposes this is currently best viewed as a unitary domain.

References

Ackerman, Farrell & LeSourd, Philip. 1997. Toward a lexical representation of phrasal predicates. In Alex Alsina, Joan Bresnan & Peter Sells, eds. *Complex predicates.* Stanford: CSLI, 67–106.

Aguado, Miquel & Grzegorz Dogil. 1989. Clitics in lexical phonology. *Linguistische Berichte* 120: 99–116.

Aikhenvald, Alexandra Y. 2002. Typological parameters of clitics, with special reference to Tariana. In R. M. W. Dixon & Alexandra Y. Aikhenvald, eds. *Word: A cross-linguistic typology.* Cambridge: Cambridge University Press, 1–41.

Albrecht, Jörn. 2002. Das Wort im europäischen Strukturalismus. In D. Alan Cruse, Franz Hundsnurscher, Michael Job & Peter Rolf Lutzeier, eds. *Lexicology: An international handbook on the nature and structure of words and vocabularies*, vol. I. Berlin: Walter de Gruyter, 138–143.

Anderson, Stephen. 1992. *A-morphous morphology.* Cambridge: Cambridge University Press.

Anderson, Stephen R., Lea Brown, Alice Gaby & Jacqueline Lecarme. 2006. Life on the edge: There's morphology there after all. *Lingue e Linguaggio* 1: 1–16.

Aronoff, Mark. 1994. *Morphology by itself.* Cambridge, MA: MIT Press.

Aronoff, Mark & Kirsten Fudeman. 2005. *What is morphology?* Oxford: Blackwell.

Atoyebi, Joseph Dele. 2010. *A reference grammar of Oko*. Cologne: Rüdiger Köppe Verlag.

Aydemir, Yasemin. 2004. Are Turkish preverbal bare nouns syntactic arguments? *Linguistic Inquiry* 35: 465–474.

Bally, Charles. 1913 [1965]. *Le langage et la vie.* Geneva: Droz.

Bauer, Laurie. 1988. *Introducing linguistic morphology.* Edinburgh: Edinburgh University Press.

Bauer, Laurie. 2000. Word. In Geert E. Booij, Christian Lehmann & Joachim Mugdan, eds. *Morphology: An international handbook on inflection and word-formation*, vol. 1, 247–257. Berlin: Walter de Gruyter.

Bauer, Laurie & Rodney Huddleston. 2002. Lexical word-formation. In Rodney Huddleston & Geoffrey K. Pullum et al. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press, 1621–1721.

Bazell, C. E. 1958. *Linguistic typology: An inaugural lecture delivered on 26 February 1958*. London: School of Oriental and African Studies.

Bermúdez-Otero, Ricardo & John Payne. Forthcoming. There are no special clitics. To appear in: Alexandra Galani, Glyn Hicks & George Tsoulas, eds. *Morphology and its interfaces*. Amsterdam: Benjamins.

Bickel, Balthasar, Goma Banjade, Martin Gaenszle, Elena Lieven, Netra Paudyal, Ichchha Purna Rai, Manoj Rai, Novel Kishor Rai & Sabine Stoll. 2007. Free prefix ordering in Chintang. *Language* 83: 43–73.

Bisetto, Antonietta & Sergio Scalise. 1999. Compounding: Morphology and/or syntax? In Lunella Mereu, ed. 1999. *Boundaries of morphology and syntax*. Amsterdam: Benjamins, 31–48.

Blevins, Juliette. 2001. *Nhanda: An aboriginal language of Western Australia*. Honolulu: University of Hawaii Press.

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Boas, Franz. 1911. Introduction. In Franz Boas, ed. *Handbook of American Indian languages*, part 1 (Bureau of American Ethnology, Bulletin 40). Washington: Government Printing Office, 1–83.

Booij, Geert. 1985. Coordination reduction in complex words: A case for prosodic phonology. In Harry van der Hulst & Norval Smith, eds. *Advances in nonlinear phonology*, vol. 7. Dordrecht: Foris, 143–160.

Booij, Geert. 2009. Lexical integrity as a morphological universal: A constructionist view. In Sergio Scalise, Elisabetta Magni & Antonietta Bisetto, eds. *Universals of language today*. Berlin: Springer, 83–100.

Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.

Börjars, Kersti. 1998. *Feature distribution in Swedish noun phrases*. Oxford: Blackwell.

Bowe, Heather J. 1990. *Categories, constituents, and constituent order in Pitjantjatjara: An aboriginal language of Australia*. London: Routledge.

Bresnan, Joan. 1974. On the position of certain clause-particles in phrase structure. *Linguistic Inquiry* 5, 614–619.

Bresnan, Joan & Sam M. Mchombo. 1995. The lexical integrity principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13: 181–254.

Broadwell, George Aaron. 2008. Turkish suspended affixation is lexical sharing. In Miriam Butt & Tracy Holloway King, eds. *Proceedings of the LFG08 Conference*. Stanford: CSLI Publications. http://www.csli-publications.stanford.edu/.

Bybee, Joan L., William Pagliuca & Revere D. Perkins. 1990. On the asymmetries in the affixation of grammatical material. In William Croft, Keith Denning & Suzanne Kemmer, eds. *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th Birthday*, 1–42. Amsterdam: John Benjamins.

Chaves, Rui P. 2008. Linearization-based word-part ellipsis. *Linguistics and Philosophy* 31: 261–307.

Coseriu, Eugenio. 1964. Pour une sémantique diachronique structurale. *Travaux de linguistique et de littérature* 2: 139–186.

Dahl, Östen & Viveka Velupillai. 2005. The future tense. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie, eds. *The world atlas of language structures*. Oxford: Oxford University Press.

Di Sciullo, Anne Marie & Edwin Williams. 1987. *On the definition of word*. Cambridge: MIT Press.

Dixon, R. M. W. 1977. Some phonological rules of Yidiɲ. *Linguistic Inquiry* 8: 1–34.

Dixon, R. M. W. & Alexandra Y. Aikhenvald. 2002. Word: A typological framework. In R. M. W. Dixon & Alexandra Y. Aikhenvald, eds. *Word: A cross-linguistic typology.* Cambridge: Cambridge University Press, 1–41.

Dryer, Matthew S. 2005a. Order of genitive and noun. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie, eds. *The world atlas of language structures.* Oxford: Oxford University Press, 350–353.

Dryer, Matthew S. 2005b. Order of adposition and noun phrase. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie, eds. *The world atlas of language structures.* Oxford: Oxford University Press, 346–349.

Dryer, Matthew S. 2005c. Position of pronominal possessive affixes. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie, eds. *The world atlas of language structures.* Oxford: Oxford University Press, 234–237.

Erdal, Marcel. 2007. Group inflexion, morphological ellipsis, affix suspension, clitic sharing. In M. M. Jocelyne Fernandez-Vest, ed. *Combat pour les langues du monde: hommage à Claude Hagège*. Paris: L'Harmattan, 177–189.

Evans, Nicholas, Janet Fletcher & Belinda Ross. 2008. Big words, small phrases: Mismatches between pause units and the polysynthetic word in Dalabon. *Linguistics* 46: 89–129.

Faarlund, Jan Terje. 2009. On the history of definiteness marking in Scandinavian. *Journal of Linguistics* 45: 617–639.

Fuhrhop, Nanna. 2007. *Zwischen Wort und Syntagma: Zur grammatischen Fundierung der Getrennt- und Zusammenschreibung.* Tübingen: Niemeyer.

Fukushima, Kazuhiko. 1999. Bound morphemes coordination and bracketing paradox. *Journal of Linguistics* 35: 297–320.

Fulmer, S. Lee. 1990. Dual-position affixes from Afar: An argument for phonologically driven morphology. In Aaron I. Halpern, ed. *WCCFL* 9. Stanford: CSLI, 189–203.

Garvin, Paul L. 1954. Delimitation of syntactic units. *Language* 30: 345–348.

Greenberg, Joseph H. 1957. Order of affixing: A study in general linguistics. In Joseph Greenberg, ed. *Essays in linguistics*. Chicago: University of Chicago Press, 86–94.

Gurevich, Olya. 2006. *Constructional Morphology: The Georgian version.* Berkeley: Ph.D. dissertation.

Guthrie, Malcolm. 1948. *Bantu word division: A new study of an old problem.* (International African Institute Memorandum XXII). Oxford: Oxford University Press.

Haebler, Claus. 2002. Zur Etymologie von Benennungen für "Wort": Beispiele aus indogermanischen Sprachen Europas. In D. Alan Cruse, Franz Hundsnurscher, Michael Job, Peter Rolf Lutzeier, eds. *Lexicology: An international handbook on the nature and structure of words and vocabularies*, vol. I, Berlin: Walter de Gruyter, 78–83.

Halpern, Aaron. 1995. *On the placement and morphology of clitics.* Stanford: CSLI Publications.

Hankamer, Jorge. 1973. Unacceptable ambiguity. *Linguistic Inquiry* 4: 17–28.

Hankamer, Jorge. 1989. Morphological parsing and the lexicon. In William Marslen-Wilson, ed. *Lexical representation and process.* Cambridge, MA: MIT Press, 392–408.

Harris, Alice C. 2000. Where in the word is the Udi clitic? *Language* 76: 593–616.

Harris, Alice C. 2006. Revisiting anaphoric islands. *Language* 82: 114–130.

Haspelmath, Martin. 1993. *A grammar of Lezgian.* Berlin: Mouton de Gruyter.

Haspelmath, Martin. 2000. Periphrasis. In Geert Booij, Christian Lehmann & Joachim Mugdan, eds. *Morphology: A handbook on inflection and word formation*, vol. 1. (Handbücher zur Sprach- und Kommunikationswissenschaft) Berlin: Mouton de Gruyter, 654–664.

Haspelmath, Martin. 2002. *Understanding morphology*. London: Arnold.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86: 663–687

Himmelmann, Nikolaus P. 2006. Prosody in language documentation. In Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel, eds. *Essentials of language documentation*. Berlin: Mouton de Gruyter, 163–181.

Hockett, Charles F. 1958. *A course in modern linguistics*. New York: MacMillan.

Hualde, José Ignacio & Jon Ortiz de Urbina, eds. 2003. *A grammar of Basque* (Mouton Grammar Library). Berlin: Mouton de Gruyter.

Iggesen, Oliver. 2005. Number of cases. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie, eds. *The world atlas of language structures*. Oxford: Oxford University Press, 202–205.

Jacobs, Joachim. Forthcoming. Grammatik ohne Wörter? To appear in: *Sprachliches Wissen zwischen Lexikon und Grammatik*. (Jahrbuch des Instituts für Deutsche Sprache). Berlin: Mouton de Gruyter.

tionist perspective. *Word Structure* 2: 254–271.

Mathieu-Colas, Michel 1996. Essai de typologie des noms composés français. *Cahiers de lexicologie* 69: 71–125.

Matthews, Peter H. 2002. What can we conclude? In R.M.W. Dixon & Alexandra Y. Aikhenvald eds. *Word: A cross-linguistic typology*. Cambridge: Cambridge University Press, 266–281.

Matthews, Peter H. 1972. *Inflectional morphology*. Cambridge: Cambridge University Press.

Matthews, Peter H. 1974. *Morphology*. Cambridge: Cambridge University Press.

Matthews, Peter H. 1991. *Morphology*. 2nd edition. Cambridge: Cambridge University Press.

McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5: 125–166.

McWhorter, John H. 2005. *Defining creole*. Oxford: Oxford University Press.

Meillet, Antoine. 1921 [1982]. *Linguistique historique et linguistique générale*. Reprint, Paris: Champion, 1982.

Mel'čuk, Igor A. 1988. Syntactic, or lexical, zero. In Igor Mel'čuk, ed. *Dependency syntax: Theory and practice*. Albany: State University of New York Press, 303–337.

Mel'čuk, Igor A. 1993. *Cours de morphologie générale (théorique et descriptive)*. Vol. I. Montréal: Presses de l'Université de Montréal.

Milewski, Tadeusz. The conception of the word in languages of North American natives. *Lingua Posnaniensis* 3: 248–268.

Milićević, Jasmina. 2005. Clitics or affixes? On the morphological status of the future tense markers in Serbian. In Wolfgang Dressler, Dieter Kastovsky, Oskar E. Pfeiffer & Franz Rainer, eds. *Morphology and its demarcations: Selected papers from the 11th Morphology Meeting, Vienna, February 2004*. Amsterdam: John Benjamins, 39–52.

Miller, Philip H. 1992a. *Clitics and constituents in phrase structure grammar*. New York: Garland.

Miller, Philip H. 1992b. Postlexical cliticization vs. affixation: Coordination criteria. *Chicago Linguistic Society* 28: 382–396.

Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60: 847–894.

Monachesi, Paola. 1999. *A lexical approach to Italian cliticization*. Stanford: CSLI Publications.

Mugdan, Joachim. 1994. Morphological units. In R. E. Asher, ed. *The encyclopedia of language and linguistics*. Oxford: Pergamon Press, 2543–2553.

Nespor, Marina & Irene Vogel. *Prosodic phonology*. Dordrecht: Foris.

Noyer, Rolf. 1994. Mobile affixes in Huave: Optimality and morphological well-formedness. *West Coast Conference on Formal Linguistics* 12: 67–82. Stanford: CSLI.

O'Connor, Rob. 2002. Clitics and phrasal affixation in Constructive Morphology. In Miriam Butt & Tracy Holloway King, eds. *Proceedings of LFG02, National Technical University of Athens*. Stanford: CSLI Online Publications, 315–332 <http://csli-publications.stanford.edu/LFG/7/lfg02.html>.

Peterson, John. 2008. *Kharia: A South Munda language*. Vol. I: *Grammatical analysis*. Osnabrück: University of Osnabrück habilitation thesis.

Pike, Kenneth L. 1967. *Language in relation to a unified theory of human behavior* (Janua Linguarum, Series Maior 24). The Hague: Mouton.

Plank, Frans. 1994. Inflection and derivation. In R. E. Asher, ed. *Encyclopedia of language and linguistics*. Oxford: Pergamon Press, 1671–1678.

Poser, William J. 1985. Cliticization to NP and Lexical Phonology. *West Coast Conference on Formal Linguistics* 4: 262–272. Stanford: CSLI.

Postal, Paul. 1969. Anaphoric islands. *Chicago Linguistic Society* 5: 205–239.

Radford, Andrew, Martin Atkinson, David Britain, Harald Clahsen & Andrew Spencer. 1999. *Linguistics: An introduction*. Cambridge: Cambridge University Press.

Reichling, Anton. 1935. *Het woord: Een studie omtrent de grondslag van taal en taalgebruik*. Nijmegen: Berkhout.

Robins, R. H. 2002. The word in American structuralism. In D. Alan Cruse, Franz Hundsnurscher, Michael Job & Peter Rolf Lutzeier, eds. *Lexicology: An international handbook on the nature and structure of words and vocabularies*, vol. I. Berlin: Walter de Gruyter, 138–143.

Sadock, Jerrold M. 1986. Some notes on noun incorporation. *Language* 62: 19–31.

Saenger, Paul. 1997. *Space between words: The origins of silent reading*. Stanford: Stanford University Press.

Sampson, Geoffrey. 2009. A linguistic axiom challenged. In Geoffrey Sampson, ed. *Language complexity as an evolving variable*. Oxford: Oxford University Press, 1–18.

Samvelian, Pollet. 2007. A (phrasal) affix analysis of the Persian Ezafe. *Journal of Linguistics* 43: 605–645.

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt Brace.

Schachter, Paul & Fe T. Otanes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.

Schiering, René, Kristine A. Hildebrandt & Balthasar Bickel. 2010. The prosodic word is not universal, but emergent. *Journal of Linguistics* 46: 537–771.

Schwegler, Armin. 1990. *Analyticity and syntheticity: A diachronic perspective with special reference to Romance languages*. Berlin: Mouton de Gruyter.

Selkirk, Elisabeth O. 1982. *The syntax of words*. Cambridge, MA: MIT Press.

Smith, George. 2000. Word remnants and coordination. In Rolf Thieroff, M. Tamrat, Nanna Fuhrhop & Oliver Teuber, eds. *Deutsche Grammatik in Theorie und*

Praxis. Tübingen: Niemeyer, 57–68.

Spencer, Andrew. 1991. *Morphological theory: An introduction to word structure in generative grammar*. Oxford: Blackwell.

Spencer, Andrew. 2001. The paradigm-based model of morphosyntax. *Transactions of the Philological Society* 99: 279–314.

Spencer, Andrew. 2004. Morphology: An overview of central concepts. In Louisa Sadler & Andrew Spencer, eds. 2004. *Projecting morphology.* Stanford: CSLI Publications, 67–109.

Spencer, Andrew. 2006. Morphological universals. In Ricardo Mairal & Juana Gil, eds. *Linguistic universals*. Cambridge: Cambridge University Press, 101–129.

Sproat, Richard. 1988. On anaphoric islandhood. In Michael Hammond & Michael Noonan, eds. *Theoretical morphology*. San Diego: Academic Press, 291–301.

Stevens, A. M. 1971. Fixed morpheme order. *Linguistic Inquiry* 2: 420–421.

Stump, Gregory T. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press.

Sweet, Henry. 1913. Word, logic, and grammar. In *Collected Papers of Henry Sweet*. Oxford: Clarendon, 1–33. First publ. 1875–76.

Tesnière, Lucien. 1932. Synthétisme et analytisme. In *Charisteria G. Mathesio quinquagenario a discipulis et circuli linguistici Pragensis sodalibus oblata.* Prague: Cercle linguistique de Prague, 62–64.

Thráinsson, Höskuldur. 2007. *The syntax of Icelandic*. Cambridge: Cambridge University Press.

Togeby, Knud. 1949 [1978]. Qu'est-ce qu'un mot? In Michael Herslund, ed. *Knud Togeby: Choix d'articles 1943–1974.* Copenhagen: Akademiske Forlag, 51–65.

Trager, George L. & Henry L. Smith. 1951. *An outline of English structure* (Studies in Linguistics Occasional Papers, 3). Norman, OK: Battenberg Press.

Tseng, Jesse. 2003. Phrasal affixes and French morphosyntax. In Gerhard Jäger, Paola Monachesi, Gerald Penn & Shuly Wintner, eds. *Proceedings of Formal Grammar 2003*. Vienna: Technische Universität Wien, 177–188.

van Wyk, E. B. 1968. Notes on word autonomy. *Lingua* 21: 543–557.

Wälchli, Bernhard. 2005. *Co-compounds and natural coordination*. Oxford: Oxford University Press.

Ward, Gregory, Richard Sproat & Gail McKoon. 1991. A pragmatic analysis of so-called anaphoric islands. *Language* 67: 439–474.

Weber, David J. 1983. The relationship of morphology and syntax: Evidence from Quechua. *Work Papers of the Summer Institute of Linguistics* (University of North Dakota) 27: 161–181.

Weinrich, Harald. 1963. Ist das Französische eine analytische oder synthetische Sprache? *Lebende Sprachen* 8: 52–55.

Wurzel, Wolfgang Ullrich. 1984. *Flexionsmorphologie und Natürlichkeit*. Berlin: Akademie Verlag.

Zedler, Johann Heinrich, ed. 1749. *Grosses vollständiges Universal-Lexicon Aller Wissenschafften und Künste*, vol. 59: *Wor–Wuq.* Halle & Leipzig: Johann Heinrich Zedler.

Zwicky, Arnold M. 1987. Suppressing the Zs. *Journal of Linguistics* 23: 133–148.

Zwicky, Arnold M. 1992. Some choices in the theory of morphology. In Robert D. Levine, ed. *Formal grammar: Theory and implementation.* Oxford: Oxford University Press, 327–371.

Zwicky, Arnold M. & Geoffrey Pullum. 1983. Cliticization vs. inflection: English *n't. Language* 59: 502–513.

Author's address:
Max-Planck-Institut für evolutionäre Anthropologie
Deutscher Platz 6
D-04105 Leipzig, Germany
e-mail: haspelmath@eva.mpg.de