

# What's in a word?

Paolo Ramat, Pavia

*A good definition of 'what is a word' is preliminary to every word-formation theory, and also every analysis of compounds. The article contains a short overview of the search for a "good", viable definition of word from Ferdinand de Saussure to André Martinet and Edward Sapir. Finally, also the Construction Grammar approach is dealt with. Three possible criteria for a prototypical definition are suggested: autonomy, mobility, and cohesion. Not all criteria are applicable to the same degree to lexemes and/or constructions, and some counterexamples are discussed. Every definition based on the notion of 'prototype' entails gradualness of the concerned items. Accordingly, in linguistics, gradualness has important consequences for every word-formation theory. For instance, the absence of water-proof word classes enables speakers to consider suffixes as compound second members and vice versa compound second members as word suffixes. Via discussion of appropriate examples it is shown that words appear as disposed along a continuum of higher or lower 'wordiness'. This situation entails that, typologically speaking, there are two morphosyntactic poles without clear-cut boundaries among the intermediate types. One pole is represented by an ideal totally isolating language where the morpheme-per-word ratio is 1:1, which means that each word contains a single morpheme. At the other end of the continuum we find the fusional, agglutinative and polysynthetic types: all of them make use of concatenative strategies, more or less extensive and extendable. On the basis of the comparison of the different linguistic types, we may define the prototypical word as a string of sounds that does not have any necessary relation to its semantic content and does not contain any morphological or syntactic relational sign. Consequently, opaqueness and symbolism are characteristic of the prototypical word, whereas iconism and transparency are characteristic of the (poly)synthetic languages and their concatenative strategy.*

**Keywords:** *definitional criteria, prototypical word, typological diversity*

## 1. Preliminaries

This article is the enlarged version of the opening lecture at the 'Conference on Word-Formation Theories', II (Košice, June 2015). Its aim is not to present new researches but to give a preliminary overview of recent and on-going discussions (for a general overview of the many aspects of word-formation one may now consult Müller et al. 2015). The main points of the article are summarized in the abstract.

If we want to have a meaningful discussion of word-formation phenomena, we are in need of a proper definition of 'word'. Therefore, my starting point will be some thoughts on the very nature of 'word'. I shall touch upon some crucial points and discuss various examples capable of illustrating these points (in spite of the risk of repeating matters well-known to specialists). I have

paraphrased in the title the famous Juliet's question 'what's in a name?' to 'what's in a word?', i.e. what is a word?, what do we mean by *word*, *slovo*, *parole*, *léxi*, *vorbă*, and further *kelime* or *söz(cük)* in Turkish, *kelma* in Maltese, *szó* in Hungarian, *xiāoxi* or *dānci* in Chinese, etc.? Are we sure that all these lexemes have the same referential meaning? Already Quintilian in his *Institutionum libri*, 1.5.2.3, noted that *uerbum* "*duplex intellectus est*", i.e. it has a double meaning:

*alter qui omnia per quae sermo nequitur significat [...] alter in quo est una pars orationis*  
"the one means all the parts of which discourse is composed, in the other meaning it refers to a part of speech"

More or less at the same time as Quintilian, *uerbum* started to be used in the Christian sense as translation of Greek *lógos*: Ἐν ἀρχῇ ἦν ὁ λόγος, *In principio erat uerbum*. I will not enter into the discussion of how *lógos* and *uerbum* have to be understood from a theological point of view. Let's stick to Quintilian and his linguistic approach. It is clear that we have to grasp which the properties of a word are, what makes a sound chain a word. Given the difficulty of finding an all-embracing, satisfying definition, some linguists such as André Martinet (1985: 84) suggested giving up the very notion of 'word' as not useful in any serious syntactic analysis. But one has to remember Saussure's words in his *Cours de linguistique générale* (p. 154):

le mot, malgré la difficulté qu'on a à le définir, est une unité qui s'impose à l'esprit, quelque chose de central dans le mécanisme de la langue [the word, in spite of the difficulties of its definition, is a unity that is evident for the mind, something pivotal in the language mechanism].

Moreover, anthropologists note that almost all written traditions have developed systems for separating the words in the phonic uninterrupted chain (see Cardona 1985: 77). Everyone who has studied archaic Greek or Latin inscriptions will remember how difficult it is to read them fluently because words are not separated. This means that speakers have an intuitive idea of what constitutes a word. Sapir spoke (1921: 33) of the psychological reality of the word in the speaker's mind. I'll come back to this crucial point later (see § 4.).

### *1.1. The prototypical properties of a word*

Many years ago I tried to define the prototypical properties of what can be considered to be a 'word' (Ramat 1990) and I picked out three main criteria. I started from the well-established, though not universally accepted, definition of 'word': a unit of meaning that can stand alone in a sentence and can be uttered in isolation. The three properties span from semantics to (phono)morphology and syntax. Though morphology seems to play the major role in defining what a word may be, a sharp division cannot be drawn between morphology and syntax, as we will see in the next examples (cp. Coulmas 1988: 316). According to (Radical) Construction Grammar the lexemes that belong to different classes of speech (or categories, i.e. the traditional *partes orationis*) derive their appurtenance from constructions (see Croft 2000: 84f.). This is true insofar as the functions and meanings of a word can be different according to different constructions. We know, however, that, for instance, a Latin word ending in *-ibus* (like *omnibus* or *hominibus*) will never be a verb and that the Czech suffix *-eme*, *-áme*, *-ime* unambiguously indicates 1<sup>st</sup> pers. plur. pres. We may conclude that, especially in fusional languages, morphology

plays a major role in individuating words, much more than in isolating languages where the syntactic position of a word establishes its function. There exists a continuum between the isolating and the polysynthetic strategy which we could label ‘phonomorphosyntactic’ since phonological, morphological and syntactic factors play a relevant role, as regards the semantic value of the concerned sound string.

The criteria I have suggested are autonomy, mobility, and cohesion. Certainly, these criteria are not brand-new (see now Mugdan’s summary in Mugdan 2015). For example, Booij (2009: 97) considers cohesiveness or non-interruptability as the defining property of the notion ‘word’. But the three criteria, if taken together, are useful since they embrace both morphology and syntax (plus phonetics as well, inasmuch as both morphological complex formations and syntactic constructions can impact on the phonological shape).

Let’s start with the autonomy property and take a very simple case:

(1a) Germ. *Ich liebe dich sehr*, lit. I love you much.

The Subject inversion in the sentence (1b)—with strong (contrastive) emphasis on the Object—is mandatory:

(1b) *DICH, liebe ich sehr (und nicht Erika)* vs. *\*DICH ich liebe sehr*.

In (1a,b) we observe that to a certain extent personal pronouns may be autonomous and movable even in non-pro-drop languages like German. Clitics, on the contrary, cannot be autonomous though they are movable in some languages:

(2) Ital. *Glie-lo dici?*<sub>INTERR.</sub> *Dí-glie-lo*<sub>IMPRT.</sub> *Sì, glie-lo dico.* *\*Sì glie-lo*  
 to.him/her-do.you.say? say-to.him/her-it. Yeah, to.him/her I.say \*Yeah, to.him/her-it  
 “Do you say it to him/her?” “Say it to him/her” “Yes, I’ll say it to him/her” -----

*Glielo* cannot appear isolated and its position depends on the morphosyntax of the sentence: preverbal in interrogative and affirmative sentences, postverbal in imperatives: *\*Glielo dí* is not acceptable, and also *\*Sì glie-lo* is ungrammatical: the verb is mandatory also in the reply. Finally, even among pronouns there are forms which are more ‘wordy’, than others. See the French difference between *je, tu, il/elle* and *moi, toi, lui*. The former cannot appear in isolation, the latter can:

(3a) *Qui va venir demain? Moi ou lui?*  
 “Who’s coming tomorrow? Me or he?”  
*versus*

(3b) *Qui va venir demain? \*Je ou il?*

We see from the examples (1)-(3) that the three properties alluded to above are only partially shared by pronouns: they can be movable; but usually they are not autonomous. We may say that *moi* and *lui* are more ‘wordy’ (or ‘wordier’) than *je* and *il* since they can stand alone. This fact hints at a gradualness of the concept of word, as we will see later on with more details: there are

prototypical words and less prototypical ones and there are even cases where one may wonder whether the phonetic string can be considered a word at all. Crucially, this gradualness has important consequences for every word-formation theory.

### 1.2. Cohesion

Cohesion refers to the fact that nothing can be inserted into a word, except infixes that have grammatical function as morphs ('grams', endowed with their own meaning/function, like the already quoted Lat. suffix *-ibus*). In ancient IE languages it was thus possible to have a nasal infix in the root which represented the basis of a word family: the root *leik<sup>w</sup>*- 'to leave, abandon' had in Latin a present form *li-n-quō* 'I leave' but a perfect *liqu-ī* 'I left', and the Semitic languages use vowel insertion in the three consonant root to distinguish different meanings: see the often-cited example of the Arabic root *k-t-b* whose general meaning has to do with the concept of 'write, writing': *kitāb* 'book', *kutub* 'books', *kātib* 'writer', *katib-a* 'written document', *kuttāb* 'writing school', *kataba* 'he wrote', *yaktubu* 'he writes', etc. We speak in such cases of inflectional allomorphy (see Dressler 2015: § 2.2). However, in order to distinguish meanings in a 'word family' there exist other strategies than stem/affix allomorphy: for example, to avoid homophonies Modern Mandarin has developed many bisyllabic compounds like *shígāo* 'plaster' (lit. stone cream), *shíkū* 'grotto' (lit. stone cave), *shíhuī* 'lime' (lit. stone dust) etc. But in compounds such as *shuì-jiào* 'to sleep' from "sleep<sub>VB</sub>+sleep<sub>N</sub>" the second element cannot appear isolated, i.e. it is cohesive, strictly united with the first one; consequently, it cannot be considered as an autonomous word, just as Engl. suffixes like *-hood* or *-dom* can't, though etymologically they derive from full words. Nevertheless, *wǒ shuì le jiào* 'I slept' with insertion of the perfectivizing morpheme *le* is perfectly grammatical (but, of course, it cannot be considered as a compound).

In the NP *the bottle* we could consider the article *the* as a prefix with the function to indicate the status of Noun of the following *bottle* (vs. *to bottle<sub>VB</sub>*). Since *the* cannot stand alone it is not a word according to the autonomy criterion, though, contrary to the cohesiveness principle, something can be inserted in the NP: e.g. *the green bottle*. This possibility does not apply to languages with postposed article like Romanian (*lup* 'wolf' and *lupul* 'the wolf', never \**lup griul* lit. 'wolf gray-the') or Macedonian (*grad* 'town' and *gradot* 'the town', never \**grad golemot* lit. 'town big-the'). We may conclude that, according to the autonomy criterion, Romanian and Macedonian definite articles are more strongly bound to Nouns than the English counterpart *the*, and therefore they must be located at a lower level of the autonomy scale. These examples confirm what I said before, namely that words are disposed along a continuum of higher or lower wordiness. Note also that *the* is an invariable morph but in other languages the article can be inflected: see French *l'enfant* 'the child' vs. *les enfants* 'the children'. However, in French, too, the article cannot appear isolated: \**les* alone does not make any sense.

### 1.3. Mobility

As for mobility, a word can change its position in a sentence:

(4a) *John has probably lost the key of his house*

(4b) **Probably**, John has lost the key of his house

(4c) John has lost the key of his house, **probably**.

Clearly, the three sentences do not have exactly the same meaning, as the adverb is differently focalised. In (4b) it has an epistemic value and the entire sentence is in its scope, whereas in (4c) it represents a kind of an afterthought. In (4a) we have the unmarked position of the adverb. Note that in (4) the NP has a rigid order: ?? *of his house the key* sounds very poetic and unusual, while *\*the house of key his* is ungrammatical. Languages with a rich array of inflections have a much freer word mobility:

(5a) *Suae domus fortasse amisit Johannes clavem*

(5b) *Johannes clavem domus suae fortasse amisit*, etc.

In strongly inflectional languages like Latin the word is, so to say, self-sufficient and capable to express ‘per se’ its ties with the rest of the sentence. On the contrary, the poorer a language is in inflection, the more crucial the syntactic order is in the sentence construction. Strongly isolating languages like Chinese show a very low degree of mobility. Modern Mandarin Chinese is an analytic language that depends on syntactic word order and sentence structure.

We have already seen examples where suffixes and grammatical particles are neither autonomous nor movable. The already quoted Engl. *-hood*, originally a noun meaning ‘condition, quality’, is just a suffix and cannot appear in isolation: *\*the/a hood* as well as the inversion *\*hoodchild* are impossible. This does not mean that speakers do not analyze *childhood* as ‘child’ plus a suffix to build an abstract noun expressing ‘the quality/state of being X’, as is proved by nonce-formations like *doghood* or *soulhood* (on the nonce formations created by the speakers according to the WFRs of their own language see Gaeta & Ricca 2015: 843, 847). Actually, some exceptions to the non-isolation principle of affixes do exist: they are represented by a few prefixes such as *extra*, *super* (see *extraordinary*, *extrasensory* versus constructs as *meals are extra*, i.e. ‘not included in the price’; *superhuman*, *supernatural* versus constructs as *that film was super*, i.e. ‘very very good’). And from words like *socialism*, *capitalism*, *naturalism* etc. the suffix *ism* has been extracted becoming a noun with the meaning ‘abstract general notion’: *I have enough of all your isms!* At this point of its development we may suspect that *ism* is felt as a compound second member and no longer as a suffix, similar to *-gate* that from Nixon’s *Watergate*, to Clinton’s *Sexgate*, and further to *Petrolgate* etc. has become the second member of compounds with the meaning of ‘scandal, dirty affair’. Note that in English and in many other languages compounding is a highly productive strategy for word-formation. Thus, it is quite understandable that English native speakers may analyze the suffix *-hood* as compound second member, as is proved by the above-mentioned nonce-formations like *dog-hood*. In the same way, from *Bronchitis*, *Gastritis*, *Nephritis* etc. the suffix *-itis* meaning ‘disease’ has been used in German for compounding new words like *Telefonitis* ‘odd habit to use the telephone, telephone addicted’ (Ramat 1992).

These examples are important: on the one hand they show the metalinguistic competence of the speakers related to properties of their cognitive endowment, on the other —more strictly relevant to the linguistic system proper— they show via categorial transformations {suffix noun; noun suffix} that the boundaries between words and compounded formations are not waterproof.

For example, the category of ADJ may include verbs or nouns whereby verbs or nouns can be used with an adjectival function: Turk. *güzel* means both ‘beauty’<sub>NOUN</sub> and ‘beautiful’<sub>ADJ</sub>, and it may be used even as an ADV: *güzel konuştu* ‘s/he spoke beautifully’ (see Gaeta 2014: 230).

The dependency relation between noun-bases and morphemes is a basic Word-Formation Rule (WFR) in every language and in general the order of the elements forming a word is fixed: *inter-nation-al-iz-ation-s* (and never *\*inter-nation-s-iz-al-ation*!) where the order of derivational affixes reflects the order of semantic operations and the inflectional suffix *-s* has the most peripheral position (Štekauer 2015: §3.6). The same holds for the elements forming a compound: in German *Feldbau* ‘agriculture’ is not equivalent to *Baufeld* ‘ground for construction’! However, Joan Bybee (1985: 96) quoted Eskimo as a language having the possibility of different suffix orders:

- (6a) *ino-*            *-rssu-*            *-anguag*  
       person- -big- -little, i.e. ‘little giant’  
       *versus*  
 (6b) *inunguarssuag*, lit. person-little-big, i.e. ‘big midget’

The suffix occurring closer to the root affects the inherent meaning of the root (‘giant’ or ‘midget’), while the second suffix has a more adjectival function. With other morphological means, namely adjectives, we obtain the same semantic effect in English or Italian, where moral connotations play an important role:

- (7a) *big little man* and *grande piccolo uomo*

have a totally different meaning than

- (7b) *little big man, piccolo grande uomo.*

Semantically, *little man* and *big man* come close to a compound, and in many linguistic traditions compounds behave like frozen phrases, the classical example being *Vergissmeinnicht*, *forget-me-not* ‘myosotis’ (note that English uses hyphens!).

Phrases like *little man, piccolo uomo*, have been defined as MWEs (‘Multi Word Expressions’), i.e. lexical unities formed by more words but referring to a unitarian concept, just as ‘midget’ (see Hüning & Schlücker 2015). The suggested criteria for individuating MWEs are twofold: semantic compositionality degree and syntactic rigidity: in an MWE the binding relation (BR) between the anaphoric pronoun and its antecedent is missing (cp. Germ. *Wer<sub>i</sub> seinerzeit<sub>j</sub> / zu seiner<sub>i</sub> Zeit ein Wappen trug, war zugleich ein Waffenträger : seinerzeit / zu seiner Zeit* means ‘in those times’ and does not necessarily refer to the specific, particular time of the person referred to by *wer*).

Also very usual expressions such as *sort of / kind of* do not strictly concern the WFRs since they are still not unverbated phrases. Voghera (2013) considers the Italian equivalent of *sort/kind of*, namely *tipo* in, for example, *tipo di pittura* ‘type of painting’, and further in constructs as *caffè espresso tipo Africa* ‘express coffee type Africa’, *una scuola tipo Università popolare* ‘a school of the popular University sort’ as a “non noun”: *tipo*<sub>[N]</sub> X . Such constructs

with *espèce, sort, type, tipo*, are called by Simone and Masini (2014b) ‘Light Nouns’ whose referential force is more or less reduced, “with a scalar classification of nouns” according to their referential force (*ibid.*, p.53). In any case, we have to distinguish between compounds and these binominal constructs. Compounds are not NPs. The former do not allow insertions, the latter do: *un grosso colpo inaspettato di fortuna*, but not *\*une pomme rouge de terre*.

At the lexical level MWEs may be subject to ‘univerbation’, thereby giving rise to new words. Univerbation conflates MWEs into full-fledged words belonging to a definite class (Simone & Masini 2014a: 4). This is the case of adverbs such as *perhaps* or ital. *forse*, Germ. *heute*, etc. whose etymology remains totally obscure to the native speaker (respectively, from *per haps* ‘by accident’ —cp. *happen, happening*—; from Lat. *fors sit* ‘be the chance’, and from Old High Germ. *\*hiu dagu* ‘an diesem Tag, today’). Moreover, in Alb. *kinse*, Serbo-Croat. *morda*, Czech *možda* lit. (it) can that > ‘perhaps’ the complementizer has been incorporated into the verbal form giving rise to the new adverb.

## 2. Compounds and word formation rules

The above examples and observations lead to the definition of ‘word’ in prototypical terms, already alluded to in §1.1: a prototypical word is autonomous, movable and cohesive.

Under this point of view compounds such as *wardrobe* and *skyscraper* are words at the same rate as *ward, robe, sky* and *scraper* (though *scraper* has a more complex formation, namely a basis *scrap* + the ‘nomen agentis’ suffix *-er*). But, contrary to the just quoted construct *un grosso colpo inaspettato di fortuna*, you cannot have insertions like *\*ward-safe-robe* or *\*sky-strong-scraper* instead of *safe wardrobe* and *strong skyscraper*: under the viewpoint of cohesiveness they are OK.

Zero-marked compounds with simple juxtaposition of its members (like *skyscraper*) but referring to superordinate-level concepts have been called ‘co-compounds’ (see Arcodia, Grandi & Wälchli 2010; Wälchli 2015): e.g. Tok Pisin *man-meri* lit. man-woman > ‘people’, Sanskr. *mātā-pitarau*<sub>DUAL</sub> lit. mother-father<sub>DUAL</sub> > ‘parents’; Mand. Chin. *papa-mama* lit. daddy-mom > ‘parents’, Erzya Mordvin *t’eta-t=ava-t* ‘father<sub>PL</sub>-mother<sub>PL</sub> > parents’ (lit. ‘fathers-mothers’, see Wälchli 2015: 712). Chin. *dāo-qiāng*, lit. sword-spear refers to the superordinate concept of ‘weapons’ and is different from Span. *lanza-espada* ‘spear-sword’ which indicates a special kind of spear with a blade and therefore establishes a dependency connection between the two terms. *Dāo-qiāng*, contrary to *lanza-espada*, represents a ‘dvandva’ compound where the two terms are not in a dependency relation: it lies on the border between words and phrases. On the other hand, compounds such as Germ. *Regierungs-präsident* ‘president of the government’ show a dependency relation between the two compound elements. Moreover, *Regierungs-präsident* has an *-s-* which does not belong to the inflection of the *-ung-* names and appears just in compounds as marker of composition, a so-called ‘Fugenelement’ or ‘Fugenmerkmal’. This *-s-* derives via analogy from the genitive of words such as *König-s* ‘of the king’ as in *Königsberg*, a place name, along with *Berg des Königs*, and *Bund-es* ‘of the league/union’ as in *Bundeskanzler*. This example shows how compounds may be influenced by inflection. In fact, *Bundeskanzler* is traditionally alluded to as ‘improper compound’, like Ancient Greek *Dióskuri* (Διόσκουροι, the ‘sons [κούροι] of Zeus [Διός<sub>GEN</sub>]’; *kúri Diós* [κούροι Διός] would also be fine. Actually, we

could consider *Bundeskanzler* and *Dióskuri* more as syntactic constructs than as compounds. Actually, in the written tradition of the Homeric texts the two members of the ‘compound’ are written separately: Διὸς κοῦροι. But even in real compounds such as *riverbank* it is possible to refer to the first member (i.e. the determiner) of the compound, *river*, as a separate unit, i.e. as *(the) river’s bank* (Coulmas 1988:324):

- (8a) *The riverbank was damaged when it overflowed after three days of heavy rain,*  
 (8b) *The river’s bank was damaged when it overflowed after three days of heavy rain.*

Clearly, the anaphoric *it* refers both in (8a) and (8b) to the river and not to its bank.

In Ancient Greek and Latin compounds –mostly, exocentric (*‘bahuvrīhi’*) compounds– were often formed with an *-i-*: *argīpus* [ἀργίπου<sub>ADJ</sub>] ‘who has rapid [ἀργός<sub>ADJ</sub>] feet [πόδες]’, *kydiáneira* [κυδιάνειρα<sub>ADJ.FEM</sub>] ‘glorious’, lit. (who gives) glory [κῦδος] to the men [ἄνδρες], *agricola* ‘farmer’, lit. one who takes care of/lives in the fields, *silvicola* ‘one who lives in the forest’. This *-i-* does not belong to any of the declensions of the concerned compound first members: the isolated forms *kudi* [\*kūdi] or *\*silvi* do not exist. Contrary to the above quoted Germ. *-s-*, the origin of the inserted *-i-* is uncertain. It has been suggested that it derives from an ancient locative: this could apply to *agricola* ‘[one who] lives in the fields (*ager*)’ and *silvicola* ‘[one who] lives in the forest (*silva*)’, but certainly it does not apply to ἀργίπους or κυδιάνειρα. Be that as it may, we note that compounds may have a dedicated marker to indicate their nature of compounded words, i.e. they show a particular word-formation rule which is more than the bare juxtaposition of the two elements we find in numerals such as Ital. *ventidue*, Engl. *twenty two*, or Turk. *yirmi iki*. The last two forms are also graphically divided into two words. Consequently, we have to consider *skyscraper* and *wardrobe* as prototypical compound forms, of the type  $N_{\phi}+N_{\phi}$ , where  $\phi$  means absence of any morph, while *Bundeskanzler* and *Dióskuri* are not prototypical compounds.

In a recent contribution two young researchers have studied, also from a historical point of view, the emergence of compounded words in the Syriac language, that, as other Classical Semitic languages, strongly resist compounding, which was basically limited to numerals (Ciancaglini & Alfieri 2013; according to the philological tradition, in the following examples transliterations are in bold characters, while transcriptions are in italics). Ciancaglini and Alfieri have shown that Syriac makes use of both matter replication, i.e. loanwords (as in **plwpywn** = *polipódion* [πολιπόδιον]), and pattern replication (i.e. calquing as in **sgy regl’** *saggī reglā* a calque from Gk. πολιπόδιον ‘polypode’) compounded by *saggī* ‘much, many’ + *reglā* ‘foot’(traditionally). Calquing involves a higher degree of linguistic consciousness, as it means “identifying a structure that plays the pivotal role in the model language and matching it with a structure in the replica language” (Matras & Sakel 2007: 829).

Phonological adaptations happen in many languages during the borrowing process and borrowed compounds are no longer understood according their components: thus, a new word arises respecting the phonology of the borrowing language: no one recognizes in Ital. *bistecca* the English compound *beefsteak*, nor in *stoccafisso* the (Old) Dutch *stoc visch*; and adaptations to the morpho-phonology of the target language may also occur, as shown by Ciancaglini & Alfieri 2013 in the case of Syriac: this language may adopt derivative suffixes from another language, as for instance in the case of the Middle Iranian relational suffix *-agān* (= **qn’**: e.g. **ḥmrqn’**



‘donkey-driver’ from **ḥmr** ‘donkey’, **qyṭwnqn** ‘chamberlain’ from **qyṭwn**, a loanword from Gk. *koitōn* [κοιτῶν] ‘bedroom’). The *-agān* (**qn**) suffix has the same function as the Engl. *-er* we have already seen in *-scrap-er*. The ‘nomina agentis’ refer to a person/a tool that has something to do with the idea expressed by the basic term: a *farmer* is someone who has to do with a farm, a *driver* is someone who has to do with the action of driving, and we could translate the Syriac **ḥmrqn** by something like ‘donkey-er’.

### 2.1. Concatenation and word formation rules

Derivation via affixes has the same concatenative strategy we find in compounds and inflection. The difference consists in the fact that derivational and inflectional morphemes cannot appear in isolation while the members of compounds can: *sky* and *scrapper* are autonomous words, whereas *-er* is not; *child* is an autonomous word, *-hood* is not. The same holds for heavily inflecting languages:

(9) Turk. *ev-* *ler-* *im-* *iz-* *-de*  
 house-PLUR-1PERS- PLUR- in “in our houses”:

*ev* is an autonomous and unchangeable word, followed by a series of morphological suffixes that have a fixed order (*\*evimlerdeiz* would be impossible). Swahili has both prefixes and suffixes, but the technique is again concatenative:

(10) *wa-* *na-* *pig* *-w* *-a*  
 3PLUR- DEF.TIME- hit -PASSV -IND  
 ‘They are being hit’

These observations raise a problem I have not yet touched upon: how are so-called polysynthetic languages to be considered? What is their position in a word-formation model? Notoriously, they are called also ‘incorporating languages’.

In Yupik (Alaska, Eskimo-Aleut) we have ‘sentence-words’ such as

(11) *tuntu-ssur-qatar -ni-ksaite-ngqiggte-uq*  
 reindeer-hunt-FUT-say-NEG-again-3SG.INDIC.  
 ‘He had not yet said again that he was going to hunt reindeer’

Only *tuntu* ‘reindeer’ can appear in isolation; all the other members of this ‘sentence-word’ cannot, just as *ev* ‘house’ versus the other morphemes in the Turkish example (9). We begin to suspect that inflectional and polysynthetic languages have much in common.

On the other hand, incorporating languages like Chukchy may unite more lexemes as, for instance, in the classical German example *Rheindampfschiffahrtgesellschaft* ‘company for the steam navigation on the Rhine’, which theoretically could be endlessly continued *-kapitänstochter* ‘the daughter of the captain of the company...’, the essential difference being that the long German compound which contains five lexemes, ‘per se’ does not build a sentence, just as the simple two-members compounds like *watchman*, *watchdog* do not, while Chukchy

does —as in the following example:

- (12) *t-ə- meyh-ə-levt-ə-pəyt-ə-rkən*  
I<sub>SUBJ</sub>-ə-great-ə-head-ə-hurt-ə-PRES 1.SG (-ə- represents a connecting vowel)  
“I have a fierce headache” (Comrie 1989: 45)

This example has three incorporated lexemes ‘great’, ‘head’ and ‘hurt’ with a verbal morpheme -*rkən*: therefore it is a full sentence. Let’s now consider the following example from Vietnamese:

- (13) *KhÓ tÛi □ Ęn nhđ bạn tôi, chúng tôi bắt □, u lđm bđi*  
when I came house friend I, PLUR I begin do lesson  
‘When I came to my friend’s house, we began to do lessons’ (Comrie 1989:43),

Vietnamese is a heavily isolating language, whereas English is only partially isolating:

- (14) *I do not say it to her*  
1SG AUX NEG say PRO PREP PRO.F.OBL

Compare now (14) to the corresponding

- (15) Ital. *Non glielo dico*  
NEG to.her/him.it I.say<sub>1SG.INDIC.PRES</sub>

English is obliged to use a personal pronoun, while the person is included in the final *-o* of the Italian verb *dic-o*; negation is expressed in English by a periphrastic form (i.e. auxiliary + *not*) and the dative (or second object) needs the preposition *to* which is not present in Italian. We may conclude that English is less analytically isolating than Vietnamese, where ‘we’ is expressed (as in Chinese) by a plural marker (*chúng*) + ‘I’ (*tôi*). At the same time, English is more analytically isolating than Italian. The morpheme-per-word ratio is in English higher than in Vietnamese, but lower than in Italian. Our examples show that from the typological point of view concerning word structure, as well as WFRs there are two morphosyntactic poles without clear-cut boundaries among the intermediate types (cf. (8a) and (8b)). One pole is represented by an ideal totally isolating language where the morpheme-per-word ratio is 1:1, which means that each word contains a single morpheme as in Vietn. *chúng tôi* ‘we’. But we have seen that even Chinese and Vietnamese are not completely isolating: above I have quoted Chinese bisyllabic compounds and example (13) presents *bắt đầu* which is a periphrastic form for ‘begin’. More or less near to the opposite pole we find what are traditionally called fusional, agglutinative and polysynthetic types: all of them make use of concatenative strategies, more or less extensive and extendable; from the theoretically endless composition of the German type (*Rheindampfschiffahrt...*) to the Turkish inflection extremely rich in morphemes and finally to the polysynthetic type. Even in this case we do not find an absolutely synthetic language. Analysis and synthesis represent two ideal types. Skalička (1966) called them ‘Konstrukte’.

### 3. What's in a word

We are now able to propose an answer to the question asked in the title of this paper, namely to advance a definition of *word*, which, by the way, is an always debated question among linguists and particularly among typologists. Typologies such as Klimov's 'kontensivnaja tipologija' which distinguishes between ergative, active and nominative/accusative languages (Klimov 1983, 1986) are less relevant for the definition of the word concept. The same holds for the distinction between head-marking and dependent-marking (Nichols 1986), as this distinction concerns the relations among the constituents of a construction: *The man's house* marks the possession relation on the dependent element of the NP, whilst Hung. *az ember háza* marks the possession on the head of the NP (lit. the man house.his).

The prototypical word is a phonetic string that does not have any necessary relation to its semantic content (remember Saussure's *arbitraire du signe*) and does not contain any morphological or syntactic relational sign: *happy, do, boy* are 'wordier' than *happiness, did, boys* and if in *zwanzig* a German speaker would perhaps still recognize the same basis as in *zwei*, than *zwanzig* would be less wordy than Fr. *vingt* whose etymological connection (< Lat. *viginti*) remains completely obscure.

**Opaqueness** and **symbolism** are characteristic of the prototypical word, whereas **iconism** and **transparency** are characteristic of (poly)synthetic words (cp. Ramat 2005 [1990] : 119). In morphology iconism is expressed by more complex forms: plurals are usually more complex than their singular counterparts: see *boy-s* vs. *boy* or Lat. *puell-ā-rum*<sub>GEN.PL.</sub> vs. its Nomin. Sing. *puella*; the indicative is usually less complex than conditional, optative and other moods (see *I say* vs. *I would say*, Lat. *dīcō*<sub>1SG.PRES.INDIC</sub> vs. *dīcerem*<sub>1SG.IMPF.SUBJ.</sub>, etc. Accordingly, completely arbitrary brand names such as *Meriva*, a type of car produced by Opel (Ronneberger-Sibold 2015: 2199), acronyms which remain totally obscure to the hearer (e.g. FIAT from *Fabbrica Italiana Automobili Torino*) and, last but not least, proper names such as *Pablo, Pavel, Paolo, Paul* are excellent examples of prototypical words, much more than the compounded names *Cam-bridge, Bene-dikt, Gott-lieb, Miro-slav* or *Blanche-fleur*, whose etymological formation is still easily recognizable. Brand names, acronyms and (personal) nouns are according to Seiler's (1975) terminology "etikettierende Benennungen", unmotivated labels (as opposed to "deskriptive Benennungen").

#### 3.1. About word-formation (theory)

Once we have obtained a satisfying word definition (see §2 and 3), we can look more accurately at word-formation and word-formation theories. What do we mean by word-formation? In this final part of the paper I will make some cursory observations about some crucial features of word-formation rules, more or less implicitly alluded to in the previous discussion and in the corresponding examples.

WFRs do not apply to simple words, like Germ. *heute*, Ital. *oggi*. Though they historically derive from NPs (see § 2.3), they are no longer analyzable in their components, whilst Engl. *today* and French *aujourd'hui* may be still analyzed (at least partially) by native speakers: *to-day* as *to-morrow* and *-jour-* in the French complex expression. Crucially, WFRs concern derivation and composition: e.g. *childhood* and *skyscraper*. There is also zero derivation or 'conversion' as

in *clean*<sub>ADJ</sub> to *clean*<sub>VB</sub>, *bottle*<sub>NOUN</sub> to *bottle*<sub>VB</sub>, *water*<sub>NOUN</sub> to *water*<sub>VB</sub>, etc. We have here a clear example of transcategorization, that is, a categorial shift of a lexical item with no superficial marking, resulting from its employment in a new syntactic environment: cp. *the water in the bottle* vs. *to water the garden* (see Ježek & Ramat 2009). The poorer a language is in morphology, the more transcategorization is possible: the same lexeme may be ADJ and ADV, Noun and VB. This refers to the difference between what Hengeveld (1992) has called ‘rigid’ vs. ‘flexible’ languages. Flexible languages allow much more permeability between word classes than rigid languages. Rigid languages such as the old Indo-European ones have dedicated markers (morphs) for different classes and different forms inside the classes: as I said before, suffixes like the DAT./ABL. *-ibus* in Latin or the GEN.SG. *-oio* [-oio] in Homeric Greek will never be used in the verbal system, nor will *-averunt* of *amaverunt* ‘they loved’ or *-somai* [-σομαι] of *bēļsomai* [βήσομαι] ‘I will come’ ever appear as adverbial suffixes. On the contrary, English has a rather poor morphology and a lexeme such as *like* may be used as VB (*I like swimming*), as ADJ (*to be of like mind*), and as ADV (*she doesn’t prefer vivid colours, like red*). Only the syntactic environment will clarify the meaning of such a polysemous word. This is the main characteristic of isolating languages such as Mandarin or Vietnamese.

### 3.1.1. Some (important) aspects of word-formation

As is well known, affixes are the main strategy to form a new word from a lexical basis. As Creissels (2014: 90) has it, “[t]he ability to be the target of word formation processes is commonly considered a typical property of major word classes”. He quotes the case of the Bantu language Tswana that makes use of the prefix *bo-* to derive nouns from adjectives as in

(16) *boi* ‘timid’ *bo-boi* ‘timidity’, *thata* ‘strong’ *bothata* ‘strength’.

Infixes are not that usual. I have already quoted (§ 2.1) the *-n-* infix in Lat. *li-n-quo* that is typical of old Indo-European languages. This is a residual morph whose function remains unclear. It disappeared during the development of Old Indic, Greek, Latin etc. (Latin has *pingo* ‘I depict’ but *pictum* ‘depicted’ without *-n-* according the old rule. On the contrary, the perfect tense *pinxi* ‘I depicted’, with extension of the *-n-*, shows that the old rule of nasal insertion just in the present tense was no longer respected, and new analogical forms replaced the old ones). Even circumfixes are not very frequent among the languages of the world. However, they exist as morphological markers. We may quote the case of the Hungarian superlative *leg...bb*: *legnagyobb* ‘biggest’ from *nagy* ‘big’; Germ. has *ge-hab-t* as the past participle of *haben* with a circumfix around the basis *-hab-* which never stands alone, except perhaps in the imperative that in fusional languages often represents the uninflected form.

Furthermore, along the inflection-derivation continuum, i.e. both in the morphology of single lexemes and in the WFRs of compounds, there are more and less productive morphs. Thus, the suffix *-ness* of *happiness* is a frequent means to form abstract nouns from adjectival and nominal bases, i.e. ‘the state of being X’: *darkness*, *kindness* and also *oneness*, *treeness* (< *tree*). On the contrary, *-ess* as a feminine marker seems to be in a recessive state: see *princess*, *actress*, *waitress* and a few others. The suffix appears to be no longer productive.

Finally, in this sketchy overview of (some of) the main problems of WFR, a couple of words about a phenomenon which has been relatively little studied by linguists: I mean the

formations that French linguists call ‘mots-valises’ (literally ‘suitcase-words’) which are formed via fusion of two words that present partial homophony such as *motel* from *motor-h(ot)el*, or even just a blending of two words such as *eurovision* from *Europe* and *television*, *Français* from *français* + *anglais* or *Spanglish* from *Spanish* + *English*. The word *television* is scarcely used in everyday language and people prefer to speak either of *tele* or of *TV* (French [tə'və], It. [ti'vi] etc.)). Nevertheless, to my knowledge a ‘compound’ *eurotele* or *eurotəvə* has never been created. Such ‘compounds’ prove that speakers have a spontaneous metalinguistic knowledge of what is a word in their own language (see the previous reference to Sapir). We get univerbations as *motel* and *eurovision* but never *\*eursion* nor *\*tortel* nor even *\*orotel*. These ‘compounds’ are more or less recognizable in their blending, so that they cannot be considered as prototypical words as defined in § 4. The public uses and sometime abuses this new compounding possibility, which has become very popular in recent times. See, for example, the name of a large chain of restaurants and food shopping centers called *Eataly* [i:'tali]: it has been created via shortening and blending from *eat Italy* [i:t 'itali], a very shrewd word play, which is intended to mean ‘eat according to the Italian style’. Syntactically, *Italy* plays in such a new formation the unusual role of object of the transitive verb *to eat*. It is too early to foresee whether *Eataly* will be generally accepted and recognized as a new word. But *motel* has certainly become an international word and there exist dictionaries of *Français* and *Spanglish*. No doubt this relatively new domain of linguistic research has its own WFRs and deserves to be accurately studied.

## References

- Aarts, Bas. 1998. Binominal noun phrases in English. *Transactions of the Philological Society* 96. 117-158.
- Alfieri, Luca. 2014. Qualifying modifier encoding and adjectival typology. In Simone, R. & Masini, F. (eds.), 119-138.
- Arcodia, Giorgio, Grandi, Nicola & Wälchli, Bernhard. 2010. Coordination in compounding. In Scalise, S. & Vogel, I. (eds.), *Cross-disciplinary issues in compounding*, 177-198. Amsterdam & Philadelphia, PI: John Benjamins.
- Booij, Geert. 2009. Lexical integrity as a formal universal: A constructionist view. In Scalise, S., Bisetto, A. & Magni, E. (eds.), *Universals of language today*, 83-100. Dordrecht: Springer.
- Bybee, Joan. 1985. *Morphology. A study of the relation between meaning and form*. Amsterdam & Philadelphia, PI: John Benjamins.
- Cardona, Giorgio R. 1985. Dall’oralità alla scrittura: La formazione delle lingue standard. In Moreschini Quattordio, A. (ed.), *La formazione delle lingue letterarie*. Atti del Convegno della Società Ital. di Glottologia, 71-80. Pisa: Giardini.
- Ciancaglini, Claudia & Alfieri, Luca. 2013. Iranian and Greek influence on the Syriac lexicon: The

- emergence of compound words. *Incontri Linguistici* 36. 109-138.
- Comrie, Bernard. 1985. *Language universals and language typology* (2<sup>nd</sup> edition). Oxford: Blackwell.
- Coulmas, Florian. 1988. Wörter, Komposita und anaphorische Inseln. *Folia Linguistica* 22. 315-336.
- Creissels, Denis. 2014. The ‘new adjectives’ of Tswana. In Simone, R. & Masini, F. (eds.), 75-94.
- Croft, William. 2000. Parts of speech as language universals and as language-particular categories. In Vogel, P. M. & Comrie, B.(eds.), *Approaches to the typology of word classes*, 65-101. Berlin & New York, NY: Mouton de Gruyter.
- Croft, William. 2007. Beyond Aristotle and gradience. *Studies in Language* 31. 409-430.
- Daniels, Peter T. 1989. Syllabic origin of writing. In Downing Pamela *et al.* 1989, 83-103.
- Downing, Pamela *et al.* 1989. *The linguistics of literacy*. Amsterdam & Philadelphia, PI: John Benjamins.
- Dressler, Wolfgang U. 2015. Allomorphy. In Müller, P. O. et al. (eds.), II: 500-516.
- Gaeta, Livio. 2014. On decategorization and its relevance in German. In Simone, R. & Masini, F. (eds.), 227-241.
- Gaeta, Livio & Ricca, Davide. 2015. Productivity. In Müller, P. O. et al. (eds.), II, 842-858.
- Hengeveld, Kees. 1992. *Non verbal predication. Theory, typology, diachrony*. Berlin & New York, NY: Mouton de Gruyter.
- Hüning, Matthias & Schlücker, Barbara. 2015. Multi-word expressions. In Müller, P. O. et al. (eds.), I, 450-467.
- Ježek, Elisabetta & Ramat, Paolo. 2009. On parts-of-speech transcategorization. *Folia Linguistica* 43. 391-416.
- Klimov, G. A. 1983. *Principy kontensivnoj tipologii*. Moskva: Nauka.
- Klimov, G. A. 1986. On the notion of language type. In Lehmann, W. P. (ed.), *Language typology 1985. Papers from the linguistic symposium Moscow 1985*, 105-110. Amsterdam & Philadelphia, PI: John Benjamins.
- Martinet, André. 1985. *Syntaxe générale*. Paris: Colin.
- Matras, Yaron & Sakel, Jeannette. 2007. Investigating the mechanism of pattern replication in language convergence. *Studies in Language* 31. 829-865.
- Mattiello, Elisa. 2013. *Extra-grammatical morphology in English. Abbreviations, blends, reduplicatives, and related phenomena*. Berlin & New York, NY: Mouton de Gruyter.

- Meid, Wolfgang. 1967. *Wortbildungslehre*. In Krahe, Hans & Meid, Wolfgang, *Germanische Sprachwissenschaft*, Bd. III. Berlin: De Gruyter & Co.
- Mihatsch, Wiltrud. 2007. Les approximateurs quantitatifs entre scalarité et non-scalarité. *Langue Française* 165. 125-153.
- Mugdan, Joachim. 2015. Units of word-formation. In Müller, P. O. et al. I, 235-301.
- Müller, Peter O., Ohnheiser, Ingeborg, Olsen, Susan, Rainer, Franz (eds.). 2015. *Word formation. An international handbook of the languages of Europe*. Berlin & New York, NY: Mouton de Gruyter.
- Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. *Language* 62. 56-119.
- Ramat, Paolo. 1990. Definizione di “parola” e sua tipologia. In Beretta, M. et al. (eds.), *Parallela 3. Morfologia/Morphologie*. Atti del V° Incontro Italo-Austriaco della Società di Linguistica Italiana, Bergamo 2-4 ottobre 1989. 3-15. Tübingen: Narr (repr. with changes in Ramat, Paolo, *Pagine linguistiche*. Scritti di linguistica storica e tipologica. 106-121. Bari / Roma: Laterza, 2005).
- Ramat, Paolo. 1992. Thoughts on degrammaticalization. *Linguistics* 30. 549-560.
- Ramat, Paolo & Ricca, Davide. 1998. Sentence adverbs in the languages of Europe. In van der Auwera, J. (ed.), *Adverbial constructions in the languages of Europe*, 187-275. Berlin & New York, NY: Mouton de Gruyter.
- Renner, Vincent. 2015. Lexical blending as wordplay. In Zirker, A. & Winter-Froemel, E. (eds.), *Wordplay and metalinguistic/metadiscursive reflection: Authors, contexts, techniques, and meta-reflection*. Berlin & New York, NY: Mouton de Gruyter.
- Renner, Vincent, Maniez, François & Arnaud, Pierre (eds.). 2012. *Cross-disciplinary perspectives on lexical blending*. Berlin & New York, NY: Mouton de Gruyter.
- Ronneberger-Sibold, Elke. 2015. Word-creation. In Müller, P. O. et al. (eds.), I, 485-500.
- Ronneberger-Sibold, Elke. 2015. Word-formation and brand names. In Müller, P. O. et al. (eds.), XIV, 2192-2209.
- Sapir, Edward. 1921. *Language. An introduction to the study of language*. New York, NY: Harcourt, Brace and World.
- Scalise, Sergio. 1995. La formazione delle parole. In Renzi, L., Salvi, G. & Cardinaletti, A. (eds.), *Grande grammatica di consultazione della lingua italiana*, vol. III, 471-514. Bologna: Il Mulino.
- Scalise, Sergio & Ceccagno, Antonella. 2006. Facile o difficile? Alcune riflessioni su italiano e cinese. In Bosc, F., Marellò, C. & Mosca, S. (eds.), *Saperi per insegnare. Formare insegnanti di italiano*

*per stranieri*, 151-177. Torino: Loescher.

Seiler, Hansjakob. 1975. Die Prinzipien der deskriptiven und der etikettierenden Benennung. In Seiler, H. (ed.), *Linguistic workshop III*, 2–57. München: Fink.

Simone, Raffaele & Masini, Francesca. 2014a. Introduction. In Simone, R. & Masini, F. (eds.), 1-13.

Simone, Raffaele & Masini, Francesca. 2014b. On light nouns. In Simone, R. & Masini, F. (eds.), 51-73.

Simone, Raffaele & Masini, Francesca (eds.). 2014c. *Word classes. Nature, typology and representations*. Amsterdam & Philadelphia, PI: John Benjamins.

Skalička, Vladimír. 1966. Ein "typologischer Konstrukt". *Travaux Linguistiques de Prague* 2. 157-164.

Štekauer, Pavol. 2015. The delimitation of derivation and inflection. In Müller, P. O. et al. (eds.), I, 218-234.

Voghera, Miriam. 2013. A case study on the relationship between grammatical change and synchronic variation. In Giacalone Ramat, A., Mauri, C. & Molinelli, P. (eds.), *Synchrony and diachrony. A dynamic interface*, 283-311. Amsterdam & Philadelphia, PI: John Benjamins.

Wälchli, Bernhard. 2015. Co-compounds. In Müller, P. O. et al. (eds.), III, 708-727.

*Paolo Ramat*  
*Piazzetta Arduino 11. I*  
*27100 Pavia*  
[paolo.ramat@unipv.it](mailto:paolo.ramat@unipv.it)

*In SKASE Journal of Theoretical Linguistics [online]. 2016, vol. 13, no.2 [cit. 2016-09-11]. Available on web page [http://www.skase.sk/Volumes/JTL32/pdf\\_doc/06.pdf](http://www.skase.sk/Volumes/JTL32/pdf_doc/06.pdf). ISSN 1336- 782X.*