

LEXICAL DECOMPOSITION MEETS CONCEPTUAL ATOMISM

PAOLO ACQUAVIVA

PHOEVS PANAGIOTIDIS

ABSTRACT: Asking what can be a substantive word in natural language is closely related to asking what can be a basic lexical concept. However, studies on lexical concepts in cognitive psychology and philosophy and studies on the constitution of lexical items in linguistics have little contact with each other. We argue that current linguistic approaches that decompose lexical items into grammatical structures do not map naturally to plausible models of the corresponding concepts. In particular, we claim that roots, as the purported carriers of lexeme-specific content, cannot encapsulate the conceptual content of a lexical item. Instead, we distinguish syntactic from morphological roots: the former act as differential indices, and the latter are forms which may or may not correlate with a stable meaning. What expresses a lexical concept is a structure which can be of variable size. We explore the view that basic lexical items are syntactically complex but conceptually simplex, and that the structural meaning defined by a grammatical construction constrains the concept associated with it. This can lead to predictive hypotheses about the possible content of lexical items.

KEYWORDS: concepts, roots, lexical decomposition

1. CONCEPTS AND WORD STRUCTURE*

* We wish to thank the audience and organizers of the first NetWordS workshop. P. Acquaviva's research was supported by an Alexander von Humboldt fellowship, which is gratefully acknowledged. Faults and omissions are the authors' responsibility.

Taking DOG to represent the concept associated with the word *dog* seems a straightforward choice, but it presupposes a clear notion of what is a word. To see that there is an issue, and that the issue is linguistic in nature, it suffices to ask whether the two word forms present in *put up* (syntactically separable, as in *put him up*) map to one or to two concepts; or whether *break* corresponds to a single concept BREAK in all its uses, including strongly idiomaticized ones like *break wind* (contrast *wind-breaker*, in the sense of a garment). Far from being a terminological quibble, this is a substantive issue about the discrimination between simple and complex concepts. If the notion of wordhood/lexicality that is relevant for conceptualization is based on morphological criteria, *break up* is not only a complex expression but also a complex concept, assembled out of two simple concepts in the same way as BROWN COW is. If instead a simple concept corresponds to a ‘semantically simple’ word, then we must decide what counts as semantically simple—a task that seems identical to the task of deciding what counts as a concept, resulting in circularity: *a b* jointly form a simple concept because they are a concept.

For different reasons, researchers in linguistics and in cognitive psychology are not generally overly worried about the linguistic bases of concept individuation. Beginning their overview of the extensive literature on concepts, Laurence and Margolis (1999:4) note that ‘For a variety of reasons, most discussions of concepts have centered around *lexical concepts*. Lexical concepts are concepts like BACHELOR, BIRD, and BITE—roughly, ones that correspond to lexical items in natural languages.’ Correspondingly, Fodor (1998:121) states the thesis of Conceptual Atomism as the view according to which ‘most lexical concepts have no internal structure’, and makes it clear that the property of being linguistically a single word is indeed crucial, by adding that (p.122, note 3) ‘actually, of course, DOORKNOB isn’t a very good example, since it’s plausibly a compound composed of the constituent concepts DOOR and KNOB.’ Expressions like ‘concepts (/lexical meanings)’ (Fodor 2008:26) are in line with this perspective.

Laurence and Margolis (1999:4) actually acknowledge that defining simple concepts as those associated to lexical words is not entirely straightforward, and add a footnote specifying that ‘the concepts in question are ones that are usually encoded by single morphemes. In particular, we don’t worry about the possibility the one language may use a phrase where another uses a word, and we won’t worry about what exactly a word is (but for alternative conceptions, see Di Sciullo and Williams 1987).’ But of course, being monomorphemic is not the same thing as being semantically basic: witness lexemes that are morphologically but not semantically complex like *con-ceive*, or lexicalized compounds like *home run*, *french*

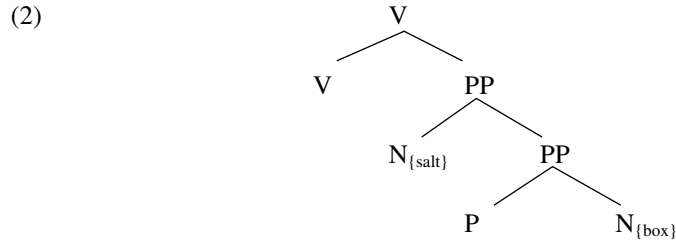
window, or indeed *cranberry*; and the reference to Di Sciullo and Williams' book merely points to the existence of a whole dimension of theoretical debate.

On the linguistic side, research into the primitives and the constituent structure of lexical meaning represents a long and richly diverse tradition of studies, but typically one that never had much direct dialogue with psychological research into the representation of concepts—even though the notion of simple concepts is linguistically rooted. Linguistic analyses of lexical concepts traditionally decompose the content of lexical words into different kinds of representations (cf. Jackendoff 1972, 1990, Pinker 1989, Pustejovsky 1995, Wunderlich 1997, among many others). Apart from differing in the primitives chosen to represent lexical structure, these approaches differ in the type of structure envisaged. Jackendoff (2002:334-339) hypothesizes two distinct types of structure, both of which are non-syntactic: Conceptual Structure, understood as 'a hierarchical arrangement built out of discrete features and functions', and Spatial Structure, understood as 'encoding the spatial understanding of the physical world'. Lieber (2004) also envisages two distinct components in her model of lexical meaning, an unstructured repository of encyclopaedic information (the 'Body') and an function-argument structure (the 'Skeleton') encoding primitive conceptual properties, neither of which corresponds to a syntactic structure. In contrast with these approaches, the influential model of Levin and Rappaport Hovav (1995, 2005) proposes a structural interpretation of their meaning by means of a single representation, which describes the argument- and event structure of a verb by means of primitive predicates (like BECOME) and constants /roots (notated BREAK, for example) forming a lexical semantic template like that illustrated in (1):

- (1) noncausative *break*: [*y* BECOME *BROKEN*]

Importantly, the same constant / root may appear in different templates, as in *to shovel* (remove) - *to shovel* (put); in other words, what identifies a verb as a lexical item is neither the semantic template (which typically generalizes over a verb class) nor the 'root' (which, understood as a label, may appear in different templates, like *shovel*), but the conjunction of both. Finally, among the approaches that decompose lexical meaning into a grammatical structure, a family of analyses explicitly take this structure to be syntactic, that is, generated accorded to the same principles that underlie sentence construction (Hale and Keyser 1993, 2002, Arad 2005, Halle and Marantz 1993, Embick and Marantz 2008, Borer 2005a,b, Ramchand 2008). As an illustration, the

syntactic structure in (2) models for Hale and Keyser (2002) the abstract structure underlying uses of *box* as a locatum verb, as in *to box the salt*.



A causative verbal head V takes as complement a PP which expresses a locative relation between an argument in specifier position (here, *salt*) and another argument in the complement position of P (here, *box*). V and P, and P and N(*box*) are in head-complement relation, but not V and N(*salt*). This makes it possible to lexicalize V-P-N(*box*), but not V-N(*salt*)-P, as a single word (the verb *to box* in its sense ‘to put into a box’), deriving the lexicalization pattern *they boxed the salt*, and correctly ruling out **they salted the box*. We will focus on approaches deriving from this type of analysis, questioning the way they deal with non-structural, idiosyncratic aspects of lexical meaning that are essential to a word’s conceptual content but apparently lack any grammatical relevance.

As Laurence and Margolis (1999) note, representing the content of a lexical item as a structural arrangement of semantic primitives enjoys widespread currency in linguistic analyses, but results problematic for the view it presupposes of lexical concepts as structural representations involving sets of precisely identified semantic primitives. In particular, the idea that linguistic word-internal structure may explain the content of lexical concepts and their mutual relations inherits the problems associated with ‘classical’ theories of concepts as structures articulated into smaller components:

- decomposition into semantic primitives faces a regress problem: what do primitives like CAUSE or THING mean, if they are not the same as the corresponding lexical words?
- if lexical meaning was analyzable into constituent parts and their relations, we would expect definitions reflecting the structural decomposition of a concept to accurately describe its content: but this typically fails, since word meaning systematically cannot be given a unique and precise definition or paraphrase;

- if lexical concepts were constituted of linguistic constructs, possession of these concepts would require being aware of their content; yet competent speakers often don't seem to know certain aspects of the meaning of the words they use, even supposedly constitutive ones;
- attested prototype effects, like the fact that a certain representation of grandmother is felt as better representing the concept than others, are unexpected if the content of GRANDMOTHER simply consists in a hierarchical arrangement of semantic primitives, defining in this case a biological relation.

This type of empirical shortcomings do not seem to have had an impact on linguistic analyses of the structure of lexical items. In part this is due to a widespread perception that such matters do not concern what speakers know about lexical items as *linguistic* representations; the content of lexical concepts certainly includes a fair deal of non-linguistic knowledge, but, it may be argued, this is irrelevant for an account of what speakers know when they know a word as a product of the language faculty. Grimshaw 1993 (cited in Laurence and Margolis 1999: 55, Jackendoff 2002: 338) has articulated this position in a particularly strong form: 'Linguistically speaking, pairs like [*break* and *shatter*] are synonyms, because they have the same structure. The differences between them are not visible to the language.'

In effect, most syntactic decomposition analyses to lexical structure largely share this view, in practice if not as an explicitly stated principle. This is a problem, however. If the semantic relations between concepts like DOG and CAT lie outside the purview of linguistic theory, as a theory of the computational capacity of the mind to represent linguistic knowledge in a way that explains the productivity and compositionality of linguistic expressions, then it is hard to see why the relation between DOG and ANIMAL should not be likewise 'not visible to the language.' And if such a canonical example of hyponymy relation as that between DOG and ANIMAL falls outside the scope of linguistic theory, then much of what speakers know about the mutual relations between word meanings becomes inaccessible to linguistic explanation, as a matter of principle. Thus, the systematic semantic deviance of comparisons like # *a dog is smaller than an animal*, involving two nouns in hyponymy relation, would have to be explained outside a linguistic theory, or not explained at all. Likewise, the fact that the difference in animacy between the senses of *hands* as 'hands of a clock' and 'hands of a person' prevents a conjunction like # *the clock's hands look funny, but mine don't*, could not be an empirical *explanandum* of linguistic analysis, even though these and similar patterns are part and parcel

of what speakers know about language, more specifically, about words and their combinatorial possibilities in sentences (as opposed to concepts and their mutual relations outside of a linguistic context). Current decompositional approaches to lexical semantics are forced to ignore these facts, which were an important part of earlier work in generative grammar (the data and discussion are from Bever and Rosenbaum 1970; cf. also Cruse 1982), and to deal exclusively with purely structural aspects of lexical meaning. The result has been a near-exclusive attention to verbal semantics, especially argument- and event structure. Apart from representing a significant empirical limitation, this development bars the way towards a linguistic theory of a possible word, psychologically and philosophically plausible theory of how words map to concepts.

By contrast, we hold that a theory of UG should have something to say about the way lexical items relate to their conceptual content. We will take as our point of departure a specific syntactic approach to lexical decomposition, which most clearly dissociates the grammatical components of a word from a non-grammatical core, and focus on the properties which can and cannot be attributed to this root element as a key locus for the relation between syntactic representation and conceptual content.

2. ROOTS IN LEXICAL DECOMPOSITION

Work in Distributed Morphology and Borer's (2005a,b) Exoskeletal approach both envisage maximally underspecified root terminals embedded inside a number of syntactic shells, which collectively define syntactic constructions that define lexical categories; a noun, adjective, or verb is for a construct, in whose innermost core lies a category-neutral root. There are many important differences between the two approaches, and indeed between the two conceptions of roots, the most apparent being that Distributed Morphology, but not Borer, mandates the presence of categorizing heads, [n], [v], or [a], immediately governing the root and categorizing it (with possible complications for complex roots). For present purposes, however, what counts is the role of the root in determining lexical semantic properties, understood as lexeme-related properties which remain constant across grammatical contexts (we undertake a fuller appraisal in Acquaviva and Panagiotidis, in preparation). Both models assume that all roots are non-categorized, so even the unique categorial determination of monomorphemic words like *fun*, *tall*, or *idiot* is inferred from the context; categorial underspecification, however, does not directly imply that roots lack the kind of semantic information which makes a difference between a noun and a verb. Analyses within Distributed Morphology, when they

address the topic, typically treat the root as a meaningful element, whose content selects a suitable syntactic context; Harley and Noyer (2000:365), for example, state that 'The speaker knows that these roots [GROW, DESTROY] denote events that may occur spontaneously, like growing, or that may be truly externally caused [...] This knowledge is part of the real-world knowledge of the speaker about the meaning of the root ... '. Importantly, however, work in this framework stresses that a root's meaning is emergent in a context. In the most comprehensive treatment of the issue in this framework, Arad (2003, 2005) develops insights by Marantz (1997) to defend a view of roots as radically underspecified but still meaningful elements which give rise to distinct interpretations depending on their immediate context. More precisely, Arad distinguishes roots with a relatively stable and well-defined meaning, from a more theoretically interesting type of roots whose semantic content cannot be stated in isolation, but emerges as a cluster of conceptually related words, giving rise to what Arad calls Multiple Contextual Meaning. Roots of the first type tend to form one or very few words only (as Hebrew nouns for animal, plants, food, or kinship terms, like *kelev* 'dog', *sukar* 'sugar', *ʔax* 'brother', *ʔaxot* 'sister'); roots of the second type give rise to larger word families, with a more or less recognizable semantic relatedness which can be very faint indeed; for example, XŠB in *xašav* 'think', *xišev* 'calculate', *hexšiv* 'consider', BXN in *mixvan* 'examination', *boxan* 'quiz', *mavxena* 'test tube', *avxana* 'diagnosis' (Arad 2005:82); but also QLT in *miqlat* 'shelter', *maqlet* 'receiver', *qaletet* 'cassette', *qalat* 'absorb, receive' (Arad 2005:97). Despite the clear statement that roots of this second type do not define a lexical sense without a context, they are unambiguously qualified as semantically contentful signs. Indeed, other analyses in this framework make crucial use of the varying semantic content of different roots, like Harley's (2005:46-50) derivation of the *Aktionsart* opposition between the unergatives *drool* (atelic) and *foal* (telic) from roots being respectively unbounded/mass and bounded/count.

In contrast, the category-free heads which correspond to roots in Borer's (2005a,b) framework lack any kind of grammatically legible information, as a matter of definition (with the exception of roots involved in idioms; cf. Borer 2005b:354). In a programmatically anti-lexicalist framework that consigns to syntax *all* grammatically relevant information of lexical words, these elements are the non-grammatical residue, which appear as listed phonological forms, or 'listemes': 'By listemes we refer to a pairing of a conceptual feature bundle with a phonological index' (Borer 2005b:25). Being void of syntactically relevant information does not mean that listemes lack content, however; on the contrary, Borer views them as encapsulating the non-syntactic information which defines a lexical item (Borer 2005b:9):

Within an XS-[exoskeletal] model, then, the particular final meaning associated with any phrase is a combination of, on the one hand, its syntactic structure and the interpretation returned for that structure by the formal semantic component, and, on the other hand, by whatever value is assigned by the conceptual system and world knowledge to the particular listemes embedded within that structure. These listemes, I suggest, function as modifiers of that structure.

In different ways, then, Distributed Morphology and Borer's Exoskeletal model posit contentful root elements at the core of their syntactic decompositions of substantive lexical items, which determine lexeme-specific and encyclopaedic aspects of lexical semantics either by themselves or as a function of their context. Our claim, now, is that roots in a syntactic decomposition sense cannot have this sort of content.

3. ROOTS ARE NOT MEANINGFUL SIGNS

In this section we will review some empirical evidence that roots do not carry any meaning and/or semantic content that could be identifiable outside of a grammatical structure, not just because they need a local context to determine a specific interpretation, but more radically because they are not signs. In fact, the evidence suggests that any sort of lexical meaning is a property of roots embedded in a grammatical structure, which can be of a rich and complex nature. The conclusion that will emerge is that there is no such thing as non-structural meaning, even at the level of 'word'.

Let us begin with some remarkable cases. It is received wisdom within the Distributed Morphology research on the systematic idiomatity of the structure below the first categorizing shell (e.g. *n*P or *v*P) that the categorizer projection acts as a sort of limit, below which interpretation is / can be / must be non-compositional (Marantz 2000; see also Marantz 2006, where inner versus outer morphology phenomena are explained in this way). In this perspective, the opposition between event nominalization and result nominal of *collection* in (3) must be due to different grammatical structures corresponding to the two readings (see Borer 2003). But since the root is the same, neither the difference in syntactic structure nor that in ontological typing (event vs. object) can be even indirectly a function of the root:

- (3) *collection*₁ 'the frequent collection of mushrooms by Eric'

*collection*₂ ‘let me show you my collection of stamps’ (result nominal)

Still, it can be argued that the two structures, while different, share a semantic core because they only differ in terms of outer morphology, above the first categorizing shell. However, as discussed in Panagiotidis (2011), we can have radically different meanings *across* the first categorizing shell. A telling example is the one below from Greek:

- (4) a. [_{VoiceP} *nom-iz-*] ‘think’
 b. [_{nP} [_{VoiceP} *nom-iz-*] *ma*] ‘coin, currency’
 c. [_{aP} *ne-* [_{VoiceP} *nom-iz-*] *men-*] ‘legally prescribed’

A large number of words relating to law, regulation and the like is derived from the root *nom-*. However, when the root is verbalized, yielding the verbal stem *nom-iz-* in (i) above, the meaning assigned is ‘think, believe’. So far there is nothing exceptional, as Marantz (2000; 2006) and Arad’s (2005) Multiple Contextual Meaning predict exactly this, roughly that roots are not assigned meaning until they are categorised.

See however what happens when we take the verbal stem, a *vP* by hypothesis, and nominalize it, using the run-of-the-mill nominalizer *-ma* in (4b). Unlike the explicit predictions in Arad (2005), and as Borer (2009) points out with similar examples, the already categorized element *nomiz-* does not keep its meaning. What happens instead is that the whole [_{nP} *n vP*] structure is (re-)assigned a new, unrelated and completely arbitrary meaning, that of ‘coin, currency’. Perhaps equally interestingly, the participle derived in (4c) from the selfsame verbal stem carries a meaning as if *nomiz-* meant ‘legislate, prescribe by law’. In other words, in (4c), the *vP* embedded within an adjectival shell also fails to keep its “fixed” meaning of ‘think, believe’ and the whole *aP* participle means ‘legally prescribed’.

The underlying question raised by such and similar examples concerns the semantic malleability of roots. Assuming that they are very underspecified semantically, one might ask how underspecified they can be before they become semantically vacuous. The most obvious example is provided by the Latinate roots like *-ceive*, *-mit*, or *-verse*, which in English underlie a variety of semantically unrelated lexemes like *con-ceive* and *re-ceive*, *ad-mit* and *per-mit*, *con-verse* and *per-verse*. In a language like Italian, this type of element does not have a special non-native or even learned status, yet roots like *mett-* which yields as diverse words as those for *di-mett-ere* ‘dismiss / resign’, *s-mett-ere* ‘quit’ and *s-com-mett-ere* ‘wager’. The likes of *mett-* can be found in a number of languages, including Greek *esth-*:

- (5) *esth-an-o-me* ‘feel’
esth-is-i ‘sense’
sin-esth-is-i ‘realisation’, ‘awareness’
esth-i-ma ‘feeling’, ‘love affair’, ‘boyfriend / girlfriend’
sin-esth-i-ma ‘emotion’
esth-is-iaz-mos ‘sensuality’
esth-an-tik-os ‘sensitive, emotional’
esth-it-os ‘perceptible’, ‘tangible’,
an-esth-it-os ‘unconscious’, ‘insensitive’
esth-it-ir-ios ‘sensory’
esth-it-ik-os ‘esthetic’, ‘beautician’

Despite the illusory affinities suggested by the Latinate English glosses (G. Longobardi, p.c.), the class of concepts words derived from *esth-* is broad enough to render impossible the task to associate the root *itself* with any kind of cognitively coherent concept, no matter how underspecified or vague, and even to the exclusion of ‘beautician’.

The problem for the hypothesis of contentful roots is not just that all too often no identifiable content appears to be there. In some cases there is evidence that the different interpretations must be linguistically visible for language-internal purposes. This happens when the same root yields interpretations of different ontological types (like (3) above), which differ for the purposes of further morphological derivations, *after* the root has been categorized, as in the following Greek example:

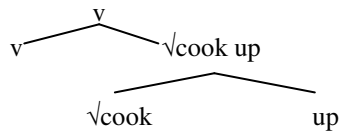
- (6) *paradosi*₁ ‘tradition’ (result / *process nominal)
*paradosi*₂ ‘delivery’, ‘surrender’ (result / process nominal)
paradosiakos relative to *paradosi*₁ (i.e. ‘traditional’), # *paradosi*₂

Acquaviva (2009a, forthcoming) discusses the related case of English deverbal nominalizations in *-ment*, which generally cannot be input to *-al* affixation (cf. *agree* - *agree-ment* - **agree-ment-al*) except when they have a non-transparent interpretation as entity-denoting nouns rather than deverbal nominalizations; for example, *argue* - *argu-ment*₁ ‘logical category’

/ *argu-ment*₂ ‘arguing event’ - *argu-ment-al*, only interpretable as ‘relative to the logical category’, not *‘relative to arguing’. Again, the root by itself cannot discriminate between the two readings, yet the readings must be linguistically visible, and are not just grammar-external.

Even clearer cases where the same root under-determines differences in linguistic behaviour are displayed in cases like the ones studied in Basilico (2008), where the same (atomic) root is compatible with different selectional restrictions, according to the grammatical environment within which it is embedded:

- (7) *the criminals cooked a meal* / *#an evil scheme* (Basilico 2008)
the criminals cooked up an evil scheme



This type of examples is particularly instructive, as it brings out an ambiguity in the notion of root: atomic element individuated morphologically (here, *cook*), or innermost category-free element, defined syntactically and possibly complex (here, *cook up*). This will play an important part on our discussion.

Finally, we can push further the empirical point that lexical meaning is not established and fixed within the first categorizing shell; in fact, we also find cases where the basic lexical predicate is determined only by the choice of inflectional morphemes, after a significant amount of structure has been built. Consider Russian, where the root *tsvet* in different inflectional paradigms (noun declensions) derives both the word for ‘colour’ and the word for ‘flower’:

- | | | | |
|-----|-----------------|----------|---------------------------|
| (8) | SINGULAR | | PLURAL |
| | <i>tsvet</i> | ‘colour’ | <i>tsvet-á</i> ‘colours’ |
| | <i>tsvet-ók</i> | ‘flower’ | <i>tsvet-’y</i> ‘flowers’ |

Even though FLOWER is a basic-level concept, the noun lexicalizing this concept is derived in the singular by the addition of the diminutive suffix *-ok* with individualizing function. There are, to be sure, an archaic form *tsvet* with the meaning ‘flower, blossom’, and a regular plural *tsvet-kí* from *tsvet-ók*; but in so far as the paradigm in (8) reflects a stable synchronic pattern, it

shows that what individuates the concept FLOWER is neither the root by itself (also appearing in *tsvestí* ‘to blossom’) nor, crucially, the root with a nominal suffix, which is absent in the plural, but the choice of one among two alternative inflectional classes, which emerge in the nominative / accusative plural. Further evidence that lexical meaning can be fully established at the inflectional level comes from the numerous idiosyncratic (specialized) interpretations for morphologically regular inflectional plurals (cf. Acquaviva 2008), like the English *brain* (count) - *brains* (count / mass), or the Cypriot Greek *nero* (‘water’), plural *nera* (‘heavy rain’).

4. TWO TYPES OF ROOTS

According to Distributed Morphology, roots are syntactically active elements (but see De Belder 2011 for an interesting alternative). Moreover, they are:

- (9) i. category-neutral and categorized in the course of the derivation;
- ii. meaningful, although there is no consensus on how much content they have;
- iii. phonologically identified as *forms*.

We have a number of objections on these (see also Acquaviva 2009b, Borer 2009; Harley 2012). The first is of a conceptual nature: if roots are indeed meaningful, then they are equivalent to verbs, nouns and adjectives except for a categorial label. This in turn raises serious concerns on the nature, purpose and inescapability of categorization in natural language (see Panagiotidis 2011 for discussion). The second objection concerns two interlinked facts: on the one hand, there exists unconstrained variation between roots that appear to be very specified (e.g. *sugar*), extremely impoverished (e.g. *mett-* in Italian or *mit-* in English) and all the in-between shades. Moreover, even if we argue for impoverished and semantically underspecified roots, we are still left with the empirical problems adumbrated in the previous section, namely that roots too often do not capture a coherent meaning (what connects, for instance, the noun *book* to the verb *to book*? what logical or ontological type should the root *book* have?). This renders unlearnable the purported ‘common semantic denominator’ the roots are supposed to express.

It seems, then, that roots in the technical sense this term has in Distributed Morphology cannot have all the three properties attributed to them. Taking into account also the recent contributions by Borer (2009) and Harley (2012), we propose an alternative which abandons (9ii) and crucially

qualifies (9iii) (see also Acquaviva, forthcoming, Panagiotidis 2011, Acquaviva and Panagiotidis, in prep.).

First, we think that it is necessary to distinguish between roots as morphological objects and roots as elements of syntactic computation. In doing so, we embrace generalized Late Insertion, not just for non-root syntactic material, as in Galani (2005: Ch. 5-6); Siddiqi (2006: Ch. 3); Haugen (2009). Thus, syntactic roots will be associated with different morphological roots (Vocabulary Items, essentially: *forms*) in particular syntactic contexts, as sketched below:

- (10) $\sqrt{\text{CAT}} \longleftrightarrow \text{cat}$
 $\sqrt{\text{GO}} \longleftrightarrow \text{go}$
 $\sqrt{\text{GO, [Tense: Past]}} \longleftrightarrow \text{went}$

Given this dissociation, we can use the notion of morphological roots to account for the multiple ‘radicals’ or ‘stems’ that occur, for instance, in French or Latin inflection and derivation (Aronoff 1994, Bonami, Boyé and Kerleroux 2009). Thus conceived, morphological roots *may* display categorial specifications like being exclusively nominal or verbal, as has been reported for some languages (cf. Hale and Keyser 2002:254) and we expect there to exist constraints on their form (a case in point being the Semitic three-consonant skeleton). Moreover, the same Vocabulary Item (form) that spell out roots, may also spell out functional terminals and semi-lexical categories, as is the case of *will* (the future marker or the noun); see also De Belder (2011). In fact, a notion of morphological root related to, but distinct from, that of syntactic root correctly predicts the existence of such ‘semilexical’ categories.

The consequence of the above dissociation is that we can now conceive syntactic roots, as distinct from morphological ones, as *abstract indices* (cf. Acquaviva 2009b, Harley 2009; 2012). By this we mean purely formal objects internal to the faculty of language in the narrow sense; that is to say, elements that are defined only as constituents of a formal syntactic representation, but have no grammar-external status—in particular, not definable, independently of a syntactic structure, as sound-meaning mappings, or even as abstract instructions to ‘fetch’ or ‘activate’ concepts (contrast Pietroski 2008:319, Pietroski and Hornstein 2009, Boeckx 2010:28-29). What we are essentially claiming is that a syntactic root is a syntax-internal criterion of lexical identity, so that two otherwise identical syntactic constructions count as different formal objects if they differ in the root, and as identical (that is, tokens of the same type) if the root is the same.

Given this characterization, there is no semantic variation to explain between root types, nor learnability problems raised by some elusive conceptual content detached from one or another lexical item; because there is no semantic content in roots. Instead, we argue, roots act as labels to identify (UG-internally) the structures which correspond to lexical words, and it is *these* which support conceptual content. The following section will make explicit the implications of our proposal for the relation between lexical conceptual content and syntactic structure.

5. MAPPING WORD STRUCTURE AND CONCEPTUAL CONTENT

It seems a truism that if lexical items are grammatically complex, then the corresponding lexical concepts are also complex. If the hypothesis we put forward can be substantiated, however, the structural complexity of a word at the linguistic level does not necessarily correspond to complexity in its conceptual content. To see why, recall that syntactic decompositional approaches aim at representing in syntactic terms the grammatically relevant information encapsulated in a lexical word, by means of a structure generated by the same principles that also generate the productive construction of sentences. Now, lexical words also have a non-grammatical content, idiosyncratic and encyclopaedic, which cannot be associated to a word-invariant grammatical shell. It seems natural to associate this irreducibly lexical residue to a root element. But if independent empirical and conceptual arguments make it problematic to associate with roots even this type of content, the question where idiosyncratic lexical meaning is represented must receive a different answer.

The answer we suggest is that a word's conceptual content does not correspond to one particular piece of the syntactic construction, but corresponds to a construction as a whole. Syntactic heads express content regimented into grammatical features, and collectively determine a grammatical interpretation; say, count noun, or unaccusative change-of-state verb, or causative verb. A root at the core of such a construction merely *labels* it; for that purpose, it does not matter whether it is a single node, realized as an invariant phonological form, or a complex node like *cook up* in (7). Assuming that pairs like *break* and *shatter*, *black* and *white*, or *dog* and *cat* have identical structural representations, what we claim is that they are differentiated, at the abstract syntactic level, before lexicalization by different morphological roots, by syntactic roots. What distinguishes morphological roots from each other is their phonological content and

possibly the alternations they determine (think of *-ceive* / *-cept*). Syntactic roots do not differ by virtue of their semantic content, because they don't have content, but by a differential marking, like subscripts. It is by virtue of having different subscripts that the structures corresponding to *dog* and *cat* count as different syntactic objects, independently of semantic interpretation. This means that they are criteria of lexical identity across syntactic representations. Simple concepts, *qua* lexical concepts, map to lexical items that are identified by roots. In a way, therefore, our conception of roots provides a UG-internal signature for lexical concepts. Something like this notion would appear to be independently necessary, according to Margolis and Laurence (2007:583), who argue for a notion of *mental orthography* as shorthand for 'the formal properties that allow the cognitive system to reidentify tokens of the same representation type'.

Consider again Borer's (2005b:9) statement quoted above in section 2, to the effect that a lexical item consists of 'its syntactic structure and the interpretation returned for that structure by the formal semantic component, and [...] whatever value is assigned by the conceptual system and world knowledge to the particular listemes embedded within that structure.' Instead of claiming that the second, conceptual component is regularly associated to grammatically inert listemes (a conclusion that has no independent motivation, even though it may appear natural as a null hypothesis), we claim that an empirically more satisfactory solution consists in taking the structural-grammatical meaning as a semantic template which *constrains* the conceptual content associated with the structure. If the syntax of a verb involves a causative *v* head, the lexical concept associated with it should be a causative verb (like *kill*); but a semantically causative verb does not have to decompose semantically into a non-causative part and a CAUSE predicate definable independently of this particular lexical concept. In essence, then, we argue that there exist morphological and syntactic roots, defined on morphological and syntactic criteria, but that there are no semantic roots as distinct from basic lexical concepts; in particular, not as the semantic content of syntactic or morphological roots.

Of course, it is at best insufficient, and at worst circular, to say that a concept may map to 'whatever' grammatical construct defines a lexical word (N. Hornstein, p.c.); but the claim that concepts do not map to fixed-size syntactic pieces is coherent and compatible with the observed data. As cases like the Russian *tsvet-ók* show, a single lexical concept can be expressed by a noun with different structures in the singular and plural; and especially a category like number may easily be an intrinsic component of the lexical concept, as is clear in 'collective' plurals like the Spanish *padres*, which shifts the meaning of *padre* / *padres* 'father / fathers') to that of 'parents',

but only denoting mother-father pairs (so, a grandmother and her daughter cannot be *padres*, despite being parents; see section 3 above and Acquaviva 2008). It seems plausible that the domain of conceptual lexicalization cannot extend beyond a nominal or verbal extended projection, probably definable as a syntactic Phase corresponding to a DP or a vP; in fact, this is expected if we take seriously the notion of Phase as derivational cycle whose output is consigned to interpretation (Acquaviva and Panagiotodis, in prep.).

5. CONCLUSION: COMPLEX WORDS, SIMPLE CONCEPTS

Lexical decomposition, as a hypothesis on the constituency of words as linguistic representations, captures fundamental aspects of lexical competence. On the other hand, it is problematic as a hypothesis on the internal constituency of lexical concepts. Our main point is that decomposition becomes problematic even from a linguistic perspective, as soon as we ask where a lexical grammatical structure hosts non-grammatical conceptual content; resorting to roots, in particular, proves empirically inadequate. The result of our linguistically motivated alternative hypothesis is that a word can be linguistically complex but conceptually simplex.

Conceptual atomism, as defined by Fodor (1998:121), holds that ‘most lexical concepts have no internal structure’. Since we still claim that the grammatical structure of words comprehends meaningful elements, we do not take this thesis to mean that lexical words are semantically unanalyzable as linguistic objects (in particular, they are not semantic atoms in a *Mentalese*; contrast Fodor 2008). What we claim is rather that a word’s conceptual content is not on a par with such grammatically encoded meaning, as the content of one syntactic piece among others, but belongs outside UG-generated representations and is mapped to them in such a way as to respect the semantic templates defined by grammar. It remains possible, then, to envisage syntax-external atomic concepts mapped to complex syntactic structures.

The difference we envisage between lexical concepts and the UG-internal content of syntactic lexical representations does not mean that the two have nothing to do with each other. On the contrary, a principled relation between the two can lead to predictive hypotheses on what can be a possible lexical word in a natural language. For instance, Fodor (1998:164-165) argues that while REDSQUARE is conceivable as a primitive concept, without having RED and SQUARE, there can be no *primitive*, atomic concept ROUNDSQUARE, as opposed to the complex ROUND SQUARE (as the conceptual content of the phrase *round square*). Such a hypothetical basic concept could never identify anything at all, and would therefore lack

all semantic content on principled grounds; while contradictory properties can be entertained, and the predicate *round square* is perfectly well-formed, this could not exist as a basic concept ('there can be no *primitive* concept without a corresponding property for it to lock to'). If that is correct, it amounts to a prediction about language: namely, that no simplex noun in any natural language can have this conceptual content. In a similar vein, further assumptions and hypotheses about the conceptual bases of lexical nouns can lead to rule out the existence of words meaning 'number of planets' or 'undetached rabbit part' as simplex lexical concepts (Acquaviva, forthcoming).

Finally, it bears emphasizing that the thesis of conceptual atomism, and our contention that syntactic lexical decomposition is compatible with it, does not deny the cognitive complexity of concepts. The content of a word enters in a complex network of relations with the content of other words, as CAT and ANIMAL. But inferences can be necessary though not constitutive: taking *water contains hydrogen* to be necessarily true, it is still possible to have the concept WATER without having the concept HYDROGEN; cf. again Fodor (1998:74):

It's perfectly consistent to claim that concepts are individuated by the properties they denote, and that *the properties* are individuated by their necessary relations to one another, but to deny that knowing about the necessary relations between the properties is a condition for having the concept.

Word meaning, in conclusion, is indeed cognitively complex, but not as a reflex of grammatical complexity. We take it to be a strength of our analysis that it makes linguistically motivated decompositions of lexical items (more) compatible not only with conceptual atomism, but also with views that, without embracing conceptual atomism, emphasize the lack of one fixed structure for lexical concepts; cf. Murphy (2002:441): 'Thus, it can be very difficult to know where to draw the line between what is part of the word meaning "per se" and what is background knowledge. It is not clear to me that drawing this line will be theoretically useful.' A linguistic analysis of the construction of lexical content which could be related to a psychologically and philosophically plausible view of lexical concepts is certainly a desirable goal. Our proposal is a contribution towards that goal.

REFERENCES

- Acquaviva, P. & P. Panagiotidis. In preparation. Roots and lexical semantics. University College Dublin - University of Cyprus.
- Acquaviva, P. (2008). *Lexical Plurals*. Oxford: Oxford University Press.
- Acquaviva, P. (2009a). Roots, categories, and nominal concepts. *Lingue e linguaggio* 8, 25–51.
- Acquaviva, P. (2009b). Roots and Lexicality in Distributed Morphology . In A. Galani, D. Redinger & N. Yeo (Eds.) *York-Essex Morphology Meeting 5*, 1–21.
- Acquaviva, P. Forthcoming. The roots of nominality, the nominality of roots. In A. Alexiadou et al. (Eds.), *The Syntax of Roots and the Roots of Syntax*. Oxford University Press.
- Arad, M. (2005). *Roots and Patterns: Hebrew Morpho-Syntax*. Berlin: Springer.
- Arad, Maya. (2003). Locality Constraints on the Interpretation of Roots: The Case of Hebrew Denominal Verbs. *Natural Language and Linguistic Theory* 21(4), 737–778.
- Aronoff, M. (1994). *Morphology by Itself*. Cambridge, MA: MIT Press.
- Basilico, D. (2008). Particle verbs and benefactive double objects in English: High and low attachments. *Natural Language and Linguistic Theory* 26, 731–729.
- Bever, T. & P. Rosenbaum. (1970). Some lexical structures and their empirical validity. In R. A. Jacobs and P. S. Rosenbaum (Eds), *Readings in English Transformational Grammar* (pp. 3–19). Waltham, MA: Ginn and Company.
- Boeckx, Cedric (2010b). Defeating Lexicocentrism. Ms., CLT/UAB. Available from <http://ling.auf.net/lingBuzz/001130>
- Bonami, O., G. Boyé, & F. Kerleroux. (2009). L'allomorphie radicale et la relation flexion-construction In B. Fradin, F. Kerleroux, & M. Plénat (Eds), *Aperçus de morphologie du français* (pp. 103–125). Paris: Presses universitaires de Vincennes.
- Borer, H. (2005a). *In Name Only*. Oxford: Oxford University Press.
- Borer, H. (2005b). *The Normal Course of Events*. Oxford: Oxford University Press.
- Borer, H. (2009). *Roots and Categories*. Paper presented at the 19th Colloquium on Generative Grammar. University of the Basque Country, Vitoria-Gasteiz.
- Cruse, D. (1982). *Lexical Semantics*. Cambridge: Cambridge University Press.
- DeBelder, M. (2011). *Roots and affixes: eliminating lexical categories from syntax*. LOT: Utrecht.
- Di Sciullo, A.M. & E. Williams. (1987). *On the Definition of Word*. Cambridge, MA: MIT Press.
- Embick, D. & A. Marantz. (2008). Architecture and Blocking. *Linguistic Inquiry* 39(1), 1–53.
- Fodor, J.A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.
- Fodor, J.A. (2008). LOT 2: *The Language of Thought Revisited*. Oxford: Oxford University Press.
- Galani, A. (2005) *The morphosyntax of verbs in Modern Greek*. Unpublished PhD thesis, University of York

- Grimshaw, J. (1993). Semantic structure and semantic content in lexical representation. Ms, Rutgers University. Published in Grimshaw, J. (2005). *Words and Structure* (pp. 101–119). Chicago: CSLI.
- Hale, K. & S. J. Keyser. (1993). On argument structure and the lexical expression of syntactic relations. In K. Hale & S. J. Keyser (Eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger* (pp. 11–41). Cambridge, Mass: MIT Press.
- Hale, K. & S. J. Keyser. (2002). *Prolegomenon to a Theory of Argument Structure*. Cambridge: MA: MIT Press.
- Halle, M. & Alec Marantz. (1993). Distributed Morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger* (pp. 111–176). Cambridge, Mass: MIT Press.
- Harley, H. & R. Noyer. (2000). Formal versus Encyclopedic Properties of Vocabulary: Evidence from Nominalisations. In B. Peeters (Ed.), *The Lexicon-Encyclopedia Interface* (pp.349–374). Amsterdam: Elsevier.
- Harley, H. (2005). How do verbs get their names? Denominal verbs, manner incorporation, and the ontology of verb roots in English. In N. Erteshik-Shir & T. Rapoport (Eds.), *The Syntax of Aspect* (pp. 42–64). Oxford: Oxford University Press.
- Harley, H. (2009). *Roots: Identity, Insertion, Idiosyncracies*. Paper presented at the Root Bound workshop, USC, February 21, 2009
- Harley, H. (2012). *On the Identity of Roots*. Unpublished ms., University of Arizona.
- Haugen, J. (2009). Hyponymous objects and Late Insertion. *Lingua* 119, 242-262
- Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.
- Jackendoff, R. (2002). *Foundations of Language*. Oxford: Oxford University Press.
- Jackendoff, Ray. (1990). *Semantic Structures*. Cambridge, MA: MIT Press.
- Laurence, S. & E. Margolis. (1999). Concepts and cognitive science. In S. Laurence & E. Margolis (Eds.), *Concepts: Core readings*. Cambridge, MA: MIT Press.
- Levin, B. & M. Rappaport Hovav. (1995). *Unaccusativity*. Cambridge, MA: MIT Press.
- Levin, B. & M. Rappaport Hovav. (2005). *Argument Realization*. Cambridge: Cambridge University Press.
- Lieber, R. (2004). *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.
- Marantz, A. (1997). No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon. In A. Dimitriadis et al. (Eds.). *University of Pennsylvania Working Papers in Linguistics* 4(2), 201–225.
- Marantz, A. (2000). *Words*. Unpublished ms. MIT.
- Marantz, A. (2006). *Phases and words*. Unpublished ms. NYU.
- Margolis, E. & S. Laurence. (2007). The ontology of concepts—abstract objects or mental representations? *Noûs* 41, 561–593.
- Murphy, G. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

- Pietroski, P. (2008). Minimalist meaning, internalist interpretation. *Biolinguistics* 2, 317–340.
<http://www.biolinguistics.eu/index.php/biolinguistics/article/view/70/87>
- Pietroski, P. & N. Hornstein. (2009). Basic operations: Minimal syntax-semantics. *Catalan Journal of Linguistics* 8, 113–139
- Pinker, S. (1989). *Learnability and Cognition*. Cambridge, MA: MIT Press.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.
- Ramchand, G. (2008). *Verb Meaning and the Lexicon*. Cambridge: Cambridge University Press.
- Siddiqi, D. (2006). *Minimize exponence: economy effects on a model of the morphosyntactic component of the grammar*. Unpublished PhD thesis, University of Arizona.
- Wunderlich, D. (1997). Cause and the structure of verbs. *Linguistic Inquiry* 28, 27–68.
-

Paolo Acquaviva
University College Dublin
Newman Building, Belfield, Dublin 4
Ireland
e-mail: paolo.acquaviva@ucd.ie

Phoevos Panagiotidis
University of Cyprus
Dept. of English Studies
Kallipoleos 75
1678 Nicosia
Cyprus
e-mail: phoevos@ucy.ac.cy