

On the nature of roots: content, form, identification¹

Phoevos Panagiotidis – *University of Cyprus*

ABSTRACT:

This paper offers a review of a current understanding of the content and the form of linguistic roots. It first updates and buttresses the case against semantic content of uncategorised roots and for Late Insertion of roots; then it investigates how native speakers identify roots. More specifically, the idea that roots may be polysemous or may encode the shadow of a denotation, namely the common denominator of the denotations of words derived from it, is refuted on the basis of conceptual and empirical arguments from a number of languages. Subsequently, the existence of a spectrum of content to which roots belong, with roots ranging from contentless to semantically specific and concrete, is also shown to be illusory, and to result from the actual productivity, hence diversity, of the words derived from it. Arguments for Late Insertion of roots are then reviewed and updated, divorcing roots from the forms that realise them. These arguments are systematically combined with the semantic contentlessness of roots in support of Acquaviva's analysis of them as abstract indices, i.e. as the syntax-internal criteria of lexical identity. This account is taken to its logical conclusion in the final section: if roots are indeed abstract indices, then they cannot be identified either by the semantic content they realise within grammatical structures or by their forms. An account is therefore advanced according to which roots are identified just once by native speakers over their lexicon at a given moment and on the basis of three heuristic principles: one form-based, one based on the feature content and the exponence of the structures in which roots are embedded, and one taking care of root suppletion.

KEYWORDS: root, denotation, index, Late Insertion, suppletion, root learning.

1 *Introduction: why roots?*

The study of linguistic roots is one of the first matters Historical Linguistics of the 19th century turned its attention to. In the historical tradition roots were initially conceived differently from what they are today. For the Indo-European research programme, roots amounted to the idealised common denominator of individual related words: a reconstructed ancestral word form stripped off from all language-specific morphology and 'corrected' for language-specific and family-specific sound laws (*Lautgesetze*). For instance, Old English *nefa* ('nephew, male cousin, grandson'),

¹ I am grateful to Hagit Borer, Richard Larson, Vitor Nóbrega, Vasilis Spyropoulos and audiences in Queen Mary, UA Barcelona, Peking University, and Göttingen for their comments, criticism, and feedback. I remain the owner of all errors and misconceptions.

Old High German *nevo* (‘nephew’), Latin *nepōs, -tis* (‘grandson, nephew’), Homeric Greek *népodes* (‘descendants’), Avestan and Sanskrit stem *napāt-* (‘grandson, descendant’), Old Lithuanian *nepotis* (‘grandson’) and so on were all argued to historically derive from the Proto-Indo-European root **nepot-*. The reason this brief historical introduction is made here is because I feel that when it comes to investigating roots in the contemporary sense of the term, we are somehow fixated to the 19th century methods and analytical concepts, especially in our understanding roots as forms and as ‘common denominators’.

With the emergence of synchronic linguistics and of more formal approaches to language in the 20th century, roots remained exclusive stock of morphology: entities completely opaque to syntactic processes due to the lexicalist assumptions, tacit or express, of most work in grammatical theory until the 1990s (Bruening 2018 for a succinct critical survey of Lexicalism). Roots in morphological theory may not be nearly as intuitively prominent as words, but still they are pre-theoretically identifiable as the basic building blocks of words. The irony here is that the intuitively straightforward words remain definitionally elusive, whereas for roots morphologists at least possess a valuable rule of thumb in Aronoff’s (1994, 40): “a root is what remains after all the morphology has been wrung out of a form”.

Since the advent of separationist / realisational theories (Ralli 1988; Beard 1995; Halle and Marantz 1993; Marantz 1997 among others), roots have become a *syntactically relevant* class and they have found a place among the actual building blocks of syntactic structures. Talking about the nature of roots here I will assume some version of separationist / realisational approaches to be true; for concreteness I will loosely couch my discussion within Distributed Morphology (Halle and Marantz 1993; Harley and Noyer 1999), one of the few frameworks with concrete hypotheses about roots, along with that of Borer (2009; 2013; 2014).

Roots in Distributed Morphology were assumed to be exceptional among the building blocks of linguistic structures in three ways:

- a. They were understood to intrinsically associated with concepts, concrete or vague. This made roots quasi-*signs* in the Saussurean sense in that they were conceived as structure-less combinations of a form (the signifier) and a (vague) signified. This understanding is all more peculiar as ‘meaning’ is otherwise assigned late in both Distributed Morphology and several contemporary syntactic frameworks.
- b. They were phonologically identified. This entailed that Late Insertion, the matching of abstract morphemes with Vocabulary Items, took place only for grammatical feature structures, with roots ‘carrying’ their phonological interpretation throughout the combinatorial process (“the syntactic derivation”). Again, this amounts to an anomaly in the Distributed Morphology architecture: Late Insertion in separationist / realisational approaches is necessary because phonological features have no business in the process structure building.

- c. They were syntactically active, occupying syntactic nodes. This is of course what took roots out of the lexicalist ‘word shell’ into the open of syntactic derivation.

The discussion that follows is informed by Harley (2014) and interacts closely with Acquaviva (2009; 2014), Borer (2009; 2013; 2014), Acquaviva & Panagiotidis (2012), and Panagiotidis (2011; 2014a; 2014b). I take for granted the Distributed Morphology claim that roots are uncategorised elements that are *categorised* by grammar, yielding nouns, verbs, and adjectives.

In the first part of this paper I will make the by now well-established case against root denotation (what is informally known as ‘root content’) and against the phonological identification of roots, such approaches carrying a 19th century bias with them; subsequently, I will rehearse the arguments for roots as indices. In the last section I will discuss how roots can be identified if not on the basis of form and/or content.

2 *The content(lessness) of roots*

Root content was until fairly recently a poorly studied matter. When in the late 20th century roots began to be considered elements manipulated by syntax, they were treated as something like ‘small words’, i.e. as Saussurean signs; even more peculiarly roots were treated, in practice if not in theory, as ‘small nouns’. Acquaviva & Panagiotidis (2012, 7) point out that roots are often “unambiguously qualified as semantically contentful signs” and they bring as an example the account in Harley (2005, 46–50), which derives the Aktionsart opposition between unergatives *drool* (atelic) and *foal* (telic) as a result of the semantic properties of roots. Accordingly, the verb *drool* is thought to be atelic because it is derived from an unbounded / mass root DROOL, whereas the verb *foal* is telic because it is derived from a bounded / count root FOAL. Their observation is that not only are roots understood as meaningful, but they are essentially treated as *nominal*, with a mass / count distinction actually constituting an intrinsic part of their semantic content. Acquaviva & Panagiotidis (2012) and Acquaviva (2014) systematically deconstruct this approach on the basis of empirical arguments regarding the variety of denotations words derived from the same root display, a matter to be revisited below.

The fact that roots are impossible to treat as ‘small words’ had of course already been dealt with. Levinson (2007) argued that roots are polysemous, with Marantz (2012) speaking of root *allosemy*. At the same time, Arad (2005, chap. 3), argued that roots are semantically impoverished / underspecified, while Borer (2009; 2014) openly claimed them to be contentless.

But why would we think either that roots are polysemous or that root content is impoverished / underspecified? The general underlying idea is that free (=uncategorised) roots are not either appropriately or adequately specified to stand on their own as legitimate LF-objects: therefore *roots must be categorised* (as nouns, verbs etc.). Now, this is an assumption independent of the stand one takes towards how much semantic content roots possess and it is known as the *Categorisation*

Assumption (Embick and Marantz 2008, 6). Of course, as it stands here, the Categorisation Assumption is an empirical generalisation. For the purposes of the discussion here let us just take it to be true and go with a modest proposal that Syntax does not use some special operation to legitimise roots but simply embeds them within a categoriser projection (Panagiotidis 2011): roots are categorised by a categorising projection merged on top of them.

Backtracking a bit, we can probably discern that the above correlation between root content (a ‘semantic’ matter) and categorisation (a ‘syntactic’ matter) is elusive. Do we know that roots must be categorised and that they cannot stand on their own (i.e. uncategorised) as legitimate LF-objects *because* root content is impoverished / underspecified? Or do we know that root content is impoverished / underspecified *because* roots must be categorised? In other words, on a purely conceptual level we risk positing that categorisation is a formal requirement for roots due to their impoverished / underspecified content, which we assume to be the case exactly because roots formally require categorisation. Empirical arguments are therefore needed, such as those hinted at above, which will be discussed in the rest of this section.

2.1 For content

Probably in all languages, casually looking at words derived from the same root one will reach the conclusion that they seem to share a common conceptual core, cf. Hale & Keyser (1993), (2002); Rappaport Hovav & Levin (1998); Levin & Rappaport-Hovav (2005). This is the case with some straightforward examples in English, where roots can be identical with free morphemes:

(1) *Well-behaved roots in English*

_Nbutter _vbutter
_Nland _vland
_Ared _Nredness _vredden

This is also the case in languages where roots are never anything like free morphemes, including Semitic languages with their typically triconsonantal ineffable roots; consider the Hebrew root QLT (Arad 2005, 97):

(2) *Well-behaved roots in Hebrew: QLT*

Nouns:

miqlat (‘shelter’)
 maqlet (‘receiver’)
 taqlit (‘record’)
 qaletet (‘cassette’)
 qelet (‘input’)

Verbs:

qalat (‘absorb’, ‘receive’)
 hiqlit (‘record’)

All the words derived from QLT have very different denotations, ranging from *qaletet*, the learned word for cassette, to the verbs ‘absorb’ or ‘shelter’. Intuitively they nevertheless seem to share a (very) abstract common meaning core, along the lines of ‘keep, preserve’. Judging from the likes of (1) and (2) we would be led to the conclusion that roots are assigned *Multiple Contextual Meaning* (Arad 2005, 65) and that an uncategorised root contains the ‘common semantic denominator’ (Arad 2005, 4–6, 55–59, 271–74) of the words derived from it. This common semantic denominator can be quite concrete, as in (1), or semantically impoverished, as in (2).²

Of course, detecting a common denominator among words derived from the same root is hardly a straightforward matter in many cases most of the times.

2.2 Content? Really?

It turns out that not all roots are like QLT. Aronoff (2007, 819) presents the words derived from the Hebrew root KBŠ, among which a common meaning core, however abstract, is impossible to detect:

(3) *Less well-behaved roots in Hebrew: KBŠ*

Nouns:

keveš (‘gangway’, ‘step’, ‘degree’, ‘pickled fruit’)

kviš (‘paved road’, ‘highway’)

kviša (‘compression’)

kivšon (‘furnace’, ‘kiln’)

Verbs:

kibeš (‘conquer’, ‘subdue’, ‘press’, ‘pave’, ‘pickle’, ‘preserve’)

kavaš (like *kibeš* plus ‘store’, ‘hide’)

It is quite telling that in order to explain the disparity among the word meanings in (3), arguments from diachrony will be customarily invoked: pickles were made by applying pressure, paving involves applying pressure, and whatever story one would have for degrees and gangways and furnaces or kilns. Clearly, just-so stories are not of interest in our case, as what we are interested in is not etymologies and word origins but speakers’ tacit knowledge of a root as bearing semantic content. Because of the way roots work in Hebrew, arguing for root homophony would not work here, although root polysemy would probably do. Still, this is hardly an argument for root polysemy: unlike allomorphy, where allomorphs are inserted according to some version of the Elsewhere Condition (Halle and Marantz 1993; Halle 1997), purported root polysemy is seemingly unrestricted, with allosemes getting inserted wherever they end up inserted. In other words, arguing for root allosemy in case like (3) can be

² The whole exercise of course presupposes that concepts are not atomic (Fodor 1998; 2008) but constructed by grammar on the basis of a ‘denotational core’ supplied by the root, whatever that might be. Incidentally, the behaviour of roots like QLT goes some way towards explaining why roots cannot just stand for “predicates”, as an anonymous reviewer suggests: if QLT were a predicate, what would be its extension?

a problem because the morphosyntactic environment in which each alloseme ('meaning') is 'inserted' is hardly definable.

Roots not expressing anything like a common semantic denominator exist in other languages, too. Modern Greek is a case in point, where roots typically behave like bound morphemes. The case of ESTH is quite telling (adapted from Panagiotidis 2014b, 417–18):

(4) *Less well-behaved roots in Greek: ESTH*

Verbs:

esth-an-o-me	'feel'
ði-esth-an-o-me	'intuit, sense'

Nouns:

esth-is-i	'sense'
sin-esth-is-i	'realisation'
en-sin-esth-is-i	'empathy'
esth-i-ma	'feeling', 'love affair'
sin-esth-i-ma	'emotion'
esth-i-ma-t-ias	'romantic'
esth-is-az-mos	'sensuality'
esth-it-ik-os	'beautician'

Adjectives:

esth-an-tik-os	'sensitive, emotional'
esth-it-os	'perceptible', 'tangible', '
esth-it-ik-os	'esthetic'
esth-it-ir-ios	'sensory'

The range of meanings words derived from this root display is very diverse, adumbrating a picture similar to that in (3). However, the breadth and the internal incoherence of this range "is obscured by our Latinate bias" because words denoting such different concepts as *sensory*, *sensitive*, and *sensual* share the same root in English. The interesting aside here is that the words in (4) all belong to the learned stock of the Modern Greek vocabulary, which was partly created in the 19th and 20th centuries as translation loans of (Latinate) vocabulary from French: in Ancient Greek the stem *αἰσθ-* (aisth-) yielded words purely related to perception and sensation.³ Whatever the historical facts, even a most vague common semantic denominator for the words in (4) is impossible to find, including something as uselessly vague as 'perceptual or emotional feel' (as a reviewer suggests for Latinate SENS) would be inadequate and or inaccurate: how to bring together beauticians and realisations? Moreover, such a vague 'denotation' for ESTH would also chaotically overlap with the 'common denominators' of numerous other roots in a non-principled way.

³ With the single exception in Aristotle's *Rhetorics* 1386a32, where *αἴσθησις* (aesthesis) is used to mean 'display of feeling' according to the Liddell-Scott-Jones dictionary of Greek.

Of course, ‘less well-behaved’ roots like those in (3) and (4) can be found in all languages we care to look at. Consider e.g. English conversion noun-verb pairs like *egg* (what non-mammals lay vs. ‘encourage’), *book* (the object or the work vs. ‘reserve’), *object* (the mental category vs. ‘disagree’) and many others, see Clark and Clark (1977). Let me just add the Italian root *METT*- here (Panagiotidis 2014a, 293–94)

(5) *Less well-behaved roots in Italian: METT*⁴

mett-ere	‘put’
am-mett-ere	‘admit’
com-mett-ere	‘commit’
di-mett-ere	‘dismiss / resign’
pro-mett-ere	‘promise’
s-mett-ere	‘quit’
s-com-mett-ere	‘wager’

Or even hand-picked examples like the following, all suggested by Paolo Acquaviva, p.c. (cf. Acquaviva and Panagiotidis 2012, 110–13):

(6) *More instances of unruliness*

Swahili

class 11/4:	u-siku	‘night’	
class 9/10:	siku	‘day’	
class 3/4:	m-ti	‘tree’	mi-ti ‘trees’
class 7/8:	ki-ti	‘chair’	vi-ti ‘chairs’

Latin

malus	‘apple tree’ (fem.)	malum	‘apple’ (neut.)
-------	---------------------	-------	-----------------

Italian

man-ic-o	‘handle’ (masc.)	man-ic-a	‘sleeve’ (fem.)
----------	------------------	----------	-----------------

Spanish

libr-o	‘book’ (masc.)	libr-a	‘pound’ (fem.)
--------	----------------	--------	----------------

2.3 Some roots are thicker than others?

Summarising, there appears to exist a picture in which roots seem to carry a varying degree of content *in isolation*: some seem to be very specific, giving off the impression that they are indeed Saussurean signs (consider *DOG*), others only carry some content, others no content at all. Before we continue, note that from a semantic viewpoint an argument based on the specificity or vagueness of the concepts roots in

⁴ In a bid to explain away examples like *METT*, both reviewers invoke the well-known observation that Romance prefixation yields non-compositional results, while one of them thinks that *METT* itself has a “perfectly well-defined core meaning”. There are at least two issues here: First, the non-compositional products of Romance prefixation are an explanandum, not an explanans: do prefixes *make* roots behave ‘like idioms’ (as one of the reviewers points out) or do they attach to a meaningless element to form a composite structure that will have to be matched with a denotation later? See also sections 2.4 and 4. Second, what is the purported “perfectly well-defined core meaning” of *METT*? If it is something like ‘abstract movement’ then we run into the problems raised in the discussion of (3) and (4).

isolation would be supposed to be associated with is non-trivial: by which metric is the ‘dog’ concept more precise and/or more specific than the ‘keep, preserve’ one exemplified in (2)?

Setting this problem aside, perhaps we could go along with Saab (2016) and argue that roots are arranged along a ‘content scale’ according to the ‘amount’ or ‘specificity’ of denotation they possess before categorisation, i.e. in isolation. On the one end of this scale we would have roots like *METT* or *KBŠ*, which are completely meaningless in isolation. On the other end of the scale we would have roots like *DOG* or, perhaps, *SUGAR*, which would possess rich and specific content. In the middle of such a scale one would find *QLT* and perhaps even *ESTH*: roots that either i) bear a shadow of content, associated with a vague ‘common denominator’, or ii) are polysemous until they are categorised (Levinson 2007; Marantz 2012).

Maybe we could follow up on the idea of this content scale along which uncategorised roots are arranged to claim that less specified roots (like *METT*) give rise to “crazier”, i.e. more idiosyncratic, word meanings whereas more specified and contentful roots (like *DOG*) lend themselves to yielding more ‘compositional’ ones. This is essentially the claim in Alexiadou & Lohndahl (2017), according to which languages like English have basically a repertory of *QLT* and *SUGAR* type roots, whereas Semitic has *KBŠ* and *QLT* type roots. The descriptive generalisation here would be that Hebrew roots mostly belong to the least specified side of the purported root content spectrum, with English roots being distributed mainly towards the opposing end.

Alexiadou & Lohndahl (2017), like Saab (2016), also draw a correlation between concreteness and morphological unboundedness: the fact that English tends to have contentful roots correlates with English roots surfacing as free morphemes; the fact that Hebrew roots tend to be vague or empty correlates with them being always bound. This more specific correlation could accommodate differences in root content within the same language: *DOG*, surfacing as a free morpheme, also has pretty concrete content, whereas *MIT*, which is always bound in verbs like *commit*, *submit*, *admit* etc., is as empty as its Italian counterpart *METT*.

There are of course at least two serious issues regarding this assumed correlation: The first has to do with the architecture of the language faculty: why would morphological behaviour, a matter of realisation and linearisation and, consequently, of the interface with the Articulatory-Perceptual systems, would correlate with the ‘amount’ and type of denotation an uncategorised root is associated with? The second issue is of course asking ourselves if this is indeed a valid correlation. Before the necessary statistics is run, one can have a cursory look at synthetic languages like Italian or Greek where the vast majority of roots are bound, but they are nevertheless scattered throughout the purported spectrum.

An elementary concern underlies whatever implementation of an idea that uncategorised roots can bear a variety of denotational content, also reminding ourselves of the obvious concerns on the conceptual side, as in footnote 2: how is a

native speaker (or a linguist) going to decide *how much* content a root possesses? They will obviously have to inspect words derived from it. However, which words will the speaker have to inspect? All of them? This is perhaps not a possibility: vocabularies are not learned in their totality. To contribute some anecdotal evidence, as a native speaker of Modern Greek I only encountered the word αποδρόμή ('apoðromi') some ten years ago. Although I could easily decompose it as containing a root ĐROM, I didn't know its meaning (it turns out it means 'remission'). Did this new word change the 'amount of content' of the root ĐROM inside my lexicon? How many words derived from a single root must one know before they can decide on its 'content'? When does one stop? This question is something we will turn to in Section 5 of this paper.

Still, the correlation Alexiadou & Lohndahl (2017) and Saab (2016) are drawing is not groundless. However it is not morphological boundedness that appears to correlate with how contentful the root appears but, rather *productivity*: more productive roots in the languages in question are usually morphologically bound, like MIT, and they seem to possess less content than the least productive roots, like DOG, which tend to be realised as free morphemes and to pass as the most denotationally concrete ones.

Consider the case of the word *laser*, derived from a root LASER. This is of course originally an acronym, dating back to 1957: "Light Amplification by Stimulated Emission of Radiation". Now, the root LASER looks like it has a *very* concrete and rich meaning; there are however no other words derived from it, so we cannot really know: hence the nouns *laser* and the uncategorised root LASER give the illusion of having identical denotations. Still, through metaphorical use the range of meanings associated with the noun *laser* seems to be expanding, possibly leading to the root becoming "less contentful" or "concrete" in the near future. urbandictionary.com defines *laser* also as "a synonym for 'cool'", whereas google searches yield *a laser stare* (this being a sharp direct stare, like the ones that metaphorically 'burn holes') and *throwing a laser* (this being a strong direct shoot in American football).

This kind of diachronic opening up of or 'diminishing' of 'root content', once the root becomes more productive, is witnessed in the case of the Greek root ZAXAR (Panagiotidis 2014a, 300–301). The word *zaxari* ('sugar') was introduced into the Greek language some time in the 19th century, along with the substance it names. We can presume that initially, just like the English root LASER, the Greek root ZAXAR would indeed look like one with a concrete and rich meaning, also deriving just one noun. Fast forwarding to the early 21st century the picture has developed into one partially captured in (7) below:

(7) *Sugar?*

zaxar-i	‘sugar’
zaxar-o	‘diabetes’, ‘blood sugar’
zaxar-en-ios	‘made of sugar’
zaxar-ux-o	‘dulce de leche’ (a substantivized adjective: ‘having sugar’)
zaxar-on-o	‘crystallize (for edibles)’, ‘leer at something’, ‘get turned on’

By now the root ZAXAR looks more like QLT in (2) or even ESTH in (4), with a much vaguer ‘content’. The emerging generalisation is then not one about morphological boundedness of roots, but about overall productivity. Simply put, the greater the number of words derived from a root, the vaguer, and indeed the more elusive, the ‘semantic content’ of this root will tend to be. This can be subsumed in the following generalisation:

- (8) The illusion of semantic concreteness of a root is created by the fact that very few words are derived from it.

This is an issue we will return to in Section 4. Until then, the following passage encapsulates the way we should in principle think about roots:

- (9) “Words have morphological structure even when they are not compositionally derived, and roots are morphologically important entities, [even] though not particularly characterized by lexical meaning.” (Aronoff 2007, 819)

2.4 Radically against root content

At this point it must have become evident that we will opt for the radical thesis that *all roots are completely meaningless in isolation*. In other words, there is no semantic content whatsoever in roots before they are inserted in grammatical structures, where they are categorised (Acquaviva 2009; Borer 2009; Acquaviva and Panagiotidis 2012; Harley 2014; Panagiotidis 2014a). Hence,

- (10) Roots don’t identify word-specific, non-structural meaning.

Root meaning is an illusion of their having a small number of words derived from them, usually nouns. The illusion of *inherent* semantic content of roots can be enhanced in cases where a particular lexical item yields a number of derived words in its turn, as is the case with all the words in (1), also including their compounds etc. What is argued here is that roots can be associated with a denotation only inside a particular grammatical structure also consisting of category, affixes, particles etc. The Greek root NOM illustrates this (Acquaviva and Panagiotidis 2012, 110–11):

(11) *Laws, beliefs, currency.*

[_{nP} nom-os]	‘law’
[_{vP} nom-iz-]	‘think’
[_{nP} [_{vP} nom-iz-] ma]	‘coin, currency’
[_{aP} ne- [_{vP} nom-iz-] men-]	‘legally prescribed’

The above picture can be captured both by Levinson and Marantz’s root polysemy, and by Arad’s Multiple Contextual Meaning (Arad 2005, 65) – although it would be hard for the latter to decide on the inherent ‘common denominator’ denotation of the root NOM. One of the reasons we will have to subscribe to the more radical thesis that roots do not identify *any* “word-specific, non-structural meaning” in (10) is that often roots acquire meaning within grammatical structure, but not necessarily the structure immediately projected by the categoriser. As Acquaviva & Panagiotidis (2012, 110–11) point out, in (11) the already categorized element *nomiz-*, a *vP*, when nominalized using the run-of-the-mill nominalizer *-ma* 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 [...] from the selfsame verbal stem carries a meaning as if *nomiz-* meant ‘legislate, prescribe by law’. In other words, in [(11)], 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’.

Interestingly, this is hardly some sort of exception. There are quite a few cases suggesting that root ‘content’ can be renegotiated even after categorisation at the First Phase has taken place and after the categoriser structure has been matched with a denotation (Marantz 2006; Panagiotidis 2011). The interested reader could refer to Acquaviva (2009; 2014), Acquaviva & Panagiotidis (2012), Borer (2014) and Harley (2014) for a discussion of cases like (11), where recategorisation thoroughly renegotiates the denotation that the root supports.

Such renegotiation is very interesting because root content is supposed to become fixed within some domain of idiosyncrasy and is understood not to be amenable to further tinkering by superimposed structures. This domain of idiosyncrasy can be either the First Phase (Arad 2003; 2005; Marantz 2006; Embick and Marantz 2008), typically the categoriser phase, or something like the topmost lexical projection (Borer 2009; 2013; 2014). The ‘First Phase’ version of this idea is already falsified by the likes of (11), where the nominalisation of the *vP nomiz-* introduces a new and unrelated denotation.⁵ It also looks like that Borer’s version of how far the idiosyncrasy domain may extend also fares less well in the face of cases where lexical

⁵ Marantz (2012) attempts to explain away examples like this along the lines of *something* like phase extension. It however turns out that phase extension would be relevant in the derivation of the adjective category in general (Mitrović and Panagiotidis 2018; 2020).

meaning may be expressed through inflectional means (Acquaviva and Panagiotidis 2012):⁶

(12) *Inflection as defining lexical meaning*

Number:

membro (MASC)	‘member’	<i>Italian</i>
membri (PL.MASC)	‘members’	
membra (PL.FEM)	‘limbs’	
nero	‘water’	<i>Cypriot Greek</i>
nera	‘waters’, ‘rain’	

Aspect & Tense

è mancato	‘was missed’, ‘died’	<i>Italian</i>
mancava	‘was missing’, ‘*was dying’	

The conclusion to be drawn here is that roots possess no content in isolation, i.e. outside the structures that categorise them. Roots are definitely nothing like either ‘shadows’ of words or stem-like ‘partial’ words that are associated with a denotation, vague or not. This means that uncategorised roots are completely contentless and contain neither a fragment nor a shadow of lexical meaning. Making (10) even more explicit, we reiterate what is already clarified in Arad (2005, 57–71):

- (13) We cannot do lexical semantics with roots: the unit of lexical semantics is the ‘word’, never the root.⁷

3 *Late Insertion for roots, too*

As already mentioned in the introductory remarks of this paper, the original formulation of ‘root’ is morphological, with roots standardly understood as ‘forms’. Hence, Aronoff (1994, 40), as already seen, informally describes a root as what remains after all the morphology has been wrung out of a form; Borer (2009; 2014) also argues that roots can be identified on the basis of their form, also arguing that they have no content. The idea that roots *are* forms is quite persistent, even in separationist frameworks like Distributed Morphology, where roots are understood to be inserted early in the derivation, either complete *with* their form or *as* forms. This, as also mentioned, stands in pronounced contrast with Vocabulary Items (i.e. all other forms), which undergo Late Insertion and get matched with feature structures.

Conceptually, inserting roots early is a problem already noticed by Emonds (1985; 2000).⁸ Indeed, if roots are identified by their form and/or are inserted as forms, this

⁶ I agree with an anonymous reviewer that this is a matter crucial for our understanding of how structures are matched to denotations, one that possibly require bold solutions such as the one in Preminger (in progress). This is where I will have to leave the discussion for the time being, sadly. See also footnote 11.

⁷ Whatever words are.

entails that they carry their phonological features throughout the derivation, i.e. throughout the ‘syntax-all-the-way-down’ structure building (Harley and Noyer 1999). Of course, phonological features are irrelevant in syntactic processes, that is in structure-building as opposed to structure-realising processes; this was one of the concerns that motivated the emergence of separationist / realisation morphological frameworks and the general movement against lexicalism in the first place (cf. Bruening 2018).

Roots as forms introduced together with their phonological features would be identifiable solely on the basis of their form (they are contentless anyway), as in Borer (2009; 2014). In Semitic this is more than plausible, as roots are typically triconsonantal ‘skeletons’; it is also plausible in languages like English, where roots vary in form and where root homophony is very limited. However, the claim that roots can be identified only on the basis of their form creates an empirical problem in synthetic languages where it is not *prima facie* evident which roots are involved (Ralli 1988; 2003). By way of illustration, here are some Modern Greek words containing a stem *nom-*:

(14) *Of sewers and laws*

<i>nóm-os</i>	‘law’
<i>nom-ós</i>	‘prefecture’
<i>nom-í</i>	‘distribution’, ‘grazing’
<i>ypo-nom-os</i>	‘sewer’
<i>nom-arx-is</i>	‘prefect’
<i>astr-o-nom-ia</i>	‘astronomy’
<i>nom-ik-os</i>	‘legal ₁ , juristic’
<i>nom-im-os</i>	‘legal ₂ , lawful’
<i>para-nom-os</i>	‘illegal, outlawed’
<i>nom-o-thet-o</i>	‘legislate’
<i>nom-iz-o</i>	‘think, believe’
<i>nom-iz-ma</i>	‘coin’, ‘currency’

Based purely on form, how many roots are involved in the derivation of the above (and more) words? Even more importantly, if form is the criterion to identify roots for native speakers and/or linguists, then

- a. there is possibly no such thing as root homophony, and
- b. true root suppletion cannot exist.

Regarding root suppletion, there are some oft-cited cases like the ones below:

(15) *Root suppletion?*

be / was / were / is / are / was / were
go / went

⁸ In the guise of ‘lexical material’ of course.

Embick & Marantz (2008) refute such cases are instances of root suppletion, recasting them as instances of functional allomorphy, where no roots are involved. To wit, Emonds (1985, chap. 4) has already identified verbs like *be*, *go*, *get*, lexical *do* etc. as ‘grammatical verbs’, which could be also understood as feature structures without any roots inserted (Panagiotidis 2011; 2015, 98–100). Still, as Harley (2014) convincingly demonstrates, true root suppletion *does* exist. Besides well-known root allomorphs such as the French *oeil* / *yeux* (‘eye’, ‘eyes’) there is also extensive verb stem allomorphy in a number of languages (Harley 2014 and references therein). It then looks like that roots are not pure forms and this understanding has prompted Galani (2005, chaps. 5 & 6), Siddiqi (2006, chap. 3), and Haugen (2009) among others to argue for generalised Late Insertion of morphological and phonological features (i.e. of Vocabulary Items) to also include roots.

If Late Insertion applies to roots as well, then roots are not their forms and genuine root suppletion exists. This will be e.g. the case of two (or more) forms (Vocabulary Items) being matched with a single root; these forms will be allomorphs and their distribution will be governed by those principles governing allomorphy in general, such as the Elsewhere Principle:

(16) *Root allomorphy*

EYE, [Num: Plu]	—>	<i>yeux</i>
EYE	—>	<i>oeil</i>

The consequences of extending Late Insertion to roots are far reaching and should be at the heart of a theoretical treatment of roots: once more, we need to move away from the analytical reflexes inherited from 19th century Historical Linguistics, which conceives roots as quasi-stems or ‘small words’, and towards understanding them as syntactic formatives participating in the derivation.

As Acquaviva and Panagiotidis (2012, 113–15) discuss, once we distinguish roots from their morphological realisation, viz. the Vocabulary Items that serve as their exponence, we can work towards capturing how it is possible for morphological forms of roots to display *apparent* categorial specifications (e.g. noun only), as discussed in Don (2004), for example. Similarly, we can better frame generalisations about morphological constraints on root form, such as roots matching with particular declension classes, their being realised as three-consonant skeletons, and so on. Finally, if morphological forms matched with roots are not roots themselves, they may also spell out functional terminals and semi-lexical categories (De Belder 2011): a case to this point would be the form *will*, either spelling out the (archaic) verb and the noun derived from a root WILL, or spelling out just a [Future] feature.

4 *Roots as indices*

Still following the discussion in Acquaviva and Panagiotidis (2012), let us finally identify a root as “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” (Acquaviva and Panagiotidis 2012, 114). This is going to be revisited in this section and put to work in order to demonstrate how roots are and used for the purpose of marking lexical identity in Section 5.

If roots are ordinary grammatical elements, we expect them to be feature structures; these feature structures would be interpreted at the interface with the Conceptual-Intentional systems and would be matched with a Vocabulary item, a form, at a ‘Morphology level’ and be realised phonologically. By hypothesis the latter is true: Vocabulary Items matching with roots are (also) inserted late, at the ‘Morphology level’. But what kind of features would make up roots? Phonological features have been excluded on principled grounds. Would then roots bear features interpretable at the interface with the Conceptual-Intentional systems? Given the radical emptiness of roots this is unlikely, as per (10) and (13): there exist no features as instructions to the interfaces they could be composed of. We could of course conceive each root as a feature, this would however trivialise the very notion of feature, not to mention that features qua roots would be uninterpretable.

The above considerations open up the possibility that roots are *featureless* indices that as such act as the syntax-internal criterion of lexical identity. As Noam Chomsky (p.c. with Vitor Nóbrega) noted, positing featureless elements participating in syntactic operations “makes no sense”. At the same time, there exist no features roots could reasonably be decomposed into, unless we abandon features as instructions to the interfaces – an undesirable decision. A way to deal with this paradox is to think of roots as numeric indices, as addresses (Saab 2016), an idea that, to the best of my knowledge, originates in Acquaviva (2009): “the idea is that the address serves as the linkage between a set of instructions for phonological realization in context and a set of instructions for semantic interpretation in context” (Harley 2014, 226).

Harley (2014, 242) develops this idea further and posits that “The root terminal node elements [...] can thus be notated as $\sqrt{279}$, $\sqrt{322}$, $\sqrt{2588}$, etc.”. Perhaps we could take this numerical notation more seriously or even literally, thinking of roots as “the hijacking of the successor function by the Faculty of Language” (Richard Larson, p.c.). In other words, our abstract ability to count would meet the combinatorial system, the Language Faculty, with numerical indices handled by the system as roots.

Here is a sketch of how this would work: different roots, different indices, enable the same syntactic structure, say nP , as a whole to be associated with different concepts.

Recall that it is syntactic structures, not roots in isolation, that are associated with denotations, i.e. that are ultimately matched with concepts. Therefore, it is exactly the existence of different roots that enables the association of grammatical structures with concepts (also in Borer 2013). If roots were not available to it, the Language Faculty would be limited to expressing only what can be composed out of UG-features, such as “This is her”, “I got that”, “It is here” etc. (cf. Emonds 1985, chap. 4; Panagiotidis 2011, 376; De Belder 2011). By inserting different roots in otherwise identical structures, these structures are distinguishable at the interface with the Conceptual-Intentional systems and thus can be matched with different concepts; e.g. [_{NP} *n* √626] will be matched with ‘cat’ and [_{NP} *n* √4711] will be matched with ‘dog’. Hence, roots are the syntax-internal criteria of lexical identity (Acquaviva and Panagiotidis 2012, 11), see also the discussion of (23) below.

Concluding, we recast roots as “elements that are defined only as constituents of a formal syntactic representation, but have no grammar-external status” (Acquaviva and Panagiotidis 2012, 114).

5 *How do we identify roots? Beyond grammar*

We moved away from a conception of roots as ‘small words’, i.e. forms carrying a (thicker or thinner) shadow of denotation, a conception inherited from the tradition of Historical Linguistics. This move comes with the advantages we have already reviewed: roots themselves as featureless indices are exceptional, as expected also from an evolutionary perspective as well (Nóbrega 2018). Nevertheless, although they are exceptional in their featurelessness they are exactly like all other elements inside the combinatorial process in that the structures they participate in are matched with form and meaning *late*, at the interfaces.

The truly challenging question is then the following: if roots are abstract indices, they cannot be identified on the basis of ‘content’ (they have none) or on that of form – also in the face of root suppletion. How then do native speakers identify roots? Recall that the answer could be deceptively easy in languages like English (pace Pullum 2010), with its multitude of roots that may surface as free morphemes, and even easier in Semitic, with its triconsonantal root skeletons. However, as pointed out above things look more complex in synthetic languages like Greek. Consider again the stem *nom-* (Panagiotidis 2014a, 295), as presented in (14). Let us this time try dividing words containing this stem according to diachronic facts.

(17) *One root for ‘law’, one for ‘distribution’?*

I

nóm-os	nom-ik-ós	nóm-im-os	nom-o-thet-ó	pará-nom-os
‘law’	‘legal ₁ , juristic’	‘legal ₂ ’	‘legislate’	‘illegal’

II

nom-ós	nom-í	ypó-nom-os	asty-nom-ia	astro-nom-ia
‘prefecture’	‘distribution’, ‘grazing’	‘sewer’	‘police’	‘astronomy’

Historically, the words in the **I**-list above derive from a root yielding words about law and rules, whereas **II**-list words derive from a root yielding words about distribution and sharing. So, historically, in (14) and (17) we are dealing with root *homophony*. What about the modern language? How many NOM roots are there in Modern Greek? Two, as in Ancient Greek or just the one? What are the actual criteria for deciding between root homophony, i.e. two roots happening to have the same exponence (*nom-*), and root identity?

Criteria cannot be about meaning, even if one still believes in semantic content for roots in isolation: looking at list II in (17) what is the possible relation between ‘grazing’ and ‘prefecture’? Aren’t police and astronomy closer to law than they are to sewers?⁹ If we turn to form as our sole criterion we can of course claim that there is but a single Modern Greek root deriving all the words in (14) and (17), but does this entail that no such thing as root homophony exists? This would be an empirically indefensible claim: if anything, there used to be two NOM roots in the ancient language. In any case, our decision on how many roots are involved in (14) and (17) must be a principled one.

I believe that principles governing the identification of roots must be very basic and UG-external, while making reference to both grammatical structure and morphological exponence. After all *root identification is where grammatical acquisition meets word learning*: roots will inevitably have to be extracted from the lexical stock as it interacts with grammatical formatives.

5.1 Preamble: figuring out the root

As already mentioned, in languages like Greek root identification is not always a straightforward business, but this is hardly an exclusive problem of systematically synthetic languages, as Pullum (2010) points out. In synthetic languages however an even more fundamental issue might need be resolved, namely *how much* of a given form is part of the root, synchronically speaking: it turns out that “wringing out the morphology” (Aronoff 1994, 40) so that we are left with the root is not always a straightforward job. Observe the following Greek triplet:

(18) *What is the root?*

lepi	lepiða	leptos
?-NEUT.SG.NON-OBL	?-FEM.SG.NON-OBL	?-?.MASC.NOM
‘(fish) scale’	‘blade’	‘thin’

Those that are philologically minded will analyse *leptos* (‘thin’) as *lep-tos*, *-tos* being an adjectival ending. However, in Modern Greek *-tos* is only used to derive adjective-like forms from verbal stems (Αναστασιάδη-Συμεωνίδη 1994; Anagnostopoulou and Samioti 2014) and *lep-* is clearly *not* a verbal stem in the modern language. Hence, despite appearances, we have (at least) *two* roots in the triplet above: *lep-* (for *lepi* and

⁹ On top of this, native speakers tend to supply folk etymologies along the lines of ‘celestial laws’ when it comes to *astronomia* (‘astronomy’).

lepida) and *lept-* (for *leptos*), which is a quite productive root, too (Ralli 2005, 154–57).

Let us now go over the proposed principles according to which native speakers identify roots. These will be called *Form*, *Morphosyntax*, and *Suppletion*; they all abide by a ‘superprinciple’: *Time*.

5.2 Principle I: Form

The first principle is quite straightforward.

(19) *Form*

Same form → same root

This is intended as an elementary general learning principle: if two words contain the same root *form* then they are derived from the same root. Root identification is then based on form, as we cannot rely on meaning (Arad 2003; Borer 2009; Acquaviva and Panagiotidis 2012; Harley 2014). Two clarifications are in order:

First, according to the Form principle in (19) and given that there is no ‘meaning bias’ in root identification, *very* different words will be derived from the same root, as already anticipated in Borer (2005, chaps. 1–3):

(20) *Of kings and sunsets*

vasil-ias	vasil-ema	Vasil-is
‘king’	‘sunset’ ‘*reign’	‘Vassilis (a name)’

Second, speakers will be strongly biased to identify roots based on form when they encounter a novel word, e.g. in *vasil-o-manitar-o* (‘cep mushroom’).

5.3 Principle II: Morphosyntax

Despite its primacy, the form principle is not sufficient. Consider for instance the duo *kori* (‘daughter’) and *kori* (‘pupil’); these nouns form one of many homophonous pairs and triplets in Modern Greek (Panagiotidis 2014b, 419). Both nouns *kori* and *kori* are feminine and have identical exponents of the various ϕ -feature combinations throughout their respective paradigms; in other words, they are *indistinguishable in form* and they mark *identical ϕ -features*.

(21) ‘*Daughter of my eye*’? No.

kor-i	kor-i
KOR ₁ -FEM.SG.NOM	KOR ₂ -FEM.SG.NOM
‘daughter’	‘eye pupil’

If we blindly follow the Form principle in (19), then both nouns in (21) would be derived from the same root. This would entail cancelling the role of the root as the syntactic criterion of lexical identity: two identical *nP* structures built on identical roots and having exactly the same morphological exponence would be matched with different denotations, ‘daughter’ and ‘pupil’ respectively, at the interface with the Conceptual-Intentional systems. Hence, the roots in (21) might look the same but must be differentiated, making this a case of and a case for *root homophony*.

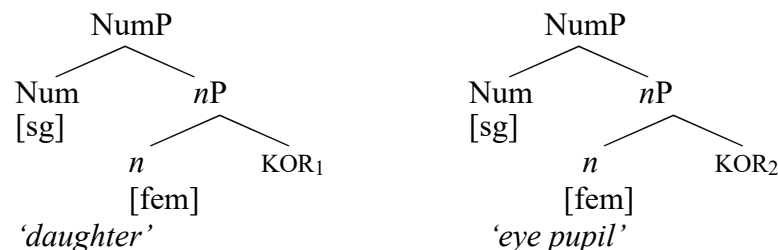
Recall that this is generally motivated by the fact that roots do not exist in uncategorised isolation but embedded within linguistic structures that are morphologically and phonologically realised. We are therefore forced to posit a second principle for the identification of roots:¹⁰

(22) *Morphosyntax*

Different denotation, same form, in the same morphosyntactic environment → different roots

Note that ‘morphosyntactic environment’ means both the morphosyntactic features (e.g. ϕ -features) and the structure, *and* their exponence. This is sketched in the diagram below:¹¹

(23) *Roots as indices of lexical identity*



Root homophony in a synthetic language like Greek is hardly an isolated curiosity. Below are more examples of root homophony from Modern Greek illustrating this second principle:

(24) *Root homophony*

kom-a kom-a kom-a neuter, singular, non-oblique case
‘party’ ‘comma’ ‘coma’

lir-a lir-a feminine, singular, non-oblique case
‘pound’ ‘lyre’

fil-o fil-o neuter, singular, non-oblique case
‘leaf’ ‘sex’

In the above cases, we must be dealing with root homophony because in each line the morphosyntactic environment (*both* ϕ -features *and* exponence) is identical.

Having said that, when exponence is *not* identical then, we are dealing with an identical root, as per the Form Principle in (19). This is true for all those *nom-* words in (14) and (17), with (14) repeated below for convenience:

¹⁰ An anonymous reviewer correctly points out that this principle “only works upon inspection of both form and meaning”. This is expected because it is a learning principle.

¹¹ The underlying assumption here is, of course, that the Saussurean sign is always structured (Borer 2013; Acquaviva and Panagiotidis 2012). See also Preminger (in progress) for a view that there are no Saussurean signs in structured linguistic utterances, only mappings from syntax to morphology and mappings from syntax to semantics; these mappings may sometimes coincide or only partially overlap.

(25) *Of sewers and laws*

<i>nóm-os</i>	‘law’
<i>nom-ós</i>	‘prefecture’
<i>nom-í</i>	‘distribution’, ‘grazing’
<i>ypo-nom-os</i>	‘sewer’
<i>nom-arx-is</i>	‘prefect’
<i>astr-o-nom-ia</i>	‘astronomy’
<i>nom-ik-os</i>	‘legal ₁ , juristic’
<i>nom-im-os</i>	‘legal ₂ , lawful’
<i>para-nom-os</i>	‘illegal, outlawed’
<i>nom-o-thet-o</i>	‘legislate’
<i>nom-iz-o</i>	‘think, believe’
<i>nom-iz-ma</i>	‘coin’, ‘currency’

The above words can be claimed to derive from a single root NOM because no two words derived from it have identical exponence. The first two words on the list are particularly illuminating: although *nóm-os* (‘law’) and *nom-ós* (‘prefecture’) have identical ϕ -features (masculine, singular, nominative), they are differentiated as *forms* by their different lexical stress: ‘law’ is stressed on the stem, ‘prefecture’ on the suffix (on Greek lexical stress, see Ralli 1988; 2005; Spyropoulos and Revithiadou 2009; van Oostendorp 2012). A similar state of affairs is the one in (26), where a single root KUR derives very different words, albeit words that are differentiated in form.

(26) *Root identity: verb stems*

<i>kur-ev-</i>	‘shear’, ‘give a haircut’
<i>kur-az-</i>	‘tire’
<i>kur-ar-</i>	‘treat as a patient’

In this case we have identical *v* heads but different exponents thereof: *-ev-*, *-az-* and *-ar-* respectively (Spyropoulos, Revithiadou, and Panagiotidis 2015; Panagiotidis, Revithiadou, and Spyropoulos 2017), so speakers can still use the same root KUR in different environments as a criterion of lexical identity.

5.4 Principle III: Suppletion

A third principle is the following:

(27) *Suppletion*

Complementary distribution → root suppletion

Harley (2014) covers the topic in sufficient detail, so the interested reader is referred there. This principle is again simple: if there is complementary distribution between forms belonging to the same word (say, *lexeme*), then speakers will assume root suppletion. Consequently, it is no accident that root suppletion is much more common with verbs (Veselinova 2006; Harley 2014), as verbs tend to offer more

morphological environments for the speaker to observe clear cases of complementary distribution.¹²

5.5 Superprinciple: Time

A final ‘superprinciple’ by which the above principles abide is time. Roots are identified on the basis of words that we learn in real time, not by running corpus searches.

(28) *Time*

All principles apply once and on the existent ‘lexicon’.

The ‘superprinciple’ in (28) answers the question in Section 2.3 on how many words one must learn before they can decide on root properties, raised in the context of Arad’s Multiple Contextual Meaning. Here the actual question is more straightforward: how many words must one know before they decide that a particular form corresponds to a root? Two interesting cases follow below:

(29) *Cows and Messiahs come first*

<i>ayel-i</i>	<i>ayelað-a</i>
‘pack, flock’	‘cow’

<i>Xrist-os</i>	<i>xris-ma</i>	<i>xri(z)-o</i>
‘Christ’	‘anointment, nomination’	‘anooint, nominate’

Both words for *cow* and *Christ* are potentially synchronically analysable as derived from the roots AYEL and XRI(Z) unlike the case of *lept-os* in (18). As explained in Panagiotidis (2019), *ayelaða* (‘cow’) is readily analysable as involving a root AYEL and a nominalising affix *-að-*, whereas *Xristos* (‘Christ’) would be readily analysable as a deverbal adjective or participle of the verb *xri(z)-o* (‘anooint’), itself containing a root XRI(Z)¹³. These morphological decomposition analyses would be supported, for what is worth, by the etymology of these words: diachronically this is how the Hellenistic forms *agelas* and *Christos* were derived in Koine Greek.

The reality of the native speakers learning words and acquiring the grammar of Modern Greek in their early childhood, blissfully ignorant of the diachronic facts, must be very different though; this is a reality that shapes how they identify the roots deriving the words for ‘cow’ and ‘Christ’. It would be very reasonable to argue that the word for cow (*ayelaða*) is learned long before the purportedly co-radical one for packs and flocks (*ayeli*). Not only is *ayeli* part of the learned vocabulary, it also collocates with wild dogs, wolves or jackals, not peacefully grazing domesticated bovines. Moreover, the word for ‘cow’ is very frequent and it would make sense to think that children would identify a root like AYELAD. What would happen when they learn the word for pack (*ayeli*)? Would they go back and reanalyse the word for cow as now involving a root AYEL?

Similar issues arise for the word for Christ – and some high frequency words derived from it, including that for Christmas: *Xristos* is encountered before learned terms

¹² Detecting complementary distribution can be done à la Yang (2016).

¹³ It would be the kind deverbal adjective or participle that *leptos* is not – see the discussion of (18).

about anointing and nominating, with such terms again being low frequency and of the learned stock. Upon encountering *xris-ma* (‘anointment, nomination’) and/or *xri(z)-o* (‘anoint, nominate’), would speakers backtrack and reanalyse the word for Christ as a deverbal adjective or a participle of a co-radical verb *xri(z)-o*?

If the superprinciple in (28) is valid, the answer is negative. Children encountering words for *cow* and *Christ* early in their lives as language learners will readily analyse them as derived from roots AYELAD and XRIST. Later encountering the similar words in (29) will lead to the identification of new roots AYEL and XRI(Z), without any retroactive effect on roots AYELAD and XRIST.¹⁴ At this point I would think that this happens only once, when a root as a criterion for lexical identity is fixed, but this is certainly an empirical matter open to experimental investigation.¹⁵

5.6 Conclusion: learning meets language acquisition again

Roots are extracted from the lexical stock via the interaction of simple learning principles that make reference to both feature structures and language-specific forms. These principles are applied on the existing lexical stock at any given time, possibly only once.

6 References

- Acquaviva, Paolo. 2009. “Roots and Lexicality in Distributed Morphology.” In *York-Essex Morphology Meeting 5: Special Issue of York Working Papers in Linguistics*, edited by Alexandra Galani, Daniel Redinger, and Norman Yeo, 2:1–21. York: University of York. <http://ling.auf.net/lingbuzz/000654>.
- . 2014. “The Roots of Nominality, the Nominality of Roots.” In *The Syntax of Roots and the Roots of Syntax*, edited by Artemis Alexiadou, Hagit Borer, and Florian Schäfer, 33–56. Oxford Studies in Theoretical Linguistics. Oxford ; New York: Oxford University Press.
- Acquaviva, Paolo, and Phoevos Panagiotidis. 2012. “Lexical Decomposition Meets Conceptual Atomism.” *Lingue e Linguaggio* XI (2): 105–20.
- Alexiadou, Artemis, and Terje Lohndal. 2017. “On the Division of Labor between Roots and Functional Structure.” In *The Verbal Domain*, edited by Roberta D’Alessandro, Irene Franco, and Ángel Gallego, 85–102. Oxford Studies in Theoretical Linguistics. Oxford ; New York: Oxford University Press.

¹⁴ An anonymous reviewer points out that this state of affairs must be true for Italian *vacca* (‘cow’) and *vaccino* (‘of cow origin’, ‘vaccine’).

¹⁵ Another question that must be left for future investigation, of the experimental sort, is what happens in cases like the one below: According to the Morphosyntax principle in (22), there should exist two roots FIL in Modern Greek because of the identical neuter nouns *filo* (‘sex’) and *filo* (‘leaf’). According to the Form principle in (19) other words with FIL would be coradical with either ‘sex’ or ‘leaf’. Which one? I have no answer here.

FIL ₁ -	fil-os	MASC, SG, NON-OBL.	<i>friend</i>
	fil-i	FEM, SG, NON-OBL.	<i>tribe</i>
	fil-i	NEUT, SG, NON-OBL.	<i>kiss</i>
	fil-o	NEUT, SG, NON-OBL.	<i>leaf or sex?</i>
FIL ₂ -	fil-o	NEUT, SG, NON-OBL.	<i>leaf or sex?</i>

- Anagnostopoulou, Elena, and Yota Samioti. 2014. "Domains within Words and Their Meanings: A Case Study." In *The Syntax of Roots and the Roots of Syntax*, edited by Artemis Alexiadou, Hagit Borer, and Florian Schäfer, 81–111. Oxford ; New York: Oxford University Press.
- Arad, Maya. 2003. "Locality Constraints on the Interpretation of Roots: The Case of Hebrew Denominal Verbs." *Natural Language and Linguistic Theory* 21: 737–78.
- . 2005. *Roots and Patterns Hebrew Morpho-Syntax*. Dordrecht: Springer Verlag.
- Aronoff, Mark. 1994. *Morphology by Itself: Stems and Inflectional Classes*. Linguistic Inquiry Monographs 22. Cambridge, Mass: MIT Press.
- . 2007. "In the Beginning Was the Word." *Language* 83: 803–30.
- Beard, Robert. 1995. *Lexeme-Morpheme Base Morphology: A General Theory of Inflection and Word Formation*. SUNY Series in Linguistics. Albany: State University of New York.
- Borer, Hagit. 2005. *In Name Only*. Structuring Sense, volume I. Oxford; New York: Oxford University Press.
- . 2009. "Roots and Categories." Talk presented at the 19th Colloquium on Generative Grammar, University of the Basque Country, Vitoria-Gasteiz, Spain, April.
- . 2013. *Taking Form*. First edition. Structuring Sense, volume III. Oxford: Oxford University Press.
- . 2014. "The Category of Roots." In *The Syntax of Roots and the Roots of Syntax*, edited by Artemis Alexiadou, Hagit Borer, and Florian Schäfer, 112–48. Oxford Studies in Theoretical Linguistics. Oxford ; New York: Oxford University Press.
- Bruening, Benjamin. 2018. "The Lexicalist Hypothesis: Both Wrong and Superfluous." *Language* 94 (1): 1–42. <https://doi.org/10.1353/lan.2018.0000>.
- Clark, Eve V., and Herbert H. Clark. 1977. "When Nouns Surface as Verbs." *Language* 55: 767–811.
- De Belder, Marijke. 2011. *Roots and Affixes: Eliminating Lexical Categories from Syntax*. Utrecht: LOT.
- Don, Jan. 2004. "Categories in the Lexicon." *Linguistics*, 931–56.
- Embick, David, and Alec Marantz. 2008. "Architecture and Blocking. *Linguistic Inquiry*." *Linguistic Inquiry* 39: 1–53.
- Emonds, Joseph E. 1985. *A Unified Theory of Syntactic Categories*. Studies in Generative Grammar 19. Dordrecht, Holland ; Cinnaminson, U.S.A: Foris Publications.
- . 2000. *Lexicon and Grammar: The English Syntacticon*. Studies in Generative Grammar 50. Berlin ; New York: Mouton de Gruyter.
- Fodor, Jerry A. 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Clarendon.
- . 2008. *LOT 2: The Language of Thought Revisited*. Oxford: Clarendon.
- Galani, Alexandra. 2005. "The Morphosyntax of Verbs in Modern Greek." PhD thesis, University of York.
- Hale, Kenneth L., and Samuel Jay Keyser. 1993. "On Argument Structure and the Lexical Expression of Syntactic Relations." In *The View from Building 20. Essays in Honor of Sylvain Bromberger*, edited by Kenneth L. Hale and Samuel Jay Keyser, 53–109. Cambridge, Mass.: MIT Press.
- . 2002. *Prolegomenon to a Theory of Argument Structure*. Linguistic Inquiry Monographs 39. Cambridge, Mass: MIT Press.
- Halle, Morris. 1997. "Distributed Morphology: Impoverishment and Fission." *MIT Working Papers in Linguistics* 30: 425–49.
- Halle, Morris, and Alec Marantz. 1993. "Distributed Morphology and the Pieces of Inflection." In *The View from Building 20. Essays in Honor of Sylvain Bromberger*, edited by Kenneth L. Hale and Samuel Jay Keyser, 111–76. Cambridge, Mass.: MIT Press.

- Harley, Heidi. 2005. "How Do Verbs Get Their Names? Denominal Verbs, Manner Incorporation and the Ontology of Verb Roots in English." In *The Syntax of Aspect: Deriving Thematic and Aspectual Interpretation*, edited by Nomi Erteschik-Shir and Tova Rapoport, 42–64. Oxford ; New York: Oxford University Press.
- . 2014. "On the Identity of Roots." *Theoretical Linguistics* 40 (3/4): 225–76.
- Harley, Heidi, and Rolf Noyer. 1999. "State-of-the-Article: Distributed Morphology." *GLOT International* 4 (4): 3–9.
- Haugen, Jason. 2009. "Hyponymous Objects and Late Insertion." *Lingua* 119: 242–62.
- Levin, Beth, and Malka Rappaport Hovav. 2005. *Argument Realization*. Research Surveys in Linguistics. Cambridge [England] ; New York: Cambridge University Press.
- Levinson, Lisa. 2007. "The Roots of Verbs." PhD thesis, NYU.
- Marantz, Alec. 1997. "No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon." *U. Penn Working Papers in Linguistics* 4: 201–25.
- . 2006. "Phases and Words." Unpublished ms. MIT.
https://files.nyu.edu/ma988/public/Phase_in_Words_Final.pdf.
- . 2012. "Locality Domains for Contextual Allophony in Words." Unpublished ms. New York University.
- Mitrović, Moreno, and Phoevos Panagiotidis. 2018. "The Categorical Anatomy of Adjectives." Submitted ms. Nicosia. <http://ling.auf.net/lingbuzz/003887>.
- . 2020. "Adjectives Exist, Adjectivisers Do Not: A Bicategorical Typology." *Glossa*.
- Nóbrega, Vitor. 2018. "Isomorphic Approach: Articulating the Lexicon and Syntax in the Emergence of Language." Unpublished PhD thesis, São Paulo: São Paulo.
- Oostendorp, Marc van. 2012. "Stress as a Proclitic in Modern Greek." *Lingua* 122 (11): 1165–81. <https://doi.org/10.1016/j.lingua.2012.05.006>.
- Panagiotidis, Phoevos. 2011. "Categorical Features and Categorizers." *The Linguistic Review* 28: 325–46.
- . 2014a. "A Minimalist Approach to Roots." In *Minimalism and Beyond: Radicalizing the Interfaces*, edited by Peter Kosta, Steven Franks, Lilia Schürcks, and Teodora Radeva-Bork, 287–303. Amsterdam; Philadelphia: John Benjamins Pub. Co.
- . 2014b. "Indices, Domains and Homophonous Forms." *Theoretical Linguistics* 40 (3/4): 415–27.
- . 2015. *Categorical Features: A Generative Theory of Word Class Categories*. Cambridge Studies in Linguistics 145. Cambridge; New York: Cambridge University Press.
- . 2019. "Ρίζες: Μεταξύ Λέξεων Και Γραμματικής." In *Γλωσσική Ποικιλία: Μελέτες Αφιερωμένες Στην Αγγλική Πάλη*, edited by Argyris Archakis, Nikos Koutsoukos, George Xydopoulos, and Dimitris Papazachariou, 437–51. Athens: Kapa Ekdotiki.
- Panagiotidis, Phoevos, Anthi Revithiadou, and Vassilios Spyropoulos. 2017. "Little v as a Categorizing Verbal Head: Evidence from Greek." In *The Verbal Domain*, edited by Roberta D'Alessandro, Ivana Franco, and Ángel Gallego, 29–48. Oxford Studies in Theoretical Linguistics. Oxford ; New York: Oxford University Press.
- Preminger, Omer. in progress. "Non-Semiotic Grammar: A Symmetric, Distributed Model for Syntax and Its Interfaces." Unpublished ms. University of Maryland.
- Pullum, Geoffrey. 2010. "Isms, Gasms, Etc." *The Language Log* (blog). February 8, 2010. <http://languagelog.ldc.upenn.edu/nll/?p=2108>.
- Ralli, Angela. 1988. "Éléments de La Morphologie Du Grec Moderne: La Structure Du Verbe." Montreal: Université du Montréal.
- . 2003. "Morphology in Greek Linguistics: The State of the Art." *Journal of Greek Linguistics* 4: 77–129.
- . 2005. *Μορφολογία [Morphology]*. Athens: Patakis.

- Rappaport Hovav, Malka, and Beth Levin. 1998. "Building Verb Meanings." In *The Projection of Arguments: Lexical and Compositional Factors*, edited by Miriam Butt and Wilhelm Geuder, 97–134. Stanford, Calif.: CSLI Publications.
- Saab, Andrés. 2016. "No Name: The Alloosemy View." Unpublished ms. National Scientific and Technical Research Council, Buenos Aires, Argentina.
<http://ling.auf.net/lingbuzz/003154>.
- Siddiqi, Daniel. 2006. "Minimize Exponence: Economy Effects on a Model of the Morphosyntactic Component of the Grammar." PhD thesis, University of Arizona.
- Spyropoulos, Vassilios, and Anthi Revithiadou. 2009. "The Morphology of Past in Greek." *Studies in Greek Linguistics* 29: 108–22.
- Spyropoulos, Vassilios, Anthi Revithiadou, and Phoevos Panagiotidis. 2015. "Verbalizers Leave Marks: Evidence from Greek." *Morphology* 25 (3): 299–325.
<https://doi.org/10.1007/s11525-015-9260-5>.
- Veselinova, Ljuba N. 2006. *Suppletion in Verb Paradigms: Bits and Pieces of the Puzzle*. Typological Studies in Language, v. 67. Amsterdam ; Philadelphia: J. Benjamins.
- Yang, Charles D. 2016. *The Price of Linguistic Productivity: How Children Learn to Break the Rules of Language*. Cambridge, MA: The MIT Press.
- Αναστασιάδη-Συμεωνίδη, Άννα. 1994. "Το Τεμάχιο -Τος Στα Ρηματικά Επίθετα Της Νεοελληνικής." *Studies in Greek Linguistics* 15: 473–84.