Special Number

Selected papers from the Word-Formation Theories III / Typology and Universals in Word-Formation IV conference

Table of Contents

Bernd Heine 2

On the grammaticalization of some processes of word formation in Africa

Susan Olsen 19

Interconnectedness and variation of meaning in derivational patterns

Makiko Mukai 35

Productivity of recursive compounds

Elizaveta Tarasova 49

Productivity of form and productivity of meaning in N+N compounds

Hans Götzsche 70

Compounds as idiom. A case study of a 'meta-trend'

Pius ten Hacken 77

Relational adjectives between syntax and morphology

Lior Laks 93

Competing suffixes: feminine formation of Hebrew loanwords

Ora (Rodrigue) Schwarzwald 109

Linear and nonlinear word formation in Hebrew – words which end with -on

Ken Sasahara 121

Derivational suffixes -ar and -nik in Upper Sorbian

Michał Rzepiela 130

Word-formation type, its reinterpretations, and possible equivalents

Petr Kos 146

Metaphor and metonymy as a means of economy of expression

Richard Skultéty Madsen 162

Word-formation preferences of non-natives

On the grammaticalization of some processes of word formation in Africa

Bernd Heine, University of Köln, Germany

The paper is concerned with linguistic data suggesting that one and the same lexical source of grammaticalization can give rise to different morphological processes, leading not only to compounding and lexicalization but also to derivation, and even to inflection. Based on data from African languages for which little or no earlier written documents are available, the paper argues that even in the absence of historical records it is possible to reconstruct some features of earlier processes of word formation.

Keywords: *compounding*, *derivation*, *diminutive*, *grammaticalization*, *inflection*

1. Introduction

It is probably a general property of languages that they show multifunctionality, that is, they dispose of morphological units belonging simultaneously to more than one grammatical category. For some linguistic schools of thought this fact poses a problem since it is at variance with a presumed one-form-one-meaning principle, according to which monosemy is more natural than polysemy or, more generally, multifunctionality. For other schools again it is more of the expected case and in need of explanation. The latter applies in particular but not only to students of transcategoriality,[1] who account for multifunctionality of linguistic units either in terms of synchronic principles of discourse processing or else in terms of diachrony (Robert 2003a; 2003b; 2004; Do-Hurinville & Hancil 2015; see also Enfield 2006).

The present paper is concerned with a case of multifunctionality as it relates to word formation processes. To this end, an example from an African language is discussed in more detail. The remainder of this section is concerned with methodological issues. The subsequent sections propose an account of multifunctionality in terms of grammaticalization, where §2 focuses on derivation and §3 on inflection. While these two sections are largely restricted to data from the !Xun language of southwestern Africa, the final §4 then draws some conclusions from the analysis presented in the paper.

1.1 *An example*

In the Khoisan language !Xun of southwestern Africa there is a morphological unit which provides a paradigm case of a multifunctional category. As the examples in (1) show, this unit, *m̀hè*, is associated simultaneously with several different morpheme types.

[1] The term *transcategoriality* refers to a structure where in a given language one and the same kind of linguistic expression is used simultaneously on two or more different planes of linguistic organization based on some regular pattern; cf. English *well*, which serves, on the one hand, as a manner adverb and, on the other hand, as a discourse marker (Robert 2004).

(1) The status of the morpheme *m̩̀hè* of !Xun (W2 dialect, Kx'a family; König & Heine 2001; 2008; Heine & König 2015)

	Expression	Meaning	Morphological status
a.	mí **m̩̀hè** (my child:PL)	'my (own) children'	Noun
b.	!xō **m̩̀hè** (elephant child:PL)	'elephant calves'	Head noun in modifying N-N compounds
c.	n!āō- **m̩̀hè** (house-DIM:PL)	'small houses'	Derivational diminutive suffix
d.	xā- **m̩̀hè** (old.man-PL)	'old men'	Inflectional plural suffix

In (1a), *m̩̀hè* is a plural noun, a relational noun meaning '(one's own) children', where the corresponding singular form is *mà* '(one's own) child'.[2] In (1b), *m̩̀hè* forms the head of a modifying compound construction, productively taking animate nouns as modifiers and meaning 'children of X' or 'young X'. With inanimate nouns as modifiers, *m̩̀hè* expresses 'small quantity or quality', that is, it has the structure and function of a productive diminutive plural suffix. Finally, in (1d), *m̩̀hè* functions as a plural form of the noun *xāmà* 'old man', having features of a plural inflection (König & Heine 2001; 2008; Heine & König 2015).

In sum, one and the same element, *m̩̀hè*, occurs in four different constructions, instantiating four kinds of morphological expressions, extending from lexical to derivational and inflectional uses. This raises the question of how the presence of such a set of multifunctionality can be explained. Judging from what has been written on multifunctionality of this kind, the hypotheses in (2) are perhaps the ones that come into one's mind.

(2) Lines of explanation

a. There is no reasonable explanation and, hence, no need to search for an explanatory account.
b. Since speakers of the language consistently distinguish the four different uses of the set, it must serve some purpose; hence, it should be explained with reference to the motivations that speakers have when using the set.
c. The presence of the set is a result of diachronic processes and, hence, can be explained with reference to these processes.

These hypotheses have been looked at in some form or other in the literature on transcategoriality (Robert 2003a; 2003b; 2004; Do-Hurinville & Hancil 2015). To my knowledge, the null hypothesis in (2a) has never been seriously proposed and will not be pursued here any further. An account in terms of (2b) is proposed by Robert (2003a, 2003b). Building on the notion of schematic form, Robert suggests explaining transcategoriality (*transcatégorialité*) in terms of economic motivation, which she views as a means for the optimization of linguistic systems since it allows having a maximum of functions with a minimum of forms. In a related fashion, Enfield (2006: 297) invokes *rule economy* as a way

[2] This noun differs from the non-relational !Xun noun *dàbà*, pl. *dèbē* 'child' (König & Heine 2001: 149).

of dealing with certain kinds of heterosemy, that is, of multifunctionality having a diachronic base.[3]

Economic motivation is an attractive notion, but so far the empirical evidence presented in its support is limited. The present paper will be restricted to (2c), and more specifically to grammaticalization to account for transcategoriality of the kind illustrated in (1).[4]

1.2 *Grammaticalization*

There is a wide range of definitions of grammaticalization (e.g. Campbell 2001; Heine & Kuteva 2002: 2; Hoper & Traugott 2003: 18; Kuteva et al. 2018) but in their major concern, most of them are essentially in accordance with the following classic definition by Kuryłowicz (1975 [1965]): "Grammaticalization consists in the increase of the range of a morpheme advancing from a lexical to a grammatical or from a less grammatical to a more grammatical status, e.g. from a derivative formant to an inflectional one" (Kuryłowicz 1975 [1965]: 52)

In order to identify processes of grammaticalization, a wide range of criteria have been proposed (see e.g. Lehmann 1982; Hopper 1991; Heine & Kuteva 2002; Hopper & Traugott 2003). In the present framework it is the four parameters listed in (3), which, as we argue, take care of most of the relevant criteria that have been proposed in other frameworks. Henceforth, these parameters are used as a tool for identifying instances of grammaticalization.

(3) Parameters of grammaticalization (Heine & Kuteva 2007: 33–46)
   a. Extension: linguistic expressions are extended to new contexts that invite the rise of grammatical functions (context-induced reinterpretation),
   b. desemanticization (or semantic bleaching): loss (or generalization) of meaning content,
   c. decategorialization: loss of morphosyntactic properties characteristic of lexical or other less grammaticalized forms, and
   d. erosion (or phonetic reduction): loss of phonetic substance.

Each of these parameters concerns a different aspect of language structure or language use; (3a) is pragmatic in nature, (3b) relates to semantics, (3c) to morphosyntax, and (3d) to phonetics. Except for (3a), these parameters all involve loss of properties. But the process cannot be reduced to one of structural "degeneration". There are also gains: in the same way as linguistic items undergoing grammaticalization lose in semantic, morphosyntactic and phonetic substance, they also gain in properties characteristic of their uses in new contexts – to the extent that in some cases their meaning and syntactic functions may show little resemblance to their original use.

---

[3] The term *heterosemy* was first proposed by Persson (1988) and subsequently modified and popularized by Lichtenberk (1991), for whom heterosemy obtains when within a single language "two or more meanings or functions that are historically related, in the sense of deriving from the same ultimate source, are borne by reflexes of the common source element that belong in different morphosyntactic categories" (Lichtenberk 1991: 476).

[4] This procedure is in accordance with Robert (2004: 119–20) when she observes that "grammaticalization is the diachronic aspect of the more general phenomenon of transcategoriality".

The ordering of these parameters reflects the diachronic sequence in which they typically apply: grammaticalization tends to start out with extension, which triggers desemanticization, and subsequently decategorialization, and finally erosion. Erosion is the last parameter to come in when grammaticalization takes place, and in many of the examples to be presented below it is not (or not yet) involved. Paradigm instances of grammaticalization involve all four parameters but, as we will see below, there are also cases where not all of the parameters play a role.

The theoretical status of the distinction between lexical and grammatical forms is questioned in some studies of Construction Grammar, and Hüning & Booij claim that "[I]t is especially the dichotomy of "the lexical" vs. "the grammatical" that turns out to be inadequate for a proper account of word formation phenomena [...]" (Hüning & Booij 2014: 599–600). That this claim must be taken with care is demonstrated in Heine et al. (2016). The transition from lexical to grammatical is not only transitional but also overlapping. Nevertheless, as a rule, there is cross-linguistically no problem distinguishing lexical from prototypical grammatical forms.

1.3 *Reconstructing in the absence of written documents*

Work on grammaticalization is based on historical reconstruction, and the safest way to achieve reconstruction is by drawing on historical documents that provide information on earlier states of language use. However, restricting the study of grammaticalization to written languages would mean that the vast majority of the world's languages would have to be excluded. I therefore adopt also an alternative, but well-established methodology of reconstruction that has been employed mostly for unwritten but also for written languages to reconstruct grammaticalization. This methodology relies mainly on three components, namely diachronic reconstruction, e.g. by means of the comparative method (CM), internal reconstruction (IR), and typological generalizations (TG).

The following example may illustrate this methodology (see also Heine 2003: 580). The Bantu language Swahili of eastern Africa has a future tense prefix *-ta-*, which is hypothesized to be historically derived from the volition verb *-taka* 'want' on the basis of the following evidence. By using the CM it is possible to establish that the verb must be older than the future tense marker: The application of the comparative method shows that the verb *-taka* is a modern reflex of the Proto-Bantu verb **-càk-a* 'desire' while it is not possible to reconstruct the future tense marker back to Proto-Bantu (Guthrie 1967–71). Internal reconstruction (IC) suggests, for example, that the earlier form of the tense marker is likely to have been *-taka-* since the form *-taka-* is still retained in relative clauses.[5] TG allow for two kinds of generalization: First, it establishes that verbs of volition ('want', 'desire') quite commonly give rise to future tense markers in the languages of the world, the English *will*-future being a case in point (see Kuteva et al. 2018, WANT > FUTURE). And second, processes of this kind tend to involve a specific type of semantic, morphosyntactic and phonological changes: Loss of lexical in favor of grammatical meaning (desemanticization), loss of morphosyntactic properties such as loss of word status (decategorialization), and loss of phonetic substance (erosion).

[5] Note that subordinate clauses tend to be more conservative in grammatical change than main clauses; for example, the English item *will* developed into a future auxiliary but retained its lexical meaning in subordinate clauses as, e.g. in *Do as you will!*

On the basis of these methodological tools it is possible to formulate a strong hypothesis to the effect that the Swahili future tense marker *-ta-* is the result of a common grammaticalization process, having lost its lexical meaning of volition (desemanticization), its status as an independent verb (decategorialization), and part of its phonetic substance, being reduced from *-taka* to *-ta-* (erosion). To conclude, while it is always desirable to search for historical records, such records are not a requirement for the reconstruction of grammaticalization processes.

1.4 *Compounding, derivation, and inflection*

Since the paper is concerned with word formation, a note on the key concepts that will figure below seems in order. Compounding and derivation are commonly classified as word formation, that is, as the creation of new lexemes (e.g. Lieber & Štekauer 2014: 3). The former is defined by Bauer (2003: 40) as "the formation of a new lexeme by adjoining two or more lexemes", cf. English *football*. Compounding may take on a number of quite divergent forms and, accordingly, has been used for a range of different kinds of meaning (e.g. Bauer 1978; Bisetto & Scalise 2005; Wälchli 2005; Lieber & Štekauer 2009).

My interest here is exclusively with right-headed modifying or endocentric compounds having two constituents, C_1 and C_2, where C_1 is a modifier and C_2 the head. Derivation is more difficult to define, I am not aware of any concise definition that is likely to be acceptable to the majority of the linguistic community (cf. the contributions in Lieber & Štekauer 2014); I therefore follow Booij (2010: 454) in using a negative definition, namely: "The common denominator for all word formation processes except compounding is derivation".[6]

Derivation shares with compounding that it belongs to word formation, and with inflection that it typically, though not necessarily, involves affixation. But in the same way as the distinction between compounding and derivation, that between derivation and inflection is complex, having been portrayed as being either problematic, essentially undefinable, or even as non-existent (see the discussions in Bybee 1985; Anderson 1992: 72ff.; Carstairs-McCarthy 1992). This issue is immediately relevant to the subject matter of the present paper, but we will not be able to deal with it in as much detail as might be desirable.

2. From compounding to derivation

2.1 *From 'child' to diminutive suffix in !Xun*

Given the unsatisfactory state of defining compounding, derivation, and inflection one can characterize the subject matter of this paper in more general terms as one that is concerned with a transition from one morphological type to another, as sketched in (4).

[6] Hence, one is tempted to follow other authors in characterizing derivation with reference to clear instances of it. Thus, English suffixes such as *-ness, -hood* and *-ation* are instances of derivational markers while units such as *foot-* in *football* or the plural marker *-s* or the past tense marker *-ed* are not.

(4) Compounding > derivation > inflection
(Bybee 1985: 82; Heine et al. 1991: 17–8; Brinton & Traugott 2005: 85–7)

The first part of the chain of grammaticalization sketched in (4) (compounding > derivation) is well established (see Heine et al. 2016) while the second part (derivation > inflection) is far from uncontroversial, since a number of counterexamples have been identified (e.g. Norde 2009). As we will see in §3, the present paper nevertheless is in support of this pathway, even if it does not seem to be a canonical process of diachronic change.

This section is restricted to the example mentioned in (1) of §1.1, that is, my concern is with the !Xun language. !Xun, also called Ju, is a traditional hunter-gatherer language spoken by approximately 15000 people in Angola, Namibia, and Botswana. The language, classified by Greenberg (1963) as forming the Northern branch of the Khoisan family, has recently been re-classified as forming one of the two branches of the Kx'a family (Heine & Honken 2010). The language is unusually context-dependent, showing fairly substantial analytic-isolating morphology; there is only a small pool of items having exclusively grammatical functions (Heine & König 2005). Typological characteristics include contiguous serial verb constructions and a noun class system having four genders, distinguished in pronominal agreement but not on the noun. The basic word order is SVO, although there is a minor SOV order, and a modifier-head construction in nominal possession. It is a tone language distinguishing four tone levels (*á* = high, *ā* = mid, *à* = low, *ȁ* = extra-low), and with its five different click types (/ = dental, *!* = alveolar, *!!* = retroflex, ǂ = palatal, and ǁ = lateral) and well over 100 phonemes it belongs to the phonologically most complex languages in the world (Heine & König 2015). Our interest here is, unless indicated otherwise, only with the W2 dialect, which belongs to the northwestern branch of !Xun.[7]

Nouns are essentially transnumeral in !Xun, that is, they are unspecified for number. This applies to the vast majority of nouns, but not to all of them. The predominant means of obligatorily marking number distinctions are listed in (5).

(5) Obligatory number marking of nouns in !Xun (W2-dialect)

	Kind of marking	**Example**		**Type frequency**
a.	No marking	*g!áùn*, pl. *g!áùn*	'tree'	Predominant
b.	Suppletion	*n!hùnwà,* pl. *cūwā*	'footprint';	Few nouns
		mà, pl. *m̏hē*	'child (of), offspring'	
c.	Plural suffix *-m̏hè*	*xāmà,* pl. *xā-m̏hè*	'old man'[8]	Few nouns

There are, however, two plural enclitics that can be added to nouns. These enclitics, which are *hŋ́* with kinship terms and *hŋ̏* elsewhere, are used optionally to emphasize plurality. They can be added when the noun is already marked for plural, as in (6b). Since the concern of this paper is with affixal morphology these enclitics are henceforth ignored.

[7] According to Heine & König (2015), !Xun has eleven dialects. Subsequently, Heine & König (2016) found yet another dialect, the total number now being twelve.

[8] It might also be possible to analyze *xā-m̏hè* 'old man' in (7c) as a suppletive form, as noted by an anonymous reviewer. But since *-m̏hè* shows some productivity (see (11)), there is reason to classify it as a suffix.

(6) !Xun (W2 dialect; König & Heine 2001: 63)

a. *mí dàbà* 'my child (not my own)'
my child

b. *djù dèbē (hŋ̀)* 'our children (not our own)'
our child:PL (PL)

The suppletive noun *dàbà*, pl. *dèbē* is not the only noun for 'child'; there is a second suppletive noun, namely *mà*, pl. *m̀̀hē* which, unlike the former, is an inalienable noun denoting 'child (of), offspring' (cf. (7a)). The latter is used productively as a head noun in modifying compounds where the modifier is an animate noun (X) and the meaning is 'young or small X', as exemplified in (7b). But the use of *mà/m̀̀hè* was extended further to inanimate nouns, and in this case, and in this case it has given rise to a fully productive diminutive suffix denoting 'a small X', as the examples in (7c) show.

(7) !Xun (W2 dialect; König & Heine 2001: 61)

	Head noun	**Singular**	**Plural**	**Meaning**	**Literally**
a.	Lexical	*mà*	*m̀̀hè*	'child (of), offspring'	
b.	Animal ('child of')	*!xō-mà*	*!xō-m̀̀hè*	'young or small elephant'	'elephant-child'
c.	Inanimate	*tc'āō-mà*	*tc'āō-m̀̀hè*	'small tooth'	'tooth-child'

We observed above that with few exceptions, nouns in !Xun do not distinguish number. Obviously, this does not apply to suppletive nouns and nouns taking the diminutive suffix *-mà/-m̀̀hè*, which are obligatorily marked for number by means of this derivational suffix, as can be seen in (7c).

The case just discussed is not an isolated phenonemon in !Xun. As can be seen in (8), there is a small range of other cases that appear to have followed the same pathway from free noun via head noun of a modifying compound to derivational suffix.[9]

(8) Lexical sources of !Xun derivational suffixes (König & Heine 2001)

	Nominal source		**Derivation**	**Function**
a.	*mà,* pl. *m̀̀hè*	'child (of)'	*-mà,* pl. *-m̀̀hè*	'small', diminutive
b.	*g‖ò̀q,* pl. *n‖àē*	'man'	*-g‖ò̀q,* pl. *-n‖àē*	'male'
c.	*dē*	'mother'	*-dē*	'female'
d.	*kx'àò*	''owner'	*-kx'àò*	'agent of an action'

[9] All four derivational suffixes appear to have evolved from endocentric noun-noun compounds, with one exception: The *-kx'àò* suffix must have originated from verb-noun compounds, that is, the modifying conjunct is always a verb, not a noun, e.g. *tc'à* 'to steal' > *tc'à-kx'àò* 'thief'.

2.2 *The mechanism of change*

A development from a noun meaning 'child (of)' to diminutive affix like the one sketched in !Xun in §2.1 is not unheard of in other languages. In fact, according to a quantitative survey of 99 African languages, nouns for *child* are used productively as derivational forms in half of the languages (50.5 %) to express diminutive meaning (Heine & Leyew 2008; Heine & Kuteva 2009). That this process is cross-linguistically fairly common has been demonstrated in a number of studies (Heine & Hünnemeyer 1988; Heine et al. 1991: 79–97; Jurafsky 1996; Heine & Kuteva 2009).

Two examples from genetically unrelated African languages, presented in (9), may suffice to illustrate the pattern of change involved, where one is from Ewe, a West African language spoken in Ghana and Togo, and the other from Ik, an East African language spoken in northeastern Uganda.

(9) Lexical sources of derivational suffixes in Ik and Ewe (Heine et al. 2016: 153)

	Language and reference	**Lexeme**	**Meaning**	**Diminutive suffix**	**Example**
a.	Ik (Kuliak, Nilo-Saharan; Schrock 2014: 180)[10]	*ím (imá-)*	'child'	*-ima-*	*kɔfɔ́-ima-* (gourd-child) 'small gourd'
b.	Ewe (Kwa, Niger-Congo; Heine et al. 1991: 79)	*vi´*	'child (of)'	*-ví*	*kpé-ví* (stone-child) 'small stone'

In both languages of (9) the general process leading from free noun to derivational suffix appears to have been the same as in !Xun, and this process was in accordance with the parameters listed in (3). First, all three languages use a pattern of endocentric noun-noun compounding where first noun (C_1) constitutes the modifier and the second noun (C_2), which is the word for 'child', the head of the construction. Second, in all three languages the construction was extended to contexts where C_1 was an inanimate noun (extension, (3a)). In this context, the lexical meaning of 'child' made little sense and was lost; what was highlighted is an inference typically associated with the concept of 'child', namely smallness (desemanticization, (3b)), and this inference was conventionalized in the resulting derivational suffix. Third, having lost its lexical meaning, C_2 also lost most of the morphosyntactic features defining it as a noun (decategorialization, (3c)): It turned into an affix of C_1, its erstwhile modifier, and as an affix it is no longer able to take modifiers or to be placed in positions other than the one immediately after C_1. And fourth, it also tended to lose phonetic features. For example, being an affix of C_1, it lost the ability to occur as a distinct prosodic unit, as its lexical counterpart does: As an affix it is part of the intonation contour of the noun. These changes are summarized in Table 1.

[10] Inverted commas signal that the genetic classification of the family concerned is controversial.

Table 1. Changes commonly observed in the grammaticalization from compounding to derivation (C_1 = modifier, C_2 = head of a lexical compound; Heine et al. 2016: 153)

Parameter	**Changes likely to occur**
Context extension	The use of C_2 (the head) is extended to a range of different C_1 constituents
Desemanticization	As a result, C_2 loses part or all of its lexical meaning, gradually acquiring a generalized and a grammatical meaning
Decategorialization	C_2 loses morphosyntactic properties characteristic of its lexical category of nouns, gradually turning into an affix
Erosion	C_2 tends to lose phonological properties

Note that for none of the languages there are historical records to assist reconstruction. Hence, reconstruction of pathways of grammaticalization in such languages is based on diachronic reconstruction as sketched in §1.3. We may illustrate the approach used with the Ewe example in (9). The Ewe noun *vi* 'child' can be reconstructed back to an earlier Proto-Niger-Congo noun *-*bí* 'child' while this is not possible for the derivational suffix -*ví* (cf. Mukarovsky 1976: 18–9; Heine et al. 2016, fn. 25). Accordingly, we hypothesize that the lexical meaning preceded the grammatical one in time and that there was a development from the former to the latter rather than the other way round.

This general process had the effect that the languages concerned now dispose of a productive means of word formation, namely a derivational suffix which is -*mà/-m̄hè* in !Xun, -*ví* in Ewe, and -*ima*- in Ik. Salient morphological characteristics of these suffixes are listed in (10).

(10) Morphological features of the diminutive derivational suffixes -*mà/-m̄hè* of !Xun, -*ví* of Ewe, and -*ima*- of Ik
   a. The suffixes are part of the paradigm of nominal derivation.
   b. Their function is to form diminutive nouns, even in specific contexts they may also assume other, derived functions (Heine et al. 2016: 155–9).
   c. They are essentially fully productive morphemes.
   d. They can create new nouns.
   e. Their use is optional.

The process from compounding to derivation that was looked at in the present section is cross-linguistically common, and it is a regular one leading from one kind of word formation to another. This is different with the process from derivation to inflection, which is the subject of the next section.

3. From derivation to inflection

While being clearly less common than the evolution dealt with in §2, processes whereby derivational markers gradually give rise to inflectional markers can be observed in all major regions of the world. Suffice it to mention a couple of examples from Eskimoan and Iroquoian languages in North America. In the Eskimoan language Yup'ik, some derivational

nominalizers have evolved into inflectional mood suffixes, and in Cherokee of the Iroquoian language family, an earlier derivational instrumental suffix, whose ultimate source can be traced to a verb root meaning 'use', evolved into an inflectional infinitive marker (Mithun 2000: 252).

The present section, however, is restricted to the language that is the focus of discussion in the present paper, namely the Khoisan language !Xun.

3.1 *!Xun*

As was observed in §2.1, nouns in !Xun are essentially transnumeral, that is, they are unspecified for number. Thus, the noun *!xō* can mean 'elephant' or 'elephants'. But there is a small number of exceptions: Some frequently used nouns, typically denoting human beings, follow a suppletive pattern, in that singular referents use a different form than plural referents. We had one of these nouns above: *mà* 'child (of), offspring' has the suppletive plural form *m̄̀hè* 'children'. Accordingly, we saw in §2.1 that in nominal compounds having this noun as their head there is an obligatory singular/plural distinction. Thus, whereas *n!āō* 'house' is not number-sensitive, in combinations with *-mà* as its head it is – hence *n!āō-mà* 'small house' has an obligatory plural form *n!āō-m̄̀hè*, that is, whenever *mà* is a derivational suffix there is an obligatory number distinction.

Now, with a number of nouns, *-mà/-m̄̀hè* has been lexicalized as a part of new nouns. In some cases, the non-derived noun still exists even if its meaning is not exactly the same as that of the derived noun. For example, *tc'ámà* in (11a) denotes a prototypical wild bird which can fly. The non-derived noun *tc'ám*, by contrast, denotes birds as an abstract life form and as such it also includes birds that cannot typically fly, such as ostriches and chickens.[11] Thus, *tc'ámà* and *tc'ám* are closely related but distinct nouns. In a similar fashion, the noun *‖hāmà* 'animal' in (11b) can be traced back to the non-derived noun *‖hā* 'meat' and, again, the two are semantically closely related but distinct nouns.

(11) Examples of !Xun nouns taking the inflectional suffix *-m̄̀hè* (König & Heine 2008)

	Singular	**Plural**	**Meaning**	**Non-derived source**
a.	*tc'ámà*	*tc'á-m̄̀hè*	'bird'	*tc'ám* 'bird as a life form'
b.	*‖hāmà*	*‖hā-m̄̀hè*	'animal'	*‖hā* 'meat'
c.	*xāmà*	*xā-m̄̀hè*	'old man'	**xā*
d.	*!!'ùìmà*	*!!'ùì-m̄̀hè*	'caracal, lynx'	**!!'ùì*
e.	*dàhmà*	*dèh-m̄̀hè*	'woman'	**dàh, *dèh*

In other cases, such as (11c-e), however, the earlier non-derived noun is no longer retrievable and the singular form is an unanalyzable noun, and the erstwhile diminutive plural suffix *-m̄̀hè* has been reduced to and reinterpreted as a plural suffix, e.g. *xāmà*, plural *xā-m̄̀hè* 'old man', having the appearance of an inflectional number marker. In short, with the grammaticalization of the noun *m̄̀hè* 'children (of)' in modifying compounds and its subsequent lexicalization, a lexical free form has turned into a *bona fide* inflectional suffix.

[11] In addition, *tc'ám* denotes 'aeroplane' (König & Heine 2008).

The development from lexical noun to plural marker is well documented, but rather than nouns for 'children' it is in most cases nouns for 'people' that serve as the input of grammaticalization (Heine & Kuteva 2002; Kuteva et al. 2018).

In (10) above, some morphological features of diminutive derivational suffixes such as !Xun *-m̄hè* were listed. Table 2 provides an overview of how this suffix differs from the inflectional plural suffix *-m̄hè*. In accordance with this table, the two are structurally clearly different. First, the derivational suffix belongs to a paradigm of nominal derivational suffixes, which include those listed in (8). The inflectional suffix, by contrast, does not belong to a paradigm since it is the only inflectional suffix on nouns.[12] Second, whereas the derivational suffix can be applied productively to essentially any noun, the inflectional suffix is restricted to a small number of nouns such as the ones in (11c-e). Third, whereas the derivational suffix is used to build new nouns, the inflectional one is restricted to number marking, that is, the latter does not change the status of the noun concerned. And fourth, unlike the derivational suffix, the inflectional one is an obligatory feature of specific nouns.

Table 2. Features distinguishing derivational *-m̄hè* from inflectional *-m̄hè* in !Xun

Features	**Derivational suffix**	**Inflectional suffix**
a. Is part of the paradigm of nominal affixes	+	−
b. Is productive	+	−
c. Can create new nouns	+	−
d. Its use is optional	+	−

To conclude, the two suffixes exhibit contrasting features, even if not all of these features are typically expected on the basis of conventional observations on the distinction between derivation and inflection (see, e.g. Stump 1998). For example, inflectional affixes have been found to be more likely to form morphological paradigms than derivational ones, and to be more productive than derivational affixes.

While the process from compounding to derivation like the one described above is cross-linguistically fairly widespread (Heine & Kuteva 2009; Heine et al. 2016), that from derivational to inflectional suffix appears to be less common. It involved, on the hand, lexicalization in that a few compositional forms of noun-affix combination such as the ones in (11c-e) turned into frozen, non-compositional new nouns. On the other hand, it also must have involved grammaticalization in accordance with two of the parameters in (3): When the erstwhile derivational plural suffix was reinterpreted as a new plural marker, the suffix lost its diminutive function, being reduced to a plural marker (desemanticization). And there was also decategorialization in that the plural marker lost its earlier derivational function, namely its ability to form new diminutive nouns.

[12] We are restricted here to the W2 dialect of !Xun. The situation is different in other dialects (see Heine & König 2015). Note further that according to an anonymous reviewer "some people would claim that only inflection shows paradigms". This is not the position adopted here: unlike the derivational suffix, inflectional *-m̄hè* clearly does not serve word formation – like in languages such as English, French, or Swahili, its function is restricted to expressing the plural of nouns.

3.2 *Discussion*

A process from derivation to inflection does not seem to be cross-linguistically very common, but it has been observed in a number of languages, as the following examples may illustrate.

The first case is taken from the Ik language that was mentioned already in §2.2, and it provides a direct parallel to the case of !Xun looked at in the preceding sections, as the examples in (12) suggest. That the process from free noun (cf. (12a)) to derivational diminutive suffix (12c) was apparently the same as that in the genetically and a really unrelated Ik language was demonstrated in §2.2. Note that in both languages there must have been an intermediate stage, illustrated for !Xun in (7b) above and for Ik in (12b), where the noun for 'child' appears to have been extended first to animate nouns with the meaning 'young or small offspring of X' before being further extended to inanimate nouns as a diminutive suffix, as exemplified in (7c) for !Xun and (12d) for Ik.

(12) From noun to inflectional suffix in Ik: The suppletive noun *ím*, pl. *wik* 'child' (Schrock 2014: 180–1)

	Head noun	**Plural**	**Meaning**	**Literally**
a.	Lexical	*im (imá-)*, pl. *wik (wicé-)*	'child'	
b.	Animal ('child of')	*dǒdo-im*, pl. *dǒdo-wik*	'lamb'	'sheep-child'
c.	Inanimate	*emútí-ím*, pl. *emút-íka-wik*	'little story'	'story-child'
d.	Lexicalized	*dúnéim*, pl. *dúné-ík*	'old woman'	'age-child'

And (12d) illustrates the final stage of lexicalization: There are a number of Ik nouns, including *dúnéim* 'old woman', which are frozen forms, but the earlier lexical or derivational meaning of *im/wik* (or its contextual variants *imá-/wicé-*) was reinterpreted as one signaling a distinction of number – that is, *-(w)ik* in (12d) can be analyzed reasonably only as a plural inflection in such nouns (see Schrock 2014: 180–1).

The second case is of a different nature and involves a European language (Nikos & Ralli n.d.). Griko, is a Greek variety of Southern Italy, spoken by approximately 20000 speakers in nine villages in the center of the Salentine peninsula. Griko has a productive derivational morpheme, *-idz-*, whose function it is to derive verbs from nouns. But the morpheme was extended to a number of verbs, and in this context it lost its derivational function, nowadays serving only to mark the verbs as belonging to Inflection Class 1. Thus, *katalo* and *katal-idz-o* both mean 'destroy', but the latter belongs to Inflection Class 1, which is the most productive of the Griko inflectional verb classes.

The processes in !Xun and Ik, on one hand, and in Griko, on the other, were strikingly different. First, the input was provided by the semantics of diminutiveness in !Xun and Ik. In Griko, by contrast, it was a morphosyntactic function, namely that of deriving one word class from another. Second, whereas lexicalization played a crucial role in the transition from derivation to inflection in !Xun and Ik, this role was played in Griko by context extension (see (3a)), in that the use of *-idz-* was extended from nouns to verbs. And finally, while the output in !Xun and Ik is a number marker, in Griko it is a marker of morphophonological classification. From a grammaticalization perspective, however, the processes were similar: All involved the desemanticization of a lexical or a derivational function resulting in the rise

of an inflectional affix.[13]

4. Conclusion

The observations made in this paper take care of just a small segment of the issues that word formation is concerned with. And the paper was restricted to one theoretical perspective, namely that of grammaticalization. The scope of grammaticalization theory is limited, it has little to contribute to many processes of word formation. Nevertheless, the observations are on the whole compatible with the generalizations on derivation as proposed in Lieber & Štekauer (2014).

The main goal of the paper was to account for why one and the same linguistic form can be associated simultaneously with the status of a lexeme, a conjunct of a compound, a derivational, and an inflectional form. The development from head of modifying compounds to derivation has been observed in a number of African languages, but it does not seem to be intrinsically different from corresponding developments in other languages as, for example, that of the Old English lexeme *hād* 'status, office, rank' developing into the derivational suffix -*hood* of Present-Day English (Traugott & Trousdale 2014).

The development from derivation to inflection, by contrast, does not seem to represent a very common pattern across languages. In the !Xun example examined the development was what one may portray as a parasitic product of a frozen derivational pattern. Nevertheless, this development is overall in accordance with general parameters of grammaticalization.

Acknowledgements

I wish to thank an anonymous reviewer for many valuable comments on an earlier version of the paper. Furthermore, my gratitude goes to the participants of the conference on *Word-Formation Theories III / Typology and Universals in Word-Formation IV*, held in Košice, June 27–30, 2018, for valuable comments on an earlier version of this paper, as well as to the organizers Pavol Štekauer, Livia Körtvélyessy and Slávka Tomaščíková. Furthermore, my thanks are due to Guangdong University of Foreign Studies and Haiping Long, and the University of Cape Town and Matthias Brenzinger for the academic hospitality I received while working there on topics leading to this paper.

References

Anderson, Steven R. 1992. *A-morphous Morphology.* Cambridge: Cambridge University Press.

Bauer, Laurie. 1978. *The grammar of nominal compounding, with special reference to Danish, English and French*. Odense: Odense University Press.

[13] Conceivably, the direct input of grammaticalization in !Xun and Ik was not provided by the derivational but rather by the lexical use of the form concerned. More reconstruction work is needed on this issue.

Bauer, Laurie. 2003. *Introducing Linguistic Morphology*, 2nd edition. Washington, DC: Georgetown University Press & Edinburgh University Press.

Bisetto, Antonietta & Scalise, Sergio. 2005. The classification of compounds. *Lingue e Linguaggio* 4(2). 319–332.

Booij, Geert. 2010. Morphological analysis. In Heine, Bernd & Narrog, Heiko (eds.), *The Oxford Handbook of Linguistic Analysis*, 2nd edition, 449–71. Oxford: Oxford University Press.

Brinton, Laurel J. & Traugott, Elizabeth Closs. 2005. *Lexicalization and Language Change* (Research Surveys in Linguistics). Cambridge: Cambridge University Press.

Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form* (Typological Studies in Language 9). Amsterdam: John Benjamins.

Bybee, Joan L. 2002. Sequentiality as the basis of constituent structure. In Givón, T. & Malle, Bertram F. (eds.), *The evolution of language from pre-language*, 109–34. Amsterdam: John Benjamins.

Campbell, Lyle. (ed.). 2001. Grammaticalization: A critical assessment. Special issue of *Language Sciences* 23(2/3). 113–161.

Carstairs-McCarthy, Andrew. 1992. *Current Morphology (Linguistic Theory Guides).* London: Routledge.

Do-Hurinville, Thành & Hancil, Sylvie. 2015. *La transcatégorialité à travers les langues.* University of Rouen, Typescript.

Enfield, N. J. 2006. Heterosemy and the grammar-lexicon trade-off. In Ameka, Felix & Dench, Allan & Evans, Nicholas (eds.), *Catching Language*, 297–320. Berlin: Mouton de Gruyter.

Greenberg, Joseph H. (ed.). 1963. *Universals of Language*. London: MIT Press.

Guthrie, Malcolm. 1967–71. *Comparative Bantu: An introduction to the comparative linguistics and prehistory of the Bantu languages*. Farnborough: Gregg.

Heine, Bernd. 2003. Grammaticalization. In Joseph, Brian D. & Janda, Richard D. (eds.), *The handbook of historical linguistics*, 575–601. Oxford: Blackwell.

Heine, Bernd & Claudi, Ulrike & Hünnemeyer, Friederike. 1991. *Grammaticalization: A conceptual framework.* Chicago: University of Chicago Press.

Heine, Bernd & Honken, Henry. 2010. The Kx'a family. *Journal of Asian and African Studies* 79. 5–36.

Heine, Bernd & Hünnemeyer, Friederike. 1988. On the fate of Ewe ví 'child' – the development of a diminutive marker. *AAP* (Afrikanistische Arbeitspapiere, Cologne) 16. 97–121.

Heine, Bernd & König, Christa. 2005. Grammatical hybrids: Between serialization, compounding and derivation in !Xun (North Khoisan). In Dressler, Wolfgang U. & Kastovsky, Dieter & Pfeiffer, Oskar E. & Rainer, Franz (eds.), *Morphology and its demarcations: Selected papers*

from the 11th Morphology Meeting, Vienna, February 2004, 81–96. Amsterdam: John Benjamins.

Heine, Bernd & König, Christa. 2015. *The !Xun language: A dialect grammar of Northern Khoisan* (Quellen zur Khoisan-Forschung 33). Cologne: Rüdiger Köppe.

Heine, Bernd & König, Christa. 2016. Observations on observations: On typological features of hunter-gatherer languages. In Shah, Sheena & Brenzinger, Matthias (eds.), *Khoisan Languages and Linguistics: Proceedings of the 5th International Symposium, July 13–17, 2014, Riezlern/Kleinwalsertal* (Research in Khoisan Studies 34), 127–43. Cologne: Rüdiger Köppe.

Heine, Bernd & Kuteva, Tania. 2002. *World lexicon of grammaticalization.* Cambridge: Cambridge University Press.

Heine, Bernd & Kuteva, Tania. 2007. *The genesis of grammar: A reconstruction*. Oxford: Oxford University Press.

Heine, Bernd & Kuteva, Tania. 2009. The genesis of grammar: On combining nouns. In Botha, Rudie & de Swart, Henriette (eds.), *Language Evolution: The View from Restricted Linguistic Systems*, 139–77. Utrecht: LOT (Netherlands Graduate School of Linguistics).

Heine, Bernd & Leyew, Zelealem. 2008. Is Africa a linguistic area? In Heine, Bernd & Nurse, Derek (eds.), *A Linguistic Geography of Africa*, 15–35. Cambridge: Cambridge University Press.

Heine, Bernd & Narrog, Heiko & Long, Haiping. 2016. Constructional Change vs. grammaticalization: From compounding to derivation. *Studies in Language* 40(1). 137–75.

Hopper, Paul J. 1991. On some principles of grammaticization. In Traugott, Elizabeth C. & Heine, Bernd (eds.), *Approaches to Grammaticalization*, Vol. 1. 17–35. Amsterdam: John Benjamins.

Hopper, Paul J. & Traugott, Elizabeth C. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

Hüning, Matthias & Booij, Geert. 2014. From compounding to derivation: The emergence of derivational affixes through "constructionalization". *Folia Linguistica* 48(2). 579–604.

Jurafsky, Daniel. 1996. Universal Tendencies in the Semantics of the Diminutive. *Language* 72(3). 533–578.

König, Christa & Heine, Bernd. 2001. The !Xun of Ekoka: A demographic and linguistic report. *Khoisan Forum* 17. Cologne: University of Cologne.

König, Christa & Heine, Bernd. 2008. *A Concise Dictionary of Northwestern !Xun* (Quellen zur Khoisan-Forschung 21). Cologne: Rüdiger Köppe.

Kuryłowicz, Jerzy. 1975 [1965]. The evolution of grammatical categories. Reprinted in Kuryłowicz, Jerzy (ed.), *Esquisses linguistiques*, Vol. 2. 38–54. Munich: Finck.

Kuteva, Tania & Heine, Bernd & Hong, Bo & Long, Haiping & Narrog, Heiko & Rhee, Seongha.

2018. *World lexicon of grammaticalization*. 2nd revised edition. Cambridge: Cambridge University Press.

Lehmann, Christian. 1982. *Thoughts on Grammaticalization: A Programmatic Sketch* (Arbeiten des Kölner Universalien-Projekts 48), Vol. 1. Cologne: Universität zu Köln, Institut für Sprachwissenschaft.

Lichtenberk, Frantisek. 1991. Semantic change and heterosemy in grammaticalization. *Language* 67(3). 475–509.

Lieber, Rochelle & Štekauer, Pavol (eds.). 2009. *The Oxford Handbook of Compounding.* Oxford: Oxford University Press.

Lieber, Rochelle & Štekauer, Pavol (eds.). 2014. *The Oxford Handbook of Derivational Morphology.* Oxford: Oxford University Press.

Mithun, Marianne. 2000. The reordering of morphemes. In Gildea, Spike (ed.), *Reconstructing Grammar: Comparative Linguistics and Grammaticalization* (Typological Studies in Language 43), 231–55. Amsterdam: John Benjamins.

Mukarovsky, Hans. 1976. *A Study of Western Nigritic*, Vol. 2. [Beiträge zur Afrikanistik, Band 2]. Vienna: Afro-Pub.

Nikos, Koutsoukos & Ralli, Angela. (n.d.). From derivation to inflection: A process of grammaticalization. (https://www.meertens.knaw.nl/cms/files/Morfologiedag/abstract_morphology%20meeting_Kouts-Ralli.pdf)

Persson, Gunnar. 1988 Homonymy, polysemy and heterosemy: The types of lexical ambiguity in English. In Hyldgaard-Jensen, Karl & Zettersten, Arne (eds.), *Symposium on Lexicography III: Proceedings of the Third International Symposium on Lexicography, Copenhagen, 14–16 May*, 269–80. Tübingen: Max Niemeyer.

Robert, Stéphane. 2003a. Polygrammaticalisation, grammaire fractale et propriétés d'échelle. In Robert, Stéphane (ed.), *Perspectives synchroniques sur la grammaticalisation* (Afrique et Langage 5), 85–120. Louvain-Paris: Peeters.

Robert, Stéphane. 2003b. Vers une typologie de la transcatégorialité. In Robert, Stéphane (ed.). *Perspectives synchroniques sur la grammaticalisation* (Afrique et Langage 5), 255–270. Louvain-Paris: Peeters.

Robert, Stéphane. 2004. The challenge of polygrammaticalization for linguistic theory: Fractal grammar and transcategorial functioning. In Frajzyngier, Zygmunt & Hodges, Adam & Rood, David S. (eds.), *Linguistic Diversity and Language Theories* (Studies in Language Companion Series 72), 119–42. Amsterdam: John Benjamins.

Schrock, Terrill B. 2014. A Grammar of Ik (Icé-tód): Northeast Uganda's Last Thriving Kuliak Language. Utrecht: LOT. (Ph.D. dissertation.)

Stump, Gregory. 1998. Inflection. In Spencer, Andrew & Zwicky, Arnold M. (eds.), *The Handbook of Morphology*, 13–43. Oxford: Blackwell.

Traugott, Elizabeth C. & Trousdale, Graeme. 2014. Contentful constructionalization. *Journal of Historical Linguistics* 4(2). 254–82.

Wälchli, Bernhard. 2005. *Co-compounds and natural coordination* (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press.

Bernd Heine
Institut für Afrikanistik
Universität zu Köln
Nonnenwerthstr. 48
50937 Köln
Germany
bernd.heine@uni-koeln.de

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/01.pdf. ISSN 1336-782X.

Interconnectedness and variation of meaning in derivational patterns

Susan Olsen, Humboldt University, Berlin

1. Introduction

Rochelle Lieber's recent book of (2016) entitled *English Nouns: The Ecology of Derivation* discusses the wide range of meanings displayed by the nominalizing suffixes of Modern English and how they are related.[1] Concentrating on just the deverbal suffixes, the following table taken from Lieber (2016) gives a slightly abbreviated summary of their potential readings (for the full table, cf. Lieber 2016: 60–61, Table 4.1):

(1) Abbreviated summary of deverbal suffixes with potential readings from Lieber (2016):

	Event	Res	Ag	Instr	Pat	Loc	…
-al	1	1			2		
-ance	1	1		2	2	2	
-ment	1	1		2	2	2	
-ure	1	1			2	2	
-ation	1	1	2	2	2	2	
-ing	1	1	2	2	2	2	
…							
-er			1	1	2	2	
…							

A first group of suffixes – termed *event suffixes* by Lieber – derive in their primary function complex nouns that refer to the event or the result of an event (indicated by the 1 in the table). The deverbal suffix *-er* belongs to a second group of suffixes, termed *participant* and *personal suffixes*.[2] It derives in its primary sense nouns that carry the meanings of the most salient participants in an event: agent and instrument.

What is of interest to Lieber (2016) is that these suffixes extend their primary meanings to take on secondary functions (indicated by the 2 in the table). The secondary meanings of the event suffixes include reference to the participants in an event (agent, instrument and patient, as well as a few other minor cases not listed here) and to the location of the event. The secondary meanings of the main deverbal participant suffix *-er* include reference to the patient of the event or to its location.

[1] I am grateful to Claudia Maienborn, the participants of the conference and an anonymous reviewer for their constructive comments on the discussion in this article.

[2] The personal suffixes not listed in (1) as well as the collective and inhabitant groups of suffixes are not primarily deverbal and hence have been left out of consideration here. An account of denominal formations that lack a verbal base (cf. *whaler*, *Londoner*, etc.) depends on the inference of an implicit predicate for their interpretation and, thus, presupposes different principles, cf. Ryder (1999), Olsen (2012), Bierwisch (2015a). They, as well as derivation via conversion, will be left to a later discussion.

It is well-known that derivational patterns can subsume a broad range of meanings. Word-formation studies often speak in this regard of "polysemy" or "affixal polysemy". Lieber herself uses the term *polysemy* in her book. *Polysemy* is a pre-theoretical term suggesting the relatedness of lexical meaning in general. The aim of this article is to investigate whether it is possible to characterize in more exact terms the sense (or senses) in which the meanings found in these patterns are related and how they vary. Success in this matter depends on the choice of an appropriate theoretical framework. Therefore, the discussion in the following will concentrate equally on finding a theory of word-formation that best captures the relevant generalizations.

The discussion proceeds by first considering the major findings of some earlier analyses of nominal derivations by means of the suffix *-er* that were carried out within different syntactically oriented frameworks. The discussion then turns in §3 to a consideration of the primary event suffixes and focuses on a major property of productive derivation, namely the compositionality of the derived meaning including the inheritance of arguments. In §4, it is shown that the secondary readings of event nominals lack the compositionality of meaning that is characteristic of the primary readings. This is seen most clearly by their inability to realize the arguments of their verbal base. §5 applies the same reasoning to the participant suffix *-er* discovering with the help of the criterion of compositionality that the two putative primary readings of agent and instrument actually arise via two semantically distinct variants of the formal suffix *-er*. §6 offers additional evidence for this assumption from the history of the Romance and Germanic languages. The final §7 summarizes the discussion as a whole.

2. Earlier analyses of the suffix *-er*

Early on, Rappaport Hovav & Levin (1992) identified two major properties of *-er* nominals:

i) Although they are frequently labeled agent and instrument nouns, *-er* nominals refer more generally to the external argument of the base verb, irrespective of its thematic role. This observation has since become known as the *external argument generalization*.

ii) The ability to inherit arguments from the verbal base of the complex formation and to realize them in a syntactic configuration is not consistent across the class of deverbal *-er* nominals.

The examples in (2) illustrate the wide range of referents that fall under the denotation of *-er* nominals as predicted by i). What ii) refers to is the fact that – as opposed to the other options – nominalizations in *-er* understood as instruments do not permit the expression of the complement of the verbal base in an accompanying *of* phrase (first brought to attention by Roeper (1987: 281–297) and also recognized by Fanselow (1988: 106)).

(2) Range of referents of the external argument of a nominal in *-er*:

a.	*signer of the contract*	agent
b.	*admirer of talent*	experiencer
c.	*owner of the car*	possessor
d.	*receiver of the package*	goal

e.	*contributer of money*	source
f.	*heater* (**of the room*)	instrument

What is the source of the difference documented in example (2f) as opposed to the examples in (2a-e) in the ability to inherit arguments of the base verb? Rappaport Hovav & Levin (1992) propose that the difference is bound to the eventive vs. non-eventive interpretation of the nominal. *Mower of the lawn* (allowing a syntactic complement) is eventive because it implies that someone is participating in or has participated in the event of mowing. *Mower* and *lawn mower* can also be understood similarly as eventive. In the eventive reading, therefore, the verbal arguments are available and the internal argument of the verbal can be realized as a syntactic complement. However, its expression is not obligatory. But *mower* and *lawn mower* also have an interpretation in which they don't imply an event but rather label a category of instruments. This is the non-eventive reading in which the argument structure of the underlying verb can be accessed, cf. **mower of the lawn* (as an instrument).

A number of linguists have attempted to account for this generalization within a syntactic framework (cf., e.g., Keyser & Roeper 1984; Roeper 1987; van Hout & Roeper 1998; Borer 2003). Recently, Alexiadou & Schäfer (2010) have rejected Rappaport Hovav & Levin's (1992) explanation in terms of an eventive vs. non-eventive sense of the noun and have claimed that all deverbal *-er* nominals are eventive, irrespective of whether they are understood as an agent or an instrument.

Their main argument centers on the non-intersective interpretation of adjectives. The examples in (3) and (4) are intended to show that in both the *eventive* and the *non-eventive* readings of *-er* nominals an attributive adjective can modify the event expressed by the verbal base of the complex noun. In this non-intersective interpretation, the adjective functions like an adverb modifying the underlying verbs *dance* and *calculate*. *Beautiful dancer* in (3) has an intersective reading as well, in which the adjective modifies the referent of *dancer*, yielding the reading 'x is dancer & beautiful'.[3]

(3)	*beautiful dancer* 'x dances beautifully'	= eventive
(4)	*fast calculator* 'x calculates in a fast manner'	= non-eventive

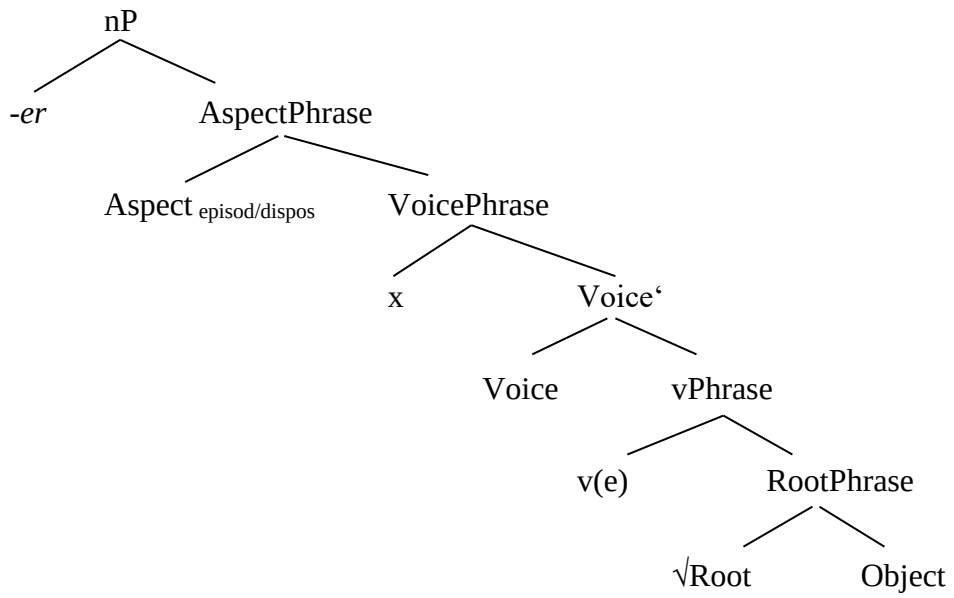
Alexiadou & Schäfer (2010) conclude from the non-intersective reading of the attributive adjectives in cases like (3) and (4) that both *eventive* and *non-eventive -er* nominals share a configuration in which an event is present and can be targeted by the adjective. They propose, in other words, that a nominal head embeds a verbal projection under it with its complete functional structure.

Working within the framework of Distributed Morphology, they assume the structure shown in (5) for a nominalization by means of the suffix *-er*. In Distributed Morphology, vocabulary items are not marked for a category but enter into an enumeration as a category-neutral root. They assume a category label only when they are inserted into a syntactic

[3] For a more thorough discussion of the distinction the intersective and non-intersective readings of adjectives see Larson (1998).

configuration in which they are dominated by a functional item that determines the categorial features of the projection in which they occur.

(5) Structure of *-er* nominalizations according to Alexiadou & Schäfer (2010: 13)



In this analysis, a category-neutral *√Root* merges with a potential object to yield a RootPhrase. The functional element *v* (little v) merges with the RootPhrase, categorizing it as a vPhrase which induces the category *verb* upon the √Root that heads the phrase. The vPhrase then merges with the functional category *Voice* that projects to a VoicePhrase in whose specifier an external argument is projected. The functional category *Aspect* merges with VoicePhrase yielding an AspectPhrase that is sister to the functional morpheme *-er* that carries the category *n*. Hence, *-er* categorizes the structure dominated by AspectPhrase as an nPhrase, or a nominal projection. Rappaport Hovav & Levin's (1992) external argument generalization is accounted for by the occurrence of the external argument *x* in the specifier of the VoicePhrase. It is bound by the suffix *-er*, determining the referent of nP as whatever thematic role is assigned to the external argument *x* by the head of vP.

So contrary to Rapport Hovav & Levin (1992), the difference between the two classes of nominals in Alexiadou & Schäfer's (2010) analysis is not a matter of the presence vs. absence of an event; the event variable *e* is provided by little v in both cases. The difference lies, rather, in the feature structure of Aspect: the so-called *eventive* nominals have an episodic aspect, while the *non-eventive* nominals are dispositional. The aspectual feature is picked up by the verb as it moves from its underlying position in the RootPhrase to join with the nominalizer *-er* in nP.

Under these assumptions the question now arises as to why dispositional *-er* nominals are unable to realize syntactic complements. That is, why can't **mower of lawns* be understood as an instrument. Alexiadou & Schäfer (2010) assume that dispositional nominals are similar to generic and habitual sentences in that they generalize over an unlimited number of instances. To take their example given in (6), *cut* is an obligatorily transitive verb (cf. **The instructor cut.*). But as shown in (6), in a habitual context the direct object can be omitted:

(6) *The sewing instructor always cuts Ø in straight lines.*

Alexiadou & Schäfer (2010) assume that in habitual contexts, unquantized objects can remain implicit, thus yielding dispositional events.

There are two points of concern with this analysis. First, non-intersective modification seen in the derived nominals in (3) and (4) applies to underived instruments as well as to those derived from a verb. For example, the instrument nouns in (7) are simple, underived nouns. Yet, when modified by an adjective like *fast*, *good*, or *slow*, the adjective takes on an adverbial function in the same manner as with derived instrument *-er* nominals in that it modifies an implicit event associated with *train*, *knife* and *gun*.

(7) a. *fast train* = travels quickly
b. *good knife* = cuts well
c. *slow gun* = shoots slowly

In fact, the non-intersective meaning is available in underived agents as well as the examples in (8) show.

(8) a. *good pilot* = good at flying planes
b. *fast cook* = cooks quickly
c. *old friend* = friendship has existed for a long time

What this shows is that the non-intersective meaning must be available in both agent and instrumental nouns without having to postulate a full sentence structure in their representation, including verb movement up to the nominal suffix.

Secondly, dispositional *-er* nominals can indeed realize an inherited verbal argument just like episodic *-er* nominals and when they do, there is no difference in their behavior. In the presence of a syntactic complement, both the episodic nominal in (9) and the dispositional nominal in (10) enforce a personal actor reading on the *-er* noun and exclude an interpretation as an instrument:

(9) Episodic
a. *The <u>mower of the lawn</u> just finished his job.*person
b. **The <u>mower of the lawn</u> was just turned off.* *instrument

(10) Dispositional
a. *I am looking for a good <u>mower of lawns</u>.* person
b. **I want to buy a good <u>mower of lawns</u>.* *instrument

From this we can conclude that the property regulating access of the derived nominal to the argument structure of its verbal base is neither an eventive vs. non-eventive interpretation in the sense of Rappaport Hovav & Levin (1992), nor can it be explained by an episodic vs. dispositional aspect as Alexiadou & Schäfer (2010) propose.

How does Lieber (2016) view the *polysemy* of *-er* with respect to the agent and instrument meanings? She assumes a basic underspecified representation for *-er* nominals that is the same regardless of whether the nominal denotes an agent or instrument. The basic skeleton for the participant suffix *-er* within her Lexical Semantic Framework is shown in

(11). The suffix *-er* carries the feature [+material] and will adopt the value for the variable of the feature [dynamic] that is present in the verb to be embedded in the place holder position of the skeleton, namely <base>. The referential argument of the affix [$_R$] will then be coindexed with an argument of the verbal base, giving the reference of the nominal.

(11) Basic skeleton for the participant suffix *-er*:
[+material, β dynamic ([$_R$], <base>]

The skeleton provides the underlying representation for both the agent and instrument readings of a nominal such as *shooter*. In context, the potential meanings of the nominal *shooter* will be disambiguated by taking on further specification, namely by adding the feature [+animate] for agent or [-animate] for instrument to the representation of the affix in the basic skeleton, cf. my underlining in the examples of (12). In both (12a) and (12b), the positive value for the feature [+dynamic] of the affix has been carried over from the corresponding feature of the base verb, i.e. *shoot*, and the referential argument R of the affix is coindexed by means of the subscript *i* with the external argument of *shoot*, giving the meanings ‘person that shoots’ and ‘thing that shoots’.

(12) a. *shooter* (agent)
[+material, +dynamic, <u>+animate</u> ([$_{R\text{-}i}$], [+dynamic ([$_i$], [ ])])]
-er *shoot*

b. *shooter* (instrument)
[+material, +dynamic, <u>-animate</u> ([$_{R\text{-}i}$], [+dynamic ([$_i$], [ ])])]
-er *shoot*

Thus, the difference in meaning between an agentive and an instrumental meaning of a *-er* nominal is captured by the addition of the feature [+animate] or [-animate] to the basic skeleton in context. And *shooter* can indeed be understood as either the agent or the instrument of the shooting (in the latter case, equivalent to ‘gun’). But the problem that arises for this analysis is that the internal argument of *shoot* (cf. the unindexed second argument [ ] of *shoot* in (12a) and (12b)) is open for satisfaction in both interpretations of *shooter*. Consequently, in this framework nothing prohibits the phrase **shooter of the bear* from taking on an instrumental reading.

Since the syntactic frameworks of Rappaport Hovav & Levin (1992) and Alexiadou & Schäfer (2010) as well as the formal lexical semantic framework of Lieber (2016) fail to explain why *mower of the lawn* or *shooter of the bear* refer to a person and not to an instrument, let us turn our attention to another type theoretical framework in the next section.

3. The event suffixes in their primary readings

Bierwisch (1989, 2015) has proposed a lexicalist theory that is based on a two-level theory of semantics in which meaning is separated into two levels of representation (cf. Bierwisch 1983, 1988, 2007, 2011, 2015a; Bierwisch & Lang 1989; Lang & Maienborn 2011; Maienborn 2017). There is, on the one hand, a highly articulated, complex level of conceptual structure (CS) that reflects our conceptual knowledge and can be enriched by contextually relevant features. On the other hand, a level of lexical-semantic structure (semantic form: SF)

is conceived of as a condensed version of CS. It represents the interface between CS and the system of grammar in that it encodes only the aspects of the more comprehensive conceptual meaning that are needed to establish the categories of grammar. Consider as an example the entry for the verb *construct* in (13).

(13)	[construct]	[V]	λx λy λe	[e : [y CONSTRUCT x]]
	PF	Cat	AS	SF

The lexical semantic meaning of the verb *construct* is given in its semantic form (SF) which is the level of meaning visible to the grammar. SF characterizes the invariant aspects of meaning bound to the language system and is strictly compositional. The verb's argument structure (AS) is derived from its SF by lambda abstracting over the syntactically active variable positions in SF and prefixing the corresponding lambda operators to the SF in inverse order of occurrence in SF. The verb's arguments are thus assembled in a hierarchy in which the outermost argument is progressively assigned to the lowest complement in the syntactic configuration.

Affixes have similar lexical entries, but they are bound morphemes that combine with a lexeme, and not a complete phrase, as their argument. The suffix *-ion*, for example, selects a verb as its lexical argument and creates a noun (cf. its grammatical category [N]). Therefore, its AS in (14a) is made up of a predicational argument λP with the annotation [V] and an argument vector $\vec{v}$ that represents the verb's unsaturated arguments to be taken over by the derived. In the case of the verb *construct* in (14b) these will be the referential, external and internal arguments (= e, y und x, respectively). The result is *construction* in (14c), cf. the discussion in Bierwisch (2015: 1062–1082).[4]

(14)	a.	[-ion]	[N]	λP λ$\vec{v}$ Gen e' [P($\vec{v}$ (e')]	
				[V]	
		PF	Cat	AS	SF
	b.	[construct]	[V]	λx λy λe [e : [y [CONSTRUCT x]]]	
	c.	[construct-ion]	[N]	λx λy λe [e : [y [CONSTRUCT x]]]	

The highest argument in the AS of the verb *construct* in (14b) (= λe) is referential and as such can be anchored in time and modified by aspect, etc. The second highest argument (= λy) is the designated argument of the verb and the lower arguments (in this case only λx) are its internal arguments. The internal arguments of a verb are in principle obligatory unless they are marked as optional. The AS of the derived noun *construction* in (14c) is similar to the verb from which it is derived. The difference lies in the fact that *construction* is a nominal. The highest argument λe in the nominal AS takes on a nominal referent and the two lower arguments are internal arguments which, in the case of a noun, are optional, cf. Bierwisch (1989, 2015a). What this framework allows us to see is that the deverbal noun *construction* is the compositional product of the meaning of the verb *construct* and the suffix *-ion*.

As already mentioned, the event variable of the underlying verb is carried over in the derived noun as the nominal referent. So, *construction* as a noun refers to the event of

[4] For the lexical entry of the suffix, I have used a formally equivalent variant of Bierwisch's actual proposal (for the discussion of which I am indebted to Claudia Maienborn).

constructing, allowing it to be assigned the feature [+/- definite], to be quantified, etc.

(15) *the construction of the airport* (*by the city*)

As many linguists have pointed out, a deverbal event noun allows different aspects of its event reference to be discerned (cf. Grimshaw 1990; Borer 2003; Bierwisch 1989, 2015):

(16)	a.	*The construction of the airport was approved by the city*.	Event
	b.	*The construction of the airport is taking many years*.	Process
	c.	*The construction of the airport cost in total around $3.6 billion.*	Result

How are these different readings to be accounted for? Bierwisch (2015) debates and then rejects the idea of assigning these different aspects (event, process, result) directly to the SF of the nominalizing suffixes because they are actually predicted by the architecture of the lexical system within the theory of two-level semantics. The lexical-semantic (SF) representation of a verb is more abstract than its representation in terms of conceptual structure (CS). SF reduces the highly complex knowledge encoded in CS to just the information relevant to the grammar, cf. again (14c) repeated here for convenience:

(14c) [construct-ion] [N] $\lambda x\ \lambda y\ \lambda e$ [e : [y [CONSTRUCT x]]]

The event variable (= e) is an example of the condensed information in the elements of SF: it abstracts over and abbreviates the conceptual options bound to the concept of an event. That is, it represents in reduced form our conceptual knowledge of an event with its different facets. This allows the SF variable e denoting an event to be flexible in its conceptual interpretation. The different facets of an event can remain indeterminate. Or, if the context requires it, they can be distinguished, cf. (17):

(17)	a.	*acquisition of the painting*	(<u>event</u>, process, result)
	b.	*accumulation of wealth*	(event, <u>process</u>, result)
	c.	*confession of guilt*	(event, process, <u>result</u>)

What is noteworthy is that the different aspects of the subordinate category *event* all allow the realization of the verbal argument as documented in the *of* phrases of (17). This indicates that they are part of the compositional meaning of the derived noun.

In accordance with these assumptions, the two primary readings of event nouns in Lieber's (2016) table 4.1 can be conflated into a single derivational pattern Event/Process/Result in a lexicalist grammar. The revised table would thus take on the shape given in (18):

(18) Revised abbreviated table from Lieber (2016: 60):

	Event/Proc/Res	Ag	Instr	Pat	Loc
-al	1			2	
-ance	1		2	2	2
-ment	1		2	2	2
-ure	1			2	2
-ation	1	2	2	2	2
-ing	1	2	2	2	2

4. The event suffixes in their secondary readings

What about the secondary readings of the event suffixes? The agent, instrument, patient and location readings of event nouns (termed *referential* or *R* readings by Lieber) don't allow access to the argument structure of the underlying verb. Lieber (2016) finds no argument inheritance with the secondary meanings in her large corpus. This is also the distinction that Grimshaw (1990), Borer (2003) as well as many other linguists make between argument structure nominals and referential" nominals.

The lack of argument inheritance indicates a lack of compositionality because the full meaning of the verb with its AS is unavailable to the nominal in its secondary meaning. The examples in (19) contrast the readings of a small sample of deverbal nouns in their event readings that permit the realization of the internal argument with a possible secondary reading (such as agent, instrument, patient, or location) that can be assigned to these nouns. The secondary readings do not allow the expression of a verbal argument and, hence, are not a product of compositional derivation:

(19) a. *prosecution* (*of the criminal* = event/ *ag)
b. *illustration* (*of the book* = event/ *instr)
c. *annexation* (*of the wing* = event/ *pat)
d. (*in the*) *refrigeration* (*of the flowers* = event/ *loc)

So, the question now emerges as to how these secondary meanings arise. They are obviously the result of a transferred (i.e. non-literal) meaning on the basis of the primary reading of the noun, cf. Bierwisch (1989: 40–42) and (2015: 1113–1116). A semantic operation shifts the reference from the event of the regularly derived deverbal noun to a participant in the event (agent, instrument, patient) or to the location of the event in the shifted variant.[5] As a result of the shift, the arguments originally inherited from the verbal base are suppressed.

(20) Shift operation
a. λP $\lambda z\ \exists e'\ [[z\ \text{AGENT}\ e'] \ \&\ P(e')]$
$[N_F]$
b. *prosecution*$_f$: $\lambda z\ \exists e'\ [[z\ \text{AGENT}\ e'] \ \&\ [\text{PROSECUTION}\ e']]$

[5] The formalization of the operation is my responsibility; it differs from Bierwisch's (2015) proposal and is to be understood only as an approximation of what the shift accomplishes.

The predicative argument λP in (20a) is annotated with a feature N_F that picks out the (restricted) class of derived nouns that are subject to the shift. The variable z identifies the derived entity as the agent of P. For example, *prosecution* can be understood as an actor as well as an event. The more precise meaning of the shifted noun is given in (20b) where the derived event noun *prosecution* has substituted for the predicate variable P in the formula which shifts its reference from the event to an actor in a *prosecution* event.

Shift operations of this type presuppose the existence of the derived nominal. For example, the *agent* produced by the shift differs semantically from the referent of a productively formed agentive *-er* noun (cf. *prosecutor*). The *prosecution* is not understood as 'one who prosecutes', but as 'person (or persons) carrying out the prosecution event'. The same goes for other such derived agents, cf. (21):

(21) *administration, government, resistance, meeting*

Interchanging the predicate constant *agent* with i*nstrument*, *result* or *location* will produce the other secondary meanings. In the case of i*nstrument*, for example, a different (and quite restricted) class of derived nouns enters into the shift and, again, the instrumental meaning derived from the shift is not equivalent to productively formed instruments in *-er*. The nominals in (22) refer in their regularly derived primary reading to an event in the broadest sense. But if subject to the shift, they can also be construed as the means of the event: *insulation*, *decoration*, etc. denote in addition to the event also the means of insulating, decorating, etc.

(22) *insulation, decoration, adornment, illustration*

5. The participant suffix -*er*$_{\text{actor}}$

Turning now to the participant suffix *-er*, it takes over the meaning of its verbal argument just like the event suffix *-ion* does but, in the course of doing so, it suppresses the event argument of its verbal complement by binding it with a generic operator Gen:

(23)	a.	[-*er*$_{\text{actor}}$]	[N]	λP λ$\tilde{v}$ Gen e' [V]	[P($\tilde{v}$) (e')]
		PF	Cat	AS	SF
	b.	[*sell*]	[V]	λx λy λe	[e : [y [SELL x]]]
	c.	[sell-er]	[N]	λx λy Gen e'	[e' : [y [SELL x]]]

With the event variable e bound by the generic operator in (23c) (= Gen e'), it is no longer active and cannot be assigned. The highest argument in the AS of *seller*, originally the external argument of the verb, now becomes the referent of the derived nominal, thus capturing the external argument generalization of Rappaport Hovav & Levin (1992) and others, *seller* 'one who sells'.

Although the variable e' is blocked grammatically, it is present in SF and is part of our conceptual knowledge. So, we could ask: What type of event is implicit in a nominal that refers – not to the event itself (like *construction*) – but to the actor in the event as in *seller*?

With personal referents, *-er* nominals can imply the following types of activity, cf. Rainer (2015: 1310):

(24) Implicit activity types of *-er* nominals

a.	a fleeting occasional activity:	*protester, voter, winner*
b.	a habitual activity:	*gambler, smoker, complainer*
c.	an occupation:	*designer, preacher, writer*
d.	often all types are possible:	*hunter, dancer, seller …*

In stark contrast to this, the referents of instrumental *-er* nouns don't imply an activity at all. In fact, what they denote is in no way dependent on an activity being carried out (cf. also Rappaport Hovav & Levin (1992); Alexiadou & Schäfer (2010), among others). They simply denote an instrument that has been constructed for an intended purpose, *cf. heater, grater, shredder*. In contrast to a *protester* or a *gambler*, a *heater* is not identified by virtue of any activity taking place or having taken place. It may never have been put to use to heat anything. A heater is a heater by virtue of its design and the purpose for which it was constructed. So, for instance, a *shredder* understood as an agent is identified by an activity: a person is or has implemented the activity denoted by the base verb. But the instrument *shredder* is a thing identified by its design, not by an activity.

In a compound, the instrument noun can be modified by its purpose, cf. the left column in (25).[6] But as head of a DP, an instrument noun cannot express a verbal argument, as shown in the righthand column.

(25)

	compound	DP
a.	*water heater*	**heater of water*
b.	*cheese grater*	**grater of cheese*
c.	*paper shredder*	**shredder of paper*

How is this lack of argument inheritance with instrumental nouns to be explained? How do instrumental *-er* nominals relate to agentive *-er* nominals? Bierwisch (2015) suggests two possibilities using the examples in (26):

(26) a. (*piano*) *player*
b. (*record*) *player*

First, *player* could be derived as a personal noun from the verb *play* to which a coercive shift applies to yield a non-personal noun. Or, alternatively, the verb *play* has one reading with a personal and one reading with a non-personal subject. As for the first option, it is hard to see non-personal *-er* nominals deriving directly from personal *-er* nominals by shifting the referent from a person to a thing. The reason for this is because not all instruments depend on the existence of personal actors. The nominals in (27), for instance, resist an interpretation as an agent:

[6] Rappaport Hovav & Levin (1992) considered the first constituent of a compound a modifier in contrast to the complement expressed in the corresponding phrase. This view actually contradicts a prominent assumption in the literature that the first constituent in such a compound is assigned a thematic role by the deverbal head, cf. Olsen (2017) for discussion of such synthetic compounds. Olsen (2012) argues explicitly however that the putative complement is a modifier as does Bierwisch (2015).

(27) Primarily instruments
computer, adapter, fertilizer, humidifier, thruster, blower, recliner, heater, trailer, freezer, feeder, condenser, muffler, bumper, beeper, tranquilizer, multiplier, refrigerator, vibrator, simulator, projector, calculator, ventilator, duplicator, monitor

As for the second option, it is unnecessarily redundant. This assumption necessitates the presence of two verbs in the lexicon, one with a personal, another with a non-personal subject. Besides the unwanted implication that the verbs are ambiguous in meaning, the problem still remains of how the verb with the non-personal subject would block the realization of its argument in the derived noun. If the verb *clean*, for example, has two lexical entries, the one with the personal subject would be chosen for the linguistic environment in (28a), while the second entry with the non-personal subject would be the basis for the sentence in (28b). But the construction in (28c), presumably resulting from the non-personal version in (28b), prohibits the realization of the argument that is expressed in (28b):

(28) a. *The janitor cleaned the floor.*
b. *This device cleans floors.*
c. **cleaner of floors*

Let us therefore entertain a third option, namely that there are two (homonymous) *-er* suffixes:

i) Personal actor $\text{-}er_{\text{pers}}$ derives compositional formations that permit realization of the inherited argument, cf. *player of the piano*.

ii) Non-personal, $\text{-}er_{\text{instr}}$ derives instruments with no access to the verbal argument structure, cf. **player of the record*. They are instead characterized by a purpose. Under these assumptions, the instrumental $\text{-}er_{\text{instr}}$ suffix characterizes a class of instruments with a variable purpose supplied by the generically bound verbal event argument, cf. (29):

(29) [$\text{-}er_{\text{instr}}$] [N] $\lambda P \ \lambda z$ Gen e' [INSTR(z) & [z PURPOSE e'] & P(e')]
[V]

The representation of the $\text{-}er_{\text{instr}}$ suffix in (29) requires a one-place verbal predicate to substitute for the predicate variable P. Thus, all arguments of the verbal predicate apart from its referential event argument must be existentially bound before entering the formula. The result of applying the $\text{-}er_{\text{instr}}$ suffix to the verb *shred* in (30a) is (30b):

(30) a. *shred*: $\lambda x \ \lambda y \ \lambda e$ [e : [y [SHRED x]]]
b. *shredder:* λz Gen e' [INSTR(z) & [z PURPOSE e'] SHRED(e')]

The class of underived instruments, on the other hand, have lexicalized their purpose, cf. *knife* in (31):

(31) *knife:* λz Gen e' [INSTR(z) & [z PURPOSE e'] & CUT(e')]

Hence, instrumental *-er* nominals are not the result of transferred meaning (a coercive shift) as Bierwisch's first option and Lieber's analysis imply, nor do they result from Bierwisch's second proposal of an additional entry with a non-personal subject. Rather, they are derived directly by a second non-personal suffix $-er_{instr}$ that is homonymous to the personal actor suffix $-er_{pers}$.[7]

6. Diachrony of the agentive and instrument suffixes

The analysis just proposed receives support from the historical facts. It can be shown that the apparent *polysemy* of the agent and instrument suffixes in Romance and Germanic can be traced back to two independent suffixes stemming from Latin *-tor* (agentive) and *-torium* (instrumental). In Old High German these suffixes emerged as *-ārius* and *-ārium*. The two distinct suffixes became conflated with time into a single form due to phonological weakening during the borrowing process, cf. Müller (2011).

The development of the suffixes in Romance was much more complex, taking different courses in each of the Romance languages. Rainer (2015: 1313) first explains that Latin *-tor* originally was limited to an agentive meaning. However, agent nouns in Latin *-tor* could give rise to relational adjectives in *-torius* so that phrases such as *opus tectorium* 'work of covering; i.e. plaster' were created. With deletion of the head noun opus, *tectorium* remained the sole carrier of meaning and took on the interpretation 'that which serves for covering; i.e. plaster'. This was the basis for the reanalysis of *-torium* as an instrumental suffix, distinct from the agentive suffix *-tor*.

The interplay between borrowing, reanalysis, ellipsis, conflation, and analogy that took place following this development is meticulously documented for the standard languages and many of their primary dialects in Rainer (2011). And remnants of the original distinction agent vs. instrument can still be seen in the present-day Romance languages, although a clear-cut distinction in meaning no longer exists primarily due to borrowings from the dialects that conflated the two suffixes, cf. Rainer (2011: 10):

(32)		Language	Agent/ instrument
	a.	Spanish	*-dor/ -dero*
	b.	French	*-eur/ -oir*
	c.	Italian	*-tore/ -toio*
	d.	Portuguese	*-dor/ -douro*

[7] Rappaport Hovav & Levin (1992) cite *defuser* in *Anger is a geat defuser of pentup emotions* as a non-agentive nominalization that shouldn't be able to realize the verbal argument. *Defuser* is surely understood here as a metaphor/personification that would fall under the actor suffix. *Inducer* in *A protein that is an inducer of blood vessel growth*, on the other hand, is meant to show that instruments can realize arguments. *Inducer* here, however, does not fall under the definition of instrument given in this section. And could be understood as an actor in the same manner as *defuser* in the first example.

7. Conclusion

The aim of this discussion was to examine the inter-connectedness and variation of meaning in the deverbal nominal patterns of English. In her book of (2016), Lieber speaks of "a complex web of polysemy". The question posed at the outset of this discussion was: Can the nature of this polysemy be determined with more precision? It has been argued that Bierwisch's (2015) lexicalist framework provides a clear definition of compositional formations. This criterion enables us to distinguish compositional derivational patterns (Lieber's primary functions, cf. the eventive *construction*) from non-compositional semantic shifts resulting in transferred meaning (i.e. Lieber's secondary functions, cf. *prosecution* understood as an agent).

The criterion of compositionality has also exposed an ambiguity found in the *-er* pattern. While both the agent and instrument meanings of *-er* nominals are compositional formations, their meaning results from two distinct suffixes, a personal *-er* that inherits the argument structure of its verbal base and a non-personal *-er* that characterizes a class of things, i.e. instruments, that have a variable function whose actual content is spelled out by the predicational argument with which the suffix combines, serving as a modifier of the PURPOSE constant in the SF of the suffix.

In summary, the primary meanings of the deverbal nominal suffixes isolated by Lieber in her (2016) study are compositional derivations; their secondary readings result from the primary meanings via meaning transfer. In addition, there are two homonymous *-er* suffixes, each with its own primary meaning.

(33) Summary of discussion given as a revised table

	Ev/proc/res	Ag	Instr	Pat	Loc	…
-al	1			2		
-ance	1		2	2	2	
-ment	1		2	2	2	
-ure	1			2	2	
-ation	1		2	2	2	2
-ing	1		2	2	2	2
-er$_{pers}$		1				
-er$_{instr}$			1			

References

Alexiadou, Artemis & Schäfer, Florian. 2010. On the syntax of episodic vs. dispositional *-er* nominals. In Alexiadou, Artemis & Rathert, Monika (eds.), *Syntax of Nominalizations across Languages and Frameworks*, 9–38. Berlin: Walter de Gruyter.

Bierwisch, Manfred. 1983. Semantische und konzeptuelle Repräsentation lexikalischer Einheiten. In Motsch, Wolfgang & Ruzicka, Rudolf (eds.), *Untersuchungen zur Semantik*, 61–99. Berlin: Akademie-Verlag.

Bierwisch, Manfred. 1988. On the grammar of local prepositions. *Syntax, semantik und lexicon*, 29. 1–65.

Bierwisch, Manfred. 1989. Event nominalizations: Proposals and problems. In Motsch, Wolfgang (ed.), *Wortstruktur und Satzstruktur*, 1–73. Berlin: Akademie der Wissenschaften.

Bierwisch, Manfred. 2007. Semantic form as interface. In Spät, Andreas (ed.), *Interfaces and Interface Conditions*, 1–32. Berlin: de Gruyter.

Bierwisch, Manfred. 2011. Semantic features and primes. In Maienborn, Claudia, & von Heusinger, Klaus & Porter, Paul (eds.), *Semantics: An International Handbook of Natural Language Meaning 1*, 322–357. Berlin: de Gruyter.

Bierwisch, Manfred. 2015. Word-formation and metonymy. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-Formation: An International Handbook of the Languages of Europe*, 1099–1128. Berlin: De Gruyter Mouton.

Bierwisch, Manfred & Lang, Ewald. 1989. *Dimensional Adjectives: Grammatical Structure and Conceptual Interpretation*. Berlin: Springer.

Borer, Hagit. 2003. Exo-skeletal vs. Endo-skeletal explanations: Syntactic projections and the lexicon. In Moore, John & Polinsky, Maria (eds.), *The Nature of Explanation in Linguistic Theory*. Chicago: CSLI and Univesity of Chicago Press.

Fanselow, Gisbert. 1988. German word order and universal grammar. *Natural language parsing and linguistic theories*. 317–355. Dordrecht: Springer.

Grimshaw, Jane. 1990. *Argument Structure*. Cambridge, MA: MIT Press.

Keyser, Samuel Jay & Roeper, Thomas. 1984. On the middle and ergative constructions in English. *Linguistic Inquiry* 15. 381–416.

Lang, Ewald & Maienborn, Claudia. 2011. Two-level semantics: Semantic form and conceptual structure. In Maienborn, Claudia & von Heusinger, Klaus & Porter, Paul (eds.), *Semantics: An International Handbook of Natural Language Meaning* 1, 709–740. Berlin: de Gruyter Mouton.

Larson, Richard K. 1998. Events and modification in nominals. In Strolovitch, Devon & Lawson, Aaron (eds.), *Proceedings of Semantics and Linguistic Theory* 8, 145–168. Ithaca, NY: Cornell University.

Levin, Beth & Rappaport Hovav, Malka. 1992. The perspective from unaccusativity. *Thematic structure: Its role in grammar* 16. 247.

Lieber, Rochelle. 2016. *English Nouns: The Ecology of Nominalization*. Cambridge: Cambridge University Press.

Maienborn, Claudia. 2017. Konzeptuelle Semantik. In Staffeldt, Sven & Hagemann, Jörg (eds.), *Semantiktheorien: Lexikalische Analysen im Vergleich*, 151–188. Tübingen: Stauffenburg.

Müller, Peter O. 2011. The polysemy of the German suffix *-er*. Aspects of its origin and development. *STUF – Language Typology and Universals* 64(1). 33–40.

Olsen, Susan. 2012. Semantics of Compounds. In Maienborn, Claudia & von Heusinger, Klaus & Portner, Paul (eds.), *Semantics. An International Handbook of Natural Language Meaning*, Vol. 3, 2120–2150. Berlin: De Gruyter Mouton.

Olsen, Susan. 2017. Synthetic Compounds from a Lexicalist Perspective. *Zeitschrift für Wortbildung/ Journal of Word Formation* 1(1). 15–43.

Rainer, Franz. 2011. The agent-instrument-place polysemy of the suffix *-tor* in Romance. *STUF – Language Typology and Universals* 64(1). 8–32. Berlin: Akademie-Verlag.

Rainer, Franz. 2015. Agent and instrument nouns. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-Formation: An International Handbook of the Languages of Europe*, 1304–1316. Berlin: De Gruyter Mouton.

Rappaport Hovav, Malka & Levin, Beth. 1998. *-er* nominals: Implications for a theory of argument structure. In Stowell, Tim & Wehrli, Eric (eds.), *Syntax and Semantics 26: Syntax and the Lexicon*, 127–153. Academic Press: New York.

Roeper, Thomas. 1987. Implicit arguments and the head-complement relation. *Linguistic Inquiry* 18. 267–310.

Ryder, Mary Ellen. 1999. Bankers and blue-chippers: an account of *-er* formations in present-day English. *English Language & Linguistics* 3(2). 269–297.

van Hout, Angelika & Roeper, Thomas. 1998. Events and aspectual structure in derivational morphology. (https://www.umass.edu/linguistics/member/tom-roeper) (Accessed 2018-08-31.)

Susan Olsen
Deparment of English and American Studies
Humboldt University, Berlin
susan.olsen@rz.hu-berlin.de

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/02.pdf. ISSN 1336-782X.

Productivity of recursive compounds

Makiko Mukai, University of Kochi, Japan

Recursion at word-level is productive in many languages across the world, just as it is at phrase-level (Roeper et al. 2002; Bisetto 2010). The standard assumption is that left-branching recursive compounds (e.g. [[student film] society]) are more productive than right-branching recursive compounds (e.g. [student [film society]]) (Mukai 2008, 2017; Tokizaki 2011; Pöll 2015). However, this assumption has hardly ever been tested empirically in more detail. Using Corpus of Contemporary American English and National Web Japanese Corpus for Japanese and native speaker's judgements on the semantic interpretation of the data, we found that the prediction is borne out. In other words, recursive compounds are preferably interpreted as left-branching. After presenting representative data from each language I will propose that left-branching recursive compounds [[A B] C] are easier to parse, since a constituent can be formed earlier than in [A [BC]] structures (Pöll 2015).

Keywords: *recursion*, *left-branching recursive compounds*, *right-branching recursive compounds*, *parse*, *corpus study*

1. Introduction

The standard assumption is that LEFT-BRANCHING RECURSIVE COMPOUNDS [[AB] C] are easier to parse, since a constituent can be formed earlier than in RIGHT-BRANCHING RECURSIVE COMPOUNDS [A [BC]] (Mukai 2008, 2014; Pöll 2015). However, this has hardly been tested empirically. In this paper, the main aim is to test the hypothesis that right-branching recursive compounds are more restricted in frequency in English and Japanese, with the assumption that the higher the token frequency of word the higher the productivity of the word is (Baayen & Lieber 1997). At the end of this paper, some potential reasons behind the restriction of right-branching recursive compounds (Pöll 2015) will be discussed.

RECURSION is said to be a fundamental property of human language that potentially differentiates it from both other human cognitive domains and known communication systems in animals (Hauser et al. 2002; Corballis 2011). So, to reveal the characteristics of recursion at word-formation can reveal some aspects of human language. Recursion has been questioned in one of the Amazonian languages, Pirahã (Everett 2005). Everett (2005: 5) gives an example of embedded clauses in English where in Pirahã it is expressed without embedding. However, this question will not be discussed in this paper, as this is not the focus of the paper.

Before discussing the main objectives of the paper, let us define what recursion is. Summarising definitions of recursion by several authors (Chomsky 1965; Hauser et al. 2002; Roeper 2007; Bisetto 2010; Corballis 2011; Ralli 2013), the author defines recursion as a phenomenon of cyclic fashion to create sentences, phrases or words, as complex or long as we like. Here, complex means embedding of phrases or words within phrases or words of the same kind. In principle, it is possible to construct limitless embedding structures in human language, at least, within the limitations of one's memory and processing capacity. At word level especially compounding is a word formation process where recursiveness is more widely attested.

Roeper (2007) gives examples at each level, phrase, clause and word levels. They are presented in the examples below (1)–(3).

(1) Phrase level
  a. Possessive: John's friend's car's motor
  b. Prepositional phrase: in the kitchen, in the cabinet, in the corner
  c. Conjunction: and I came and I saw and I conquered, John and Bill and Susan

(2) Clause level
  a. Infinitive: John wants to start to go to sing
  b. Finite: Mary thinks I think you think she did it

(3) Word level
  a. Prefixation: re-re-read, anti-anti-missile
  b. Adjective: big, black, strange bear
  c. Compound: student film group festival

(Roeper 2007: 108)

As (1)–(3) show, at phrase level, clause level and word level, it is possible to have embedding of phrases or words within phrases or words of the same kind. For example, (3c) shows that four nouns are merged together to create a word. In addition, unlike at phrase level, it is impossible to have another noun within this compound, or otherwise the resulting compound would be a completely different word. In other words, recursion of compounding obeys the Lexical Integrity Principle (Bresnan & Mchombo 1995). The Lexical Integrity Principle disallows insertion of an extra element inside. Thus, recursive compounds are more like words in this sense.

With the definition of recursion in mind, it is also important to define left-branching and right-branching recursive compounds. A left-branching recursive compound is one where a two-member compound is formed and another head is merged at the right-hand side, whereas right-branching is where a noun is added to modify a two-member compound. The resulting interpretation is different according to the branching direction (Roeper et al. 2002). A typical example of right-branching and left-branching recursive compound is as follows.

(4) student (*the) film committee [= 'student committee on the films']

(5) student film (*the) committee [= 'the committee on student films']

(Roeper et al. 2002: 2)

Roeper et al. argue that in both (4) and (5), overt functional heads, such as determiners, are disallowed. (4) is a right-branching recursive compound.[1] The two-member compound *film committee* is modified by *student*, whereas (5) is a left-branching recursive compound. The two-member compound *student film* is a modifier for the noun *committee.* The interpretation of the whole compound (4) and (5) is provided in the brackets.

According to Berg (2006: 197), compounding is a quite productive word-formation process in English, and its productivity extends not only to the semantic

[1] Bisetto (2010: 15) argues that recursion is said to be distinguished from iteration. What we call right-branching recursive compounds are not recursive compounds, because they are just repetition with the same head. However, in this paper, I argue that right-branching compounds are also recursive compounds, since at phrase level and clause level it is possible to have recursion of right-branching structure in many languages.

relationships holding between the constituents of compounds, but also to the number of compounds. The example (3c) given by Roeper (2007) is a representative example of this type of compounds. In this paper, this is assumed to be true, not only in English but also in Japanese. Some more examples of four-member compounds in English, some taken from Berg and others from my observations (7)–(9) in English and those (10)–(13) in Japanese.

(6) a. child language acquisition research
b. child language acquisition research group
c. child language acquisition research group member
(Berg, 2006: 198)

(7) timeshare property ownership[2]

(8) UN Security Council Resolutions

(9) muscle contraction mechanism calcium

(10) nihongo-kyouiku-nouryoku-kennteishiken
Japanese-teaching-competency-examination
'Japanese Language Teaching Competency Examination'

(11) omelet rice kaido project
omelet rice highway project
'Omelet-Rice Highway Project[3]'

(12) kakuheiki-haizetsu-kokusai-campaign
nuclear-abolishment-international-campaign
'Treaty on the Prohibition of Nuclear Weapons'

(13) Kikoh-hendoh-wakugumi-johyaku
Climate-change-framework-treaty
'United Nations Framework Convention on Climate Change'

The data above clearly shows that recursivity of compounds with more than four members is also observed in English and Japanese. However, this paper focuses on three-member compounds. Ambiguity of four-member compounds is in need of further research.

This paper is organised in the following way: the next section will explain a corpus study of the compounds with native speakers' judgements of each language. This section is followed by a discussion on some potential reasons for restriction of right-branching recursive compounds. In the conclusion, the findings of the studies will be summarised.

2. Corpus study of recursive compounds

As discussed in the previous section, right-branching recursive compounds are more restricted than left-branching recursive compounds in English and Japanese. In this

[2] This data is from *British National Corpus*.
[3] This is a name of a project conducted by a village, Hidaka, in Kochi Prefecture, Japan.

section, I will first present data in these languages from my observations. This is followed by the corpus study of recursive compounds with native speakers' judgements.

2.1 *Recursive compounds in English and Japanese*

Recursive compounds can be observed in languages across the world, and English and Japanese are not exceptions. They are usually constituted of three nouns and denote product names, names of organisations, descriptions of phenomena, restaurant menus. They are also productively produced by children. An alternative strategy for the speakers would be using the corresponding phrases. The following examples are taken from the literature and my observations in everyday life in English and Japanese.

(14) a. [[air traffic] control]
b. # [air [traffic control]]

(15) a. [[student film] group]
b. [student [film group]]

(16) a. [[peanut butter] sandwich][4]
b. # [peanut [butter sandwich]]

(17) a. [[toilet paper] tissue]
b. [toilet [paper tissue]][5]

(18) a. [[logpile] house]
b. *[log [pile house]]

(19) a. [[underground] house][6]
b. *[under [ground house]]

(20) a. [[potato chip] plant]
b. # [potato [chip plant]]

(21) a. [[student film] committee]
b. [student [film committee]]

(Roeper, et al. 2002: 2)

(22) a. [mail [delivery service]]
b. [[mail delivery] service]

(23) a. [rail [network]][7]
b. *[[rail net] work]

The examples (a) are left-branching recursive compounds, whereas (b) are the corresponding right-branching recursive compounds. The original examples found are left-branching recursive compounds in (14)–(21), and (22)–(23) are right-branching ones. The examples (14)–(20) are preferably interpreted as left-branching, not right-

[4] A menu in a café and Google UK.
[5] My observation – toilet.
[6] (19) and (20) are from the children's book *The Gruffalo* (Donaldson 1999)
[7] (22) and (23) are from Google UK.

branching, by native speakers. In addition, the lexicalised compounds such as *logpile* or *network* cannot be separated into two different words. These examples show that there are more left-branching recursive compounds than right-branching recursive compounds in English.

Let us observe some phonological characteristics of these recursive compounds. According to the Compound Stress Rule (Lieberman & Prince 1977), the branching-direction is responsible for stress assignment in NNN compounds. In left-branching recursive compounds, in (14a)–(21a) the left most constituent is assigned highest prominence whereas in (14b)–(21b), (22a) and (23b), the second constituent is the most prominent one. For example, (14a) the stress is placed on *air*, whereas in (22a), *delivery* has the most prominent stress.

In addition, the children's data from CHILDES are all left-branching recursive compounds, not right-branching recursive compounds. The following are some representative examples. From the interpretation, the brackets on each compound are added to show the branching direction.

(24) a. [[Christmas tree] cookie] = 'a cookie shaped of Christmas tree'
b. [[peanut butter] sandwich] = 'a sandwich with peanut butter'
c. [[baby doll] napkin] = 'a napkin for a baby doll'
d. [[nursery school] book] = 'a book at nursery schools'
(Roeper, et al. 2002: 3)

What about in Japanese? Left-branching recursive compounds are more productive than right-branching recursive compounds, like in English.

(25) a. [[doitsu bungaku] kyoukai]
Germany literature association
'association for German literature' (Nishiyama 2015: 79)
b. [doitsu [bungaku kyoukai]]
'literature association in Germany' (Nishiyama 2015: 79)

(26) a. [[nise danuki] shiru]
fake badger soup
'soup with fake badger (either chicken or fox)' (Tokizaki 2011: 5)
b. [nise [takuki ziru]]
'plastic model of badger soup (presented in restaurant for customers)' (Tokizaki 2011: 5)

(27) a. [[anpan man] myuuziamu]
Anpan man museum
'museum of a character Anpanman'[8]
b. *[anpan [man myuuziamu]]

(28) a. [[ushiro muki] kaidan]
back direct-FINITE stairs
'stairs for people going backward'[9]
b. # [ushiro [muki kaidan]]

[8] This museum is in Kochi, Japan.
[9] My son produced this, when he was 3 years old, walking backward on a staircase.

(29) a. [[kodomo hon] kurabu]
child book club
'club for children's book'
b. # 'book club for children' (cf. kodomo no hon kurabu, with GEN *no*)

(30) a. [bikini hibaku] sosho]
Bikini hibaku lawsuit
'lawsuit for survivor of hydrogen bomb test at Bikini atoll'[10]
b. #'hibakusha lawsuit for Bikini atoll'

The above examples of Japanese recursive compounds collected from observations and the literature on compounding also show that left-branching recursive compounds are more productive than right-branching recursive compounds, both in number and interpretation. Let us consider some characteristics of these recursive compounds. Note that the contrast between (25a) and (25b) is not only the interpretation, but also the accent. According to Nishiyama (2015: 79), (a) has word accent on the vowel of *kyo*, since the whole compound is phonologically one word. In contrast (b) has phrasal accents on both *do* and *kyo*. So, the right-branching phonologically behaves more like a phrase. Similarly, (26a) and (26b) have different accents.

Notice that in the example (26), when the interpretation is left-branching (26a), the initial sound of *tanuki* /t/ becomes voiced when combined with another word, whereas in (26b), the initial sound of *soup* /s/ becomes voiced. This is a typical example of RENDAKU, Japanese sequential voicing. This phenomenon appears in the initial sound of the second constituent of a compound, so sequential voicing is a sign of a compound word.[11]

In addition, it is possible to have the right-branching interpretation, represented in (29b), not only the left-branching, represented in (29a). However, it is better to have the genitive case marker between the first and second constituents. In summary, this sub-section has shown that right-branching are more restricted than left-branching recursive compounds from my own observations and the literature on compounding in English and Japanese. It is not impossible to have a right-branching recursive compound, but it is somehow more restricted, as show in this sub-section. In addition, some phonological characteristics of recursive compounds were observed.

2.2 *Corpus studies in English and Japanese*

2.2.1 *Previous studies*

Kösling & Plag (2009) conducted an empirical study of phonological characteristics to see if NNNs overall conform to the *Compound Stress Rule's* (Lieberman & Prince 1977) predictions or not (See §2.1 for more details). To test if this rule is right, an empirical study was conducted. When taking acoustic data from the Boston University Radio Speech acoustic data, out of 448 NNN compounds, 326 were classified as being left-branching, 122 as right-branching. Their aim of the study was not to see the difference of productivity of left-branching in contrast to right-branching recursive

[10] Politics/Economics, Kochi Newspaper, 21st July, 2018.

[11] Otsu (1980) argues that Rendaku applies only to elements that are on a right branch at the lowest level of compound structure (Right Branch Condition, Otsu 1980). For example, (26a), *[[nise tanuki] shiru]*, the word *tanuki* is at the lowest level of compound structure. Here, the initial sound /t/ becomes voiced according to the Right Branching Condition. In contrast, (26b), *[nise [tankuki shiru]] shiru* is at the lowest level of compound structure. So, the initial sound becomes voiced.

compounds. However, from their sampled data, it is possible to see left-branching recursive compounds are more frequent than right-branching.

Another study of triconstituent was conducted by Lauer (1995) and it was found that left-branching recursive compounds are twice as frequent right-branching recursive compounds in English. He also proposes that compounds with more than three nouns are syntactically ambiguous. From these studies, phonological characteristics and systematic computer analysis were conducted. However, the productivity of recursive compounds in English and Japanese has not been conducted. In addition, the interpretations are assumed to be correct and have not been empirically tested by native speakers of English. For these reasons, the present study needs to be conducted.

2.2.2 *The National Web Japanese Corpus and the Corpus of Contemporary American English*

The Japanese data used in this study is taken from the *National Web Japanese Corpus* (NWJC). The *National Institute for Japanese Language and Linguistics*, Japan (NINJAL) has compiled a ten-billion-word scale Japanese web corpus named *NINJAL Web Japanese Corpus* (hereafter NWJC) (Asahara et al. 2014).

The *National Web Japanese Corpus* was chosen for this study because of the following reasons. First, the corpus contains data from a number of resources, including the product names, names of organizations, descriptions of phenomena, restaurant menus, which we expected to contain a fair number of NNN compounds. Second, it is a representative corpus of Japanese. Another study of recursive compounds in Japanese was conducted by Koyama (2009). Using a Japanese domain corpus, they obtained hierarchical relations of recursive compounds. The present study explores interpretations and structures of recursive compounds in Japanese, more than specialised terms.

For the English data, the *Corpus of Contemporary American English* was used. 100 million words and which is freely available from Prof. Mark Davies's website, among other online services. The corpus was used for this study, because this is the largest corpus of American English and contains a number of samples, including news genres and many others (§2.1), which are expected to have recursive compounds. It was not possible to find tags for recursive compounds, so I asked the founder of the corpus, Prof. Mark Davies to get NNN strings and to rank them in order.

The 40 highest frequency words were the data for the experiment. The researcher conducted a control experiment for each language to see if the data collected were words that average educated speakers would be familiar with. For the first aim of this paper, the following experiment was conducted.

2.2.3 *Procedure*

For the experiment for English and Japanese recursive compounds, eight monolingual native speakers of English and Japanese, aged between 30 and 80, were asked to provide their intuitive interpretation for each target NNN string. For English, all the participants grew up in the UK and five of them had been living there all their lives and three had left the UK to live in another country and one in Japan for the last 5 years at the time of the experiment.[12] For Japanese, all grew up in Japan and three had left Japan

[12] The participants are British English native speakers, not American English. This is because the researcher has easier access to British native speakers. As discussed later, the place where the participants grew up or lived does not seem to affect their knowledge of the target NNN strings. However, for future research it would be interesting to see the difference in the interpretation between the regional differences of the participants.

to live in another country. The researcher conducted pre-study on the participants who had lived abroad, and it was clear that the place where they had lived did not affect their interpretation. However, it was made sure that they all had at least higher education and/either read newspapers regularly to have enough data for their interpretations. Before the experiment, the participants were asked to read the instructions provided to them on a sheet of paper. During the experiment they were given as long as they needed to write the interpretations after they had read the word. Depending on the subjects' thinking and writing speed, each session took between 15 and 25 minutes. They had time to ask questions to the researcher if necessary. In addition, they were asked to write down their immediate interpretation for each word when they first read the word. In total 320 interpretations for each language were collected (40 NNN strings x 8 native speakers).

After obtaining the data from the participants, the researcher observed their interpretations and paraphrasing carefully, and divided them into right-branching and left-branching interpretations. The intuition of the researcher was used for the categorisation. Right-branching is where a two-member compound is modified by a modifier, whereas left-branching is where a two-member compound modifies another noun. For example, for the NNN string of *vanilla ice cream*, the participants' interpretation was 'vanilla-flavoured ice cream', so it is a right-branching recursive compound. The compound *ice cream* is modified by *vanilla*.

In summary, left-branching recursive compounds and right-branching recursive compounds were observed. In addition, it was summarised that left-branching recursive compounds are more frequent than right-branching recursive compounds. Next, the details of the present study were discussed including the descriptions and reasons for using the two corpora. The next section will be the results and discussions of the study.

3. Results and Discussions

3.1 *Results of English native speakers' judgements*

In both English and Japanese, the forty NNN strings were tested for native speakers' judgements. Japanese left-branching recursive compounds are 9 times as productive (frequent) as right-branching in terms of the interpretation given by the participants. In English, left-branching recursive compounds are 3 times as productive as right-branching in terms of interpretation. Table 1 summarises this. 1 percent of the Japanese participants and 4 percent of the English ones said they did not know.

Table 1: Interpretations of the data (%)

	Japanese	**English**
Right-branching	11	20
Left-branching	88	76
I don't know	1	4

The first aim of the paper: the hypothesis that right-branching recursive compounds in Japanese and English are more restricted in frequency than left-branching is borne out. As discussed in §1, this is based on the assumption that Baayen & Lieber (1997) are right in that the higher the token frequency of word, the higher the productivity of the

word. The 40 highest frequency NNN strings from these corpora were interpreted as left-branching more frequently. To show this is true in Japanese, typical examples of Japanese with the native speakers' interpretations are shown below.

(31) [[omu raisu] bento]
omelette rice lunchbox
'lunch-box which only has omelette rice'

(32) [[zidoh hukushi] center]
children welfare center
'center for children's welfare'

(33) [[zinken hooritsu] sohdankai]
human-rights law advisory-meeting
'advisory meeting for human rights law'

(34) [[zidoo kurabu] setsumeikai]
child club orientation
'orientation for children's club'

(35) [[nihon bungaku] kyookai]
Japan literature association
'association for Japanese literature'

(36) [shinbun [yoron choosa]]
newspaper public-opinion research
'public opinion research organised by newspaper company'

All the eight participants interpreted the data in the examples as represented in the brackets. Two of the participants said they did not know two of the target strings. However, these data clearly show that the participants prefer interpretation of left-branching recursive compounds than that of right-branching recursive compounds. This is especially true when the NNN strings are ambiguous, like (34) and (35). For example, (34) can be interpreted as 'club orientation for children'. This string might be more easily interpreted as 'orientation for children's club', since the term *zido club* is used recently to describe an organisation for school children before/after school to be minded by city-employees. So, *zido club* is now a compound in native speakers' lexicon. Similarly, (35) can be interpreted 'literature association in Japan' as well as 'association for Japanese literature'. However, all the participants preferred the latter, the left-branching interpretation.

For English native speakers' judgements, typical examples are represented below.

(37) [health care] reform]]
'Health care reform can occur when government policies call for change in practices and ideas surrounding a nation's health system'

(38) [world health] organization]]
'an organization to help the health of the people of the world'

(39) [public opinion] poll]]
'A system of gathering information from the public that determines a national viewpoint on topics through local voting system'

(40) [reader service] card]]
'a notification or advertisement in a library'

(41) [[vanilla [ice cream]
'a dessert made from blending and freezing cream, milk, eggs, sugar, and vanilla'

The interpretation written underneath for each string is the same for all the participants, although some of the wordings are different. Two of the participants said they did not know and were not sure about the interpretations for two of the strings. Interestingly, in English, names of organisations are observed to have the preposition of in-between the constituents. This is more so than in Japanese with a genitive case marker. However, from the observations and the corpus, recursive compounds are not unproductive. In fact, it is more productive and observed, at least in the corpus used for this study.

It is important to note the following fact. In (41) it is easier to interpret *ice cream* as a compound and vanilla is its flavour than cream of vanilla-ice. There are some examples like this, where it is easier to have the interpretation that the participants wrote. However, the main point here is that the frequency of left-branching interpretation by the native speakers is higher than that of right-branching. So the researchers hypothesis for English is borne out.

3.2 *Potential reasons for restrictions*

In the previous sub-section, it was found that in both English and Japanese, there are more left-branching recursive compounds than right-branching recursive compounds with native speakers' judgements. Let us discuss some of factors that could be responsible for restricted recursion in right-branching recursive compounds in contrast to left-branching recursive compounds. The factors discussed by Pöll (2015) are represented in this paper.

First, he cites Haider (2001) who argues that the direction of headedness, i.e. left-headedness in Romance languages vs. right-headedness in Germanic languages is the reason behind productivity of recursive compounds.

(42) a. French: *[poisson [chat [bébé]]]
'lit. fish cat baby; baby cat fish' (Haider 2001: 1)
b. Italian: *[gatto [capo stazione]]
'lit. cat head station; stationmaster cat' (Pöll 2015: 9)

Haider (2001) and Pöll (2015) argue that in Romance languages, it is not possible to translate into English the counterpart of the data, fish cat baby, for example. In (42a) and (42b), the left-most constituent, *poisson* or *gatto*, is the head of the whole string. However, it is not possible to have these strings in these languages, and Haider argues this is due to the left-headedness of these languages. However, this reason, as also Pöll argues, is not valid when one looks at other languages with left-head compounds, like Vietnamese (Dang 2013: 21) or Malay (Post: personal communication). Pöll also argues that the above examples are in fact productive in the languages.

Another factor Pöll (2015) discusses is the existence of a linking element. This reason is argued by Mukai (2008). In Swedish, according to Josefsson (1997), a linking element needs to exist in left-branching recursive compounds, or otherwise the strings would be ungrammatical. Some typical examples are represented below.

(43) a. [barn [bogklubb]]
child book-club
'book club for children'
b. [[barnbok]-s-klubb]
children-book club
'club for children's book'

The example above shows the difference in the interpretation, and to have the left-branching interpretation, *club for children's book*, it is obligatory to have a linking element. With this in mind, Mukai (2008) argues that the existence of a linking element is the reason behind the productivity of left-branching recursive compounds in English, Japanese and Mainland Scandinavian. However, Pöll argues that this is not true, since there is no linking element in English or Japanese. In Norwegian or Danish, the element is not obligatory[13]. In addition, in Dutch or German where the linking element does exist in recursive compounds, it is not obligatory either. As a result, the existence of a linking element does not seem to be true for the reason, either.

In this paper, the fact that processing is the most possible reason for the restriction, at least in English and Japanese, is argued. According to Pöll, compounds of the type [[AB] C] are easier to parse, since a constituent can be formed earlier than in [A [BC]] structures. In left-branching compounds, the speaker forms a constituent out of adjacent roots earlier than in right-branching compounds, where a constituent cannot be formed until the last root is pronounced (Hawkins 1990; Sugioka 2008). For instance, upon hearing or speaking it is easier for the speaker/hearer to form the AB as a two-member compound (left-branching recursive compound).

In this section, from the study of the corpus with native speakers' judgements in Japanese and English, left-branching recursive compounds are more frequently interpreted than right-branching recursive compounds. Thus, the hypothesis is borne out for this empirical study. In addition, the author argued that Pöll's reason of processing is the most possible behind the restriction of right-branching recursive compounds in English and Japanese.

5. Conclusion

The aim of this paper has been to test the hypothesis that left-branching recursive compounds are more easily interpreted than right-branching recursive compounds in English and Japanese. To test the hypothesis, empirical studies with corpus and native speakers' judgements were conducted. From the studies, it was clear that the hypothesis is borne out, i.e. left-branching recursive compounds are more preferably interpreted than right-branching recursive compounds. This means that the two-member compound can be a modifier of a noun. It is possible to have another modifier for a two-member compound (right-branching), but this is less frequent. However, when presenting the data to native speakers for their judgements of interpretation, it is necessary to consider the influence of already-existing/lexicalized compounds (e.g. *vanilla ice cream*, where

[13] I would like to thank Professor Anders Holmberg for his comments on this subject.

ice cream is already a lexicalized compound in native speakers' minds).

In addition, some potential factors for the restrictions of right-branching recursive compounds have been discussed along with Poll's (2005) reasons, and the reason is that it is easier for the speaker/hearer to parse left-branching recursive compounds, just like Pöll argues.

However, it is necessary to consider the following facts to prove that the reason of parsing is right. First, more data needs to be collected with native speakers' judgements in English and Japanese. In addition, it is important to consider the difference of productivity of recursive compounds between different languages across the world.

Acknowledgements

I would like to thank Professor Mark Davies for obtaining the data from the corpus. I also would like to thank the eight native speakers of Japanese and English. All the errors are mine.

References

Asahara, Masahara & Maekawa, Kikuo & Imada, Mizuho & Kato, Sachi & Konishi, Hikari. 2014. Archiving and Analysing Techniques of the Ultra-large-scale Web-based Corpus Project of NINJAL, Japan. *Alexandria* 2(12). 129–148.

Baayen, Harald & Lieber, Rochelle. 1997. Word frequency distributions and Lexical Semantics. *Computers and Humanities* 30. 281–191.

Berg, Thomas. 2006. The internal structure of four-noun compounds in English and German. *Corpus Linguistics and Linguistic Theory* 2(2). 197–231.

Bisetto, Antonietta. 2010. Recursiveness and Italian compounds. *SKASE Journal of Theoretical Linguistics* 7(1). 7–35.

Bresnan, Joan & Mchomo, Sam. 1995. The Lexical Integrity Principle: Evidence from Bantu *Natural Language & Linguistic Theory* 13(2). 181–254

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Corballis, Michael. 2011. *The Recursive Mind: The Origins of Human Language, Thought, and Civilization.* Princeton: Princeton University Press.

Dang, Zimmer. 2013. *Does prosodic word recursion cause phonetic initial strengthening?* California State University, Fresno. (MA dissertation.)

Donaldson, Julia. 1999. *The gruffalo*. Illustrated by Scheffler, Axel. Macmillan: Macmillan Children's books.

Everett, Daniel. 2005. Cultural constraints on Grammar and cognition in Pirahã. *Current Anthropology* 46(4). 621–646.

Haider, Hubert. 2001. Why are there no complex head-initial compounds? In Schauer-Wolles, Chris & Rennison, John & Neubarth, Friedrich (eds.), *Naturally!*, 165–174. Torino:

Rosenberg & Sellier.

Hauser, Marc D. & Chomsky, Noam & Fitch, W. Tecumseh. 2002. The Faculty of Language: What is it, Who has it, and How did it Evolve? *Science* 98.1569–1579.

Hawkins, John. A. 1990. Parsing theory of word order universals. *Linguistic Inquiry* 21. 223–261.

Josefsson, Gunlög. 1997. *On the Principles of Word Formation in Swedish*. Lund: Lund University Press.

Kosling, Kristina & Plag, Ingo. 2009. Does branching direction determine prominence assignment? An empirical investigation of triconstituent compounds in English. *Corpus Linguistics and Linguistic Theory* 5(2). 201–239.

Koyama, Teruo. 2009. Composite Term Extraction from Japanese Text. *Joho Chishiki Gakkaishi* 1(4). 306–315.

Lauer, Mark. 1995. Corpus statistics meet the noun compound: Some empirical results. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, 47–54. Stroudsburg, PA: Association for Computational Linguistics.

Liberman, Mark & Prince, Alan. 1977. On stress and Linguistic Rhythm. *Linguistic Inquiry* 8(2). 249 –336.

Mukai, Makiko. 2008. Recursive compounds. *Word Structure* 1(2). 178–198.

Mukai, Makiko, 2017. Semantic characteristics of recursive compounds. In Sheehan, Michelle & Bailey, Laura R. (eds.), *Order and structure in syntax II: Subjecthood and argument structure*, 285–288. Berlin: Language Science Press.

Nishiyama, Kunio. 2015. Nihongo no kukozo (Structure of phrases in Japanese). In Nishihara, Tetsuo & Shinichi, Tanaka (eds.), *Gendai no keitairon to onseigaku・oninron no shiten to ronten* [Aspects and discussions in current morphology and phonetics-phonology], 78–95. Tokyo: Kaitakusha.

Otsu, Yukio. 1980. Some aspects of rendaku in Japanese and related problems. In Otsu, Yukio & Farmer, Anne (eds.), *Theoretical issues in Japanese linguistics* (MIT Working Papers in Linguistics 2), 202–278. Cambridge, MA: MIT Department of Linguistics and Philology.

Pöll, Bernhard. 2015. Restricted recursion in N-N compounding: some thoughts on possible reasons. *Linguistische Berichte* 242. 141–163.

Ralli, Angela. 2013. Compounds and its locus of realization: Evidence from Greek and Turkish. *Word Structure* 6(2). 181–200.

Roeper, Thomas. 2007. *The Prism of Grammar: How child language illuminates humanism.* Cambridge, MA: MIT Press.

Roeper, Thomas & Snyder, William & Hiramatsu, Kazuko. 2002. Learnability in a Minimalist Framework: Root Compounds, Merger, and the Syntax-Morphology Interface. In Lasser, Ingeborg (ed.), *The Process of Language Acquisition: Proceedings of the 1999 GALA Conference*, 25–35. Frankfurt: Peter Lang Verlag.

Sugioka, Yoko. 2008. Remarks on asymmetry and recursion in compound formation. In Sano, Tetsuya & Endo, Mika & Isobe, Miwa & Otaki, Koichi & Sugisaki, Koji & Suzuki, Takeru. (eds.), *An enterprise in the cognitive science of language: A festschrift for Yukio Otsu*, 65–78. Tokyo: Hituzi Shobo.

Tokizaki, Hisao. 2011. The Nature of Linear Information in the Morphosyntax-PF Interface. *English Linguistics* 28 (2). 227–257.

Makiko Mukai
Department of Cultural Studies
University of Kochi
2-22 Eikokuji
Kochi 780-8515
Japan
mukai@cc.u-kochi.ac.jp

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/03.pdf. ISSN 1336-782X.

Productivity of form and productivity of meaning in N+N compounds

Elizaveta Tarasova, IPU New Zealand

This paper reports on the results of a corpus-based study of English N+N compounds that are hyponyms of their rightmost elements (Bloomfield 1933: 235), e.g. library book, hen house, coffee cup, *with the aim of investigating the relation between productivity, on the one hand, and, on the other hand, the structural pattern of N+N compound formation (with a particular reference to the relationship between the elements of the compound). The discussion of the results addresses the question of whether productivity of a noun in forming a paradigm of compounds (constituent word family) could be connected with the productivity of the relation realised in the compound paradigm containing this noun. This question has its roots in the ongoing discussion of the role of analogy in the formation of complex lexical items, and its investigation may shed light on the issue of analogy in both formal and semantic patterning.*

Keywords: *N+N compounds*, *productivity*, *constituent word family*, *analogy*, *semantic relations*

1. Introduction

Research into the formation of meaningful sequences consisting of two adjacent nouns (henceforth N+N), e.g. *door handle*, *pencil sharpener*, *biology research*, *team management*, *singer-poet*, etc. in English (as well as in other languages) has taken a new turn in the last few years with the focus of the study shifting from the argument on what formal characteristics make an N+N sequence a compound to the investigations on how the meaning of a compound is construed in the mind of an individual, and how the knowledge that native speakers possess is applied to form new sequences that follow the same pattern. This brings us to the question of morphological productivity, the question addressed in a large body of research in word-formation (e.g. Schultnik 1961; Aronoff 1976; Bauer 1983, 2001; Baayen 1992, 1993, 2003; Plag 1999; Rainer 2004, to name just a few).

Despite a considerable amount of work in the area of morphological productivity and the development of various tools (quantitative and qualitative) for measuring productivity (Baayen 1992, 1993; Baayen & Lieber 1991; Baayen & Renouf 1996; Bauer 2001; Hay 2003; Hay & Baayen 2005 Hay 2003; Fernández-Domínguez 2009, etc.), productivity in the formation of N+N compounds has not been analysed in much detail. Only a limited number of researchers focus on productivity in compound formation in different languages (e.g. Zwitserlood 1994; Krott et al. 1999; Hein & Engelberg 2017).

In this paper, I look only at one group of compounds, i.e. English subordinative N+Ns; however, the suggestions proposed here can probably be applied to other kinds of N+N sequences. What follows is the discussion of the issues of productivity in compound formation with the special focus on individual nouns when used as constituents of N+N compounds, with the aim of connecting the productivity of a noun as a compound constituent with the productivity of a relational meaning that a given constituent realises. Based on this, we can formulate two hypotheses that are tested in this study:

i. A compound constituent is used more productively in one position (modifier or head) but not the other.
ii. There is a connection between the family size of the compound constituent and this constituent's concentration on one semantic relation.

2. Limiting the scope of structures

One of the most heated ongoing debates in compounding research is devoted to the question of whether a sequence of two nouns is a unit of morphology (a lexeme) or a unit of syntax (phrase), since compounding borders on affixation and on syntax (Bauer 2017; ten Hacken 2017). In order to be able to distinguish between compounds and other phenomena, a number of different criteria have been offered and quite a few of them have been shown to have limited applicability in Bauer (1998, 2017) and Giegerich (2015). Another question in this regard concerns the importance of the ability to apply such criteria crosslinguistically, since, although it is not universal, compounding is widely spread in world languages (Štekauer et al. 2008). Ten Hacken (2017: n.p.g.), in his overview of different approaches to delimiting compounds from other phenomena, outlines the importance of the theoretical framework for the selection and formulation of criteria, because "[…] if, for instance, word-formation and syntax are strictly separated and compounding is in word formation, it is crucial to draw this borderline precisely". In the case of the present investigation, the borderline between syntax and morphology is not important, as the question of whether N+N compounds are words or phrases is not relevant for this discussion, which looks at the process of combining two noun concepts into a single meaningful unit (be it a complex word or phrase). In this research, I accept the opinion shared by some linguists (e.g. Warren 1978; Bauer 1978, 1983, 1998; Olsen 2000; Bell 2005, 2011; Bell & Plag 2012) that the evidence for a class of syntactic N+N constructions is limited. Here, the term *N+N compound* is applied (quite loosely) to any N+N construction that consists of two adjacent nouns, in which one modifies the meaning of the other, or where together they have a single meaning different from the meaning of either constituent individually. Proper names (e.g. *James Smith*), constructions with an appositive modifier (e.g. *lady Jane*), and sequences like *letter A*, *number nine* are excluded.

Thus, this investigation is concerned with English endocentric N+N compounds, in which a concept denoted by the lexical complex N1+N2 is an instance of the concept denoted by N2. In my research I follow the commonly accepted definition offered by Bloomfield (1933: 235), who identifies such compounds semantically, by means of hyponymy, i.e. the referents of such compounded structures are hyponyms of their head elements. For example, *apple tree* is a kind of tree, *cookie jar* is a kind of jar, and *sun glasses* are a kind of glasses. I focus on analytical structures, also called *root compounds* or *primary compounds* (Harley 2009: 139), in which the head element is not derived from a verb, and the whole sequence does not yield a Verb-Object interpretation, i.e. *synthetic compounds* in Bloomfield's (1933) terms, or *verbal-nexus compounds* in Marchand's (1969) terms. This means that expressions like *taxi-driver* and *language teacher* are not included into this analysis.

It is important to note a certain fuzziness in terms of categorizing compounds into different types. Fabb (1998: 67), in the discussion of the matter gives the example of *greenhouse* whose being endocentric or exocentric depends on whether you think of it as a kind of house or not. This makes it obvious how the interpretation (and also the opinion of the analyst) may completely change our understanding of whether one and the same

compound should be analysed as an endocentric or an exocentric one. Due to the limitations of space, I will not focus on the matters of exocentricity, as well as other issues regarding categorising nominal compounds into different types here and will refer to Bauer (2017) for a most comprehensive overview and discussion on this issue. In this study, I follow the assumption expressed in Bauer et al. (2013: 465) that an endocentric compound is one "[…] in which the compound as a whole denotes a subset of the head element of the compound".

3. Issues of productivity

Bauer et al. (2013: 451) state that English endocentric N+N compounds are the most productive kind of nominal compounds in English. By this they seem to consider that the pattern of forming such sequences is characterised by its *availability*, i.e. the scope of a word-formation rule, and *profitability*, i.e. the actual use of this rule in the process of forming new coinages (Kastovsky 1986: 586). The dichotomy between these two notions is discussed in Fernández-Domínguez (2015: 220), who demonstrates the inter-relation between availability and profitability of a word-formation rule and the ability of developing paradigmatic and syntagmatic relations in the language respectively. Consideration of both of these two notions is important for the synchronic analysis of a given word-formation pattern (in our case N+N compounding), since it allows the combination of the qualitative (availability) and quantitative (profitability) aspects of productivity.

The issue that seems to be overlooked in the research on productivity of compounds is productivity of nouns as constituents of compounds, and there are very few publications that look at the productivity of individual members of compounded structures in English. This question, however, is important for the discussion of how compounds are produced and what mechanisms are involved. A recent quantitative study by Hein & Engelberg (2017) looks at productivity values for nouns that are commonly used as constituents of German compounds, where the authors analyse such factors as *semantic proximity* and *frequency of the head noun* in order to account for the degree of productivity with which certain nouns are used as constituents of compound words.

The question of productivity of nouns as constituents of compounds is relevant for English too. One of the opinions presented in the literature is that the distribution of nouns between modifier and head roles is relatively equal in the BNC (e.g. Maguire et al. 2010). However, a closer look at the proportion of noun occurrences in combination by modifier and head, as shown in Maguire et al. (2010: 61), reveals that nouns belonging to such categories as substance, plant, location and etc. demonstrate the preference to be used as modifiers. Thus, we can see that on the one hand, nouns in general seem to be used equally frequently in both N1 and N2 positions. On the other hand, the semantic content of certain nouns may promote their use in only one of the roles.

A quick search in the corpus for one of the frequent nouns denoting substance, i.e. *water*, supports this observation. According to the data extracted from the BNC, *water*, when used as a modifier, occurs three times more often and in a wider variety of combinations than as a head noun. To be more specific, the modifier family for the constituent *water* contains 48 types (e.g. *water authority*, *water industry*, *water services*, *water board*, *water charges*, *water temperature*, *water condition*, *water garden*, *water shares*, *water park*, etc.), whereas the head family contains only 15 types (*sea water*, *surface water*, *bath water*, *salt water*, *waste water*, *rain water*, etc.).

Baayen's (2010) study reports that out of 2200 compound constituents, 710 were used only as heads, 902 only as modifiers, and 588 both as heads and modifiers. Despite the considerable number of nouns that occur both as heads and modifiers, the regression models presented in the study demonstrate significant correlations of productivity ranks of the constituents with degrees of constituent productivity. This also means that the prevalence for one position cannot be limited to modifiers. The corpus data for the word *problem*, for example, shows that it can be found in the head position in 116 compounds (*back problem*, *crime problem*, *money problem*, *business problem*, *quality problem*, *family problem*, *disease problem*, *city problem*, *attitude problem*, *acquisition problem*, *housing problem*, etc.), but its occurrence as a modifier is very limited, with only 10 compounds where it is used in this role (e.g. *problem behaviour*, *problem children*, *problem situation*, *problem area*, *problem families*, etc.). This means that its head family size is much larger than that of the modifier.

This puts Maguire et al.'s suggestion about the equal distribution of nouns between the modifier and head roles into question and calls for further investigation of the issue with a more detailed analysis of individual nouns that are used as elements of compounds. However, consideration of the semantics of N+N compounds and their connection to the elements' preferential use in either of the positions should also be considered. Looking at the issue of productivity from the position of the constituents comprising a compound as well as the relations that these constituents realise, may help shed light on the question of how compounds are formed.

4. Issues of semantics

It is a distinguishing feature of English N+N compounds that as linguistic forms, they combine two or more discrete lexemes into a semantic whole, despite the fact that there are no grammatical markers to indicate what the relation between these two (or more) parts is, or how this relation has been obtained. It has been noted that endocentric compounds that have a common constituent realise different relations. Bauer (1979: 45) provides the example of the word *pill*, which, when used in combination with different modifiers, realises different meanings, e.g. a *sleeping pill* CAUSES somebody to sleep, a *seasickness pill* is used FOR seasickness, and an *antihistamine pill* CONTAINS antihistamines. The same happens if we make changes to the first constituent, e.g. *citizen army* vs. *liberation army*; *riot police* vs. *peace police*, *jewellery box* vs. *bargain box*. However, on the level of structural representation, we are still dealing with the combination of two nouns and there are no formal markers to indicate the changes in meaning.

This ability of nouns to realise different relations when they are put together into a syntagm has been the central topic of discussion in the research on the semantics of compounds for over half a century. The question of the semantics of N+Ns has been considered in a number of theoretical frameworks bringing up a number of important issues, the analysis of which helps shed light on the nature of this phenomenon (Lieber & Štekauer 2009). However, there is still no unanimous opinion on how many of such relations there are in language or how to classify them (for a brief discussion of the existing problems see Bauer & Tarasova 2013: 1–3). The number of semantic relations in different classifications ranges from four (Granville Hatcher 1960) to over a hundred (Brekle 1976), with the most influential being Levi's (1978) classification. Though not without drawbacks, Levi's set of nine fundamental relations, aka recoverably deletable predicates (RDPs), provides a

manageable number of relations that can be applied to N+N compounds, as well as other nominal structures (Bauer & Tarasova 2013). Despite a number of problems with Levi's (1978) classification (see ten Hacken 1994: 44–49 and ten Hacken 2016: 4–5 for discussion), the practicality of Levi's set makes it the choice of convenience for this research as well, since it allows for demonstrating the trends important for the present discussion on the connection between productivity on the level of structural representation with the productivity on the semantic level. The list of relations with examples are given in Table 1.

Table 1: Levi's (1978) classification of RDPs

Meaning	RDP	Examples
N1 CAUSE N2	CAUSE1	*sex scandal*, *withdrawal symptom*
N2 CAUSE N1	CAUSE2	*tear gas*, *shock news*
N1 HAVE N2	POSSESSION1	*lemon peel*, *school gate*
N2 HAVE N1	POSSESSION2	*camera phone*, *picture book*
N1 MAKE N2	COMPOSITION1	*snowball*
N2 MAKE N1	COMPOSITION2	*computer industry*, *silkworm*
N2 USE N1	INSTRUMENT2	*steam iron*, *wind farm*
N2 BE N1	ESSIVE2	*island state*, *soldier ant*
N2 IS IN N1	LOCATION2	*field mouse*, *letter bomb*
N2 IS FOR N1	PURPOSE2	*arms budget*, *steak knife*
N2 IS FROM N1	SOURCE1	*business profit*, *olive oil*
N2 IS ABOUT N1	TOPIC2	*tax law*, *love letter*

As can be seen from Table 1, Levi's (1978) set of RDPs includes such central, highly abstract predicates as CAUSE, HAVE, MAKE, BE, USE, FOR, FROM, IN, ABOUT. The semantic relations in the compounds have a number marker, which signifies the directionality of the relation. Notably, at least three of the semantic relations (CAUSE, HAVE, MAKE) demonstrates bidirectionality, e.g. *tear gas* – 'gas CAUSE tears' vs. *drug deaths* – 'drug(s) CAUSE deaths' (Levi 1978: 76). The necessity for distinguishing between the directions of the relations is dictated by the difference in the readings of such compounds as *heat birth* and *problem children*, since in the first case N1 CAUSES/INDUCES N2, and in the second one N2 CAUSES N1. In the course of this study it was noticed that such oppositions are not limited to the three relations pointed out by Levi (1978: 76), and examples of compounds realizing contrasting directions of the same relation may require reconsideration of this issue, as shown in examples in (1)–(4).

(1) SOURCE: *photon energy* (N1 IS FROM N2) vs. *heart sounds* (N2 IS FROM N1)
(2) ESSIVE: *mansion house* (N1 IS N2) vs. *tower house* (N2 IS N1)[1]

[1] The distinction in the direction of the ESSIVE relation in compounds seems to be quite subtle and generally follows the principle: ESSIVE1 – every N1 is N2; ESSIVE2 – N2 which is N1. Looking at the above examples, we can see that every mansion is a house, which does not seem to be the case for every tower. However, this principle does not always seem to work. For example, for the compound *building business* the relation is

(3) TIME: *work day* (N1 IN/DURING N2) vs. *crisis decisions* (N2 IN/DURING N1)
(4) INSTRUMENT: *farm machinery* (N1 USE N2) vs. *wind farm* (N2 USE N1)

At the same time, it is necessary to point out that usually one of the directions is actively utilised, whereas the cases of the opposite direction can be marginal. For example, in the corpus collected for this research, compounds realising CAUSE2 relation occur only 22 times, whereas CAUSE1 is much more frequent, with 244 cases.

Another issue with Levi's (1978) classification concerns the necessity to distinguish between the SPACE and TIME relations (Levi views TIME as a subtype of SPACE). In the current research the relations of TIME and SPACE are separated and viewed as two distinct categories. The parallelism of the concepts of SPACE and TIME has been long discussed and the overall historical tendency of spatial expressions to develop temporal meanings is recorded across languages. In the cognitive approach, this parallelism is explained by Mapping Theory (Lakoff 1993; Lakoff & Johnson 1999; Boroditsky 2000; Radden 2003), according to which the abstract domain of time gets its structure from a more concrete domain of space. Studies on locative prepositions in English (Sandra & Rice 1995; Kemmerer 2005) suggest that language users clearly distinguish temporal meanings from spatial ones. The results of Kemmerer's (2005) research suggest that although the spatial and temporal meanings of prepositions are historically linked by virtue of the TIME IS SPACE metaphor, they can be represented and processed independently of each other in the brains of modern adults. It seems logical to suggest that the concepts of TIME and SPACE probably have separate entries in the mental lexicon and the same can be said about the corresponding semantic relations. Matlock et al. (2005) point out that although the domains of space and time share conceptual structure, with frequent use, mappings between space and time come to be stored in the domain of time and so thinking about time does not necessarily require access to spatial schemas.

5. Number of relations per noun

The next question addressed here is whether a given noun can realise any possible relation when combined with another noun to form an N+N compound, or whether the number of relations for a single noun is limited. According to data from psycholinguistic research, semantic relations by which the overall meaning of an N+N compound is characterised are conceptual rather than lexical (Gagné et al. 2009) and should be viewed as "independent representational units that can be utilised by various and dissimilar concepts" (Estes & Jones 2006: 89). This makes it possible to assume that in theory, any relation can be used with any noun, e.g. *bear paw* (POSSESSION), *bear cave* (LOCATION), *bear medicine* (PURPOSE), *bear scare* (CAUSE), etc. A number of researchers support this opinion, and argue that different interpretations are possible for a single compound (e.g. Lieber 1992; Coulson 2001). At the same time, the fact that that different relations differ in their degrees of availability for individual nouns has been demonstrated in some empirical studies. For example, in the experiment described in Štekauer (2005), the participants were presented with novel compounds and were requested to propose and rate all possible interpretations for these

defined as ESSIVE1 because it is possible to say that building (as industry) is a kind of business, but business is building is somewhat clumsy. The decision on the directionality of the ESSIVE relation was taken based on the more natural interpretation of a compound.

compounds. Interestingly, almost all compounds received only one reading that was dominant.

This tendency goes in line with predictions of Competition Among Relations In Nominals (CARIN) model (Gagné 2000; Gagné & Shoben 1997), which is based on the assumption that availability of different relations varies for different concepts. For example, when the word *mountain* is used in the modifier role, we are more likely to expect that the entity denoted by the head will be LOCATED near/around/in a mountain (e.g. *mountain cloud*), rather than telling us ABOUT the mountain (e.g. *mountain magazine*) (Gagné 2000). It has been often suggested that the modifier element has more weight than the head in determining the relational reading a compound receives (Gagné 2001; Gagné & Shoben 1997, 2002; Gagné & Spalding 2004; van Jaarsveld et al.1994). Speakers apply the knowledge of how the modifier is used in various combinations thus building a database of all possible relations that are associated with the modifier. The psycholinguistic evidence obtained by Gagné & Shoben (1997, 2002) suggests that the selection of a relation for a novel compound is affected by the speakers' linguistic and statistical knowledge about how the modifier is generally used in the language, which is directly connected to the modifier constituent word family (henceforth CWF) size. The same was noticed for compounds in German and Dutch. Krott (2009) claims that the modifier family appears to be more influential than the head family, and its influence is clearly seen in the choice of linking elements, which is shown to be determined by the semantics of the modifier rather than the semantics of the head. Based on this, Krott (2009) assumes that the semantic properties of the modifier element are more valuable for compound processing. Moreover, this importance is attributed not to the first element in the compound but to the modifier. Storms & Wisniewski (2005) performed a psycholinguistic study of compounds in Indonesian, a language in which the modifier constituent follows the head. Their results suggest that the position of the modifier does not influence compound processing. This means that the subcategorisation process and meaning construal takes place on the conceptual level and may not be connected with the structural peculiarities of the language.

While the arguments regarding the role of the modifier in the process of interpretation are more than persuasive, the same is not necessarily true for the process of creation of compounds. The number of works that focus on the process of compound formation is limited and the evidence collected in empirical studies is inconclusive at this stage. The present investigation follows the assumptions of the slot-filling models (Murphy 1988; Wisniewski 1997), which suggest that the process of compound formation is head-oriented. This means that the semantics of the head noun serves as the schema whose properties are responsible for attracting certain modifiers based on the slots that are more available for filling. The availability of slots is determined primarily by the semantic properties of the head noun and only then by other factors, including our previous linguistic and non-linguistic experience (Tarasova 2013).

6. Data collection

In order to investigate the issue of productivity of a noun as a constituent of a compound and the issue of productivity of the semantic relation realised by this constituent, a database of CWFs was collected. This section provides information on the stages of the data collection

process and limitations of the search, and presents the details of the statistical analysis that was performed to check the general trends in relation to the issues above.

500 examples of non-lexicalised nominal N+N compounds were picked from New Zealand printed media. The 500 examples were analysed and classified in accordance with Levi's (1978) set of RDPs; therefore, it was important to choose items that are semantically transparent and do not include constituents that are extensions of the meaning of existing lexemes by metaphor or metonymy. Established compounds were included as long as they satisfied these two criteria. Such caution was dictated by the necessity for consistency in the analysis of the semantic relations, as Levi's set of RDPs is aimed at non-lexicalised compounds (1978: 8); however, some of her examples are clearly institutionalised (in Bauer's (1983) understanding). The purpose of the analysis was not to define the relations between the constituents as precisely as possible (since the interpretation of the meaning of a compound is relatively subjective), but to show that similar patterns recur and to suggest the reasons why this might be the case. The constituents of compounds from the corpus were then checked for their frequency of occurrence in the BNC, and the ones whose frequency of occurring in the corpus (irrespective of whether they are used as single words or elements of compounds) was 1000 or more occurrences per 100,000,000 were considered candidates that could form the basis for collecting a corpus of CWFs. The frequency limitation was set in order to be able to collect a representative corpus of compounds in which a given word may occur either in the head or modifier position, and a higher frequency of occurrence of a noun in the corpus was viewed as a predictor of the noun's ability to form extended CWFs.

Overall, the results of the semantic analysis of the relations connecting the elements of compounds in the collected corpus showed that the distribution of the relations in compounds is not even, and some of the categories occur more often than others. Almost two thirds of all the compounds in the collected data set realised only three semantic relations: PURPOSE (34%), POSSESSION (16.2%) and LOCATION (14.4%). Such uneven distribution of the semantic categories is assumed to be natural for English endocentric N+N compounds in general, and some semantic categories seem to occur more frequently in certain nominal constructions (Bauer & Tarasova 2013).

Such an uneven distribution of semantic relations required stratification of the sample in order to include compounds from different categories. Proportionate allocation strategy was used; i.e. a sampling fraction (the ratio of the sample size to the size of the stratum) was picked in each of the strata that was proportional to that of the total collected corpus. 50 compounds were chosen from the set of 500 items for the corpus experiment.

100 different constituents comprising the selected 50 items served as the basis for extracting the paradigms of compounds, i.e. CWFs, in which each of the constituents occurs. The semantic relations realised in compounds comprising CWFs were determined based on Levi's (1978) classification. In the process of picking the sample, the distinction into the different directions of a relation was not made, since for some categories the number of nouns was not large enough to be included in the stratified sample. However, some recent works in compounding (e.g. Plag et al. 2007; Kunter 2011) demonstrate the necessity to extend Levi's list in order to make a distinction in the direction of the relation.

The search engine of *BYU-BNC: The British National Corpus* (Davies 2004–) was used to collect N+N combinations in which every single noun of the 100 nouns that constitute the selected compounds may occur. Individual nouns were typed into the *search* field and [nn*] was typed into the *collocates* field. The collocations were limited by one noun before and one noun after the searched item. Then, where there was one, the plural form of

the same noun was put into the search engine in order not to miss compounds in which the searched constituent in the head position does not occur in the BNC in the singular form.

The search of the BNC database was targeted and the examples were hand-picked in order to avoid the combinations that would not satisfy the criteria, as in examples like *century house* and *umbrella back* in (5) and (6), which either have a different bracketing structure (as in the first case) or are not compounds at all.

(5) so we found a minister who was very excited by the idea of marrying us in a seventeen [sic] **century house**, and in fact, on the morning of the wedding he was more worried about what he was wearing than what I was!

(6) Give me my **umbrella back**.

Since the current research is concerned with the constituent family size, rather than the constituent family frequency, there were no limitations concerning the frequency of the collected items. As this study is concerned with non-lexicalised compounds, a high frequency of occurrence of all individual compounds within a CWF cannot be expected; moreover, it is generally believed that lexicalisation is strongly associated with high frequency. At the same time, lower frequency compounds, especially compound hapaxes, i.e. the types that have only one token in a given corpus, provide important information about the development of the morphological pattern at the current time (Baayen & Lieber 1991; Baayen 1992, 1993; Ricca 2010), and, hapaxes in particular, are viewed as indicators of productivity.

Manual picking of the examples was also necessary to exclude synthetic compounds. It does not seem possible to set the limitations for these using the corpus tools, as the morphological criterion offered by Lieber (1992) does not always work. For instance, if we set the limitations on nouns with the agentive endings *-er* and *-or* in the head position, then compounds with head words like *professor* and *minister* will be excluded from the analysis. In the current research I do not follow Lieber's (1992) suggestion that all compounds in which the head element is a deverbal noun should be considered synthetic. In this research compounds are considered synthetic only if the modifier functions as a direct object of the derivational base of the head noun in the interpretation, e.g. *literature teacher* – teach literature, *company management* – manage a company. Combinations like *art competition*, *tourist information*, *state prosecutor*, *weight division*, etc. are not considered to be synthetic since they require relational interpretations. The so-called non-affixal deverbal compounds, i.e. NDVCs (Lieber 2010) like *history research*, *business talk* are included in the corpus if their interpretation involves one of Levi's (1978) semantic relations.

As a result of the data collection, 7332 compounds that make up 197 head and modifier constituent word families were collected for further analysis. Although the expected number of constituent word families is 200, three of the nouns (*matter*, *future*, *figure*) appear only in one position in N+N combinations in the collected corpus. These were still included in the statistical analysis. The semantic relations realised in the collected N+N sequences were analysed and the number of occurrences of each of the compounds in a given constituent word family was checked.

7. Data analysis

The final stage of the data collection involves statistical analysis in order to obtain a picture of the general trends.

7.1 *Productivity of form*

The first hypothesis tested through statistical analysis was aimed at answering the question of whether a compound constituent is used more productively in one position (modifier or head) but not the other.

In order to test this the collected corpus data were coded on the basis of a table that listed for each compound constituent the number corresponding to the family size of this constituent as it is used in the head position (famSizeN2) and in the modifier position (famSizeN1). Table 2 gives an overview of the distribution of family sizes. Family sizes of 0 occurred only once in N2 position, and only twice in N1 position (with each position having 100 different nouns).

Table 2: Distribution of family sizes

Constituent	Minimum	Maximum	Mean	Standard Deviation
N1	0	83	31.79	20.47
N2	0	116	45.95	30.26

To reduce the skewing of the distribution of the family sizes, and in order to alleviate possibly harmful effects of extreme values on our statistical models, the family sizes were log transformed (see Baayen 2008: 38). In doing this, we also follow the standard practice with compound family size measures in other studies (e.g. Bell & Plag 2012). Since some of the frequencies were 0, we added 1 to all family sizes before log transformation. The statistical analysis at this stage was carried out using the statistical package R (Pinheiro & Bates 2000; Bates et al. 2007; R Development Core Team 2011).

A linear regression model with the log family size of N1 as the independent variable and the log family size of N2 as the dependent variable was devised. The model with the best fit included a quadratic term for the independent variable and showed highly significant correlations between the terms for family size of N1 and family size of N2. Table 3 gives the coefficients of the model.

Table 3: Model coefficients for linear model showing the relationship between N1 and N2 family sizes. N1 = 98, Adjusted R2 = 0.3796

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	3.43574	0.42932	8.003	2.65e-12 ***
logFamSizeN1	1.07457	0.32638	3.292	0.00139 **
$(\text{logFamSizeN1})^2$	-0.30143	0.06094	-4.946	3.17e-06 ***

The graph in Figure 1 below shows for each noun the size of its family in N1 position and in N2 position. The vertical and the horizontal axes represent the logs of family sizes for constituents used as heads (logFamsizeN2) and modifiers (logFamsizeN1) respectively. The dots represent the observed values for nouns that are attested both in N1 and N2 position. The solid line is the regression line, with the 95 percent confidence interval given by the two dotted lines. The negative coefficient for the quadratic term gives us the downward slope of the regression line (see, for example, Baayen 2008: Ch 4). As can be seen from the graph, the dots are not very concentrated. However, the distribution and the shape of the regression line clearly demonstrate that the larger the head family is for a given noun, the smaller the modifier family is for the same noun. This means that the more often a given noun occurs as a head of an N+N sequence, the less often it occurs as a modifier. As can be seen from the graph, the effect is not very pronounced for smaller N1 family size and becomes more pronounced with rising N1 family sizes.

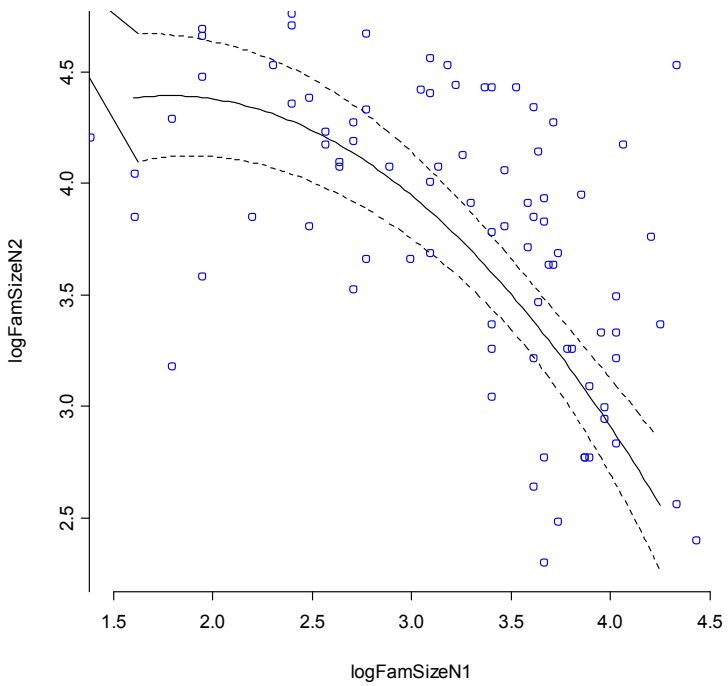


Figure 1: Relationship between family sizes

This means that it is possible to say that a noun when used as a constituent of a compound tends to be used more productively either as a modifier or as a head (but not both).

7.2 *Productivity of meaning*

Now I will turn to the next hypothesis, according to which we expect to see a connection between the family size and the constituent's concentration on one semantic relation. In order to test this, it is necessary to look at the two suggestions that it implies: (a) the constituent should demonstrate a preference for certain relations, or even one particular relation, over other relations; (b) this tendency should be stronger the larger the constituent family is.

To show that there is a preference for one semantic relation over the others within a given constituent word family, a measure called *instantiation index* is used. As a means to determine the instantiation index, the number of instantiations of different relations (DiffInstN1 and DiffInstN2) and the most frequent relation (HighestInstN1 and HighestInstN2) in each constituent family was calculated and added to the coded data table. Then the proportion of the highest instantiation among all family members was computed.

The diversity of meanings present in the family was factored in by multiplying the proportion of the highest instantiation in a CWF with the number of different relations attested in the family. For example, if we have a family size of 20 with eight different relations, and 10 compounds have the most frequent relation A, the index is (10/20)*8 = 4. If this family had only two different relations, the fact that 10 out of 20 have relation A is not so striking. This is exactly what we can see from the instantiations index, which will be much lower in this case: (10/20)* 2 = 1. So, we can assume that the higher the instantiations index is, the more concentrated the constituent family is on one relation.

In reference to suggestion (b), we should also expect the effect of the family size, i.e. the index should be higher for larger constituent families.

In order to test the above suggestions, the instantiation indices for constituents N1 and N2 were calculated and used for the analysis. The obtained measures were used for testing the family size effect. Two linear regression models (one for each constituent) were fitted in order to see if there is a connection between the family size and the constituent's concentration on one semantic relation. Log family size was used as an independent variable and log instantiation index as a dependent variable. Tables 4 and 5 below give the coefficients of the two models respectively.

Table 4: Model coefficients for linear model showing the relationship between family size N1 and instantiations of relations index for N1. N1 = 98, Adjusted R2 = 0.6503

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	0.34443	0.08427	4.087	9.06e-05***
logFamSizeN1	0.33816	0.02511	13.468	< 2e-16***

Table 5: Model coefficients for linear model showing the relationship between family size N2 and instantiations of relations index for N2. N2 = 99, Adjusted R2 = 0.4333

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	0.40984	0.11722	3.496	0.000713***
logFamSizeN2	0.27787	0.03189	8.715	7.95e-14***

As can be seen from the tables, the family size is a highly significant predictor in both models, which, together with the positive coefficients, supports the suggestion that the constituent's concentration on one semantic relation is stronger in larger constituent families. Note also that these models have a very good fit and can explain a large proportion of the variance (R^2 = 0.65 for N1, R^2 = 0.43 for N2).

Figures 2 and 3 below represent the findings in graphical form. The vertical axes in both graphs show the measures for instantiation indices of semantic relations in a constituent

family, and the horizontal axes show the log measures for the respective family size. The lines are the regression lines of the respective models.

The lines show that with increasing family sizes, the instantiation indices become higher. This also means that compounds with larger family sizes tend to be more concentrated on one relation than compounds with smaller constituent family sizes.

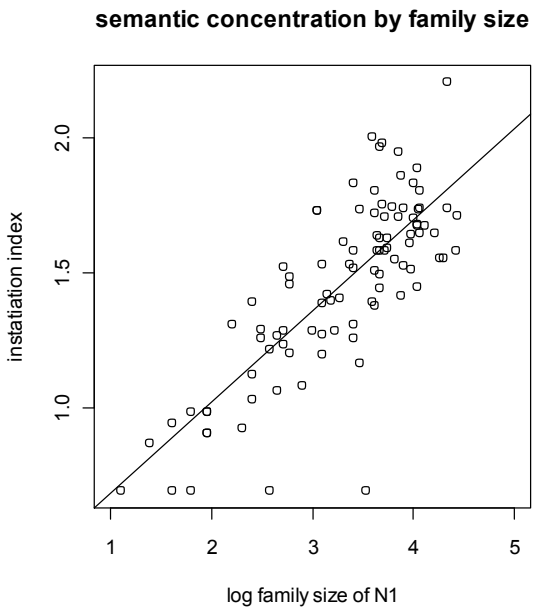


Figure 2: Relationship between N1 family size and the ratio of instantiations of different semantic relations

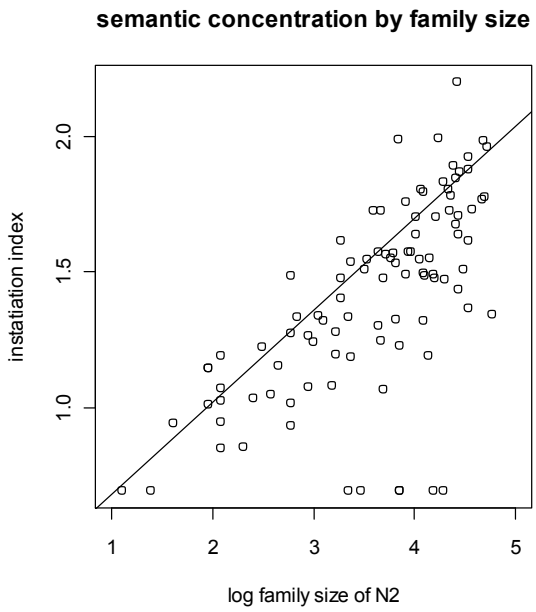


Figure 3: Relationship between N2 family size and the ratio of instantiations of different semantic relations

Thus, the obtained results demonstrate that there is a possible connection between the productive use of a compound constituent in one position and productive use of the semantic relation, which can be considered evidence of interaction between the levels of structural and semantic representation in compounds.

8. Discussion and conclusions

The first hypothesis that the analysis tested is whether, as a constituent of a compound, a lexeme is productive in one position and not the other. The obtained results clearly indicate a negative correlation between the sizes of the modifier and head families for one constituent. This means that with the increase of the head family size for a given noun, the size of the modifier family for the same noun decreases. This suggests that any given noun is used more productively either as a modifier or as the head of the compounded structure. This result is in line with Baayen's (2010) claim that the constituents of lexicalised compounds are position-bound. The fact that compounds used in the current study are not lexicalised, makes it possible to suggest that this claim holds true for non-lexicalised compounds too. This allows for a speculation that in the process of forming a new compound on the level of structural representation, for each noun or a group of semantically similar nouns, we follow a certain pattern that predetermines the use of a noun as a head or modifier. The development and

solidification of this pattern in language can be connected with speakers' experience of how a noun is used in an N+N compound, and this experience may also be considered responsible for the extension of the constituent family paradigm, which becomes a dominant for this noun. Inevitably, the speakers' experience will depend on the amount of exposure to how a noun is used. As discussed in Bauer et al. (to appear), the analysis needs to consider a combination of factors, rather than one single factor. The possible factors include (but are not limited to) high frequency of use of a limited number of established compounds within a paradigm, the number of such compounds in a paradigm, the use of compounds using a specific pattern (e.g. *water + N*, *N + problem*, etc.) in a wide range of contexts and genres, the importance of the concept denoted by a compound for the life of the language community, etc. Another preliminary speculation that can be put forward here concerns the morphological properties of the noun that demonstrates the tendency to be used in one position. For example, items that are adjective-like, e.g. *future*, *animal*, *chocolate*, *lemon*, have a higher probability of occurring in the roles typical of an adjective (attributive role in our case) and, therefore, are more often used in the modifier position than in the head position. The analysis of a larger data set might help shed light on this issue.

Another issue that needs to be pointed out here is that the degree of preference for being used in one position differs for different nouns and ranges from absolute (100%) to low (≥51%), which allows for the suggestion that morphological productivity of an element in one position should be viewed as a scalar phenomenon that may or may not change under the influence of the factors outlined above. The differences in the degree of productivity may be dependent on the number of factors that are involved in each individual case, with the productivity of a noun as a constituent of N+N compound increasing when more than one factor is involved.

This research has also looked into the semantic relations realised by noun concepts as constituents of compounds in order to check whether there is an overall preference for a limited number of semantic relations to be realised in a constituent family. The statistical analysis of the collected data demonstrates this trend. It has also been found that the size of the constituent family can be considered a strong predictor of the degree to which a constituent concentrates on one semantic relation. The results suggest that the possibility of instantiation of one particular relation over the others increases as the family size increases, notwithstanding a logical supposition that large CWFs should have a wider distribution of semantic relations. This tendency is slightly stronger for head CWFs than for modifier CWFs (See Tarasova (2013) for the discussion on possible reasons for this). The consistency in realising one relation varies for different concepts; but, as suggested by the statistical analysis, the preference for one relation is not random and is consistent in the collected corpus. This implies that there should be a connection between the productivity of a noun as a constituent of a compound and the consistent realisation of one semantic relation by this noun within a constituent family.

This finding has important implications for further research in this area. One of the areas for further exploration concerns the use of analogy in the process of coining a compound. In the cognitive framework analogy is defined by Blevins & Blevins (2009: 2) as "[…] a general cognitive process that transfers specific information or knowledge from one instance or domain (the analogue, base, or source) to another (the target). Sets of percepts, whether visual images, auditory signals, experiences, or dreams, are compared, and higher-order generalisations are extracted and carried over to new sets". It is obvious from this

definition that analogy is not limited to the structural patterns but also involves other levels, e.g. level of conceptual representation.

Analogy is widely discussed in derivation since it can be seen in affixes that are used for deriving morphologically complex words. Affixes usually alter the meaning of a derivational base in a specifically prescribed way. Extending this to compounds, Booij (2008, 2010) claims that analogical compounding is based on an individual compound (model word) with an idiosyncratic meaning. This meaning must be known for new compounds formed by analogy from the model compound to be understood. Analogy can be traced using semantic evidence when a particular idiosyncratic interpretation recurs in newly coined words (Booij 2008: 37). However, following the opinion expressed in Krott et al. (2007: 27) and Schlücker & Plag (2011: 1542), it is assumed here that analogy does not have to be driven by individual model words, but also involves paradigms that function as the basis for analogy. In this case not a single compound but a set, i.e. a paradigm (in our terms, constituent word family and its characteristics), predetermines the formation of new items. Krott et al. (2007: 27–28) use the term *paradigmatic analogy* to refer to the type of analogy in which "[…] the selection is based on the similarity of the target compound to a set (i.e., paradigm) of compounds, opposed to its similarity to a single exemplar, i.e., a single compound".

The studies to date, including the study on the use of linking elements in Dutch and German (Krott et al. 2007), the studies on stress assignment in compounds (Plag et al. 2007; Plag 2010), and the studies on the interpretation of the semantic relations (Gagné & Shoben 1997; Gagné 2001; Estes & Jones 2006, 2008), provide sufficient evidence that a paradigmatic approach to analogy provides a good account for the patterns noted in the formation of compounds. The analogical approach claims that the formation of new complex lexemes is based on the paradigms of similar existing complex lexemes and their formal properties rather than on abstract rules (Schlücker & Plag 2011: 1540). Therefore, the form of new compound coinages relies on the formal and semantic properties that the constituent words in the new combinations share with other compounds these constituents occur in.

The results obtained in the course of the current study indicate that the use of analogy in the formation of N+N compounds possibly takes place on at least two levels of representation: structural and semantic. It may also be the case that there should be interaction between the levels of structural and semantic representation in the process of forming compounds. Further research in this direction would helpful for a better understanding of the processes involved in the formation of compounds.

Acknowledgements

I would like to express my thanks and gratitude to Laurie Bauer, Natalia Beliaeva, Ingo Plag, and Gillian Claridge for their comments, advice and recommendations.

References

Aronoff, Mark. 1976. *Word formation in Generative Grammar*. Cambridge: Cambridge University Press.

Baayen, Harald. 1992. Quantitative aspects of morphological productivity. In Booij, Geert & van Marle, Jaap (eds.), *Yearbook of morphology 1991*, 109–149. Dordrecht: Kluwer.

Baayen, Harald. 1993. On frequency, transparency and productivity. In Booij, Geert & van Marle, Jaap (eds.), *Yearbook of morphology 1992*, 181–208. Dordrecht: Kluwer.

Baayen, Harald. 2003. Probabilistic approaches to morphology. In Bod, Rens & Hay, Jennifer & Jannedy, Stefanie (eds.), *Probabilistic linguistics*, 229–287. Massachusets: The MIT press.

Baayen, Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, Harald. 2010. The directed compound graph of English: An exploration of lexical connectivity and its processing consequences. In Olsen, Susan (ed.), *New impulses in word-formation* (Linguistische Berichte Sonderheft 17), 383–402. Hamburg: Buske.

Baayen, Harald & Lieber, Rochelle. 1991. Productivity and English derivation: A corpus-based study. *Linguistics* 29(5). 801–844.

Baayen, Harald & Renouf, Antoinette. 1996. Chronicling the Times: Productive lexical innovations in an English newspaper. *Languages* 72(1). 69–96.

Bates, Douglas & Maechler, Martin & Dai, Bin. 2007. lme4: Linear mixed-effects models using S4 classes. (http://lme4.r-forge.r-project.org/) (Accessed 2018-08-27.)

Bauer, Laurie. 1978. *The Grammar of nominal compounding with special reference to Danish, English and French*. Odense: Odense University Press.

Bauer, Laurie. 1979. On the need for pragmatics in the study of nominal compounding. *Journal of Pragmatics* 31(1). 45–50.

Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.

Bauer, Laurie. 1998. When is a sequence of two nouns a compound in English? *English Language & Linguistics* 21(1). 65–86.

Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

Bauer, Laurie. 2017. *Compounds and compounding*. Cambridge: Cambridge University Press.

Bauer, Laurie & Lieber, Rochelle & Plag, Ingo. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Bauer, Laurie, & Tarasova, Elizaveta. 2013. Meaning link in nominal compounds. *SKASE Journal of Theoretical Linguistics* 10(3). 12–18. (http://www.skase.sk/Volumes/JTL24/pdf_doc/01.pdf) (Accessed 2018-08-27.)

Bauer, Laurie & Beliaeva, Natalia & Tarasova, Elizaveta. (To appear). Productivity revisited. *Zeitschrift für Sprachwisenschaft.*

Bell, Melanie. 2005. Against nouns as syntactic premodifiers in English noun phrases. *Working Papers in English and Applied Linguistics* 11(1). 1–48.

Bell, Melanie. 2011. At the boundary of morphology and syntax: Noun noun constructions in English. In Galani, Alexandra & Hicks, Glyn & Tsoulas, George (eds.), *Morphology and its interfaces* (*Linguistics Today* 178), 137–167. Amsterdam: John Benjamins.

Bell, Melanie & Plag, Ingo. 2012. Informativeness is a determinant of compound stress in English. *Journal of Linguistics* 48(3). 485–520.

Blevins, James & Blevins, Juliette (eds.). 2009. *Analogy in grammar: Form and acquisition*. Oxford: Oxford University Press.

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

Booij, Geert. 2008. Construction morphology and the lexicon. In Montermini, Fabio & Boyé, Gilles & Hathout, Nabil (eds.), *Selected proceedings of the 5th Décembrettes: Morphology in Toulouse*, 34–44. Somerville, MA: Cascadilla Proceedings Project.

Booij, Geert. 2010. Compound construction: Schemas or analogy? A Construction Morphology perspective. In Scalise, Sergio & Vogel, Irene (eds.), *Cross-disciplinary issues in compounding*, 93–108. Amsterdam: Benjamins.

Boroditsky, Lera. 2000. Metaphoric Structuring: Understanding time through spatial metaphors. *Cognition* 75(1). 1–28.

Brekle, Herbert. 1976. *Generative Satzsemantik im System der englischen Nominalkomposition*. München: Fink.

Coulson, Seana. 2001. *Semantic leaps: Frame-shifting and conceptual blending in meaning construction.* Cambridge: Cambridge University Press.

Davies, Mark. *2004–*, *BYU-BNC* (Based on the British National Corpus from Oxford University Press). (http://corpus.byu.edu/bnc/) (Accessed 2018-08-27.)

Estes, Zachary & Jones, Lara. 2006. Priming via relational similarity: A COPPER HORSE is faster when seen through a GLASS EYE. *Journal of Memory & Language* 55(1). 89–101.

Estes, Zachary & Jones, Lara. 2008. Relational processing in conceptual combination and analogy. *Behavioral & Brain Sciences* 31(4). 385–386.

Fabb, Nigel. 1998. Compounding. In Spencer, Andrew & Zwicky, Arnold (eds.), *The handbook of morphology*, 66–83. Oxford: Blackwell.

Fernández-Domínguez, Jesús. 2009. *Productivity in English word-formation: An approach to N+N compounding*. Bern: Peter Lang.

Fernández-Domínguez, Jesús. 2015. The syntagmatic and paradigmatic axes of productivity: An application to the onomasiological model of word-formation. *SKASE Journal of Theoretical Linguistics* 12(3). 218–237. (http://www.skase.sk/Volumes/JTL29/pdf_doc/10.pdf) (Accessed 2018-08-27.)

Gagné, Christina. 2000. Relation-based combinations versus property based combinations: A test of the CARIN theory and dual process theory of conceptual combination. *Journal of Memory & Language* 42(3). 365–389.

Gagné, Christina. 2001. Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 27(1). 236–254.

Gagné, Christina, & Shoben, Edward. 1997. The influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 23(1). 71–87.

Gagné, Christina, & Shoben, Edward. 2002. Priming relations in ambiguous noun-noun combinations. *Memory & Cognition* 30(4). 637–646.

Gagné, Christina, & Spalding, Thomas. 2004. Effect of relation availability on the interpretation and access of familiar noun-noun compounds. *Brain & Language* 90(1–3). 478–486.

Gagné, Cristina & Spalding, Thomas & Figueredo, Lauren & Mullaly, Allison. 2009. Does snowman prime plastic snow? The effect of constituent position in using relational information during the interpretation of modifier-noun phrases. *Mental Lexicon* 4(1). 41–76.

Giegerich, Heinz. 2015. *Lexical structures: Compounding and the modules of grammar*. Edinburgh: Edinburgh University Press.

Granville Hatcher, Anna. 1960. An introduction to the analysis of English noun compounds. *Word* 16(3). 356–373. (https://doi.org/10.1080/00437956.1960.11659738) (Accessed 2018-01-22.)

Harley, Heidi. 2009. Compounding in Distributed Morphology. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford handbook of compounding*, 129–144. Oxford: Oxford University Press.

Hay, Jennifer. 2003. *Causes and consequences of word-structure*. London: Routledge.

Hay, Jennifer & Baayen, Harald. 2005. Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences* 9(7). 342–348.

Hein, Katrin & Engelberg, Stefan. 2017. Morphological variation: The case of productivity in German compound formation. In Koutsoukos, Nikos & Audring, Jenny & Masini, Francesca (eds.), *Morphological variation: Synchrony and diachrony*, 11th Mediterranean Morphology Meeting, Online proceedings 2018, 36–50. (http://mmm.lis.upatras.gr/index.php/mmm/article/view/2871) (Accessed 2018-08-27.)

Kastovsky, Dieter. 1986. The problem of productivity in word formation. *Linguistics* 24(3). 585–600.

Kemmerer, David. 2005. The spatial and temporal meanings of English prepositions can be independently impaired. *Neuropsychologia* 43(5). 797–806.

Krott, Andrea. 2009. The role of analogy for compound words. In Blevins, James & Blevins, Juliette (eds.), *Analogy in grammar: Form and acquisition*, 118–136. Oxford: Oxford University Press.

Krott, Andrea & Schreuder, Robert & Baayen, Harald. 1999. Complex words in complex words. *Linguistics* 37(5). 905–926.

Krott, Andrea & Schreuder, Robert & Baayen, Harald & Dressler, Wolfgang. 2007. Analogical effects on linking elements in German compound words. *Language & Cognitive Processes* 22(1). 25–57.

Kunter, Gero. 2011. *Compound stress in English: The phonetics and phonology of prosodic prominence.* Berlin: de Gruyter.

Lakoff, George. 1993. The contemporary theory of metaphor. In Ortony, Andrew (ed.), *Metaphor and thought* (2nd edition), 202–251. Cambridge: Cambridge University Press.

Lakoff, George, & Johnson, Mark. 1999. *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.

Levi, Judith. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.

Lieber, Rochelle. 1992. *Deconstructing morphology: Word formation in syntactic theory*. Chicago: University of Chicago Press.

Lieber, Rochelle. 2010. On the lexical semantics of compounds: Non-affixal (de)verbal compounds. In Scalise, Sergio & Vogel, Irene (eds.), *Cross-disciplinary issues in compounding*, 127–144. Amsterdam: Benjamins.

Lieber, Rochelle & Štekauer, Pavol. 2009. Introduction: Status and definition of compounding. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford handbook of compounding*, 3–18. Oxford: Oxford University Press.

Maguire, Phil & Wisniewski, Edward & Storms, Gert. 2010. A corpus study of semantic patterns in compounding. *Corpus Linguistics & Linguistic Theory* 6(1). 49–73.

Marchand, Hans. 1969. *The categories and types of present-day English word-formation: A synchronic-diachronic approach* (2nd edition). München: Beck.

Matlock, Teenie & Ramscar, Michael & Boroditsky, Lera. 2005. On the experiential link between spatial and temporal language. *Cognitive Science* 29(4). 655–664.

Murphy, Gregory. 1988. Comprehending complex concepts. *Cognitive Science* 12(4). 529–562.

Olsen, Susan. 2000. Compounding and stress in English: A closer look at the boundary between morphology and syntax. *Linguistische Berichte* 181(1). 55–70.

Pinheiro, José & Bates, Douglas. 2000. *Mixed Effects Models in S and S-Plus*. New York: Springer.

Plag, Ingo. 1999. *Morphological Productivity*. Berlin: Mouton de Gruyter.

Plag, Ingo. 2010. Compound stress assignment by analogy: The constituent family bias. *Zeitschrift für Sprachwisenschaft* 29(2). 243–282.

Plag, Ingo & Kunter, Gero & Lappe, Sabina. 2007. Testing hypotheses about compound stress assignment in English: A corpus-based investigation. *Corpus Linguistics & Linguistic Theory* 3(2). 199–233.

R Development Core Team 2011. Vienna: R Foundation for Statistical Computing.

Radden, Günter. 2003. The metaphor TIME AS SPACE across languages. In Baumgarten, Nicole & Böttger, Claudia & Motz, Markus & Probst, Julia (eds.), *Übersetzen, Interkulturelle Kommunikation, Spracherwerb und Sprachvermittlung – das Leben mit mehreren Sprachen. Festschrift für Juliane House zum 60. Geburtstag. Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 8(2/3). 226–239. (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.91.7412&rep=rep1&type=pdf) (Accessed 2018-08-27.)

Rainer, Franz. 2004. Studying restrictions on patterns of word-formation by means of the Internet. *Italian Journal of Linguistics* 15(1). 131–139.

Ricca, Davide. 2010. Corpus data and theoretical implications: With special reference to Italian V-N compounds. In Scalise, Segio & Vogel, Irene (eds.), *Cross-Disciplinary Issues in Compounding*, 237–254. Amsterdam: Benjamins.

Sandra, Dominiek & Rice, Sally. 1995. Network analyses of prepositional meaning: Mirroring whose mind – the linguist's or the language users? *Cognitive Linguistics* 6(1). 89–130.

Schlücker, Barbara & Plag, Ingo. 2011. Compound or phrase? Analogy in naming. *Lingua* 121(9). 1539–1551.

Schultnik, Henk. 1961. Produktiviteit als morfologish fenomeen. *Forum der Letteren* 2(1). 110–125.

Štekauer, Pavol. 2005. *Meaning Predictability in Word Formation: Novel, context-free naming units*. Amsterdam: Benjamins.

Štekauer, Pavol & Valera, Salvador & Körtvelyessy, Livia. 2008. *Universals, tendencies and typology in word-formation: A cross-linguistic study*. Košice: Josef Šafárik University.

Storms, Gert & Wisniewski, Edward. 2005. Does the order of head noun and modifier explain response times in conceptual combination? *Memory & Cognition* 33(5). 852–861.

Tarasova, Elizaveta. 2013. *Some new insights into the semantics of English N+N compounds*. Wellington: Victoria University of Wellington. (Doctoral dissertation.)

ten Hacken, Pius. 1994. *Defining morphology: A principled approach to determining the boundaries of compounding, derivation and inflection*. Hildesheim: Olms.

ten Hacken, Pius. 2016. Introduction: Compounds and their meaning. In ten Hacken, Pius (ed.), *The semantics of compounding*, 1–11. Cambridge: Cambridge University Press.

ten Hacken, Pius. 2017. Compounding in Morphology. In Aronoff, Mark (ed.), *The Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press. (http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-251) (Accessed 2018-08-27.)

van Jaarsveld, Henk J. & Coolen, Riet & Schreuder, Robert. 1994. The role of analogy in the interpretation of novel compounds. *Journal of Psycholinguistic Research* 23(1). 111–137.

Warren, Barbara. 1978. *Semantic Patterns of Noun-Noun Compounds*. Goethenburg: Acta Universitatis Gothoburgensis.

Wisniewski, Edward. 1997. When concepts combine. *Psychonomic Bulletin & Review* 4(2). 167–183.

Zwitserlood, Pienie. 1994. The role of semantic transparency in the processing and representation of Dutch compounds. *Language & Cognitive Processes* 9(2). 341–368.

Elizaveta Tarasova
Department of BCIS
IPU New Zealand
Palmerston North
etarasova@ipu.ac.nz

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/04.pdf. ISSN 1336-782X.

Compounds as idioms. A case study of a 'meta-trend'

Hans Götzsche, Aalborg University, Denmark

1. Introduction

When writing a manuscript for a Festschrift, I came across some compounds in Swedish that could only be interpreted as tropes. Take the word *kulturmaffia* 'cultural mafia'. It does not seem to be found in English and is confined to Scandinavia. Apart from the Swedish spelling of *mafia*, the compound is hardly understandable without some contextual information. The meaning is 'a group of people who are especially influential in the sphere of fine culture (literature, drama, paintings) in a society'. Whereas the word *culture* is polysemous, the word *mafia* cannot, in this context, be used in its original meaning; a fact that is well known. So, is its meaning metaphorical, and, in that case, since it has become conventionalised, is its combination with *culture* idiomatic?

Another case is the use of *meta-* in combination with, mainly, abstract expressions. The compound *metadata* is now ubiquitous because of the dissemination of electronic devices and one may find quite many, often technical, terms with a prepositioned *meta-*. Some dictionaries classify the expression *meta-* as a prefix, and considering the somehow peculiar origin of the accepted term *metaphysics* one may ask if the expansion of its use is some kind of 'meta-trend'.

Below I will go trough a few of the Swedish idiomatic compounds and then take up a number of word-formations with *meta-* in order to try to clarify whether you could call these formations compounds or not, and whether their semantics justify a classification as idioms. This paper is not what some of my colleagues in linguistics would call deep linguistics, whatever the metaphorical meaning of the word *deep* in that context, only an attempt to find out why people do peculiar things with words.

2. The research tradition

I shall not embark on the process of reviewing the research tradition on the topic, but some colleagues in linguistics have pointed to Benczes (2006); and the book may be relevant. But I shall draw your attention to the review by Melanie Bell in *English Language and Linguistics*, in which she says that "[h]owever, when one delves into the contents in more detail, there is a lack of precision and explicitness that makes for a frustrating read" (Bell 2008), and I have to agree. Instead, one might peruse the paper by Kooij (1968), which offers an overview over the theoretical and analytical intricacies presented by connecting the two phenomena, compounds and idioms.

3. A Swedish trend?

The frame of reference of this approach is the question if a metaphorical meaning can be comprised in a compound and, furthermore, if such a metaphorical meaning can be – maybe not overused ending up as a cliché but – integrated into language usage as an idiomatic expression.

In order to look into that, I checked a number of pages in Swedenborg (2001 [1986]) on new words in Swedish from the 1940s to the 1980s, not in a totally random fashion but somehow. I chose the entries beginning with the letter *m* (ibid.: 156–169) and then I looked for compounds that intuitively could be characterised as idioms.

In addition to *kulturmaffia* 'cultural mafia', which is spelled *-ff-* in accordance with Swedish orthography, I found the word *tjejmaffia* (lit. 'girl mafia'). Both *kulturmaffia* and *tjejmaffia* can be seen as oxymorons since the concepts of 'culture' as well as 'girl' will not normally be associated with criminal behaviour, whereas *Mafia* 'mafia' will, in general, be used as a label on some group of organised criminals. While the meaning of *kulturmaffia* is the one mentioned above, the meaning of *tjejmaffia* is 'a group of females (feminists) in the society who have an overwhelming, and unjustified, influence on what should be defined as gender', which might not have been extracted from the compound meaning. One will need some context, in this case Swedish culture in the beginning of the 21st century, in which these themes are highly controversial. But the words are commonly used, because they have found their way into the dictionary, and to the extent words are in common usage and their semantics cannot be interpreted on the basis of their parts, I shall suggest that these compounds can be classified as idioms.

By way of illustration I shall pick up one more Swedish word: *mammutfilm* 'mammoth film'; apparently not found in English. If I were to present my grandchildren with the word, I would expect them to answer 'a film about mammoths', following a procedure of semantic interpretation saying that the first word denotes a subset of the set of all films, viz. the subset dealing with the topic of mammoths. But the definition in the dictionary says that it is a film that is 'longer than usual, has a bigger format or has many actors'. Having in mind that the word was first introduced in 1965, when these features were interesting for the audience, the semantics of *mammoth*, in Swedish, in that context seems to have become vague, and the expression *mammoth* became a prefix functioning as an intensifier.

4. The origin and usage of *meta-*

What prompted my interest in the use of *meta-* as a prefix was the manuscript on *metaphilosophy* that I had been asked to review. At the outset, I was a little puzzled by the notion since, as a philosopher of language – and as a linguist – I had always conceived of philosophy proper as an intellectual enterprise dealing with the basic notions used in the sciences, thereby seeing philosophy and the sciences as complementary to each other. As such it comprises the subjects of:

i. metaphysics (the basic theoretical concepts, sometimes in the form of axioms),
ii. ontology (the ideas one has about how the world is built),
iii. epistemology (the ideas one has about how we know how the world is built), and a few other things.

But what, then, is *metaphilosophy*? At face value it is contemplations about philosophy itself. We will come back to that under the label *self-reference* below. Instead we will search for the origin of the word *meta*(-). Most of us know the word *metaphysics*, and from the context of linguistics we know, e.g. *metacognition* and *metalinguistic awareness*, and if we check the website *English Language & Usage Stack Exchange*, at the webpage

(https://english.stackexchange.com/questions/245403/how-did-meta-come-to-mean-self-referential?utm_medium=organic&utm_source=google_rich_qa&utm_campaign=google_rich_qa), we find the definition (referring to *Oxford*, i.e. *English Oxford Dictionaries*):

"(Of a creative work) referring to itself or to the conventions of its genre; self-referential: *the enterprise is inherently 'meta', since it doesn't review movies, for example, it reviews the reviewers who review movies*".

But the origin section of the entry redirects [here] and there's no apparent link with *self-reference*. It means that there is no further information on *self-reference* so we try instead *Etymonline* that has an entry on *meta-* (https://www.etymonline.com/search?q=meta-), but not on *meta* as a single word. There are 24 sub-entries with *meta-* but not all will be perceived as combinations of a prefix and a base in Modern Standard English.

At this point, it is evident that *meta-* and *self-reference* depart because the notion of *self-reference* may not be an essential part of the meaning of *meta-* for the reason that not many substitutions of *meta-* with *self-referential* will bring about meaningful expressions. The background seems to be that *self-reference* and *self-referential* are mainly used in formal logic and mathematics about a case when an expression refers to itself.

Personally, as a logician, I am not especially happy with this theoretical concept since only humans can refer to themselves by means of words, and words cannot, by nature, refer to themselves; and I would like philosophers and other scholars to find another term about recursive systems. As a linguist and, hypothetically, trying to be rude I might even say that it is a typical instance of the way philosophers misunderstand reflexive pronouns.

In contrast with this, *meta-* can actually be traced back in history in a way explaining how constructions like *metamathematics*, *metacommunication*, *meta-analysis* (*metanalysis**), *metapolitics*, *metaphrastic*, *metamorphize*, *metaphor*, *metathesis*, *metamorphism*, or *metamorphosis* have come up. In the context of the language sciences it is interesting that the term with the asterisk, *metanalysis*, has the *Etymonline* note:

"*1914, from meta- "transcending, overarching, dealing with the most fundamental matters of" + analysis. Coined by Danish philologist Otto Jespersen."

(https://www.etymonline.com/word/metanalysis)

But the maybe surprising information is that the word *metaphysics* in its current use is based on a mistake that goes back to the bibliographic ordering of Aristotle's work:

> *[...] from Greek* ta meta ta physika *"the (works) after the Physics," title of the 13 treatises which traditionally were arranged after those on physics and natural sciences in Aristotle's writings. The name was given c.70 B.C.E. by Andronicus of Rhodes, and was a reference to the customary ordering of the books, but it was misinterpreted by Latin writers as meaning "the science of what is beyond the physical."*
>
> (https://www.etymonline.com/word/metaphysics)

In due course this has lead to the

> *[...] misinterpretation of metaphysics as "science of that which transcends the physical." This has led to a prodigious erroneous extension in modern usage, with meta- affixed to the names of other sciences and disciplines, especially in the academic jargon of literary criticism.*
> (https://www.etymonline.com/word/meta-#etymonline_v_14705)

This narrative is confirmed by the entry *metaphysics* in *The Stanford Encyclopedia of Philosophy* (SEP).[1] One may wonder why *meta-* is frequent in modern technical languages, and one of the usual suspects may be Tarski who had a problem with specific truth conditions, so he suggested that (according to SEP):

> *If the language under discussion (the* object language*) is* L, *then the definition should be given in another language known as the* metalanguage, *call it* M. *The metalanguage should contain a copy of the object language (so that anything one can say in* L *can be said in* M *too), and* M *should also be able to talk about the sentences of* L *and their syntax. Finally Tarski allowed* M *to contain notions from set theory, and a 1-ary predicate symbol True with the intended reading 'is a true sentence of* L*'.*[2]

In formal logic this seems fair enough but, facilitated by what may be called *real-life language users*, the expression *meta-* appears to have spread in an out-of-control fashion, as is demonstrated by the *Wikipedia* entry:

"Meta (from the Greek preposition and prefix meta- (μετά-) meaning "after", or "beyond") is a prefix used in English to indicate a concept which is an abstraction behind another concept, used to complete or add to the latter."

(https://en.wikipedia.org/wiki/Meta)

It is hard to know what is meant by "an abstraction behind" but maybe it is the understanding, so to speak, behind the word *metadata*, i.e. 'information about data', even though both *abstraction* and *behind* for about must be seen as tropes. So, from the technical use by, e.g. Tarski the next step of entering the open market of language use is an opaque semantics, leading to the state in which any use of the expression is contextual. When a word has reached that status no one is able to interpret its meaning until one knows its specific context of use. Then the word is on the free market, meaning that anyone is authorised to use it in his/her own sense.

This seems to have happened to the prefix *meta-* in the word *meta trend*. When I made a search at *Google* for *metatrend*, *Google* asked: "did you mean: meta trend?" The search result for *meta trend* was about 102.000.000 results (0,32 seconds) hits. Well, 102M! So one could say that *meta* had disseminated a little since Aristotle's work had been mishandled some two thousand years ago. The top hit by *Google* (28/08/18) was:

[1] https://plato.stanford.edu/entries/metaphysics/
[2] https://plato.stanford.edu/entries/tarski-truth/

> *5 meta-trends underlying almost all of modern marketing*
> *The fast-paced evolution of marketing is a sea of trends, from attribution models to the Zero Moment of Truth. So many trends, in fact, that it can be hard to keep track of them all. A meta-trend runs deeper, powering more specific trends, like a tidal force that drives waves to the shore.*
> (https://chiefmartec.com/2012/10/5-meta-trends-underlying-almost-all-of-modern-marketing/)

It is hard to know what the meaning of *meta-trend* is and the only thing one can say for sure is that somebody may think that it sounds good. It may also be concluded that:

1. the expression *meta* has become the victim of a 'metatrend', viz. to attach a prefix to some word one is especially fond of and thereby using the prefix as an intensifier, and
2. the language users who do this have no idea whatsoever about the meaning of this expression.

Consequently this is idiomatic word formation based on ignorance. Contrary to the opinions of some of my colleagues in linguistics the term has been carefully chosen. The expression has, *per se*, no pejorative meaning, even though it is often used about people who deliberately ignore acknowledged information in a derogatory way, what I would prefer to call *disregard*, but if ignorance is just the state of being unaware of information it does not entail a negative attitude. One may contemplate the fallacy of *ignoratio elenchi* in classical logic (see Walton 2008: 100 et passim), i.e. drawing a conclusion from what is not known. This is just a label of the characteristics of the inference.

But one thing is marketing. Has *meta-* a mission among the academic subjects? Let us take a look at *metamodernism*. In Vermeulen et al. (2010) the authors introduce the word *metamodernism* and propose it as a substitute for *postmodernism*:

> *We will call this structure of feeling metamodernism. According to the Greek-English Lexicon the prefix "meta" refers to such notions as "with", "between", and "beyond". We will use these connotations of "meta" in a similar, yet not indiscriminate fashion. For we contend that metamodernism should be situated epistemologically with (post) modernism, ontologically between (post) modernism, and historically beyond (post) modernism.*
>
> Vermeulen et al. (2010: 2)

If we ignore the reference to a *Lexicon* translation – which may be fair enough – the use of all three translations combined with two philosophical disciplines and an historical label offers a less than transparent guide to the meaning of the concept *metamodernism*. The phrase is almost poetic, but a cynical empiricist may ask how something can be "epistemologically with" something, how something can be "ontologically between" some entities, and how something can be "historically beyond" something. In a philosophical and scientific context it does not make sense. Maybe not because of ignorance – they have checked the lexicon – but because of disregard. The authors have not taken into account that the word has a history, and historical usage, that might have been appropriately taken as a frame of reference when choosing the word.

Is *meta-*, then, an instance of *determinologization* as suggested by a colleague? If we compare it with *digital* it is a well-known story that it has been in use since the late Middle Ages, but only in certain contexts. It is, for instance, not found in *Svenska Akademiens*

Ordbok, the first volume of which was published in 1898 and which has by now arrived at the letter *v*. Only when the digital computer was invented the word appeared to be an appropriate technical term; which is no longer outside computer science. Any electronic device and its use is labelled *digital*, but the semantics of the word is not void. Contrary to this, the expression *meta-* has never been a technical term in its own right, only as a prefix of some kind, for instance in formal logic, mathematics and linguistics. Like some specific expressions that have become discourse markers (or particles), for instance *well* – also in prefix-expressions like *well-bred* – the expression *meta-* has lost any meaning outside special domains, and, in the end, may re-enter colloquial speech, like in the hypothetical: **it is just so meta!*, so I would not call it "determinologisation".

5. Conclusion

Idiomatic expressions are sometimes complicated, like the metaphors they originate from. And to the extent the constituent parts of metaphors cannot be interpreted based on their original meanings and the combination is accepted as conventional usage they may end up as idioms. When contracted they may also end up as compounds, as has been illustrated above. In the case of *meta-* one thing is the etymology of the word, another thing is its meanings in technical contexts. But, anyhow, its use in these technical contexts has increased the opacity of its semantics to the effect that it has become an intensifier like *mega-*. As mentioned above, one may speculate if it will proceed further and be used as a discourse marker, functioning the same way as *you know* in *it – was you – know rather nasty*. Hopefully we will never know; because of ignorance.

References

Bell, Melanie. 2008. Review of Benczes, Réka (2006). *English Language and Linguistics* 12(3). 549–552.

Benczes, Réka. 2006. *Creative compounding in English: The semantics of metaphorical and metonymical noun-noun combinations* (Human Cognitive Processing 19). Amsterdam: John Benjamins.

Brinker, Scott. 2012. *5 meta-trends underlying almost all of modern marketing.* (https://chiefmartec.com/2012/10/5-meta-trends-underlying-almost-all-of-modern-marketing/)

English Language & Usage Stack Exchange. (n.d.). *How did "meta" come to mean self-referential?* (https://english.stackexchange.com/questions/245403/how-did-meta-come-to-mean-self-referential?utm_medium=organic&utm_source=google_rich_qa&utm_campaign=google_rich_qa) (Accessed 2018-08-28.)

Harper, Douglas. 2001–2019. *Etymonline*. (https://www.etymonline.com)

Hodges, Wilfrid. 2018. Tarski's Truth Definitions. In Zalta, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 edition). (https://plato.stanford.edu/entries/tarski-truth/)

Kooij, J. G. 1968. Compounds and Idioms. *Lingua* 21. 250–268.

SAOB. 1898–2018. *Svenska Akademiens Ordbok*. Stockholm: Svenska Akademien. (https://www.saob.se)

Swedenborg, Lillemor. 2001 [1986]. *Nyord i svenskan från 40-tal til 80-tal*. Stockholm: Svenska språknämnden/Norstedts Ordbok.

van Inwagen, Peter & Sullivan, Meghan. 2018. Metaphysics. In Zalta, Edward N. (ed.), *The -Stanford -Encyclopedia of Philosophy* (Spring 2018 edition). (https://plato.stanford.edu/entries/metaphysics/)

Vermeulen, Timotheus & van der Akker, Robin .2010. Notes on metamodernism. *Journal of Aesthetics and Culture* 2. 1–14.

Walton, Douglas. 2008. *Informal Logic: A Pragmatic Approach*. Cambridge: Cambridge University Press.

Wikipedia contributors (eds.). 2018. Meta. In *Wikipedia: The Free Encyclopedia*. (https://en.wikipedia.org/wiki/Meta)

Hans Götzsche
CFL AAU Center for Linguistics, Aalborg University
Rendsburggade 14, 9000 Aalborg
www.cfl.hum.aau.dk/
Denmark
Phone: *+45-9940 9058*
Fax: *+45-9815 9434*
goetzsche@id.aau.dk

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/05.pdf. ISSN 1336-782X.

Relational adjectives between syntax and morphology

Pius ten Hacken, Universität Innsbruck, Austria

A relational adjective (RA) is an adjective that does not express a property, but rather a relation to a concept designated by a noun. It is a controversial issue whether RA+N combinations are compounds or syntactic phrases. If we adopt a definition of compound that takes the semantics as an important component, RA+N combinations should be analysed as compounds. If RA+N and N+N are both compounding, we might expect that languages that have both choose names for the corresponding concepts independently of each other, so that an RA+N in one language is randomly connected to the name for the same concept selected in another language. I tested this hypothesis by analysing German and Italian translations of Levi's (1978) list of 383 compounds illustrating her Recoverably Deletable Predicates (RDPs) and nominalization types. The analysis shows that there is a strong cross-linguistic correlation in the use of RA+N. This raises the question of how to explain the correlation. I argue that it cannot be explained by a translation bias or a semantic bias, but that it provides evidence for how the choice of a name in the naming process is influenced by a speaker's mental lexicon and how speech communities relate to such choices by individual speakers.

Keywords: *relational adjective*, *compounding*, *translation*, *naming*

1. Introduction

Relational adjectives (RAs) are adjectives such as *architectural*. They do not refer to a property, but express a relation to a corresponding noun. Typically, OED (2018 [1885]: *architectural*)[1] gives its meaning as 'of, relating to, or according to, architecture'. Over the years, it has been a controversial issue whether RA+N combinations such as *architectural monument* should be analysed as morphological objects (compounds) or syntactic objects (phrases). In §2, I will briefly outline the two positions and indicate why I will take the analysis as compounding as a starting point. Together with some general considerations concerning the definition of compounding, this leads to a hypothesis about the occurrence of RA+N combinations and their cross-linguistic correspondence, which serves as the main focus of this article. §3 gives some comparative data from English, German and Italian which can be used as a test for this hypothesis. As the data seem at least problematic for the hypothesis that RA+N combinations are unmarked compounds, §4 proposes a number of possible explanations. §5 summarizes the conclusions. In this article, I aim to contribute to this discussion by giving some quantitative evidence about the use of RA+N combinations in English, German and Italian.

[1] For references to the OED, I give the entry and in square brackets the year stated as the last general update for the entry. In some cases, readings or examples have been added in an OED entry without triggering a full update of the entry.

2. Relational adjectives and the limits of compounding

As an example of an RA, let us consider *architectural*. As illustrated in (1), this adjective has different meanings depending on the noun following it.

(1) a. architectural heritage
b. architectural staff
c. architectural interest

The shared meaning component of the three occurrences of *architectural* in (1) is that it expresses a relation to the concept designated by the noun *architecture*. The nature of the relation depends on the noun. As such, *architectural* contrasts with qualitative adjectives as illustrated in (2).

(2) a. rich heritage
b. new staff
c. strong interest

The contrast between RAs and other adjectives is not always as straightforward as in (1) and (2). In particular, many RAs have additional readings, as in (3).

(3) a. commercial law
b. commercial television

OED (2018 [1891]: *commercial*) gives the reading 'having reference to or bearing on commerce' for (3a), which is a typical description of the meaning of an RA. For (3b), OED gives a separate reading 'paid for by the revenue of broadcast advertisements', which only applies to radio and television. However, it is related to the meaning 'looking for financial profit', which has a more general application. In (3b) and related readings, *commercial* is not an RA.

A long-standing question is whether RA+N combinations such as (1) and (3a) should be considered compounds or not. The more traditional position on this is represented by Matthews (1974: 35), who discusses the Latin contrast in (4).

(4) a. tribunus militaris 'tribune$_{\text{nom.sg}}$ military$_{\text{nom.sg}}$'
b. tribunum militarem 'tribune$_{\text{acc.sg}}$ military$_{\text{acc.sg}}$'

The expression in (4) refers to a person of a particular rank in the ancient Roman army. The contrast between the nominative in (4a) and the accusative in (4b) is expressed both on the noun and on the adjective. For this reason, Matthews does not call such RA+N combinations *compound lexemes*, but *idioms*.

The opposite view is defended by Levi (1978: 15–48). Her argumentation is based on theoretical assumptions rooted in generative semantics. However, we can also take Levi's (1978: 38) examples of the type in (5) as an argument without referring to generative semantics.

(5) a. atomic bomb
b. atom bomb

Both expressions in (5) mean the same thing. OED (2018 [2008]) gives entries for both, with first attestations of 1914 for (5a) and 1921 for (5b). Both are quoted in the *Times* article reporting on the 1945 bombing Hiroshima. OED gives a slightly higher frequency (one class difference) for (5a), which is supported by a BNC (2007) frequency of 102 for (5a) as against 41 for (5b). Important is that both expressions are used and they refer to the same concept.

Depending on various theoretical assumptions, other arguments have been put forward for one position or the other. It is not my intention here to give a historic overview of this discussion, but it is good to consider some consequences of the persistence of both positions. The existence of two incompatible views of RA+N combinations does not mean that *compound* is a vague concept in which RA+N combinations are somewhat less typical than N+N combinations. As I argued in the case of the distinction between inflection and derivation in ten Hacken (2014), the question whether to have a terminological concept with a precise boundary depends on how linguists decide to use the relevant concept. For compounding, if we want to make claims about the concept, we need a definition that determines the extent of the concept. Such a definition is declared valid. A definition is not an empirical claim and it cannot have counterexamples. The only way to evaluate a definition is by assessing the usefulness of the concept it determines. Also the cross-linguistic validity of a definition cannot be evaluated on an empirical basis, but only argued for on the basis of the usefulness of the concept.

The choice between the two positions on the analysis of RA+N combinations can therefore not be based on empirical evidence alone. The question is whether we see *compound* as a primarily morphophonological concept or as a concept based on semantic considerations. In principle, both positions are coherent. Although the positions have here been illustrated on the basis of Matthews (1974) and Levi (1978), the choice for one of these positions does not entail a commitment to Matthews's Word-and-Paradigm morphology or Levi's generative semantics. In my view, there are at least two reasons why a more semantically oriented definition is to be preferred. First, I would argue that compounding is first of all interesting as a particular type of naming device. Taking the inflection-based criteria illustrated in (4) as more important than the semantic considerations applying to (5) would imply that compounds constitute primarily a superficial formal category. Second, cross-linguistic coherence is much bigger in semantic properties than in morphophonological properties. A well-known example of an often used criterion for compoundhood is stress. However, stress is notoriously language-specific and there is no cross-linguistic compounding stress pattern.

In ten Hacken (2013), I argue for a concept of compounding that is cross-linguistically valid and is based on its role as a naming device. The form and meaning of a compound arise from the interaction of two processes. One is the application of the compounding rule. The compounding rule takes two elements as input, the head and the non-head. The head is a lexeme and the non-head is categorially unspecified. The relationship between the two is not specified. Any restrictions at this stage are due to the meanings of the head and the non-head. The other process is onomasiological coercion. This is what happens in the confrontation between a potential name resulting from a word formation process and a concept for which a name is necessary. The outcome is that the concept to be named determines the meaning of the output of the word formation rule. In the case of

compounding, onomasiological coercion typically determines the semantic relation between the head and the non-head. However, it can also shift the meaning of a compound beyond what can be predicted on the basis of the compounding rule. For *spaceship*, the fact that it designates a vehicle that is not a ship is the result of onomasiological coercion.

In a definition that is based on semantic considerations, RA+N combinations are equally good as compounds as other compounds, such as N+N combinations. This implies that the choice between the two forms in (5) depends only on accidental properties of the use of these forms, not on the concept. This leads to the hypothesis in (6).

(6) In two languages which have RA+N combinations and other compounding constructions, there is no correlation between the names chosen for the same concept.

In (6), I use *construction* in the pretheoretical sense of ten Hacken (1994), without any implications that I adopt Goldberg's (1995) construction grammar or Booij's (2010) construction morphology. The lack of correlation mentioned in (6) is of the type illustrated in the pairs in (7).

(7) a. unleaded unverbleit
b. lead-free bleifrei

In (7), English and German adjectives are given for the meaning '(of petrol or other fuel) without added lead'. The English adjective is followed by the German adjective with the same morphological structure. (7a) and (7b) are the result of competing naming processes for the same concept. In English, *unleaded* is the standard form. BNC (2007) gives a frequency of 235 as against 56 for *lead-free*. Several of the latter but none of the former refer to a property of paint instead of fuel. OED (2018) gives a separate entry for *unleaded* [2017], whereas *lead-free* is only given among the combinations in the entry *lead* [1902]. In German, by contrast, *bleifrei* is the unmarked form. DeReKo (2018) gives 2105 occurrences for *bleifrei* as against 36 for *unverbleit*. Anyone who has used petrol stations in the UK and in German-speaking countries will have seen the resulting contrast in the designation on the pump. In view of (6), cases such as (7), where different strategies are chosen in different languages, are expected to occur approximately with the same frequency as cases where the strategies for two languages coincide, provided that the strategies are in general equally frequent in the two languages.

3. Relational adjectives in English, German, and Italian

In order to test the hypothesis in (6), we need data about names for the same concepts in different languages. The most obvious method for obtaining such data is collecting compounds in one language and translating them. In translation theory, the translation of individual words is controversial. However, the problems with such lexical translations arise especially where no lexical equivalent is available. Overviews of bilingual lexicography such as Adamska-Sałaciak (2013) and Fontenelle (2015) take the idea that bilingual dictionaries give translations as a starting point. The focus of their overviews is the question how bilingual dictionaries should treat cases where no direct equivalent exists or where additional information is required.

For the translation method to work, we need a monolingual list of compounds to start with. There are various ways to compile such a list. It would be problematic to choose the compounds individually. In order to test (6), we need a list of RA+N combinations and of other compounding constructions. Ideally, the list should neither be biased to individual heads or non-heads nor to individual semantic relations between them.

Given these constraints, the list Levi (1978: 279–284) gives as an appendix to her discussion of complex nominals is a reasonable choice. Levi's list contains 383 compounds. The list was compiled independently of the hypothesis to be tested here, so that there is no bias in this respect. One of Levi's theoretical points is that RA+N combinations are compounds. Therefore, the list contains both RA+N and N+N compounds. The purpose of the list is to illustrate the mechanism by which she proposes to account for the semantic relation between the head and the non-head. This mechanism involves different types of nominalization and nine Recoverably Deletable Predicates (RDPs). Nominalizations are appealed to when the head of the compound is deverbal. They are classified by the meaning of the result (act, product, agent, patient). RDPs account for all other non-lexicalized compounds. They are specified as CAUSE, HAVE, MAKE, USE, BE, IN, FOR, FROM and ABOUT. The first three RDPs can have either the head or the non-head as the subject. For several RDPs, Levi distinguishes different senses. Thus for FROM, one sense is 'produced from' (e.g. *olive oil*), another 'originating from' (e.g. *home remedy*). Each of these is illustrated with a number of compounds.

For our purposes, it is not necessary to commit to Levi's theoretical framework or to her analysis in order to use her list of compounds. Elsewhere, I argued against using systems for the characterization of the semantic relation between head and non-head of a compound as a basis for the classification of compounds (ten Hacken 2016: 211–214). This argument applies to any such system. The side effect of illustrating the semantic relations is, however, that we have a good degree of variation in these semantic relations. This, together with the extensive but still manageable size of the list, makes it a good basis for drawing conclusions about the validity of hypothesis (6).

For the research reported on in ten Hacken & Muigg (in press), Levi's list was translated into German and Italian. In the translation of the compounds, particular care was taken to consider the status of the expression in the target language. A central question in this context was to what extent there is an established name for the concept in German and in Italian. Where several possibilities exist and none is established, competing names are included. Determining whether an expression is established in the language is of course not an empirical question, because languages such as German and Italian are not empirical objects. However, standard dictionaries, large corpora and native speaker judgements were used to arrive at a reasonable approximation.

In Levi's list, approximately two thirds (257) are N+N and one third (126) RA+N combinations. There is one RDP for which all examples are N+N, FROM (e.g. *olive oil*). All other characterizations of semantic relations are divided between RA+N and N+N.

In German, compounding in the sense of the formation of N+N combinations is notoriously frequent. It is not surprising, therefore, to find that 289 of the items from Levi's list are translated by an N+N compound. With more than three quarters of the compounds translated as N+N, this construction is significantly more frequent in German than in English. Four semantic classes determined by RDPs are entirely covered by N+N compounds in German. They are illustrated in (8).

(8)	a.	Nasenspray	nasal mist	FOR
	b.	Rohrzucker	cane sugar	FROM
	c.	Honigbiene	honey bee	*y* MAKE *x*
	d.	Blumenkranz	floral wreath	*x* MAKE *y*

In (8), the German example is followed by the English item from Levi's list and the RDP characterizing the semantic class. In (8c-d), *x* is used for the non-head and *y* for the head of the compound.

Apart from the N+N compounds, there are also 40 RA+N combinations and 65 other translations. The numbers do not add up to 383, because there was no single translation for each of Levi's examples. In some cases, there was no equivalent established German translation as an N+N or RA+N. An example is *bird reproduction*. Although the components can be translated, *Vogel* ('bird') and *Fortpflanzung* ('reproduction'), actual German translations would use a paraphrase involving these two words rather than a compound. The compound *Vogelfortpflanzung* is a possible word, but it is not used. This is an example of an other translation. There are also cases in which we counted more than one German equivalent. An example is (9).

(9)	a.	Polizeieinsatz	'police operation'
	b.	polizeiliches Eingreifen	'$\text{police}_{\text{ADJ}}$ intervention'

Both German expressions are established translations of *police intervention*, but they have slightly different meanings. In (9b), the unit of police is presented more as an agent, whereas in (9a), it is rather an instrument. There is also a difference in syntactic use, because (9a) is countable and (9b) not. In such cases, we counted both translations. (9b) also provides an example of an RA+N construction in German. Some more examples are given in (10).

(10)	a.	historisches Drama	historical drama
	b.	künstlerischer Leiter	artistic manager

Even though in the sample determined by Levi's list, German has more N+N compounds than English, there are also cases where an English N+N compound is translated by a different construction in German. Two examples are given in (11).

(11)	a.	rührselige Geschichte	sob story
	b.	Professorinnen	women professors

In (11a), a non-relational adjective is used in a lexicalized A+N expression. In (11b), a derivational rule adding the suffix *-in* is used to express what in English is expressed by the non-head *women*.

In order to test the hypothesis in (6), we do not only need the totals of the expressions according to the German mechanisms used, but also the extent to which they correspond to English RA+N and N+N expressions. Table 1 summarizes these figures.

Table 1: Translation of English N+N and RA+N compounds into German

German	N+N	RA+N	other	total
English N+N	226 (87.9%)	4 (1.6%)	27 (10.5%)	257
English RA+N	78 (60%)	38 (29.2%)	14 (10.8%)	130

Table 1 gives absolute figures and percentages. The percentages are relative to the totals given in the last column. As explained, there are more German translations than English originals. No obvious correlation between the use of other mechanisms in German and the opposition between N+N and RA+N in English can be observed, but the correlation between RA+N combinations in German and in English is striking. Whereas only a tiny minority of English N+N compounds is translated as RA+N in German, more than a quarter of English RA+N compounds is translated in this way. This casts doubt on (6).

Let us now turn to Italian. The position of compounding in Italian is very different to its position in German. Konecny & Autelli (2015) give an overview of the debate about this issue among Italian linguists. A common type of compound in Italian is illustrated in (12).

(12) a. apribottiglie ‘open-bottles’, i.e. bottle opener
b. portagioielli ‘carry-jewels’, i.e. jewellery box

In both examples in (12), the Italian expression has a V+N structure in which the N is interpreted as the object of the V and the entire expression designates an instrument for V-*ing* N. In ten Hacken (2010) I point out the similarity of this construction to English synthetic compounds, as indicated by the idiomatic translation in (12a). In (12b), the idiomatic translation takes a different naming motive, as *box* focuses on the nature of the object as opposed to its function, which is highlighted by the verb *portare* (‘carry’) in Italian.

When we compare the Italian construction in (12) and corresponding synthetic compounds in English, however, we find that they are not fully equivalent. In Italian, the construction requires that the noun is the object of the verb. In English, the head noun of the compound is derived from a verb. There is some pressure to interpret the non-head as the object of the verb, but this is not mandatory as in Italian. Some relevant examples are in (13).

(13) a. night sleeper
b. pressure opener

In (13a), *sleep* is intransitive so it is necessary to find a different interpretation. In fact, (13a) designates a type of train. In (13b), *open* is transitive, but *pressure* is not a possible object for semantic reasons, so that it is again necessary to find a different interpretation. In this case, it is an instrument that opens wine bottles by means of air pressure. The availability of alternative interpretations, as illustrated in (13), is typical of compounds. By contrast, for the Italian examples in (12) an analysis as phrasal conversion, as proposed in ten Hacken (2010) is more appropriate. Therefore, I do not consider the expressions in (12) compounds.

In Italian we find few N+N compounds. This is generally typical of Romance languages. Although Radimský (2015) discusses quite a significant number of examples, they are very rare among the translations of the Levi list. More common are constructions with the preposition *di* as in (14).

(14)	a.	gruppo di lavoro	‘group of work’, i.e. working group
	b.	torta di mele	‘cake of apple’, i.e. apple cake
	c.	olio d’oliva	‘oil of olive’, i.e. olive oil

The examples in (14) illustrate how the Italian N+*di*+N construction corresponds to English N+N compounds. The Italian construction is left-headed, but its semantic behaviour, i.e. the way the meaning of the expression is determined, corresponds to that of the English right-headed N+N cases. In ten Hacken (2013), I argue for a similar analysis of French N+*de*+N. The French preposition *de* can be analysed as a periphrastic genitive, corresponding to the morphological genitive found in languages such as Polish. Extending the scope of the argument to Italian seems straightforward and I will assume that cases such as (14) are compounds.

In the Italian translations of the Levi list, about 70% belong to the RA+N or N+*di*+N constructions. Whereas in German, the proportion of translations that are not compounds is barely over 10%, in Italian it is almost 30%. Some examples of these are given in (15).

(15)	a.	curry con coriandolo	coriander curry
	b.	aspirapolvere	vacuum cleaner
	c.	carillon	music box

In (15), the Italian expressions are followed by the compounds of the Levi list they translate. (15a) is a paraphrase, literally ‘curry with coriander’. As argued in ten Hacken (2013: 106–108), prepositions generally determine the relationship between the nouns they connect much more precisely than periphrastic genitives as in (14). Therefore cases such as (15a) should not be considered compounds. (15b) is a case of the type we encountered in (12), which I analyse as phrasal conversion. (15c) is a simplex noun.

Among the translations as compounds, there is no huge gap between the frequency of RA+N and N+*di*+N. The former account for 39%, the latter for 32% of translations. There are no semantic classes that fall entirely in one or the other, but there are still some preferences we can observe. For FROM, no RA+N translations were used and almost all are N+*di*+N, as in (14c). Some relations in which RA+N translations account for a clear majority are illustrated in (16).

(16)	a.	rivista sportiva	sports magazine	ABOUT
	b.	gocce nasali	nose drops	FOR
	c.	problemi familiari	family problems	IN

As in (8) for German, (16) gives the Italian expression followed by the item from Levi’s list and the RDP. As the Italian construction is left-headed, the N precedes the RA and the translation reverses the elements compared to the English original.

Our interest in the Italian data is of course triggered by the hypothesis in (6). In order to test this hypothesis, Table 2 brings together the relevant figures in the same way as Table 1 did for German.

Table 2: Translation of English N+N and RA+N compounds into Italian

Italian	N+*di*+N	RA+N	other	total
English N+N	111 (39.9%)	64 (23.0%)	103 (37.1%)	278
English RA+N	20 (15.2%)	95 (72.0%)	17 (12.9%)	132

As in Table 1, the total number of translations is higher than 383, because in cases where two translations are equally possible, both were recorded. The most striking figures in Table 2 are in the column headed *RA+N*. They show that less than a quarter of English N+N compounds were translated as Italian RA+N combinations, but almost three quarters of English RA+N compounds were translated in this way. In some way, Italian RA+N is less attractive as a translation of English N+N. This results in a higher proportion of N+*di*+N translations, but also in a much higher proportion of other translations, cases such as (15). There were a total of 10 N+N translations, e.g. *attore bambino* ('child actor'), all of them corresponding to English N+N compounds.[2] As they make up less than 4% of the total, they are included in the *other* column, but their exclusive occurrence with English N+N compounds presents another case of a cross-linguistic bias.

When we consider the use of RA+N in English, German and Italian on the basis of these data, we see first of all that there is a clear hierarchy in the sense that Italian has more RA+N items than English, which in turn has more of them than German. Secondly, there seems to be a strong correlation between the use of RA+N in the three languages. When we consider English and German, almost all German RA+N correspond to English RA+N. For English and Italian, English RA+N are three times as likely to correspond to an Italian RA+N as English N+N. Clearly, this result goes against the hypothesis in (6).

4. Possible explanations for the cross-linguistic correspondence in RA+N

The hypothesis that served as a starting point is (6), repeated here for ease of reference.

(6) In two languages which have RA+N combinations and other compounding constructions, there is no correlation between the names chosen for the same concept.

The idea behind (6) is that if RA+N combinations are compounds in the same way as N+N in English and German and N+*di*+N in Italian, the choice is independent in each language. The approximate proportions we found for the compounding types in §3 are 2:1 for English, 9:1 for German and 5:6 for Italian. Here the last figure is for the RA+N compounds and the first for N+N in English and German, N+*di*+N in Italian. What (6) predicts is that, for instance, the proportion of 9:1 for German N+N vs RA+N applies equally to translations of English N+N and of English RA+N constructions. In fact, for N+N the proportion is 56:1 and for RA+N it is 2:1. Similarly, in Italian we expected 5:6 for translations of English N+N and English RA+N, but we found 2:1 for the former and 1:5 for the latter. The evidence against (6) is so strong, that we need an explanation.

[2] Examples such as *attore bambino* and English *girlfriend* are sometimes considered coordinative compounds. However, the relation between the two components is not symmetrical. *Girlfriend* designates a type of friend, but not a type of girl. Similarly, *attore bambino* is a type of *attore* but not a type of *bambino*. Therefore, I consider them regular, headed compounds, not coordinative compounds.

A first possible explanation to consider is what I will call *translator's bias*. As the English list served as a basis for translation into German and Italian, there is a risk that not only the meaning, but also the structure of the source language expression is taken over in the translation. The issue of equivalence is a central point of discussion in translation theory. In Munday's (2016) catalogue of translation theories, the explicit discussion of equivalence is concentrated in the earlier chapters, covering earlier research, but its results are presupposed also in many later approaches.

A landmark in the discussion of equivalence is Nida's (1964) distinction between what he calls *formal* and *dynamic equivalence*. Nida (1964: 159) describes formal equivalence as concentrating "on the message itself, in both form and content" and dynamic equivalence as aiming for "complete naturalness of expression" while maintaining equivalent effect. In her rather practically oriented introduction to translation studies, Baker (2018) organizes much of the theory around the level at which equivalence is aimed for. She starts with equivalence at word level, moving to multi-word expressions, sentence structure and various aspects of text structure and the organization of meaning in a text.

In the context of the translations of compounds from Levi's list, we are dealing with structured words. The question is to what extent the structure of the word is considered in choosing a translation. The main translators involved were Christina Muigg, a graduate of the MA Translation Studies in Innsbruck, and Laura Rebosio, a student on that degree. As professional translators, they are aware of the different levels of equivalence and experienced in working with them. They were also encouraged to look for the degree to which an expression was established in the language standard, consulting dictionaries and (other) native speakers in cases of doubt. In these circumstances, a significant bias from the source language expression is hardly probable. Although Tosi (2001: 244–262) observes such influences in Italian translations produced at the EU, it should be noted that these are typically produced under high time pressure. Moreover, the awareness that the results will be used for linguistic research increases the urge to take a well-founded decision in the choice of construction.

Therefore, it seems safe to dismiss translator's bias as a major factor in explaining the difference between the predictions made in (6) and the actual figures found in §3. Although a certain degree of bias cannot be entirely excluded, it is certainly not sufficient to explain the degree of divergence.

A second possible explanation is what I will call *semantic bias*. The idea is that the RA+N construction has a meaning of its own. In the case of compounding, the meaning of the construction is very limited. Compounds are basically combinations of a head and a non-head in which the semantic relation between the two is unspecified by the construction. However, in ten Hacken (2013: 106–108), I discussed cases such as (17).

(17) piosenka o miłości
'song about love_{LOC}', i.e. love song

In (17), a Polish translation of *love song*, occurring in Levi's list in the RDP ABOUT, is given. Whereas the English compound is underspecified for the relation between its components, the Polish translation in (17) has the preposition *o* ('about') making this relation explicit. When I discussed this example with other Polish speakers, they came up with different translations that highlighted other aspects of the relation between *love* and *song* in *love song*. Clearly, whatever the semantic bias, RA+N is not specific in the sense that *o* in (17) is. A better

parallel may be the French use of *à*. This is a preposition with a wide range of meanings. Collins-Robert (1987: *à*) gives 17 main senses. However, in the translation of compounds from Levi's list, it is only used in the types illustrated in (18).

(18)	a.	frein à main	hand brake	USE
	b.	boîte à musique	music box	*y* MAKE *x*
	c.	sauce à la crème	cream sauce	*y* HAVE *x*

The article in (18c) is also typical of only the RDP illustrated in that example. However, the identification of such a range of relations is still not a good parallel for RA+N. The French preposition *à* does not have obvious equivalents in other languages. For an explanation of the effects in Tables 1 and 2 by semantic bias, we need not only a set of meanings associated with the construction, but this set should also be cross-linguistically stable.

When we consider the distribution of RA+N among RDPs and nominalization types, there seem to be certain trends identified in §3 which point in this direction. Thus, the RDP FROM is expressed by means of N+N in all English and German examples and by N+*di*+N in almost all Italian cases. In German, not only FROM but also FOR and MAKE are exclusively expressed by N+N compounds in our sample. This may suggest a kind of implicational hierarchy of semantic relations. Such a hierarchy would be a cline with relations typically expressed by N+N compounds on one end and relations typically expressed by RA+N combinations on the other. A language would then be assigned a particular point on the cline.

There is one major flaw with such a model, however. In English, FROM is invariably expressed by N+N, but all other relations can be either N+N or RA+N. This means that the generalization is only a tendency. Formulating a generalization as a tendency is not an explanation of the observations. The tendency is rather a result of the interaction of a number of different underlying factors. For a proper explanation, these factors should be identified.

Therefore, the role of a semantic bias in the choice of RA+N combinations in contrast to N+N or N+*di*+N is that of an observational generalization. It is not an explanation, but has a status similar to Tables 1 and 2.[3]

As a third option for explaining the cross-linguistic coincidence in the use of RA+N combinations as a compounding construction, I propose to turn to the naming process producing new words. In most cases, new words are formed in order to name new concepts. Here *concept* refers to any meaning that may be named. This implies that function words are not necessarily names for concepts, but lexical words generally are. *New* has to be understood in relation to a speaker. Meanings of words are not represented in performance (as is well-known in the case of corpora), but only in a speaker's competence. When a speaker has no lexical entry for a concept they want to mention, they can either use a descriptive phrase or come up with a new name. It is at this point that the choice between RA+N and N+N plays a role.

In the same way as performance, also naming is the outcome of the complex interaction of many factors. While corpus-based studies can come up with interesting

[3] Marchis Moreno (2018) observes a correlation between the choice of construction and the argument structure realization in Spanish. He gives both *pesca de ballenas* and *pesca ballenera* as possible translations of *whale fishing*, but the former has an event reading and the latter a result reading (2018: 153). This leads to different syntactic possibilities (2018: 113). This observation only applies to a highly restricted subset of the cases investigated here and assessing its applicability to German, English and Italian requires a different methodology than the one based on the translation of compounds. I will leave this for further research.

quantitative generalizations about performance, the explanation of such phenomena requires the identification of underlying factors. In the case of performance, a crucial underlying factor is the speaker's competence. However, whereas competence is individual, because it is realized in a speaker's brain, performance has in general a social, communicative aspect to it. Speakers choose from their linguistic repertoire to come up with utterances or texts reflecting their intentions and their assessment of the situation. Here *situation* includes the knowledge and expectations of the hearer(s) or the readership.

In the case of naming, the role of competence is taken, at least in part, by the word formation system. Of course, word formation is not the only naming mechanism. It competes and interacts with sense extension and borrowing. However, of these mechanisms only word formation can be described by rules. It is on the basis of such considerations that Štekauer (2009) studies context-free compound formation. The aim of such a study is to identify the rules that are used in interaction with other factors to come up with compounds.

Let us now consider how this perspective can be used to account for the remarkable cross-linguistic coincidence in the use of RA+N as a compound form. For the sake of brevity, I will speak about N+N as the alternative, assuming that it is realized as N+*di*+N in Italian. It is unlikely that there are conditions inherent in the rules that determine the choice between RA+N and N+N. If there were, we would observe a much more clear-cut division. I see three good candidates for factors influencing the choice.

First, naming takes place in a context of social interaction. Even if the new name is first written, the author will have the readership of the text in mind. Of course, different authors or speakers have different degrees of communicative talent, but there is at least an unconscious tendency to search for common ground. This tendency is first of all reflected in the choice of language, but also in the register and the background knowledge that is assumed. At this level, the choice between RA+N and N+N is influenced by language-specific preferences. English, German, and Italian are not empirical entities, so that it is not possible to measure the relative importance of RA+N and N+N objectively, but speakers belonging to a speech community will be exposed to RA+N and N+N to different degrees, which determines to what extent one or the other of the constructions is marked. This type of markedness can be studied on the basis of corpora.

A second factor influencing the choice of a name is the position a particular name assigns to the lexical unit in the lexicon. By *lexicon* I refer here to the mental lexicon of an individual speaker. Of course, within a speech community, mental lexicons of different speakers have a good deal of similarity. Otherwise, they would not belong to the same speech community. Lexical entries are connected to each other on the basis of different criteria. They include what Saussure (1916: 170–175) calls *rapport associatifs*, i.e. paradigmatic relations. For any new word, they include formal and semantic similarities to existing entries. These factors influence the choice of a name for a new concept. For a new compound, in particular the relation to other compounds with the same head, other compounds with the same non-head and other compounds with the same relation between the head and the non-head are relevant. The choice of RA+N or N+N in a particular case influences the position in the network because it strengthens the relationships with some other words while increasing the distance to others.

If these two factors were the only ones, it might be possible to predict whether a particular new concept would be more likely to have an N+N or an RA+N name, but there would be no reason to expect any cross-linguistic similarity in such decisions. In order to explain cross-linguistic tendencies, we must assume cross-linguistic influence. It is well

known that most people worldwide are not monolingual. Moreover, it can be expected that being in contact with speakers of other languages tends to increase the chance of coming across new concepts that need naming. In many cases, such concepts will have a name in another language that is known to the speaker naming it in the language they speak at that point. Borrowing is an obvious naming procedure in such circumstances. However, the name in the other language may also influence the choice of a name in more subtle ways. An example of this type of influence is (19).

(19)	a.	gorsaf drenau	train station	Bahnhof
	b.	gorsaf pŵer	power station	Kraftwerk
	c.	gorsaf heddlu	police station	Polizeirevier

In (19), three compounds in Welsh are given with their English and German equivalents. Welsh compounds are left-headed. The use of *gorsaf* as an equivalent to English *station* in all three cases seems straightforward from an English perspective. However, when we consider the German equivalents, it is much less obvious why this should be so. All of the German equivalents are compounds as well, but they have different heads. *Hof* means 'yard', *Werk* means 'plant, factory' and *Revier* means 'territory'. The use of *gorsaf* in all three examples in (19) is not surprising, because all speakers of Welsh also know English.

The influence of other languages in the choice between RA+N and N+N is not restricted to borrowing and does not have to be as direct as in (19) for all individual cases. As soon as some compounds are available in the mental lexicon of a speaker, they will influence the sense in which an RA+N or N+N name for a related concept sounds natural. That is to say, the third factor can feed into the second factor. The social interaction in the first factor listed above means that as soon as there is one speaker in a community who has such an influence from another language in their mental lexicon, they can influence other speakers even independently of their knowledge of other languages.

5. Conclusion

I started this paper with the observation that there are two different views on the position of relational adjectives (RAs) with respect to morphology and syntax. In one view, RA+N combinations are compounds (i.e. morphological constructs), in the other, they are phrases (i.e. syntactic constructs). If we adopt a definition of compounding of the type I argued for in ten Hacken (1994, 2013), which is based on semantic and high-level syntactic properties (e.g. island effects in pronominal reference) rather than on phonological and low-level syntactic properties (e.g. agreement), RA+N combinations are compounds. A plausible hypothesis based on such a classification is that the choice between different compounding constructions is cross-linguistically random. This means that in each language the choice is made independently.

On the basis of an analysis of the translations of Levi's (1978) list of 383 compounds into German and Italian, I showed that this hypothesis, though plausible, is not confirmed by the data. However, I argued that there is no reason to assume that the observed tendency for RA+N combinations to be used for the same concepts cross-linguistically is based on a translator's bias or on a cross-linguistically specified range of meanings assigned to the RA+N construction. Instead, I showed that the observed bias can be explained by an appeal

to the nature of the naming process. In the choice of a name, speakers take into account the mental lexicon they already have. Multilingual speakers will also be influenced by the lexicon of the other languages they have. As multilingual speakers are often the first in a speech community to be in contact with new concepts, they have an advantage as namers in the speech community. When some RA+N names are chosen in this way, language-internal mechanisms continue their influence on the names for related concepts.

Therefore, the hypothesis of randomness is false, but this does not show that the analysis of RA+N combinations as compounds is false. Rather it suggests interesting perspectives on naming in a speech community.

References

Adamska-Sałaciak, Arleta. 2013. Issues in Compiling Bilingual Dictionaries. In Jackson, Howard (ed.), *The Bloomsbury Companion to Lexicography*, 213–231. London: Bloomsbury.

Baker, Mona. 2018. *In Other Words: A coursebook on translation*, 3rd edition. London: Routledge.

BNC. 2007. *British National Corpus*, XML edition. Oxford: University of Oxford. (http://www.natcorp.ox.ac.uk/)

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

Collins-Robert. 1987. *Collins-Robert French-English English-French Dictionary*, 2nd edition. Atkins, Sue & Duval, Alain (eds.). London: Collins & Paris: Robert.

DeReKo. 2005–2018. *Deutsches Referenzkorpus*. Mannheim: Institut für Deutsche Sprache. (www.ids-mannheim.de/DeReKo)

de Saussure, Ferdinand. 1916. *Cours de linguistique générale*, In Bally, Charles & Sechehaye, Albert (eds.), Édition critique préparée par Tullio de Mauro, 1981. Paris: Payot.

Fontenelle, Thierry. 2015. Bilingual Dictionaries: History and development; current issues. In Durkin, Philip (ed.), *The Oxford Handbook of Lexicography*, 44–61. Oxford: Oxford University Press.

Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.

Konecny, Christine & Autelli, Erica. 2015. Scale ripide, mobili e a chiocciola – Steile, roll- und wendeltreppen: Diverse categorie di combinazioni lessicali italiane del tipo “n + agg o sprep” e i loro equivalenti tedeschi – un confronto interlinguistico. In Lavric, Eva & Pöckl, Wolfgang (eds.), *Comparatio delectat II: Akten der VII. Internationalen Arbeitstagung zum romanisch-deutschen und innerromanischen Sprachvergleich*, Innsbruck, 6–8 September 2012, Teil 2, 567–586. Frankfurt am Main: Lang.

Levi, Judith N. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.

Marchis Moreno, Mihaela. 2018. *Relational Adjectives in Romance and English: Mismatches at Interfaces*. Cambridge: Cambridge University Press.

Matthews, Peter H. 1974. *Morphology: An Introduction to the Theory of Word Structure*. Cambridge: Cambridge University Press.

Munday, Jeremy. 2016. *Introducing Translation Studies: Theories and Applications*, 4th edition. London: Routledge.

Nida, Eugene A. 1964. *Toward a Science of Translating, with Special Reference to Principles and Procedures Involved in Bible Translating*. Leiden: Brill.

OED. 2000–2018. *Oxford English Dictionary,* 3rd edition. Simpson, John (ed.). Oxford: Oxford University Press. (www.oed.com)

Radimský, Jan. 2015. *Noun+Noun Compounds in Italian: A corpus-based study*. České Budojěvice: Jihočeská Universitá.

Štekauer, Pavol. 2009. Meaning predictability of novel context-free compounds. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford Handbook of Compounding*, 272–297. Oxford: Oxford University Press.

ten Hacken, Pius. 1994. *Defining Morphology: A Principled Approach to Determining the Boundaries of Compounding, Derivation, and Inflection*. Hildesheim: Olms.

ten Hacken, Pius. 2010. Synthetic and Exocentric Compounds in a Parallel Architecture. *Linguistische Berichte Sonderheft* 17. 233–251.

ten Hacken, Pius. 2013. Compounds in English, in French, in Polish, and in General. *SKASE Journal of Theoretical Linguistics* 10. 97–113.

ten Hacken, Pius. 2014. Delineating Derivation and Inflection. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford Handbook of Derivational Morphology*, 10–25. Oxford: Oxford University Press.

ten Hacken, Pius. 2016. Three analyses of compounding: A comparison. In ten Hacken, Pius (ed.), *The Semantics of Compounding*, 211–232. Cambridge: Cambridge University Press.

ten Hacken, Pius & Muigg, Christina. (In press). Nominal Compounds in German and in Italian. In Konecny, Christine & Autelli, Erica & Abel, Andrea & Zanasi, Lorenzo (eds.), *Lexemkombinationen und typisierte Rede im mehrsprachigen Kontext*. Tübingen: Stauffenburg.

Tosi, Arturo. 2001. *Language and society in a changing Italy*. Clevedon: Multilingual Matters.

Pius ten Hacken
Institut für Translationswissenschaft
Leopold-Franzens-Universität Innsbruck
Herzog-Siegmund-Ufer 15
A-6020 Innsbruck
Austria
pius.ten-hacken@uibk.ac.at

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/06.pdf. ISSN 1336-782X.

Competing suffixes: feminine formation of Hebrew loanwords

Lior Laks, Bar-iLan University, Israel

The study examines cases of variation in feminine formation of Hebrew loanwords, where the same masculine base can take two different feminine suffixes, -it and -a. While most loanwords in Hebrew demonstrate uniformity in feminine formation and take the default feminine suffix -it, the study reveals a set of words that can take both suffixes. What triggers this variation and what blocks it? I argue that it results from the interaction of both semantic and morpho-phonological criteria, and that variation is predictable based on systematic guidelines. On the semantic dimension, words that also take -a have negative meaning, and the use of this vowel is indeed more typical of marking lexical meaning, in addition to grammatical gender. From the morpho-phonological point of view, words that take both suffixes do not have typical non-native structure and they resemble, to some extent, to native Hebrew words that take the suffix -a. The study sheds light of the factors that play a role in morphological variation and the adaptation of loanwords.

Keywords: *feminine formation*, *Hebrew*, *gender*, *variation*, *loanwords*, *morphological adaptation*

1. Introduction

This study examines the criteria that are responsible for variation in feminine formation of Hebrew loanwords. Such variation is demonstrated in (1).

(1) a. hem xošvim še-ani eyze **snob-<u>it</u>**
'they think I am some snob'
http://www.tapuz.co.il/blogs/viewentry/371153

b. ve-hu xošev še-ani eyze **snob-<u>a</u>**
'and he thinks I am some snob'
https://stips.co.il/ask/4988953/%D7%A9%D7%90%D7%9C%D7%94-%D7%9C%D7%A1%D7%A0%D7%95%D7%91%D7%99%D7%95%D7%AA-%D7%90%D7%95-%D7%9E%D7%99-%D7%A9%D7%9E%D7%91%D7%99%D7%9F-%D7%AA%D7%A8%D7%90%D7%A9

(2) hi mamaš **larj-<u>it</u>** / ***larj-<u>a</u>** be-tšekim
'she is really larje with checks'
celebs.walla.co.il/item/2968806

The loanword *snob* 'snob' takes either the suffix *-it* (1a) or *-a* (1b), and both forms are found in similar contexts. In contrast, the loanword *larj* 'large (generous)' (2) does not demonstrate variation and takes only the suffix *-it* (*larj-it* / **larj-a*). It is important that not the majority of Hebrew loanwords take the suffix *-it* in feminine formation, and cases examples like snob-a in (1b) are not very common. However, web searches reveal a set of words that take both suffixes *-it* and *-a*, while other suffixes systematically take only *-it*. Why does such variation occur and while is it blocked in other cases? Such competing forms within a single morphological slot deviate from canonicity, as defined in Corbett (2005, 2007) and they pose a challenge for models that aim to explain why and how speakers select one form and not

another. I argue that the existence of variation (and lack thereof) is predictable based on systematic guidelines, and is based on the interaction of both semantic and morpho-phonological criteria.

This paper is structured as follows. §2 provides background in gender marking in Hebrew with focus on loanwords. In §3, I provide an analysis of the criteria that are responsible for variation in feminine formation of loanwords. Specifically, I will show that loanwords that can also take the suffix *-a* has negative connotation and that the selection of *-a* in such cases marks the negative meaning in the lexicon. Further, loanwords that take *-a* do not have a typical foreign structure and could fit into the type of Hebrew native words that typically take the suffix *-a*. §4 consists of concluding remarks.

2. Feminine formation in Hebrew

2.1 *Grammatical gender and word formation*

Grammatical gender is a core feature in various areas of linguistics and it has drawn a great deal of attention from different perspectives. It is a dominant feature is some languages, while it is absent in many others. The rules if gender assignment vary from being fully transparent is some languages to representing complex systems (see Corbett 2006, 2013 and references therein). Languages with grammatical gender typically have two to four gender classes, but there are some languages with more classes. Nouns can be assigned with different gender classes and adjectives and verbs can have gender agreements, depending on the specific languages. In Hebrew, all nouns have grammatical gender, and adjectives, verbs and pronouns are morphologically inflected fore gender in order to match the gender of nouns.[1] Before turning to feminine formation in Hebrew, some background on words formation in Hebrew is in order.

Word formation in Hebrew relies highly on non-concatenative morphology (Berman 1978, 1987; Bolozky 1978, 1999a; Schwarzwald 1981, 2002; Ornan 1983, 2003; Goldenberg 1985; Ravid 1990; Aronoff 1994, among others). Hebrew verbs are formed only via non-concatenative morphology. The Hebrew verbal system consists of patterns. The pattern indicates the prosodic structure of verbs, their vocalic patterns and their affixes (if any) (Bat-El 2011). For example, the verbs *siper* 'told', *xibek* 'hugged' and *diber* 'spoke' belong to the *CiCeC* pattern and differ only in their consonantal root. Every new verb that enters the language must conform to one of the existing patterns. Noun and adjective formation in Hebrew is in general more varied in its formation strategies in comparison to verb formation. Nouns can be formed in patterns, but are also formed by affixation and other word formation strategies. For example, agent nouns can be formed in patterns like *CaCaC*, e.g. *cayar* 'painter', or by suffixation, e.g. *psanter* 'piano' – *psantr-an* 'pianist', as well as by other strategies.[2] While there are only five verbal patterns, there are between 30–35 nominal and adjectival patterns, that differ from each other in productivity and semantic function. It is important to note that the meaning of Hebrew patterns reflect tendencies, rather than

[1] In some verb forms there is no gender distinction, e.g. third person plural past forms. These cases are irrelevant to this study.

[2] Nouns and adjectives, unlike verbs, can also be borrowed directly from other languages without regarding to templatic structure (see for example, Ravid 1992; Schwarzwald 1998 and references therein).

one-to-one relations. Some patterns host nouns with typical meaning, while other host a variety of types of nouns, but none of the pattern has exclusive meaning (see Berman 1978, 1987; Ravid 1990, 2006; Bolozky 1999a; Schwarzwald 2002, 2009; Berman & Seroussi 2011; Shatil 2014, among many others).

Hebrew inflection is typically marked via affixation. Hebrew has three feminine suffixes: *-a*, *-it*, *-et*.[3] *-a* consists of the unmarked vowel *a* and can be attached to a variety of stems. *-it* is the productive default suffix in acronyms and loanwords. *-et* is the least productive and is not discussed here (Schwarzwald 1984, 2002). Suffix selection is predictable based on semantic and morpho-phonological properties of the base, but is subject to irregularity. For example, when the masculine noun is formed in the *CaCaC* pattern, different bases can take different suffixes. The examples in (3) share *CaCaC* pattern but each form takes a different feminine suffix. It is unclear, for example, why *zamar* 'singer' (3a) takes *-et,* while *sapar* 'barber' (3b) takes *-it*. However, both of them are occupational nouns and they never take *-a*. In contrast, *ravak* 'single' (3c) is not occupational and it takes *-a* and not *-et* or *-it*. These examples show that even though suffix selection is to some extent arbitrary, it also follows some systematic guidelines. In this case, there is a semantic condition that determines in which cases variation is possible. In case of occupational nouns, both suffixes *-et* and *-it* can be selected (and the selection between them is unpredictable), while in other bases *-a* is selected.

(3) <u>Feminine formation of CaCaC forms</u>

a.	zamar	–	zamer-et / *zamar-it / *zamar-a	'singer'
b.	*sapar*	–	*sapar*-it / *saper-et / *sapar-a	'hairdresser'
c.	ravak	–	ravak-a / *ravek-et / * ravak-it	'single'

However, in most cases suffix selection is based on properties of the base. Schwarzwald (1984, 1991, 2002; Faust 2013) proposes a model that predicts the selection of each suffix. This model is based mostly on phonological and morphological features of the base. For example, words in the *CaCCan* pattern systematically take *-it* (e.g. *daykan – daykan-it* 'punctual').[4] Monosyllabic words and words whose last vowel is *o*, *u* and *i* typically take the suffix *-a* (e.g. *pil – pil-a* 'elephant', *xamud – xamud-a* 'cute'). Words whose last vowel is *e* typically take the suffix *-et* (e.g. *meratek – meratek-et* 'fascinating'). Acronyms and loanwords take the suffix *-it.* For example, *mankal* 'CEO' stand for ***men**ahel **kl**ali* ('general manager') and its feminine form is *mankal-it*. These are only some of Schwarzwald's generalizations, and they provide the relevant picture of the complexity of feminine formation in Hebrew. These are all examples of strong tendencies that apply systematically, though they are subject to a few exceptions.

Cases of variation in Hebrew feminine formation are not new, and have been addressed in different studies on Hebrew native words. It is found in some present participle forms (Schwarzwald 1982, 1991), e.g. *kore – koret* / *korʔa* 'reads', in nationality related adjectives that denote humans, e.g. *yehudi – yehudit* / *yehudiya* 'Jewish', in child language (Berman 1978, 2003; Ravid 1995), e.g. *šakran – šakranit* / *šakraniya* 'liar', and in present forms of weak verbs (Schwarzwald 1977, 1984; Asherov & Bat-El 2016), e.g. *mexate – mexatet* / *mexata* 'disinfects'. Asheron & Bat-El (2016) examine the competition between *-a*

[3] The suffix *-ot* is also attached to a small set of nouns, e.g. *ax* 'brother' – *axot* 'sister'. In addition, some suffixes have allomorphs. These cases are irrelevant to the current study and therefore will not be discussed.
[4] Faust (2013) that the suffix *-it* is a complex suffix composed of two elements *-i* and *-t* and.

and -*et*, distinguishing between local default and global default. They argue that -*et* is the local default of present feminine singular forms, while -*a* is the global default in feminine formation. In cases where the assignment of -*et* is blocked, for different reasons, -*a* takes over. In general, gender marking in Hebrew has been addressed from various perspectives (see Muchnik 2015 and references therein). This study addresses a specific aspect of gender marking, namely, cases where loanwords can take more than one suffix.

2.2 *Loanwords in Hebrew*

Hebrew is subject to the influence of foreign languages in general, and English in particular (Kutscher 1956; Rosén 1956, 1977; Blanc 1957; Blau 1976; Berman 1978; Ornan 1983, 2003; Ravid 1992; Nir 1993; Kaddari 1993; Agmon-Fruchtman & Alon 1994; Schwarzwald 1995, 1998, 1999, 2013; Izre'el 2002; Eldar 2007; Rosenhouse & Fisherman 2008, among many others). Loanwords undergo two types of adaptation (Schwarzwald 1998, 2002, 2009). Phonological adaptation concerned with adaptation to the phonetic inventory and the prosodic constraint of the native language. It is to a great extent obligatory. Non-native segments in the donor language undergo adaptation in order to comply with the native phonological restrictions. Loanwords also undergo adaptation to the syllable structure of the native language in order to comply with prosodic constraint on syllable structure. Morphological adaptation is based on adaptation to word formation strategies of the native language (see Ravid 1992; Schwarzwald 1998, 2002, 2009, 2013). Ravid (1992) accounts for the adaptation strategies of three lexical categories of loanwords in Hebrew: verbs, nouns and adjectives. She offers a model that allows to predict what adaptation strategy (or lack thereof) is more likely to be selected based on the lexical category of the loanword, ranging from complete adaptation of verbs, hardly any adaptation in most nouns and partial adaptation of adjectives. Schwarzwald shows in a series of studies (1998, 2002, 2013) the systematic behavior of loanwords. Unlike native words, loan words do not undergo alternations when they are inflected for gender and number, the selection of gender and number suffixes is highly predictable and systematic and, again unlike native words, the location of stress is preserved. With respect to feminine formation, loanwords systematically take -*it* as a default suffix, e.g. *bos – bosit* 'boss', *partnter – partmerit* 'partner', *eks – ekit* 'ex- (husband, wife, etc)'. I now turn to a case study of variation in feminine formation of loan words, where the same base can take the suffix -*a* in addition to -*it*. Although such cases are less common, I will show that when they occur, they follow systematic guidelines that shed light on the criteria of word formation in general and specifically the morphological adaptation of loanwords.

3. Competing suffixes in feminine formation of loanwords

As noted in §2.2 above, the majority of Hebrew loanwords take the suffix -*it* in feminine formation. Some loanwords also take the suffix -*a*, but there are no cases where only -*a* is selected. More examples of such variation are presented in (4) and (5) below, where the loanwords *xnun* 'nerd' and *debil* 'idiot' can take both -*it* (4a, 5a) and -*a* (4b, 5b) and occur in similar contexts.

(4) a. hayiti kazot **xnun-it** in those
'I was such a nerd those days'
https://www.ted.com/talks/keren_elazari_hackers_the_internet_s_immune_system/transcript?language=he

b. bederex-klal lo hayiti kazot **xnun-a**
'I usually wasn't such a nerd'
http://hportal.co.il/index.php?act=fanfiction&showpic=4285&showchap=1

(5) a. eze **debil-it** ani, ze mamaš be-roš ha-amud
'I am such so stupid idiot, it really on top of the'
page'http://www.tapuz.co.il/forums/viewmsg/183/28634756/%D7%94%D7%9E%D7%95%D7%9E%D7%97%D7%99%D7%9D/%D7%94%D7%95%D7%A8%D7%99%D7%9D_%D7%91%D7%98%D7%99%D7%A4%D7%95%D7%9C%D7%99_%D7%A4%D7%95%D7%A8%D7%99%D7%95%D7%AA

b. at lo mevina eze **debil-a** ani
'you don't understand what an idiot I am'
http://www.kipa.co.il/community/show/9342048

Cases in which *-a* is also selected are not highly common, but I will show that they are predictable to a great extent. The paper is based on web searches in which the selection of both suffixes was compared. Let us compare the loan words in Table 1. The items in a were found mostly with *-it* and a few instances were also found with *-a*. In contrast, the items in b were found only with *-it*.

Table 1. Feminine formation of Hebrew loanwords

a. Words with both feminine suffixes *-it* and *-a*		**b. Words with the suffix *-it* only**	
snob	'snob'	larj	'large (generous)'
debil	'stupid'	bos	'boss'
xnun	'nerd'	barmen	'barmen'
farš	'not cool'	inteligent	'intelligent'
maafan	'lame'	hiler	'healer'
babun	'monkey (ugly)'	slizi	'sleazy'
šlux/šlox	'slopy'	homles	'homeless'
dos	'ultra-orthodox' (derogatory)	trabel maker	'trouble maker'
xnevec	'nerd'	kul	'cool'

Using web-searches raises the question to what extent the examples represent the actual usage. How many examples should be found in order to determine that variation and doublet formation indeed exit? While there is obviously no absolute answer to this question, it is important to emphasize that the analysis is not based on the number of instances that are found, but mainly on the contrast between what is found, to different extents, and what is not found at all. Loanwords like *snob* 'snob' and *debil* 'stupid' mostly take the suffix *-it*, while there are only a few instances with the suffix *-a*. However, other adjectives like *larj* 'large' and *kul* 'cool' take only *-it* and no instances with the suffix *-a* were found. Although the use of *-a* with loanwords is marginal, I argue that the contrast between the two groups of items in

not a coincidence and can be predicted systematically. Such contrast can be revealed mostly by conducting web searches. Relying on spoken corpus can also provide relevant instances, but the chances of finding enough examples are rather low.

Why then, can some loan words take both feminine suffixes, while most of them take only -*it*? I argue that the fact that some loan words can also take the feminine suffix -*a*, even to a small extent, is not arbitrary, but stems from the interaction of semantic and morpho-phonological criteria. I now turn to examining the factors that bring about the selection of -*a*.

3.1 *Semantic condition*

Let us compare again the loanwords that take both -*it* and -*a* (6a) with words that take only -*it* (6b) above. The picture that emerges is that almost all loan that take -*a*, in addition to -*it*, have negative meaning, e.g. *debil* 'idiot', *farš* 'jerk' and *xnun* 'nerd'. Examine for example, the loanwords *bos* 'boss' and *dos* 'ultra orthodox'. The former only takes the -*it* suffix (6a), while the latter can take both -*it* (6b) and -*a* (6c). Both words are monosyllabic and share the same vowel *o*. There seems to be no phonological or morphological factor that would explain the difference in their feminine formation. I argue that the difference is semantic and specifically, it lies in the existence of negative connotation and lack thereof. *bos* 'boss' has a neutral meaning, while *dos* is used to relate to an ultra orthodox in a derogatory way.

(6) a. kol axat meitanu hi **bos-it / *bos-it** šel acma
'each one of us is her own boss'
http://ambarchia.wixsite.com/shvilhamishpaha/single-post/2015/05/10/% D7%D7%99

b. axoti ha-gdola mamaš kazot **dos-it**
'my elder sister is really such a Dos'
https://www.askpeople.co.il/question/221009

c. kedey še-lo yaxševu še-anu kazot **dos-a**
'so that (they) don't think I am such a Dos'
https://www.kipa.co.il/community/show/4891642/

Note that not all words with negative meaning can take both suffixes. The word *luzer* 'loser', for example, is negative, but can take only -*it* (*luzer-it*/ **luzer-a*). As will be explain in 3.2, this can also be predicted systematically. The claim made here is that only loanwords with negative meaning can take -*a* in addition to -*it*, but not necessarily all of them.

The effect of the negative meaning in suffix selection is even attested in cases where the same word takes only -*it* when it has a neutral meaning, but can also take -*a* only with a negative meaning. Examine the words *babun* 'baboon'. In case it denotes a female baboon, only -*it* is used (7). In contrast, when it has a derogatory meaning of an ugly person, it can take both -*it* (8a) and -*a* (8b).

(7) sagit me-ha-safari be-ramat-gan tesaper lanu al **babun-it** / ***babun-a** jinjit xadaša
'Sagit from the Ramat-Gan safari will tell us about a new red hair baboon'
http://www.iba.org.il/zap/zap.aspx?classto=InnerKlali&type=247&entity=711321&topic=888

(8) a. hi medaberet kše-yeš la mašehu baal erex lomar, lo mekaškešet kmo eze **babun-it**
'she speaks when she has something valuable to say, doesn't babble like some baboon'
http://www.mako.co.il/nTalkbacksPage/vgnextchannel=4e4652692dddb210VgnVCM2000002a0c10acRCRD&page=18&vgnextoid=1e38a7fcd9a9d210VgnVCM2000002a0c10acRCRD&

b. ani kozot **babun-a** lo fotogenit
'I am such an unphotogenic baboon'
http://israblog.nana10.co.il/blogread.asp?blog=442924

How is the selection of the feminine suffix -*a* related to negative meaning? I claim that -*a* is more accessible as a marker of such meaning, and as a result, it is attached to other words, competing with the default suffix -*it*. When selecting -*a*, speakers lexically mark words as negative. Why is it so? In order to account for that, some background on gender marking in Hebrew is in order. In general, several studies have shown that Hebrew is a gender biased language. To put is very generally, there are many cases in which language usage and specifically gender marking reflects some aspects of the society's attitude towards women (see for example, Ariel 1982, 1988; Ariel & Giora 1998; Giora 2001; Tobin 2001; Rosental 2015; Muchnik 2003, 2007, 2015; Valden 2005; Livnat 2006, among others).

Examine the Hebrew native words *par* 'bull' and its feminine form *para* 'cow'. Only the feminine form *para* has the metaphoric meaning of 'fat', and it is used to depict overweight women in a derogatory way (Ariel 1982). In contrast, the masculine form *par* is not used to refer to overweight men. Ariel (1982) shows systematically that in many cases, words and expressions with negative connotation are used exclusively for women. This is not related to the use of one feminine suffix or another, but to the fact that Hebrew has feminine forms with negative meaning that is not shared with masculine forms. In most of Ariel's examples, the feminine forms with negative meaning end with the suffix -*a*. There are also native words with negative meaning that exist only in the feminine form, e.g. *zona* 'prostitute' (**zone*).

In addition, Hebrew has feminine loanwords with no masculine base, and the majority of them have negative meaning e.g. *pustema* 'blockhead' (**pustem*), *kunefa* 'ugly' (**kunef*). More such examples are presented in (9).

(9) Feminine loanwords with no masc. base

pustema	'blockhead'
fakaca	'shallow and stupid person'
kunefa	'ugly'
jifa	'filthy, ugly'
primadona	'prima dona'

These words in (9) were borrowed into Hebrew as is from different languages, where the ending vowel does not function as a suffix but as part of the stem. In some cases, they could also relate to men, but they are typically feminine. There are some cases of back formation, where the *a* vowel is dropped and the remaining stem is used as a masculine form, but such case are rare and have entered the language late. The point made here is that there is a set of -*a* ending loanwords are typically feminine and the majority of them has negative

meaning.[5] The fact that most of these words end with -*a* and have negative meaning makes the vowel *a* more typical of negative meaning and associated with it. Such loanwords are stored in lexicon, making the negative meaning more accessible. This echoes with Schwarzwald's (1991, 2002) findings on the role of the lexicon in feminine formation. Schwarzwald shows cases where properties of the masculine base have to be specified in the lexicon so that the right feminine suffix is selected. The case examined in this study is one in which the feminine suffix itself marks a specific semantic feature.

It follows that the feminine suffix -*a* is less semantically neutral, in comparison to -*it*, as it is more accessible to the negative meaning. I argue that as a result of the negative association of this suffix, especially in loanwords, it is also attached other loanwords with negative meaning, marking the feminine form as typically negative. I claim that such marking is assumed to apply in the lexicon, in contrast to regular gender marking which is typical inflection that is assumed to be syntactic. As noted in §2.2, all loanwords can take the suffix -*it*, which is the default suffix for loanwords. There is a set of words that can receive additional gender marking that also bears the semantic feature of negative meaning.

The tendency to select a suffix that has the vowel *a* in order to lexically mark it with additional meaning is not surprising. In general, *a* has a morpho-lexical status in Hebrew. It is the most frequent vowel in word formation processes (Plada 1959; Bolozky & Becker 2006) and it is part of various word formation processes. Bolozky (1999a, 2003), Schwarzwald (2002, 2012) and Schwarzwald & Cohen-Gross (2000) show that -*a* is the most common vowel in Hebrew patterns, and Bat-El (1994) and Bolozky (1999b) show that it is the default vowel in acronym formation. I claim here that the use of -*a* in feminine formation of loanwords does not only play a grammatical role of gender marking, but it also labels these words as typically negative.[6] Though gender is considered inflection, which typically applies in the syntax (see Perlmutter1988; Anderson 1992; Corbett 2013; Štekauer 2015; Stump 2016, among others), marking the negative meaning applies in the lexicon because of the accessibility of -*a*. Semantic marking is more typical of derivation than inflection, and assuming that derivation typically applies in the lexicon, the morphological mechanism marks loanwords as negative with a typical vowel that is used in derivation.

The picture that emerges is that the two feminine suffixes -*it* and -*a*, do not completely overlap on their function. While -*it*, which is the default suffix of loanwords, has only a grammatical role of gender marking, -*a* also has a semantic role of marking a negative meaning. The distinction between the two suffixes resembles, to some extent, the distinction between inherent and contextual inflection (Booij 1996, 2006). In general, inherent inflection is not required by the syntactic context, although it can be relevant to the syntax. Examples of inherent inflection are number and gender for nouns and tense and aspect for verbs. Contextual inflection is dictated by syntax, e.g. agreement markers for verbs and adjectives. I argue that the use of -*a* as a feminine marker for loanwords is "more inherent" than the use of -*it* in the sense that it provides a morphological marker of the negative meaning and does not only denote grammatical gender.

---

[5] There are some rare counter examples of -*a* ending loanwords without a negative meaning, e.g. *balerina* 'ballerina'.

[6] Note that this does not mean that only -*a* can contribute to the lexical meaning. The suffix -*it* can also be used as a derivational suffix with different meanings, for example, diminution (e.g. *mapa* 'map' – *mapit* 'napkin') (see Bolozky 1994; Muchnik 1996; Schwarzwald 2002; Faust 2013). However, -*a* is more typical of semantic marking.

So far, I have discussed the semantic condition for -*a* selection, i.e. the existence of negative meaning. However, not all loanwords with negative meaning take -*a*. I now turn to the morpho-phonological condition that dictates which loanwords with negative meaning can take -*a*.

3.2 *Morpho-phonological conditions*

I have shown that only loan words with negative meaning can take the suffix -*a* in addition to -*it*. This semantic condition is the primary condition for variation in suffix selection and the current study shows that it is exception free. However, not all loanwords with negative meaning can take -*a*. Loan words like *luzer* 'loser' (10) and *klules* 'clueless' (11) can only take -*it*.

(10) ani pašut margiša **luzer-it / *luzer-a**
'I simply feel like a loser'
http://hadas123.bloger.co.il/page/4/

(11) az ani dey **klules-it / *klules-a** ba-inyan
'so I am quite clueless in the matter'
http://forums.tipa.co.il/forum-10/msg-37269.html

Why do loanwords with negative meaning like *debil* 'idiot' and *snob* 'snob' can take both suffixes, while loanwords like *luzer* 'loser' and *klules* 'clueless' cannot? I argue that only loanwords with no typical non-native features can take both suffixes. Such words could be perceived as similar to Hebrew native words, and such native words can indeed take the suffix -*a*. As detailed below, there are two main morpho-phonological criteria that make loanwords typically non-native. Typical non-native structure brings about the selection of the default suffix -*it* and the selection of -*a* is blocked.

Loanwords that take only -*it* can have typical non-native structure if they consist of non-native suffixes like -*er* (12a) and -*less* (12b).

(12) Feminine formation of loanwords with non-native suffixes

a. -er			
tizer	–	tizer-it / *tizer-a	'teaser (provocative)'
maxer	–	maxer-it / *maxer-a	'kon'
luzer	–	luzerit / *luzer-a	'loser'
hipoxonder	–	hipoxonder-it / *hipoxonder-a	'hypochondriac'
b. -less			
hoples	–	hoples-it / *hoples-a	'hopleless'
klules	–	klules / *klules-a	'clueless'
homles	–	homles-it / *homles-a	'homeless'

Because of these foreign suffixes speakers identify these words as typical loanwords, and as a result they employ the default feminine suffix -*it*.

In addition, Hebrew words are typically disyllabic or monosyllabic, unless they consist of a suffix (e.g. *xašmal* 'electricity' – *xašmal-i* 'electronic'). Words without suffixes

that exceed two syllables, e.g. *paranoid* 'paranoid', are also perceived as atypical Hebrew words and again, take only the *-it* suffix (13).

(13)	pedofil	–	pedofil-it / *pedofil-a	'pedophile'
	paranoid	–	paranoid-it /* paranoid-a	'paranoid'
	socyopat	–	socyopat-it / *socyopat-a	'sociopath'
	mizantrop	–	mizantrop-it / *mizantrop-a	'misanthrope'

The picture that emerges is that loanwords with negative meaning that take only *-it* either have a non-native suffix or they exceed two syllables, which is less typical for Hebrew word. In contrast, loanwords that take *-a* do not have a particular foreign structure and some of them resemble native words. They are mostly monosyllabic words or disyllabic words with final stress. The Hebrew lexicon consists of such words that do not belong to a particular pattern, but are also perceived as part of the basic vocabulary of native speakers. For example, a loanword like *dos* 'ultra orthodox' is monosyllabic and has negative meaning, and can take both feminine suffixes (*dosa*/*dosit*) because it resembles the structure of native monosyllabic Hebrew words like *gur* 'cub' that take systematically only the suffix *-a* (*gura*), regardless of their meaning (Schwarzwald 1982, 2002). These loanwords demonstrate variation in feminine formation as on the one hand, they can take the typical suffix of loan words, and on the other hand, they can behave like similar native Hebrew words and that take *-a*. Such loanwords act as an intermediate category on a continuum between typical native words and typical non-native words. Being an intermediate category also triggers variation. In contrast, loanwords words like *socyopat* 'sociopath' and *klules* 'clueless' do not resemble any type of native words and therefore they only take the suffix that is typical for loanwords.

As shown in (14), there seems to be a continuum of types of words that behave differently with respect to feminine formation.

(14) <u>The continuum of suffix selection in feminine formation</u>

a. Typical Hebrew Structure		**b. Intermediate category**		**c. Typical non-native Structure**	
-a/-it/-et		**-it / -a**		**-it only**	
CaCiC		snob	'snob'	tiz**er**	'teaser'
ʔamic – ʔamica	'brave'	farš	'jerk'	hoples	'hopeless'
-an					
aclan – aclanit	'lazy'	debil	'stupid'	socyopat	'sociopath'
CoCeC					
boged – bogedet	'traitor'				
monosyllabic					
pil – pila	'elephant'				

On one edge of the scale (14a) we have native Hebrew words with typical Hebrew structure, mostly a pattern or suffix that select one of the feminine suffixes -*a*, -*it* and -*et* based on the system of rules proposed by Schwarzwald (1984, 2002). There is some degree of irregularity is suffix selection in this group as well, as discussed in 2.1, but in most cases it is possible to predict which suffix is selected. This groups also consists of native words with no pattern of suffix, where the feminine form is also predictable based on Schwarzwald's model. On the other edge of the scale (14c), we have loanwords with typical non-native structure: either a non-native suffix or words with more than two syllables. Words in this group take only -*it* in feminine formation, as this is the default suffix for loanwords. Between these two edges, we have an intermediate group (14b). On the one hand, these words are non-native and hence tend to take the suffix -*it*, but on the other hand they do not have a typical non-native structure. These are mostly mono-syllabic words that could be regarded as native words. As a result, they could behave like native Hebrew monosyllabic words that mostly take the suffix -*a*, and therefore they take this suffix as well in addition to -*it*. The structure of these words allows the existence of variation in feminine formation.

Note that Hebrew also have borrowed derivational suffixes like -*ist* and -*nik*. These suffixes have been established as part of the language and are attached to native Hebrew words (Schwarzwald 1998, 2002). Hebrew words with these suffixes take only -*it* as a feminine suffix (15). It is therefore not surprising that loanwords with a typical foreign suffix take only -*it*, which functions as a default feminine suffix.

(15) <u>Hebrew native words with borrowed suffixes</u>

kibuc 'Kibbutz'	–	mošav-**nik**	– mošav-**nik-it**	'Kibbutz member'
bicuʔa 'execution'	–	bicuʔ-**ist**	– bicuʔ-**ist-it**	'go getter'

It can be concluded, that typical foreign structure brings about the selection of the default suffix -*it*, while non-typical foreign structure can allow the selection of -*a* as well, subject to the semantic condition discussed in §3.1.

4. Conclusions

This study examined variation in the formation of feminine forms of loanwords in Hebrew. While the default feminine suffix of loanwords is -*it*, there are some cases of loanwords that take -*a* as well. It has been shown that such cases follow systematic guidelines based on two conditions. The first condition is semantic. Only loanwords with negative meaning take -*a*. The vowel -*a* has a lexical status in Hebrew as it plays a major role in word formation. By attaching -*a* to loanwords, speakers do not mark them only as feminine, but also mark them as negative. The second condition is morpho-phonological. Out of the loanwords with negative meaning, those that take also -*a* are mostly words with no typical non-native structure. These are mainly monosyllabic and disyllabic words, that can also be perceived as native words. Such words show partial resemblance to native words that take -*a* and therefore they can take this suffix as well.

The study sheds light on the role of status of gender with respect to morphological change. It adds to previous accounts of morphological variation and change, which result in doublet formation (Thornton 2012a, 2012b; Aronoff 2017, among others). The study enables

to reveal the motivation for such change in gender marking from both morpho-phonological and semantic perspectives. The study also intertwines variation with the degree of integration of loanwords. This provides direct access to word formation and shows how different types of criteria are taken into consideration.

References

Agmon-Fruchtman, Maya & Alon, Emanuel. 1994. *Chapters in the history of the Hebrew language, the modern division. Unit 8: the revival of Hebrew*. Tel-Aviv, Israel: Open University publishers. (In Hebrew.)

Anderson, Stephen. R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Ariel, Mira. 1982. Women in the hands of the language: Three aspects of women and language. *Noga* 7.

Ariel, Mira. 1988. Female and male stereotypes in Israeli literature and media: Evidence from introductory patterns. *Language and Communication* 8(1). 43–68.

Ariel, Mira & Giora, Rachel. 1998. Power and cooperation: An analysis of impositive speech acts: Gender biases in the Israeli cinematic discourse. In Gertz, Nurith & Lubin, Orly & Ne'eman, Judd (eds.), *Fictive aspects of Israeli movies*, 179–204. Tel-Aviv: Open University.

Aronoff, Mark. 1994. *Morphology by Itself: Stems and Inflectional classes* (Linguistic Inquiry Monographs 22). Cambridge, MA: MIT Press.

Aronoff, Mark. 2017 (submitted). Competitors and alternants. To appear in a volume of *selected papers from International Morphology Meeting 17*, Vienna, 2016.

Asherov, Daniel & Bat-El, Outi. 2016. Multiple defaults: Hebrew feminine *-et* and *-a* in the present tense. *Morphology* 26. 399–423.

Bat-El, Outi. 1994. The optimal acronym word in Hebrew. In *Proceedings of the 1994 Annual Conference of the Canadian Linguistic Association*. In Koskinen, Paivi (ed.), *Toronto Working Papers in Linguistics*. 23–37.

Bat-El, Outi. 2011. Semitic templates. To appear in van Oostendrop, Marc & C. Ewen, Colin J. & Hume, Elizabeth V. & Rice, Keren (eds.), *Blackwell Companion to Phonology*.

Berman, Ruth A. 1978. *Modern Hebrew Structure*. Tel-Aviv: University Publishing Projects.

Berman, Ruth A. 1987. Productivity in the lexicon: New-word formation in Modern Hebrew. *Folia Linguistica* xxi (2–4). 425–461.

Berman, Ruth A. 2003. Children's lexical innovations: developmental perspectives on Hebrew verb-structure. In Shimron, Joseph (ed.), *Language Processing and Language Acquisition in a Root-based Morphology*, 243–291. Amsterdam: John Benjamins.

Berman, Ruth A. & Seroussi, Batia. 2011. Derived nouns in Hebrew: Structure, meaning, and psycholinguistic perspectives. In Fiorentino, Giuliana & Comrie, Bernard (eds.), *Nouns and nominalizations cross-linguistically (Special issue of Italian Journal of Linguistics)*. 105–125.

Blanc, Haim. 1957. Hebrew in Israel: Trends and problems. *Middle East Journal* 11. 397–409.

Blau, Joshua. 1976. *The revival of Hebrew and Literary Arabic*. Jerusalem: The Hebrew Language Academy. (In Hebrew.)

Bolozky, Shmuel. 1978. Word formation strategies in MH verb system: denominative Verbs. *Afroasiatic Linguistics* 5. 1–26.

Bolozky, Shmuel 1994. On the formation of diminutives in Modern Hebrew morphology. *Hebrew Studies* 35. 47–63.

Bolozky, Shmuel 1999a. *Measuring Productivity in Word Formation: The Case of Israeli Hebrew*. Leiden: Brill.

Bolozky, S. 1999b. On the Special Status of the Vowels *a* and *e* in Israeli Hebrew. *Hebrew Studies* 40. 233–250.

Bolozky, Shmuel 2003. Phonological and morphological variation in spoken Hebrew. In Hary, Benjamin H. (ed.), *Corpus Linguistics and Modern Hebrew: Towards the Compilation of the Corpus of Spoken Hebrew (CoSIH)*, 19–156. Tel Aviv: Tel Aviv University Press.

Bolozky, Shmuel &Becker, Michael. 2006. Living lexicon of Hebrew nouns. Ms., University of Massachusetts Amherst. (http://becker.phonologist.org/LLHN)

Booij, Geert. 1996. Inherent vs. contextual inflection and the split morphology hypothesis. In Booij, Geert & van Marle, Jaap (eds.), *Yearbook of Morphology 1995*, 1–16. Dordrecht: Kluwer.

Booij, Geert. 2006. Inflection and derivation. In Brown, Keith (ed.) *Encyclopedia of Language & Linguistics.* New York: Elsevier.

Corbett, Greville G. 2005. The canonical approach in typology. In Frajzyngier, Zygmunt & Hodges, Adam & Rood, David S. (eds.), *Linguistic Diversity and Language Theories*, 25–49. Amsterdam: John Benjamins.

Corbett, Greville G. 2006. Gender, Grammatical. *Encyclopedia of Language & Linguistics*. 749–756. Oxford: Elsevier.

Corbett, Greville G. 2007. Canonical typology, suppletion, and possible words. *Language* 83(1). 8–41.

Corbett, Greville G. 2013. *The Expression of Gender*. Berlin: Mouton De Gruyter.

Eldar, Ilan. 2007. Following the language revival: The beginning of spoken Hebrew. *Ha-ivrit Ve-Ahyoteha* 6–7. 39–53. (In Hebrew.)

Faust, Noam. 2013. Decomposing the Feminine suffixes of Modern Hebrew. *Morphology* 23(4). 409–440.

Giora, Rachel. 2001. Theorizing gender: Feminist awareness and language change. In Baron, Bettina & Kotthoff, Helga (eds.), *Gender in Interaction*, 329–347. Amsterdam: John Benjamins.

Goldenberg. Gideon. 1985. On Verbal Structure and the Hebrew verb. In Bar-Asher, Moshe (ed.), *Language Studies 1*, 295–348.

Izre'el, Shlomo. 2002. The emergence of Spoken Hebrew in Israel. *Teuda* 18. 217–238. (in Hebrew)

Kaddari, Zvi. 1993. Foreignness in Hebrew. *Leshonenu La'am* 44. 99–109.

Kutscher, Edward Yechezkel. 1956. Modern Hebrew and Israeli Hebrew. *Conservative Judaism* 10(3). 28–45.

Livnat, Zohar. 2006. Gender on-line in Hebrew: New technology, old language. In Thune, Eva-Maria Leonardi, Simona & Bazzanella, Carla (eds.), *Gender, language and new literacy: A multilingual analysis*, 169–181. London and New York: Continuum.

Muchnik, Malka. 1996. A little suffix growing up. *Helkat Lashon* 23. 13–22. (In Hebrew.)

Muchnik, Malka. 2003. Gender characteristics in political writing. In Toury, Gideon & Shahar, Ben R. (eds.), *Hebrew – a Living Language* III, 207–220. Tel-Aviv: The Porter Institute. (In Hebrew.)

Muchnik, Malka.2007. Gender differences in slang expressions as social communication. *Social Issues in Israel* 3. 5–20

Muchnik, Malka. 2015. *The Gender Challenge of Hebrew*. Leiden: Brill

Nir, Refael. 1993. *Word Formation in Modern Hebrew*. Tel Aviv: The Open University. (In Hebrew.)

Ornan, Uzzi. 1983. How is the Hebrew word formed? In Bar-Asher, Moshe & Dotan, Aron & D. Tene & G.B. Sarfatti (eds.), *Language Studies*, 13–42. Jerusalem: Magnes. (In Hebrew.)

Ornan, Uzzi. 2003. *The Final Word*. Haifa: Haifa University. (In Hebrew.)

Perlmutter, David. 1988. The split-morphology hypothesis: evidence from Yiddish. In Hammond, Michael & Noonan, Michael (eds.), *Theoretical Morphology: Approaches in Modern Linguistics*, 79–99. Orlando: Academic Press.

Plada, N. 1959. Phoneme frequency in Hebrew. *Leshonenu* 23. 235–242. (In Hebrew.)

Ravid, Dorit. 1990. Internal structure constraints on new-word formation devices in Modern Hebrew. *Folia Linguistica* 24. 289–347.

Ravid, Dorit. 1992. The absorption of foreign words in the noun and verbal systems. In M. Muchnik (ed.), *Foreign Influences on Contemporary Hebrew*, 11–20. Tel-Aviv: The Open University of Israel.

Ravid, Dorit. 1995. *Language change in Child and Adult Hebrew: A Psycholinguistic Perspective*. Oxford: Oxford University Press.

Ravid Dorit. 2006. Word-level morphology: A psycholinguistic perspective on linear formation in Hebrew nominals. *Morphology* 16. 127–148.

Rosén, Haim B. 1956. *Our Hebrew*. Tel-Aviv: Am Oved. (In Hebrew.)

Rosén, Haim B. 1977. *Contemporary Hebrew*. The Hague: Mouton.

Rosenhouse, Judith & Fisherman, Haya. 2008. Hebrew: Borrowing ideology and pragmatic aspects in a modern(ised) language. In Rosenhouse, Judith & Kowner, Rotem (eds.) *Globally speaking: Motive for adopting English vocabulary in other languages*. *Multilingual Matters*, 121–144.

Rosental, Ruvik. 2015. *The reflection of military hierarchy in the Israel Army language*. Israel: Bar-Ilan University. (Doctoral dissertation)

Schwarzwald, Ora R. 1977. The weak verb representation in the Modern Hebrew lexicon. *Hebrew Computational Linguistics* 12. 25–36. (In Hebrew.)

Schwarzwald, Ora R. 1981. *Grammar and Reality in the Hebrew Verb.* Ramat Gan: Bar Ilan University Press. (In Hebrew.)

Schwarzwald, Ora R. 1982. Feminine Formation in Modern Hebrew. *Hebrew Annual Review* 6. 153–178.

Schwarzwald, Ora R. 1984. Analogy and Regularization in Morphophonemic Change: The Case of the Weak Verbs in Post-Biblical Hebrew and Colloquial Modern Hebrew. *AfroAsiatic Linguistics* 9(2). 87–100.

Schwarzwald, Ora R. 1991. Lexical weight in Hebrew inflectional feminine formation. *Semitic Studies in Honor of Wolf Leslau on the Occasion of his 85th Birthday*. 1409–1425. Otto Harrssowitz: Wiesbaden.

Schwarzwald, Ora R. 1995. The components of the Modern Hebrew lexicon: The influence of Hebrew classical sources, Jewish languages and other foreign languages on Modern Hebrew. *Hebrew Linguistics* 39. 79–90. (In Hebrew.)

Schwarzwald, Ora R. 1998. Word foreignness in Modern Hebrew. *Hebrew Studies* 39. 115–142.

Schwarzwald, Ora R. 1999. Trends of foreign influence on Modern Hebrew. In Targarona Borras, Jufit & Saenz-Badillos, Ángel (eds.), *Jewish Studies at the Turn of the 20th Century*, 81–89. Leiden: Brill.

Schwarzwald, Ora R. 2002. *Hebrew Morphology: 12 Study units*. Tel Aviv: The Open University. (In Hebrew.)

Schwarzwald, Ora R. 2009. Historical and foreign influences on the development of Hebrew. In Bar-Asher, Moshe & Cohen, Chaim E. (eds.), *Mas'at Aharon: Linguistic Studies presented to Aron Dotan*. 619–629. Jerusalem: The Bialik Institute. (In Hebrew.)

Schwarzwald, Ora R. 2012. On Vowel Polarity in Hebrew. *Leshonenu* 74. 79–92. (In Hebrew.)

Schwarzwald, Ora R. 2013. The typology of nonintegrated words in Hebew. *SKASE Journal of Theoretical Linguistics* 10(1). 41–53.

Schwarzwald, Ora R. & Dalia Cohen-Gross. 2000. The productive noun patterns in Hebrew. In Horvits, Miri (ed.), *The Language of Contemporary Press: Mina Efron's Memorial Volume*, 148–161. Tel-Aviv: The Mofet Institute. (In Hebrew)

Shatil, Nimrod. 2014. Noun patterns and their vitality in Modern Hebrew. *Hebrew Studies* 55. 171–203

Štekauer, Pavol. 2015. The delimitation of derivation and inflection. In Müller, Peter O., & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-Formation. An International Handbook of the Languages of Europe*, 218–234. Berlin: De Gruyter.

Stump, Gregory T. 2016. *Inflectional Paradigms: Content and Form at the Syntax-Morphology Interface*. Cambridge University Press.

Thornton, Anna M. 2012a. Overabundance in Italian verb morphology and its interactions with other non-canonical phenomena. In Stolz, Thomas & Otsuka, Hitomi & Urdze, Aina & van der Auwera, Johan (eds.), *Irregularity in Morphology (and beyond)*, 251–269. Berlin: Akademie Verlag.

Thornton, Anna M. 2012b. Reduction and maintenance of overabundance: A case study on Italian verb paradigms. *Word Structure* 5. 183–207.

Tobin, Yishai. 2001. Gender switch in Modern Hebrew. In Hellinger, Marlis & Bußmann, Hadumod (eds.), *Gender across languages: The linguistic representation of women and men*, 177–197. Amsterdam: John Benjamins.

Valden, Zvia. 2005. Rav, rabit, rabat, raba ve-ravraba. *Panim* 33. 77–84. (In Hebrew.)

Lior Laks
Deparment of English Literature and Linguistics
Bar-Ilan University, Israel
Lior.Laks@biu.ac.il

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/07.pdf. ISSN 1336-782X.

Linear and nonlinear word formation in Hebrew – words which end with -*on*

Ora (Rodrigue) Schwarzwald, Bar-Ilan University, Israel

Two major word formation processes exist in Hebrew (in addition to minor compounds, blends and acronyms): (a) nonlinear formation: a combination of consonantal root with template, e.g. higdil *'increased' and* migdal *'tower' are derived from the root* √gdl *using the templates hiCCiC and miCCaC; (b) linear formation: affixation to a stem, for example* balšanut *'linguistics' from* balšan *'linguist'* + -ut*, and* xidon *'quiz' from* xida *'riddle'* + -on*. The ending* -on *exhibits ambiguous cases of root and template construction as opposed to suffixed word formations. In many cases this ending is built using the nominal templates CiCaCon, CiCCon and CaCCon, the first of which usually create abstract nouns. In other cases* -on *is attached to various stems carrying the following connotations which are not always mutually exclusive, and sometimes share some of their meanings with words formed by the above templates: diminutive (e.g.* suson *'small horse'); collective (e.g.* še'elon *'questionnaire'); instrumental (e.g.* 'ecba'on *'thimble'); flora and fauna (e.g.* zeron *'harrier (bird)'); periodicals (e.g.* šavu'on *'weekly newspaper'); and division related words (e.g.* 'axuzon *'percentile'). Thus the ending* -on *creates opacity as part of a template and as a suffix for both derivational processes and meanings. One outcome of the findings is that syllabic structure is the most important factor in determining Hebrew word structure.*

Keywords: *word formation*, *suffix*, *root*, *template*, *syllabic structure*

1. Introduction

Most Hebrew words are derived from one of two major processes, linear and nonlinear (in addition to minor compounds, blends and acronyms):

i) A combination of consonantal root with vocalic template which sometimes includes additional consonants, for example *gadal* 'grew up', *higdil* 'increased', *gidel* 'raised', and *migdal* 'tower', all of which are derived from the root √*gdl* using the following templates: CaCaC, hiCCiC, CiCeC, miCCaC. This formation is nonlinear because the root is interwoven into the template, neither of which can be pronounced without the other.

ii) Linear affixation to a stem, for example, *du-mašma'i* 'ambiguous' from *du-* 'two' + *mašma* 'meaning' + -*i* (ADJECTIVAL suffix); *balšanut* 'linguistics' from *balšan* 'linguist', and + -*ut* (ABSTRACT suffix); *xidon* 'quiz' from *xida* 'riddle' + -*on*.

Most affixation in Hebrew is suffixal. Suffixes are added to various word classes for inflection and also to certain stems for derivation. The analysis of suffixes raises the issue of possible clashes between the two word formation processes when endings are involved.

The ending -*on*, which creates nouns, shows partial polysemy but at the same time demonstrates ambiguous cases of formation by root-template combination and by suffixation of -*on* to a stem. In many cases it is part of the nominal templates CiCaCon, CiCCon and CaCCon, in which case the ending -*on* is part of a template (§2.1); in other cases it is suffixed

to various nominal stems (§2.2). The distribution of this ending creates opacity with regards to derivational processes (§2.3, §2.4), and perhaps needs modification of the Word-to-Word Principle (§3) (Aronoff 1994; Stump 2001).

One outcome of the findings is that syllabic structure is the most important factor determining Hebrew words. Word-to-word formation must be modified in Hebrew to include nonlinear consonantal roots rather than existing lexical stems which include vowels.

2. The ending *-on*

There are more than 400 words that end with *on* in the *Rav Milim* Dictionary (Choueka 1997), all of which are stressed.[1] I am intentionally suggesting here that *on* is a word ending and not a suffix because, as will be explained, it is only used as a suffix in about two thirds of the cases where it is added to existing stems. This ending can be an integral part of a template (§2.1) as well as a suffix to nominal stems (§2.2). It is very rare for it to occur in words where the syllabic structure does not utilize a specific template or a stem (§2.3).

The following two groups need to be excluded from the set of words derived by the ending *on*: loan words in (1) and non-analyzable words in (2). Loan words were mostly borrowed into Hebrew during its classical periods, with the final *on* becoming part of the base in their original languages, and adjusted to the Hebrew structure (this is the case for about 30 words). The non-analyzable words in (2) include forms whose etymology is unclear, but which also include the final *on* as part of the base form (about 20 words).

(1) *doron* 'present' (Greek), *kidon* 'bicycle handlebars' (French *guidon*),[2] *limon* 'lemon' (Persian), *notarikon* 'acronym' (Lat. *notarium*), *picpon* 'tiny' (Yiddish *píce-le*), *seminaryon* 'seminar', *sirton*[3] 'sandbank, shoal' (Greek), *špiyon* 'spy' (German), *xelmon* 'egg yolk' (Aramaic *ḥelmona*), *xilazon* 'snail' (Arabic *ḥalazun*) (see (13) in the §Appendix)

(2) *'aron* 'cupboard', *'armon* 'palace; chestnut',[4] *harmon* 'harem',[5] *naxšon*[6] 'pioneer', *ciyon* 'Zion', *kilšon* 'pitchfork', *misderon* 'corridor',[7] *rimon* 'pomegranate; grenade', *šfifon* 'horned viper',[8] *širyon* 'armor; carapace' (see (14) in the §Appendix)

---

[1] I would like to thank the *Rav-Milim* team for supplying me with a list of Hebrew words that end in *on*. Some loan words were not included in the list. A loan word like *mégafon* 'megaphone' has pre-penultimate stress because it ends with the foreign element *-fon* 'phone' (cf. *télefon*, *míkrofon*). Such words are excluded from this discussion.

[2] The biblical word *kidon* is homonymic with this word, meaning 'spear, lance'. It belongs to the non-analyzable words. See (14) in the Appendix.

[3] The word *sirton* here is homophonic with *sirton* 'short film / video' in (6). These are spelled differently in Hebrew: *sirton*, as shown here, with the letter *sin*, *sirton* in (6) with the letter *samex*.

[4] The word *'armon* is homophonic: when the first letter is *aleph* the meaning is 'palace', and when it is *ayin* the meaning is 'chestnut'.

[5] In Classical Arabic, *ḥarīm* 'women', from which the word *harem* was spread into European languages. In biblical Hebrew *harmon* is a variant of *'armon* 'palace'.

[6] Named after the biblical hero Nachshon, who was first to enter the Red Sea after the Exodus from Egypt according to the Babylonian Talmud (Sota 37a).

[7] The word *misderon* 'corridor' is biblical (Jud 3:23), its etymology is unclear but it is thought to be related to the root √*sdr* which indicates order (perhaps the order of pillars or of rows). I have included this word here because its etymology is dubious.

In the following sections I will present the classifications of words ending in *on*, explain the problems which arise from these classifications, and conclude by showing that syllabic structure is the most important factor in determining the shape of a Hebrew word.

2.1 *Templatic, non-suffixed words*

In about 130 cases, the ending *on* is part of the nominal templates CiCaCon (3), CiCCon (4) or CaCCon (5),[9] which can also be realized as CeCaCon, CeCCon or CaCaCon for phonological reasons: when either *'*, *h* or *r* occurs as a second consonant or when *'* or *x* occurs sometimes as the first consonant.[10] Many of these nouns, especially those which follow the CiCaCon template, express abstractness. In such cases one cannot consider the ending as a suffix because, unlike the examples given in 2.2, there is no stem or base word that the words can be related to. The root in each example appears in many other words, none of which can be considered as the stem for suffixation.[11]

(3) CiCaCon: *bizayon* 'shame' (√*bzy*), *dera'on* 'disgrace' (√*dr'*), *dika'on* 'depression' (√*dk'*), *gizaron* 'etymology' (√*gzr*), *higayon* 'logic' (√*hgy*), *nisayon* 'trial; experience' (√*nsy*), *šiga'on* 'madness' (√*šg'*), *te'avon* 'appetite' (√*t'v*), *timahon* 'amazement' (√*tmh*), *xipazon* 'haste' (√*xpz*) (see (15) in the Appendix)

(4) CiCCon: *bidyon* 'fiction' (√*bdy*), *birkon* 'booklet containing the grace after meal' (√*brk*), *dimyon* 'imagination; similarity' (√*dmy*), *'egron* 'lexicon' (√'gr), *gilšon* 'hang glider' (√*glš*), *pitron* 'solution' (√*ptr*), *pizmon* 'short song; ditty' (√*pzm*), *riv'on* 'quarter (of a year)' (√*rv'*), *šilton* 'rule, government' (√*šlt*), *yitron* 'advantage' (√*ytr*) (see (16) in the Appendix)

(5) CaCCon: *'agmon* 'reed, bulrush' (√*'gm*? *'agam* 'lake'), *caharon* 'school activity after school hours' (√*chr*; *cohoráyim* 'noon'), *kadmon* 'ancient' (√*kdm*; *kadum* 'ancient'), *kapcon* 'cap (for a cap gun)' (√*kpc*; *kipec* 'hop.v'), *pa'amon* 'bell' (√*p'm*; *pa'am* 'beat.v'), *ra'ayon* 'idea' (√*r'y*), *xanyon* 'parking lot' (√*xny*; *xana* 'park.v'), *zaharon* 'sunfish' (√*zhr*; *zóhar* 'brightness')

2.2 *Suffixed words*

In other cases, the suffix *-on* is linearly attached to nominal stems to form nouns which carry the following meanings: a. diminutive, affectionate, derogative and augmentative nouns, as in (6); b. collection nouns, as in (7); c. instruments, devices and objects, as in (8); d. flora and

---

[8] The etymology of this word is unclear, although the consonants *š* and *p-f* occur in other Semitic languages to indicate the same kind of snail.

[9] The approximate roots from which these words were derived are listed in these examples. The source stem in §2.2 is given in parentheses.

[10] This is a residue of the historical process whereby a high vowel was lowered whenever it occurred close to a guttural consonant (Gesenius 1910: 76–79; Joüon & Muraoka 2011: 77–84).

[11] In several examples here, I have listed up to 10 items. For each category that includes more examples, an additional list, although not an exhaustive one, is featured in the Appendix.

fauna, as in (9); e. periodicals (publications), as in (10); f. division related words, as in (11).[12] Very few words which end with *-on* do not belong to one of these categories, and these are presented in (12).

(6) DIMINUTIVE, AFFECTIONATE, DEROGATIVE AND AUGMENTATIVE NOUNS: *'agamon* 'small lake' (*'agam* 'lake'), *ce'iron* 'young man (derogatory)' (*ca'ir* 'young', √*c'r*), *ganon* 'pre-kindergarten' (*gan* 'kindergarten', √*gnn*), *kalbon* 'puppy' (*kélev* 'dog'),[13] *karxon* 'glacier; iceberg' (*kérax* 'ice'), *kušon* 'Niger' (*kúši* 'Ethiopian, black'), *pic'on* 'pimple' (*péca* 'wound', √*pc'*), *sifron* 'booklet' (*séfer* 'book', √*sfr*), *sirton* 'short film/ video' (*séret* 'film', √*srt*; see Footnote 3), *tiyulon*[14] 'short trip' (*tiyul* 'trip', √*tyl*), *yehudon* '(derogatory) Jew, Yid, Kike' (*Yehudi* 'Jew', √*yhd*) (see (17) in the Appendix; Dressler & Merlini Barbaresi 1994)

(7) COLLECTION NOUNS: *'alfon* 'alphabetical index' (*'alef* 'Aleph'), *hemšexon* 'sequel' (*hemšex* 'continuation', √*mšx*), *mexiron* 'price-list' (*mexir* 'price'), *milon* 'dictionary' (*mila* 'word', √*mll*), *munaxon* 'glossary' (*munax* 'term'), *nivon* 'anthology of proverbs' (*niv* 'proverb'), *pa'oton* 'nursery school' (*pa'ot* 'small child'), *še'elon* 'questionnaire' (*še'ela* 'question', √*š'l*), *šnaton*[15] 'college/school calendar' (*šana* 'year', √*šny*) (see (18) in the Appendix)

(8) INSTRUMENTS, DEVICES AND OBJECTS: *'aviron* 'airplane' (*'avir* 'air'), *birkon* 'knee high stockings' (*bérex* 'knee', √*brk/x*), *cilon* 'window shade, blind' (*cel* 'shadow', √*cll*), *kova'on* 'condom' (*kova* 'hat'), *mapiyon* 'paper napkin dispenser' (*mapit* 'napkin', √*npy*), *maxševon* 'pocket calculator' (*maxšev* 'computer', √*xšv*), *mešivon* 'answering machine' (*mešiv* 'answering', √*šwv*), *sfaton* 'lipstick' (*safa* 'lip'), *tiyulon* 'stroller' (*tiyul* 'trip', √*tyl*), *yisumon* '(computing) applet' (*yisum* 'application', √*ysm*) (see (19) in the Appendix)

(9) FLORA AND FAUNA: *ce'elon* 'Poinciana (plant)' (*ce'el~cel* 'shadow', √*cll*), *naxšon* 'Anhinga rufa' (*naxaš* 'snake'), *nišpon* 'mouse-eared bat' (*néšef* 'late night', √*nšf/p*), *pilpelon* 'pepper tree', (*pilpel* 'pepper'), *savyon* 'groundsel' (*sav* 'old man'), *tukon* 'budgerigar (bird)' (*tuki* 'parrot'), *xasilon* 'shrimp' (*xasil* 'a kind of locust'), *ya'aron* 'field mouse' (*ya'ar* 'forest', √*y'r*), *yarkon* 'greenfinch (bird, √*yrk*)' (*yarok* 'green', √*yrk*), *zmiron* 'singing bird' (*zamir* 'nightingale', √*zmr*) (see (20) in the Appendix)

(10) PERIODICALS: (publications): *'alilon* 'comics' (*'alila* 'plot', √*'ll*), *cehubon* 'tabloid' (*cahov* 'yellow', √*chb/v*), *'iton* 'newspaper' (*'et* 'time', √*'tt*), *mekomon* 'local newspaper' (*makom* 'place', √*kwm*), *šavu'on* 'weekly newspaper' (*šavúa* 'week',

[12] Hebrew lexical entries show many morphophonemic alternations, e.g. *séfer* 'book', *sifri* 'my book', *sfarim* 'books'; *cel* 'shadow', *cili* 'my shadow', *clalim* 'shadows'. Consequently, it sometimes seems that there is a very remote connection between lexical stem and the actual realization of formed words with a suffix; however, in all these cases one of the variants is the stem.

[13] Hebrew reflects residues of the biblical Spirantization Rule realized in *b-v*, *p-f* and *k-x* alternations (Schwarzwald 2001: 10–11; Joüon & Muraoka 2011: 76–77). These are only marked in this paper when relevant.

[14] See the other meaning in (8).

[15] *Šnaton* is collective as well as a periodical, see (10).

√*šv'*), *šnaton* 'annual publication; yearbook' (*šana* 'year'), *tkufon* 'periodical' (*tkufa* 'period'), *xinamon* 'free local newspaper' (*xinam* 'free, no charge'), *yarxon* 'monthly publication' (*yérax* 'month', √*yrx*), *yomon* 'daily journal' (*yom* 'day') (see (21) in the Appendix)

(11) DIVISION RELATED WORDS: *'asiron* 'decile' (*'asara* 'ten.M', *'asiri* 'tenth', √*'sr*), *'axuzon* 'percentile' (*'axuz* 'percent', √*'xz*), *revi'on* 'quarter, quadrant' (*réva* 'quarter', √*rv'*), *šlišon* 'trimester' (*šliš* 'third', √*šlš*), *tixon* 'middle, central; median; high school' (*tox* 'inside, interior', √*twx*), *xamišon* 'quintile' (*xamiša* 'five.M', √*xmš*), *xecyon* '(statistics) median' (*xéci* 'half', √*xcy*)

(12) NON-SPECIFIED MEANINGS: *'eglon* 'carter, coachman' (*'agala* 'carriage, wagon', CNST: *'eglat-*), *purimon* 'Purim festival', *re'avon*[16] 'hunger' (*ra'av* 'hunger', √*r'v*), *rišon* 'first' (*roš* 'head'), *šabaton* 'sabbatical' (*šabat* 'Sabbath', √*šbt*),[17] *xicon* 'exterior; external' (*xuc* 'outside.N', √*xwc*)[18]

2.3 *Intermediate words*

Two words need special consideration: *'afifon* and *kdimon*, both formed using a similar syllabic structure, neither of which can be classified according to the groups described in §2.1 and §2.2.

The word *'afifon* 'kite' is related to the root √*'f* which carries the connotation of flying, yet the formation of the root is new because √*'f* is historically derived from the root structure CVC (CwC or CyC, √*'wf* in this case) rather than $C_1C_2C_2$. The reduplication of the second radical is quite common in newly formed roots. It appears that the word was formed in the same way as *svivon* 'spinning top' (from *saviv* 'around' (19)) or *yešišon* 'sweet old man' (from *yašiš* 'old man' (17)), although there is no existing word from which *'afifon* could have been derived.

The word *kdimon* 'promo' is linked to the root √*kdm* which in turn is related to many words: *kódem* 'before, early', *hikdim* 'preceded', *hitkadem* 'advanced', *hitkadmut* 'progress', *kidem* 'promoted, welcomed', *kidum* 'promotion', *hakdama* 'introduction', etc., all of which were formed by combining consonantal roots with existing templates.[19] However, no grammar book will argue that the template CCiCon exists purely to satisfy the word *kdimon*, because other than *'afifon*, no other Hebrew word is formed using this structure. If *kdimon* is a derivative of stem + *-on*, it would be beneficial to look for its stem; however, *kadim* does not exist in any similar sense: *kadim* can be related to the adverb *kadíma* 'forward, let's go', or to the compound *rúax kadim* 'an eastern hot and dusty wind'; there is no **kdim* either. The word *kdima* 'precedence' (derived from √*kdm*+CCiCa) only appears in the compounds *dmey kdima* 'down payment', or in *zxut kdima* 'priority, right of way'. It is clear that although the

[16] Based on its pronunciation, *re'avon* was formed using the CiCaCon template with the root √*r'v*. However, according to the Hebrew vocalization system, *re'avon* is a derivative of the word *ra'av* 'hunger'.
[17] See the other meaning of *šabaton* in (17).
[18] The three words, *rišon* and *xicon* here and *tixon* in (11) were formed by dissimilation in order to avoid two consecutive back non-low vowels. The words are used as adjectives and as nouns, and they all were inherited from biblical Hebrew.
[19] CóCeC, hiCCiC, hitCaCeC, hitCaCCut, CiCeC, CiCum, haCCaCa. Many other words are derived with these templates.

word *kdimon* is related to the root √*kdm* and associated to its derived words, it cannot be associated with any specific stem or template.[20]

2.4 *Inferences*

A few observations can be drawn from data presented above in (1–12):

a) The data are classified according to morphological and semantic criteria. Primary morphological classification differentiates between words derived nonlinearly by a consonantal root and template (3–5), and words derived linearly by stem and suffix *-on* (6–12). Semantic classification applies only partly to words derived nonlinearly but more systematically to linear suffixed formation.

b) Many words expressing abstractness were formed using the template CiCaCon (3), and rarely to CiCCon (4). This characteristic is not typical of the various meanings of words derived by stem and the suffix *-on* (6–12).

c) Some loan words have been adjusted to fit Hebrew syllabic structures (1, 13), e.g. *doron* 'present (Greek) looks like *'oron* 'proper name; the Hebrew name of Oranus'; *'evyon* 'poor' (Old Egyptian), *sirton* 'sandbank, shoal' (Greek), *xelmon* 'egg yolk' (Aramaic), and *picpon* 'tiny' (Yiddish) take the same syllabic structure as CiCCon words in (4); *kidon* '(biblical) spear, lance' and 'bicycle handlebars' (French), *šifon* 'chiffon' (French), and *limon* 'lemon' (Persian) (1, 13) resemble *'iton* 'newspaper' (10) and *cilon* 'window shade, blind' (8); *kinamon* 'cinnamon' (Greek) in (13) and *xilazon* 'snail' (Arabic) in (1) have the same syllabic structure as CiCaCon words in (3).

d) In fact, because of the strong semantic load of the roots, it is quite difficult to determine the distinction between nonlinear root and template formed words on the one hand, and linear stem plus suffix words on the other. Each root in the templates (3–5) and many of the stems in (6–12) originate from roots. From the root √*rv'* one can derive nonlinearly the word *riv'on* 'quarter (of a year)' in (4) and the stem *réva* 'quarter' from which *revi'on* 'quarter, quadrant' in (11) is derived linearly. The same root √*rv'* is the base for many words, for example *'arba* 'four', *'arba'im* 'forty', *ribúa* 'square', *meruba* 'quadrangle', *ribéa* 'great-grandson.N; quadruple.V', *revi'i* 'fourth', and so on. The classification into root and template was specifically determined in cases where no clear stem was available. Therefore, the word *xinamon* 'free local newspaper' was included in (10; periodicals) because it is clearly derived from the stem *xinam* 'with no charge', rather than from the CiCaCon template, as the word *xinam* is not related to any other words, hence to no consonantal root.

e) The classification of meanings suggested here in (3–12) is not mutually exclusive. Some of the words can belong to more than one category. Some instruments in (8 and 19) could have been included in the diminutives and vice versa. Although *makle'on* 'small improved machine gun', *maxševon* 'pocket calculator', *yisumon* '(computing) applet', *mesofon* 'portable terminal' and *takliton* 'diskette' are instruments, they could also have been included in the diminutive examples (6); *clavon* 'small crucifix (jewelry)' (*clav* 'cross', √*clv*) in (17) could have been included in the instruments and objects (8, 19). Some periodicals could also be considered diminutives (6), e.g.

[20] This word is discussed at length in Schwarzwald (2009).

mekomon ‘local newspaper’ in (10) has a derogatory connotation. The same applies to the other categories: *tukon* ‘budgerigar (bird)’ (9) could also be a diminutive. The name of the tree *sigalon* ‘jacaranda’ and *'iparon* ‘pencil’ which appear within CiCaCon template word examples (15), could be considered instances of flora and fauna (9), and instrument (8) respectively.

f) Some of the nouns are homonymic and fit into two categories, e.g. *šabaton* has two meanings and is therefore given in two places (12) and (17). The traditional meaning refers to the Sabbath, the seventh day of rest in a week, but it can also refer to a sabbatical year, the year a person takes off from work, which does not necessarily need to be the seventh year. The word *birkon* utilizes the CiCCon template (4) from the root √*brk* meaning a booklet containing the grace after meal, whereas the linear formation of the word meaning knee high stockings in (8) is derived from *bérex* ‘knee’ (PL. *birkáyim*). The word *caharon* means ‘afternoon newspaper’ (21) as well as ‘school activity after school hours’ (5). Although this can be attributed to the root √*chr* extracted from *cohoráyim* ‘noon’, it can also be derived linearly from the common pronunciation of the word as in *becaharey yom* ‘in the middle of the day, at noon time’. The same considerations apply to *tiyulon* ‘short trip’ (6) and ‘stroller’ (8); *šnaton* ‘college/school calendar’ (7) and ‘annual publication; yearbook’ (10); *naxšon* ‘pioneer’ (2) and ‘Anhinga rufa’ (9); *yarkon* ‘greenfinch (bird)’ (9) and ‘thigh high socks’ (19).

g) Many of the words derived linearly by stem and suffix *-on* carry the same syllabic structure as the nonlinearly templates CiCCon (4) or CaCCon (5). Thus there is no clear cut phonological distinction between discontinuous root and template derivation and linear stem and suffix formation, compare *bidyon* ‘fiction’ (√*bdy*) and *pizmon* ‘short song; ditty’ (√*pzm*) in (4), with *sifron* ‘booklet’ (*séfer* ‘book’), in (6), and *sirton* ‘short film / video’ (*séret* ‘film’; see Footnote 3 above), and *nišpon* ‘mouse-eared bat’ (*néšef* ‘late night’, √*nšf/p*) in (9). Linear *'alfon* ‘alphabetical index’ in (7), *kalbon* ‘puppy’, and *karxon* ‘glacier; iceberg’ in (6), take the same syllabic structure as nonlinear *kadmon* ‘ancient’ (√*kdm*) from CaCCon template in (5), etc.

h) When analyzing the words derived from nonlinear roots and templates (3–5, 15–17), it is clear that some of these carry the same meaning as stem and suffixed words. The following examples demonstrate this resemblance: *bita'on* ‘journal, mouthpiece’ (√*bt'*) in (15) (cf. periodicals in (10)), *'egron* ‘lexicon’ (√*'gr*) in (4) (cf. collective in (7)), and *tilyon* ‘pendant’ (√*tly*) in (16) (cf. instruments and devices in (8)).

i) Two synonymous words with the suffix *-on* mean emoticon, *parcufon* and *rigšon* (19). The first of these is derived from *parcuf* (an old loan word from Greek, *prosopon*) ‘face’ + *-on*, while the second comes from the word *régeš* (cf. *rigšotay* ‘my feelings’) + *-on*. The first instance refers to the actual image of the emoticon, while the second reflects the feelings it arouses.

j) Many of the stems onto which the suffix *-on* was added were derived from consonantal roots.

3. Conclusion

The data presented here demonstrates that no matter what their derivation is, many Hebrew words that end in *on* exhibit similar syllabic structures and stress pattern. Morphological Hebraism is actually a matter of syllabic structure. Regardless of the way in which a derived word is formed, words sound Hebrew because of their syllabic structure. The generalized form for all the examples presented here can be ***Xon*** where X is either a linear stem or a kernel of a consonantal root. The suffix -*on* is attached to nominal stems in about two thirds of the occurrences of *Xon*; in other cases it is either part of a template, or part of borrowed and base form which is hardly recognized as foreign in many cases.

The few exceptional cases such as (1), (2), or the two single examples *'afifon* and *kdimon* in §2.3, also follow this generalization. While stems with suffixed -*on* shape distinct meaning classes (6–12), root and template formation words do not carry similar connotations, although singular words occasionally have the same meanings. Abstractness typifies many words which correspond to the CiCaCon and a few CiCCon templates, but not stems which feature the suffix -*on*.

Polysemy and sometimes homonymy may exist in the affixes of other languages (e.g. Plag 1997; Beard & Volpe 2005; Lieber 2005; Scalise & Guevara 2005; Rainer 2014; Efthymiou 2016; Nagano & Shimada 2016). The English suffix -*en* can provide some examples of this, forming causative or inchoative verbs from adjectives or nouns such as *sharpen* and *heighten*, as well as adjectives such as *golden* or *woolen*. The suffix -*al* can form adjectives like *national* or *commercial*, and substantives, such as *signal* and *betrayal* (Jespersen 1965: 337 ff., 380 ff.). The Hebrew case is unique because with the exception of a few loan and base words (examples (1) and (2)), the consonantal root plays an important role in word formation. This is important not only for cases which take the templates CiCaCon, CiCCon and CaCCon, but also for the derivation of many of the stems that utilize the suffix -*on*. Therefore, it is not always a necessarily word-to-word derivation as suggested by Stump (2001, 2005) but rather also a root-to-word derivation, where root refers to the Hebrew consonantal root.

Appendix: Additional examples

(13) LOANS: *'alaxson* 'oblique line' (Greek *loxon*), *directoryon* 'board of directors', *historyon* 'historian', *kinamon* 'cinnamon' (Greek), *melafefon* 'cucumber' (Greek *melopepon*), *proletaryon* 'proletarian', *šifon* 'chiffon' (French), *simpozyon* 'symposium'

(14) BASE WORDS: *dišon* 'antelope', *hamon* 'crowd', *kamon* 'cumin', *midron* 'slope', *cnon* 'radish', *ciklon* 'bag, sack', *lacon* 'clowning, joking', *kifon* 'mullet (fish)', *kimšon* 'thorn', *xazon* 'dream, vision'

(15) CiCaCon: *bita'on* 'journal, mouthpiece' (√*bt'*), *bicaron* 'fortress' (√*bcr*), *bitaxon* 'security', *cima'on* 'thirst' (√*cm'*), *'eragon* 'strong yearning' (√*'rg*), *'eravon* 'guarantee, deposit' (√*'rv*), *gilayon* 'sheet of paper' (√*gly*), *herayon* 'pregnancy' (√*hry*), *higayon* 'logic' (√*hgy*), *hizayon* 'fantasy' (√*hzy*), *hišavon* 'drawback, restitution' (√*šwv*), *'icavon* 'sorrow' (√*'cv*), *'idašon* 'apathy' (√*'dš*), *'ikaron* 'principle'

(√*'kr*), *'ikavon* ‘(law) lien’ (√*'kv*), *'ilafon* ‘fainting’ (√*'lf*), *'imadon* ‘(medicine) stasis’ (√*'md*), *'iparon* ‘pencil’ (√*'pr*), *'ivaron* ‘blindness’ (√*'vr*), *'izavon* ‘inheritance’ (√*'zv*), *kiba'on* ‘fixation’ (√*kb'*), *kihayon* ‘darkness’ (√*khy*), *kixalon* ‘cyanosis’ (√*kxl*), *kilayon* ‘total destruction’ (√*kly*), *kimahon* ‘yearning’ (√*kmh*), *kimašon* ‘withering disease’ (√*kmš*), *kipa'on* ‘freezing; stagnation’ (√*kp'*), *kišalon* ‘failure’ (√*kšl*), *kišaron* ‘talent’ (√*kšr*), *nicaxon* ‘victory’ (√*ncx*), *nicayon* ‘quarrel’ (√*ncy*), *nikayon* ‘cleanliness’ (√*nky*), *pera'on* ‘redemption, payment’ (√*pr'*), *pikadon* ‘deposit’ (√*pkd*), *pikaxon* ‘realistic observation’ (√*pkx*), *se'avon* ‘filth’ (√*s'v*), *pitayon* ‘bait, lure; temptation’ (√*pty*), *rikavon* ‘decomposition; rotting’ (√*rkv*), *rišayon* ‘permit’ (√*ršy*), *seraxon* ‘stench’ (√*srx*), *šibaron* ‘distress’ (√*šbr*), *šigaron* ‘rheumatism’ (√*šgr*), *sigalon* ‘jacaranda’ (√*sgl*), *šitafon* ‘flood’ (√*štf*), *šizafon* ‘suntan’ (√*šzf*), *teradon* ‘obsession’ (√*trd*), *xeragon* ‘trance’ (√*xrg*), *xidalon* ‘destruction, cessation’ (√*xdl*), *xikayon* ‘expectation’ (√*xky*), *xiladon* ‘wheat rust’ (√*xld*), *xisaxon* ‘saving’ (√*xšx*), *xišaxon* ‘blackout’ (√*xšx*), *xivaron* ‘pallor’ (√*xvr*), *xisaron* ‘disadvantage’ (√*xsr*), *xizayon* ‘spectacle’ (√*xzy*), *zikaron* ‘memory’ (√*zkr*), *zikayon* ‘concession, permit’ (√zky)

(16) CiCCon: *biryon* ‘hooligan’ (√*bry*), *bitron* ‘ravine, gorge’ (√*btr*), *'elbon* ‘insult’ (√*'lb*), *'elson* ‘(literary) joy, cheerfulness’ (√*'ls*), *'elyon* ‘superior’ (√*'ly*), *gidron* ‘wren’ (√*gdr*), *'igron* ‘letter writing manual’, *kifyon* ‘epilepsy’ (√*kfy*), *kišyon* ‘(medicine) rigidity, stiffness’ (√*kšy*), *kivšon* ‘secret, mystery’ (√*kvš*), *nicron* ‘plumule’ (√*ncr*), *pidyon* ‘redemption’ (√*pdy*), *pirxon* ‘fledgling’ (√*prx*), *piryon* ‘fertility’ (√*pry*), *rifyon* ‘weakness’ (√*rfy*), *šifyon* ‘clear mind’ (√*šfy*), *šivyon* ‘equality’ (√*švy*), *tilyon* ‘pendant’ (√*tly*), *xelbon* ‘white of egg’ (√*xlb*), *xerbon* ‘(vulgar) disaster’ (√*xrb*; cf. *xurban* ‘destruction’), *xešbon* ‘calculation’ (√*xšb*)

(17) DIMINUTIVE, AFFECTIONATE, DEROGATIVE AND AUGMENTATIVE: *'avazon* ‘gosling’ (*'avaz* ‘goose’), *barvazon* ‘duckling’ (*barvaz* ‘goose’), *cnimon* ‘crouton’ (*cnim* ‘zwieback, rusk’, √*cnm*), *clavon* ‘small crucifix (jewelry)’ (*clav* ‘cross’, √*clv*), *crifon* ‘small hut’ (*crif* ‘hut’), *gišron* ‘small bridge’ (*géšer* ‘bridge’), *gvi'on* ‘small goblet’ (*gavía* ‘goblet’, √*gv'*), *gvišon* ‘small crystal’ (*gavish* ‘crystal’, √*gvš*), *'iyon* ‘islet’ (*'i* ‘island’), *ka'aron* ‘(geology) synclinorium’ (*ke'ara* ‘bowl, basin’, √*k'r*), *kficon* ‘small spring, coil’ (*kfic* ‘spring’, √*kfc*), *kimton* ‘small wrinkle’ (*kémet* ‘wrinkle’ √*kmt*), *krixon* ‘small sandwich’ (*karix* ‘sandwich’, √*krx*), *kumkumon* ‘small kettle’ (*kumkum* ‘kettle’), *me'ilon* ‘jacket, small coat’ (*me'il* ‘coat’, √*m'l*), *mifracon* ‘creek’ (*mifrac* ‘bay’, √*p/frc*), *miskenon* ‘little poor’ (*misken* ‘poor, miserable’, √*skn*), *misxakon* ‘one set of a tennis game’ (*misxak* ‘game’, √*sxk*), *misron* ‘SMS’ (*méser* ‘message’), *mitbaxon* ‘kitchenette’ (*mitbax* ‘kitchen’, √*tbx*), *nexmadon* ‘sweetie’ (*nexmad* ‘cute’, √*xmd*), *pilon* ‘young elephant’ (*pil* ‘elephant’), *pinkason* ‘small notepad’ (*pinkas* ‘small notebook’), *pislon* ‘figurine, statuette’ (*pésel* ‘sculpture’, √*psl*), *pišpešon* ‘unworthy person’ (*pišpeš* ‘bug, cimex’), *pkidon* ‘unworthy clerk’ (*pakid* ‘clerk’, √*pkd*), *šabaton* ‘(big) rest’ (*šabat* ‘Sabbath’, √*šbt*), *šedon* ‘sprite, imp’ (*šed* ‘demon’), *siflon* ‘demitasse’ (*séfel* ‘cup, mug’), *sifron* ‘booklet’ (*séfer* ‘book’, √*sfr*), *simlon* ‘small badge’ (*sémel* ‘symbol, sign’, √*sml*), *šoršon* ‘(botany) radicle’ (*šóreš* ‘root’, √*šrš*), *štixon* ‘small rug’ (*šatíax* ‘carpet’, √*štx*), *šu'alon* ‘small fox’ (*šu'al* ‘fox’), *šulxanon* ‘small table’ (*šulxan* ‘table’), *suson* ‘pony, small horse’ (*sus* ‘horse’), *taltalon* ‘small curl’ (*taltal* ‘hair curl’, √*tll*), *tanuron* ‘small stove; toaster-oven’ (*tanur* ‘oven’), *tapuxon* ‘small apple’ (*tapúax* ‘apple’, √*tpx*), *tipšon* ‘silly little fool’ (*tipeš*

'silly', √*tpš*), *tmimon* 'endearing naïve person' (*tamim* 'naïve', √*tmm*), *xaravon* '(literary) heavy heat' (*xarev* 'dry', √*xrv*), *yaldon* 'little boy' (*yéled* 'child', √*yld*), *yešišon* 'sweet old man' (*yašiš* 'old man')

(18) COLLECTION NOUN: *mat'imon* 'concordance' (*mat'im* 'fitting', √*t'm*), *meyda'on* 'information booklet' (*meyda* 'information', √*yd'*), *mišpaxton* 'pre-nursery play group' (*mišpaxa* 'family', √*špx*), *mo'adon* 'club' (*mo'ed* 'appointed time', √*y/w'd*), *naxšon* 'series of hairpin bends on the road' (*naxaš* 'snake'), *nofšon* 'package tour' (*nófeš* 'vacation', √*nfš*), *pitgamon* 'collection of proverbs' (*pitgam* 'proverb'), *ptiton* 'confetti' (*ptitim* 'flakes', √*ptt*), *rešimon* 'list of entries' (*rešima* 'list', √*ršm*), *ša'ašu'on* 'quiz show, game show' (*ša'ašúa* 'entertainment', √*š'š'*), *šimušon* '(internet) useful links' (*šimuš* 'usage', √*šmš*), *širon* 'song book' (*šir* 'song', √*šyr*), *sixon* 'conversation manual' (*sixa* 'conversation', √*syx*), *šuron* 'lined sheet' (*šura* 'line'), *tarmilon* 'didactical word list' (*tarmil* 'satchel'), *xidon* 'quiz' (*xida* 'riddle', √*xdd*), *yedi'on* 'bulletin' (*yedi'a* 'news item; knowledge', √*yd'*)

(19) INSTRUMENT, DEVICES AND OBJECTS: *'adilyon* 'medallion' (blend of *'adi* 'piece of jewelry' and *tilyon* 'pendant', √*'dy*+√*tly*), *'aslon* 'paper for covering toilet bowl' (*'asla* 'toilet seat'), *saharon* 'croissant' (*sáhar* 'moon'), *camron* 'ear stick' (*cémer* 'wool'), *cavaron* 'collar' (*cavar* 'neck'), *darkon* 'passport' (*dérex* 'way', √*drk/x*), *'ecba'on* 'thimble' (*'ecba* 'finger'), *halixon* 'walker' (*halixa* 'walking', √*hlx*), *klasteron* '(police) facial composite, Identi-Kit' (*klaster* 'portrait', of Greek or Latin origin),[21] *magafon* 'bootee' (*magaf* 'boot', √*gff*), *makbilon* '(geometry) parallelepiped' (*makbil* 'parallel', √*kbl*), *makdexon* '(carpentry) hand drill' (*makdéax* 'drill bit', √*kdx*), *makle'on* 'small improved machine gun' (*makléa* 'machine gun', √*kl'*), *malkon* 'queen (bee) cup/cell' (*malka* 'queen', √*mlk/x*), *mapuxon* 'accordion' (*mapúax* 'bellows', √*npx*), *maskoron* 'document for calculating salary' (*maskóret* 'salary', √*sk/xr*), *maxperon* 'backhoe loader' (*maxper* 'dredger', √*xp/fr*), *meguron* 'caravan' (*megurim* 'dwelling', √*gwr*), *mesofon* 'portable terminal' (*masof* 'terminal', √*swf*), *mexikon* 'Tipp-ex, eraser' (*mexika* 'erasing, deletion', √*mxk*), *miklaxon* 'shower stall' (*mikláxat* 'shower' √*kl'*), *ne'imon* 'ringtone' (*ne'ima* 'melody', √*n'm*), *pa'alulon* 'activity game for babies' (*pa'alul* 'effect (in movies)', √*p'l*), *patišon* '(music) percussion mallet' (*patiš* 'hammer'), *parcufon* '(internet) emoticon' (*parcuf* 'face', of Greek origin), *paxon* 'tin shack' (*pax* 'tin', √*pxx*), *ptilon* 'colored marker' (*ptil* 'wick', √*ptl*), *rigšon* '(internet) emoticon' (*régeš* 'feeling', √*rgš*), *ša'on* 'clock, watch' (*ša'a* 'hour'), *šalgon* 'popsicle, ice lolly' (*šéleg* 'snow', √*šlg*), *šarvulon* 'open front bolero with sleeves' (*šarvul* 'sleeve'), *svivon* 'spinning top' (*saviv* 'around', √*svv*), *ta'arixon* 'date display' (*ta'arix* 'date', loan from Arabic), *takliton* 'diskette' (*taklit* 'record', √*klt*), *tikon* 'binder, folder for tax documents' (*tik* 'bag, folder'), *tiyon* 'teapot' (*te* 'tea'), *xalifon* 'garment bag' (*xalifa* 'suit', √*xlf*), *yarkon* 'thigh high socks' (*yarex* 'thigh')

(20) FLORA AND FAUNA: *gibton* 'bunting (bird)' (*gav*, *gavnun* 'hunchback'), *tmiron* 'stilt bird' (*tamir* 'tall', √*tmr*), *zeron* 'harrier (bird)' (*zer* 'wreath'), *zer'on* 'seed, grain' (*zéra* 'seed', √*zr'*)

[21] Unlike the examples in (1) where borrowed words already include the ending *on*, in *klasteron* the *-on* is used as a suffix. The same applies to *parcufon* below.

(21) PERIODICALS: *'alon* 'leaflet, brochure' (*'ale* 'leaf', √*'ly*), *caharon* 'afternoon newspaper; school activity after school hours' (√*chr*; *cohoráyim* 'noon', √*chr*), *pnimon* 'the inside pages of the newspaper' (*pnim* 'inside', √*pnm*)

References

Aronoff, Mark. 1994. *Morphology by Itself: Stems and Inflectional Classes* (Linguistic Inquiry Monographs 22)**.** Cambridge, MA: MIT Press.

Beard, Robert & Volpe, Mark. 2005. Lexeme-morpheme base morphology. In Štekauer, Pavol & Lieber, Rochelle (eds.), *Handbook of Word-Formation*, 189–206. Dordrecht: Springer.

Choueka, Yaacov. 1997. *Rav-Milim: A comprehensive dictionary of modern Hebrew*. Tel Aviv: C.E.T, Miskal & Steimatzky. (http://www.ravmilim.co.il/naerr.asp) (Accessed 2017-08-24.)

Dressler, Wolfgang U. & Merlini Barbaresi, Lavinia. 1994. *Morphopragmatics: Diminutives and Intensifiers in Italian, German, and Other Languages*. Berlin: Mouton de Gruyter.

Efthymiou, Angeliki. 2016. On the polysemy of the Modern Greek prefix *para-*. In Körtvélyessy, Lívia & Štekauer, Pavol & Valera, Salvador (eds.), *Word-Formation across Languages*, 91–107. Newcastle upon Tyne: Cambridge Scholars Publishing.

Gesenius, Wilhelm. 1910. *Hebrew Grammar*. Kautzsch, Emil (ed.) & Cowley, Arthur Ernst (trans.). Oxford: Clarendon Press.

Jespersen, Otto. 1965. *The Philosophy of Grammar*. New York: W. W. Norton & Company.

Joüon, Paul & Muraoka, Takamitso. 2011. *A Grammar of Biblical Hebrew*. Roma: Gregorian & Biblical Press.

Lieber, Rochelle. 2005. English word-formation processes. In Štekauer, Pavol & Lieber, Rochelle (eds.), *Handbook of Word-Formation*, 375–427. Dordrecht: Springer.

Nagano, Akiko & Shimada, Masaharu. 2016. How poor Japanese is in adjectivizing derivational affixes and why. In Körtvélyessy, Lívia & Štekauer, Pavol & Valera, Salvador (eds.), *Word-Formation across Languages*, 219–240. Newcastle upon Tyne: Cambridge Scholars Publishing.

Plag, Ingo. 1997. The polysemy of -ize derivatives: On the role of semantics in word formation. In Booij, Geert & Van Marle, Jaap (eds.), *Yearbook of Morphology*, 219–242. Dordrecht: Kluwer.

Rainer, Franz. 2014. Polysemy in derivation. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford Handbook of Derivational Morphology*, 338–353. Oxford: Oxford University Press.

Scalise, Sergio & Guevara, Emiliano. 2005. The Lexicalist Approach to Word-Formation and the Notion of the Lexicon. In Štekauer, Pavol & Lieber, Rochelle (eds.), *Handbook of Word-Formation*, 147–188. Dordrecht: Springer.

Schwarzwald (Rodrigue), Ora. 2001. *Modern Hebrew*. München: Lincom Europa.

Schwarzwald (Rodrigue), Ora. 2009. Three related analyses in Modern Hebrew morphology. In Goldenberg, Gideon & Shisha-Halevy, Ariel (eds.), *Egyptian, Semitic and General Grammar: Workshop in Memory of H. J. Polotsky, 8–12 July, 2001*, 277–301. Jerusalem: The Israel Academy of Sciences and Humanities.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure* (Cambridge Studies in Linguistics 93). Cambridge: Cambridge University Press.

Stump, Gregory T. 2005. Word-formation and inflectional morphology. In Štekauer, Pavol & Lieber, Rochelle (eds.), *Handbook of Word-Formation*, 48–71. Dordrecht: Springer.

Ora (Rodrigue) Schwarzwald
Department of Hebrew and Semitic Languages
Bar-Ilan University, Israel
Ora.schwarzwald@biu.ac.il

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/08.pdf. ISSN 1336-782X.

Derivational suffixes *-ar* and *-nik* in Upper Sorbian

Ken Sasahara, Reitaku University, Japan

Upper Sorbian (West Slavonic, Indo-European) often uses suffixes to create new derived nouns. The suffixes -ar (and its phonological variant -er) and -nik are widely used, and both indicate that a male human is associated with the meaning of the source word. It is not obvious why either suffix is used in a particular case. This paper describes the two suffixes and tries to explain the motivation working behind the choice between -ar and -nik, by analysing the possible approaches of part-of-speech, phonological, and semantic motivations. Based on my analysis of the nature of these suffixes, I argue that semantic motivation offers the most promising and plausible explanation because -ar denotes a male human and -nik denotes a thing inherently. Relying on the concept of cognitive grammar reinforces the hypothesis of semantic motivation. The explanation of a semantic approach shows the increased likelihood that -ar will be preferred if a new concept denoting a male human arises in the future. On the other hand, if a new concept of a thing as a derived noun emerges, it is likely that the noun will take -nik.

Keywords: *Upper Sorbian*, *suffix*, *male human*, *inherent meaning*, *cognitive grammar*

1. Introduction[1]

Sorbian belongs to Indo-European and is a member of the West Slavonic languages of the Slavonic branch – together with Czech, Polish, Slovak, and Kashubian. It is spoken in eastern Germany and classified into two variants – namely Upper Sorbian and Lower Sorbian. Upper Sorbian is distributed in the Free State of Saxony around the upper Spree River, and Lower Sorbian is distributed in the Land of Brandenburg around the lower Spree River. Although there are no demographic statistics for the number of speakers of Upper Sorbian, it is said that there are maximally 15,000 speakers out of 40,000 Sorbs and a few thousand speakers of Lower Sorbian – as long as my understanding reaches through casual conversations with Sorbs. All native speakers of Sorbian are bilingual with German.

As I will show below, Upper Sorbian has many suffixes which produce new nouns. In this paper, I concentrate on two suffixes, *-ar* and *-nik*; I describe their distribution and try to delimit when one is used rather than the other. Throughout this paper, I use the term *Sorbian* for Upper Sorbian.

2. Derivational morpheme denoting male human

In Sorbian, there are many suffixes to create a new derived noun which will denote a male human associated with the stem. Representative ones are shown in (1).

[1] I appreciate deeply Dr. Anja Pohončowa, who reviewed my draft. Her comments and suggestions made this paper worth publishing. All the footnotes in this paper are the realization of them, although I could not include all indications of her for the discussion here. Remaining issues should be my future work. Needless to say, the possible mistakes and misunderstandings belongs of course to me.

(1)	a.	*-ar*/*-er*:	*čitar* ‘reader’	<	*čitać* ‘to read’
	b.	*-ćer*:	*dźěłaćer* ‘worker’	<	*dźěłać* ‘to work’
	c.	*-ćel*:	*darićel* ‘donator’	<	*darić* ‘to present’
	d.	-nik:	*rěčnik* ‘speaker’	<	*rěčeć* ‘to speak’

The suffixes *-ar* and *-er* are phonological variants, the latter of which is realized after a palatalized consonant, and they are in complementary distribution. Note that the verb *spisować*, which we can trace as the source form, does not seem to exist in present-day Sorbian.

The starting point of this research focuses on the question of what kind of stem takes *-ar*/*-er*, *-ćer*, *-ćel*, *-nik,* and so on. Note that all examples in (1) are from a verb-to-noun type of change, but as we will see later, there are also other types.

3. Previous studies

The body of previous study of this research area lacks comprehensive description, except Pohončowa (2017), who describes the word formation in the present-day Upper Sorbian language. Her section 4 (Pohoncova 2017: 76–81) devotes to the derivational processes. She describes thoroughly how derived noun/verb/adjective are created from base noun/verb/adjective.

If we look into the older, but comprehensive grammars, we find that Faßke (1981), on the one hand, is a comprehensive grammar written in German, but the author describes only some lines about such derivational suffixes in his section on noun gender. According to Faßke, the gender of a derived noun with any of the productive derivational suffixes is male. He offers a list of such productive derivational suffixes. On the other hand, Šewc-Schuster’s grammar (1984), written in Upper Sorbian, describes some derivational process of noun, but it is not comprehensive. Faßke’s list of male human suffixes (1981: 401; excerpted) is shown in (2).

(2) List of male human suffixes according to Faßke (1981: 780) (excerpted)

a. *-ar*/*-er*: *čitar* ‘reader’, *lětar* ‘pilot’, *spěwar* ‘singer’, *kuzłar* ‘magician’, *lijer* ‘founder’, *hrajer* ‘player’, *pěsnjer* ‘song-writer’ ...

b. *-ač*: *brodač* ‘man with beard’ ...

c. *-ak*: *čušlak* ‘snooper’, *rybak* ‘fisherman’, *čornak* ‘black horse’ ...

d. *-an*: *delan* ‘lowlander’, *krajan* ‘compatriot’, *Budyšan* ‘man from Bautzen’, *Američan* ‘American’ ...

e. *-’enc*: *młodźenc* ‘youth’, *wotrodźenc* ‘apostate’, *konjenc* ‘(horse) stable’, *kruwjenc* ‘cow stable’, *hosćenc* ‘restaurant’ ...

f. *-ćel*: *darićel* ‘donator’, *spisowaćel* ‘author’, *wěrićel* ‘creditor’, *radźićel* ‘adviser’, *knježićel* ‘ruler’...

g. *-ćer*: *dźěłaćer* ‘worker’, *knježićer* ‘ruler’, *radźićer* ‘adviser’ ...

h. *-nik*: *rólnik* ‘farmer’, *rěznik* ‘butcher’, *pinčnik* ‘waiter’, *hajnik* ‘forester’, *sudnik* ‘judge’ ...

i. *-owc*: *sportowc* ‘sportsman’, *figowc* ‘fig tree’, *seršćowc* ‘brush’, *blachowc* ‘tin pot’, *snopowc* ‘scarecrow’ …

(English translations by me)

The apostrophe in *-’enc* means that the non-palatalized consonant immediately preceding the suffix changes to a palatalized consonant (e.g. *kru**wa*** ‘cow’ > *kru**wj**enc*).

For words denoting a female human, there are some suffixes for them. One of the common suffixes is -*ka*, which is attached to the noun denoting a male human. This suffix is productive. Examples are shown in (3).

(3) a. *přećel**ka*** 'female friend' < *přećel* '(male) friend'
b. *čitar**ka*** 'female reader' < *čitar* '(male) reader'

Note that *čitarka* is derived from *čitar*, which is in turn from *čitać* 'to read' by the derivational suffix -*ar*.

Now we return to my starting point. The question precisely consists of the following three questions:

i) why do some verbs take -*ar*/-*er*, while others use -*nik* etc.?
ii) are there some specific patterns in this choice?
ii) if so, are such patterns predictable?

To answer these questions and to show possible explanations for them, this paper focuses specifically on the suffixes -*ar* and -*nik*.

4. Suffixes -*ar* and -*nik*

In this section, we examine the suffixes -*ar*/-*er* and -*nik* in general. Since, as I described above, the suffix -*er* is the phonological variation of -*ar*, we will use hereafter -*ar* as the representative form. The examples with -*ar* are shown again in (4).

(4) Examples with -*ar*
a. *čit**ar*** 'reader' < *čitać* 'to read'
b. *lět**ar*** 'pilot' < *lětać* 'to fly'
c. *spěw**ar*** 'singer' < *spěwać* 'to sing' / *spěw* 'song'
d. *kuzł**ar*** 'magician' < *kuzłać* 'to conjure, to magic'
e. *lij**er*** 'founder' < *leć* 'to pour'
f. *hraj**er*** 'player' < *hrać* 'to play'
g. *pěsnj**er*** 'song-writer' < *pěsnić* 'to write song'

We see that the noun with -*ar* denotes mainly a person associated with the stem.

The suffix -*nik*, on the other hand, denotes either a person associated with the stem or a thing associated with the stem.

(5) Examples with -*nik*
a. *rěčnik* 'speaker' < *rěčeć* 'to speak'
b. *słownik* 'dictionary' < *słowo* 'word'[2]

At first glance, -*ar* is a derivational suffix for male human, while -*nik* can denote a male human as well as a thing. The fact is, however, that the use is not entirely straightforward. In the following section, I will seek the possible principles behind and explanations for the two suffixes.

[2] The word *słownik* can be a loan word from Polish or Czech (Anja Pohončowa p.c. indicating Jentsch 1999: 241).

5. Possible principles

How can we determine the nature of the two suffixes? On what motivations are these suffix choices based? In this section I will explore likely explanations from the following three approaches, which we will analyse one by one:

i) part-of-speech motivation
ii) phonological motivation
iii) semantic motivation

The first motivation I will consider is part-of-speech motivation. It focuses on examining the stem in terms of its part of speech: that is, whether the derived noun is from a verb (stem), from a noun, or from another part of speech. In other words, this approach seeks the answer from the part-of-speech restriction. The second motivation, phonological motivation, is based on precepts such as the idea that a consonant cannot stand after specific consonant. It can be also called phonotactic motivation. On the other hand, examining semantic motivation, we assume that the morphemes *-ar* and *-nik* have a different core meaning, respectively, and that the selection of either *-ar* or *-nik* is based on that meaning. To discern the nature of these morphemes, I researched the entries in the Upper Sorbian retrograde dictionary by Meškank (2001), which is based on Völkel's dictionary (1981) and contains 44500 entries. It is interesting to observe that Meškank (2001) uses the character *ŕ* for palatalized *r*, which was used up to the early 20th century, but it is no longer used in the actual orthography.

5.1 *Part of speech motivation*

According to Meškank (2001), the suffix *-ar* is mainly attached to a verb stem, but examples of noun + *-ar* are also attested. Examples are shown in (6).

(6) a. *pluwar* 'swimmer' < *pluwać* 'to swim' (verb)
b. *blidar* 'carpenter' < *blido* 'table' (noun)

The suffix *-nik*, on the contrary, is mainly attached to a noun, but words constructed of the root of the (derived) Adjective + *-nik* also occur.[3] However, no example of Verb + *-nik* is found. Examples are given in (7).

(7) a. *słownik* 'dictionary' < *słowo* 'word' (noun)[4]
b. *pinčnik* 'waiter' < *pinca* 'cellar' (noun)
c. *kajkostnik* 'adjective' < *kajkosć* 'property' (adjective derived from interrogative *kajki* 'which')
d. *industrialnik* 'industrial person' < *industrialny* 'industrial' (adjective) < *industrija* 'industry' (noun)

Note that adjective stem of *-nik* must have always been derived from a noun.

To sum up, the analysis of the distribution of the suffixes and the part of speech of the stem suggests an extent of regularity. See Table 1.

[3] It might be better to analyze the root of the (derived) Adjective + *-ik*. (Anja Pohončowa p.c.)
[4] Anja Pohončowa (p.c.) realyzes me that in relation with fn. 3, the word *słownik* can be derived from adjective *słowny* "of word", which comes in turn from the noun *słowo*.

Table 1: Relation of stem to suffix

	verb	noun	adjective
-ar	++	+	–
-nik	–	++	+

From the point of view of the stem, a verb takes only *-ar* and an adjective always selects *-nik*, while a noun takes either *-ar* or *-nik* (although *-nik* is dominant). One might conclude that it is plausible that the part-of-speech motivation explains the selection of the suffixes *-ar* and *-nik*. The question remains as to how we explain the fact that the noun stem can take both suffixes even though the choice of suffix for the noun is determined in almost all cases. It is therefore better to find another solution in order to resolve these questions.

5.2 *Phonological motivation*

Like many Slavonic languages, Sorbian has some phonological features. Palatalization is a good example. It is therefore reasonable to seek an answer to the selection between *-ar* and *-nik* in the phonological area. One might suppose that the suffix selection is dependent on some phonological constraints, i.e. phonotactics.

As stated in §2, the suffixes *-ar* and *-er* are in complementary distribution according to the palatality and non-palatality of the preceding consonant. If the source form ends in a vowel, then *j* is added to the stem final as a sort of glide. This *j* signals palatality, so that the following suffix must be *-er*, not *-ar*. See (8).

(8) a. *hrajer* 'player' < *hrać* 'to play'
b. *lijer* 'founder' < *leć* 'to pour'

Note that in the case of *lijer,* the vowel *e* of *leć* is alternated to *i* through the phonological process. In other words, the choice between *-ar* and *-er* follows the phonotactic rules and has nothing to do with *-nik*.

The suffix *-nik*, on the other hand, can be placed after a non-palatalized as well as a palatalized consonant. No case in which *-nik* follows a vowel immediately can be found. Some examples are shown in (9).

(9) a. *pomocnik* 'helper' < *pomoc* 'help', 'path'[5]
jazyčnik 'lingual (sound)' < *jazyk* 'tongue'
b. *adresnik* 'address book' < *adresa* 'address'
próšnik 'dust bag' < *proch* 'dust'
c. *rěčespytnik* 'linguist' < *rěčespyt* 'linguistics'
pućnik 'road sign, directory' < *puć* 'way'

In each pair above, consonants of the stem final correspond to both non-palatal and palatal (*c* vs. *č*, *s* vs. *š*, *t* vs. *ć*). Since the suffix *-nik* can follow any consonant, palatality does not play a role here. There seems to be no phonotactic rule for *-nik*.

[5] *Pomocnik* can be seen as a derived noun from adjective *pomocny* "of help", which is in turn from noun *pomoc*. (see also fn. 4.)

Looking back at the question of the possible motivations for the choice of *-ar* and *-nik*, we conclude that there is no positive evidence to support the phonological motivation.

5.3 *Semantic motivation*

5.3.1 *Semantics of -ar and -nik*

In terms of semantics, the suffixes *-ar* and *-nik* require investigation of their meanings. As I described in §4, a noun with *-ar* generally denotes a male human while a noun with *-nik* denotes either a human or a thing.

If we look into *-ar* more precisely, we find that the noun typically denotes old concepts, examples of which are shown in (10).

(10) a. *ratar* 'farmer'
b. *rybar* 'fisherman' < *rybać* 'to fish, to angle'[6]
c. *kowar* 'blacksmith' < *kować* 'to forge'

Note that *ratar* has no derivational source, at least in present-day Sorbian. It could be derived from verb *ratarić* 'to farm', to whose stem the zero-suffix is attached. This point of view is, however, not clear to me. There are instead words derived from *ratar* like *ratarstwo* 'agriculture', *ratarski* 'agricultural'.

Sorbs were originally an agricultural and gathering people in the settlements located near water (Kunze 1995). The word *ratar* is old. According to Schuster-Šewc (1978–1989: 1207–1208), it originated in Proto-Slavic *órtajь whose descendant Polish *rataj* is used in the 14th-15th century. The concept which the word *kowar* expresses is relatively new – at least newer than that of *ratar* and *rybar*. Schuster-Šewc (1978–1989: 649) explains that the verb *kować*, which originated in Proto-Slavic **kovati* or **kujǫ* and from which *kowar* is derived, means 'to shoe a horse' and 'to mint coins' besides the general 'to forge'. Although the origin of *horseshoe* is controversial, one of the oldest horseshoes was found in an Etruscan tomb dated around 400 B.C. (Bates 1902). Even though the Sorbian *kować* is used in the early 17th century, it is unknown when Sorbs became familiar with horseshoes. It is, however, not hard to imagine that horseshoes were not new for Sorbs by this time. From these historical descriptions, those words in (10) were apparently in use in the Sorbian community from ancient days, though archaeological verification is still needed. Let us assume, therefore, that *human* is inherent in the meaning of *-ar*.

In contrast, a word with the suffix *-nik*, again, can denote a male human as well as a thing. Take a look at an example *wołojnik* 'pencil'. This word is derived from *wołoj* 'lead'. In other words, the word *wołojnik* means a thing, more precisely a tool or an instrument, to write with lead. Such examples are found throughout Sorbian.

(11) a. *čajnik* 'kettle' < *čaj* 'tea'
b. *wódnik* 'water tank' < *woda* 'water'
c. *popjelnik* 'ashtray' < *popjeł* 'ash'

From these examples we can assume that the suffix *-nik* primarily denotes a thing, but there is another example to reinforce this view. Take for instance the case of *časnikar*, 'clock-maker', which has the suffix *-ar* attached after *-nik*. This word is derived from

[6] The verb *rybać* is taken from Schuster-Šewc (1978–1989: 1255–1256), which seems not to exist in the present-day Upper Sorbian.

*čas**nik*** 'clock', which is originated in turn from *čas* 'time'. Example (12) shows the derivational process.

(12) *čas* 'time' > *čas**nik*** 'clock' > *časnik**ar*** 'clock-maker'

To denote a tool to measure the time (*čas*), the word *čas**nik*** is built. Then, as the noun for the craftsman for that tool, the word *čas**nikar*** is made secondarily. Although such secondary derivations occur infrequently, it is safe to assume that the morpheme *-nik* means a thing.

Based on the discussion heretofore, our working hypothesis is formulated in (13).

(13) Working hypothesis of *-ar* and *-nik*:
The suffix *-ar* denotes a male human and *-nik* denotes a thing inherently.

This hypothesis would be convincing if the semantic motivation were to work on the selection of *-ar* and *-nik*. In §5.3.2, we look at the secondary derivation, that is, the combination of *-ar-nik* and *-nik-ar*, more precisely in order to ascertain this view.

5.3.2 *Secondary derivation*

As for the secondary derivation mentioned above, there are logical possibilities for *X-nik-ar* and *X-ar-nik*. There are not so many words with the secondary derivation in the retrograde dictionary. Comparing the two combinations, the case of *-nik-ar* is very rare.

(14) *X-nik-ar*
hygienikar (= *hygienik*) 'hygenist' (but **hygienar*) < *hygiene* 'hygiene'

Note that there is a word *hygienik* 'hygienist', which has the same meaning as *hygienikar*.[7] In this case, *hygienik* is the base form and then *hygienikar* is made by attaching the suffix *-ar* to it. The form *hygienar*, on the contrary, is not found.

In the cases of *X-ar-nik*, the situation is different from *X-nik-ar*. In terms of the number of the entries in the retrograde dictionary, the former is more common than the latter. In §5.3.1, we posited that *-ar* denotes a male human and *-nik* denotes a thing generally. The semantics of *X-ar-nik* is, however, not merely the addition of the meaning of *-ar* and that of *-nik*. Adding *-nik* to a word that means a male human attached by *-ar* (i.e. the form *X-ar*), the resulting form contains a more abstract, more specific concept. See (15).

(15) *X-ar-nik*
*lěkar**nik*** 'pharmacist' < *lěk**ar*** 'doktor' (? < *lěkować* 'to cure, to heal')[8]
*rybar**nik*** 'expert fisherman' < *ryb**ar*** 'fisherman' < *ryba* 'fish'

In both examples, the derivational process to form *X-ar* is straightforward. From the (possible) source *lěkować* and *ryba*, the derived noun denotes a male human who is associated to the source word. Once *-nik* is attached to those derived nouns, however, they obtain a different meaning although the association to the meaning of the stem remains. *Lěkar**nik*** is no longer a doctor, but a pharmacist. In the same way, *rybarnik* is

[7] It is possible that the form *hygienikar* is a loan word from German *Hygieniker* (Anja Pohončowa p.c.)
[8] The derivational process of *lěkarnik* might be more complex, because it is possible that the word is derived from *lěkarnja* "hospital" – using suffix denoting place – which is in turn from *lěkar*.

not a mere fisherman. This fact suggests that *-nik* has extended the meaning beyond denoting a thing to denoting a person with special property.

Viewing the hypothesis in (13) through the lens of this analysis, we can see that semantic motivation is the most plausible in terms of the selection between *-ar* and *-nik*. In the next subsection, we will look at further examples in English as a supporting explanation from cognitive accounts.

5.3.3 *English -er*

In English, there is a derivational suffix to create a noun denoting a human which is associated with the meaning of the source word. Taylor & Seto (2008: 147) point out that the English *-er* denotes a person. According to them, *X-er* originally means 'a person doing X'. As examples, they point to *walker*, *singer*, *murderer*, and so on. As a result of the semantic development of *-er*, the resulting word can mean a thing or an instrument like *can opener* and *dishwasher*. They conclude that "agent and instrument do often an act as one entity" (original in Japanese; English translation by me).

If we apply this cognitive explanation to the Sorbian suffix, it is likely that our hypothesis in (13), that *-ar* denotes a male human and *-nik* denotes a thing inherently, is correct. In sum, of the three motivations posited in §5, the semantic motivation is the most reliable.

6. Concluding remarks

In this paper, we identified the nature of the Sorbian derivational suffixes *-ar* and *-nik* and tried to give a possible explanation for the selection of these suffixes. We assumed three motivations, i.e. part-of-speech motivation, phonological (or phonotactic) motivation, and semantic motivation. Part-of-speech motivation provides a possible explanation, but it is ultimately weak. Indeed, verb stems tend to take *-ar* overwhelmingly, and adjectives are attached by *-nik* exclusively, but nouns with *-ar* as well as with *-nik* are equally prevalent. As for phonological motivation, we cannot find evidence to support it as an explanation for suffix selection. Semantic motivation is more premising. Assuming that *-ar* denotes a male human and *-nik* denotes a thing inherently, the selection of these suffixes might be explained in a unified way. Ultimately, though, I must emphasize that the first two motivations are totally excluded as explanations for suffix choice. The derivational process is not as straightforward as we think. Considering the ways that several factors may work together, I argue here that semantic motivation works as the strongest of the three factors.

This hypothesis leads to new perspectives in terms of the distribution of *-ar* and *-nik* and in terms of the dominance of semantic factor. Indeed, if a new concept in Sorbian denoting male human should arise in the future, I would argue that it is more likely that the morpheme *-ar* will be preferred because of its inherent meaning of male human. If a new concept of a thing as a derived noun should emerge in Sorbian in the future, it is highly likely that the noun will take the suffix *-nik*. This patterning has already begun, which Pohončowa (2009) argues in relation to internationalism. The tendency will be stronger in the future, as her examples which denote new concepts like *surfowar* 'surfer', *akcionar* 'activist' (from German *Aktionar*) etc. indicate.

Acknowledgements

This research was supported by JSPS KAKENHI Grant Number JP16K02695. I would like to express my gratitude to Fabian Kaulfürst for bibliographical information and suggestions. I thank also Yui Ito for bibliographical assistance.

References

Bates, William Nickerson. 1902. Etruscan Horseshoes from Corneto. *American Journal of Archaeology* 6. 398–403.

Faßke, Helmut. 1981. *Grammatik der obersorbischen Schriftsprache der Gegenwart – Morphologie*. Bautzen: Domowina.

Jentsch, Helmut. 1999. *Die Entwicklung der Lexik der obersorbischen Schriftsprache vom 18. Jahrhundert bis zum Beginn des 20. Jahrhunderts*. Bautzen: Domowina.

Kunze, Peter. 1995. *Kurze Geschichte der Sorben*. Bautzen: Domowina-Verlag.

Meškank, Timo. 2001. *Retrogradny słownik hornjoserbskeje rěče*. Berlin: Mensch & Buch.

Pohončowa, Anja. 2009. Internacionalizmy w hornjoserbskej spisownej rěči přitomnosće: Zarys problematiki [Internationalism in the present-day Upper Sorbian written language: Outline of the problems]. *Lětopis* 56. 81–92.

Pohončowa, Anja. 2017. Zarys hornjoserbskeje słowotworby přitomnosće [Outline of the word formation in the present-day Upper Sorbian]. *Lětopis* 64. 71–86.

Taylor, John R., & Seto, Kenichi. 2008. *Ninchi bunpoo no essensu* [Essence of cognitive grammar]. Tokyo: Taisyukan.

Völkel, Pawoł. 1981. *Prawopisny słownik hornjoserbskeje rěče*. *Hornjoserbsko-němski słownik/ Obersorbisch-deutsches Wörterbuch*. Budyšin: Ludowe nakłdnistwo Domowina.

Schuster-Šewc, Heinz. 1978–1989. *Historisch-etymologisches Wörterbuch der ober- und niedersorbischen Sprache*, Heft 1–24. Bautzen: Domowina.

Šewc-Schuster, Hync. 1984. *Gramatika hornjoserbskeje rěče. 1. zwjazk – fonologija, fonetika, morfologija*. Budyšin: Ludowe nakłdnistwo Domowina.

Faculty of Foreign Studies
Reitaku University
2-1-1, Hikarigaoka, Kashiwa City
Chiba Prefecture, 277-8686
Japan
ksasahar@gmail.com

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/09.pdf. ISSN 1336-782X.

Word-formation type, its reinterpretations, and possible equivalents[1]

Michał Rzepiela, Institute of Polish Language of The Polish Academy of Sciences, Kraków

This paper discusses terminology used to classify the groups of derivatives characterized by some common formal and semantic features and especially those coined with one specific affix. Special attention is given to the concept of word-formation by Czech scholar Miloš Dokulil and the phrase word-formation type *introduced by him. One points out a strict hierarchical order of the terms relating to the products of word-formation in Miloš Dokulil's framework and demonstrates how the phrase* word-formation type *was reinterpreted by other scholars regarding the exploitation of electronic corpora by František Štichá and to the onomasiological theory by Pavol Štekauer. The terms* microstructure (lexicale) *and* semantische Nische*, employed in French and German studies of word-formation, respectively, are comparatively recalled. Finally, attention is focused on the phrase* lexico-semantic class *and its use as usually encountered in computational linguistics.*

Keywords: *word-formation terminology*, *classification of lexical items*, *word-formation type*, *lexico-semantic class*, *computational linguistics*, *Latin*

1. Introduction

The phrase *word-formation type* (Czech *slovotvorný typ*; hereafter WFT), commonly used in the studies of word-formation (WF) in Central European countries, especially in the Czech Republic, Slovakia and Poland, seems to have a rather local extension and might not be understandable to scholars formed by currents outside of the Prague school of structural linguistics. It was coined as part of an original terminological apparatus by Dokulil (1962) to describe WF rules in modern Czech. However, for me, the inspiration to examine Dokulil's terminology came not from Czech as much as from Latin. Some years ago, when I was doing my PhD thesis research on suffixation in medieval Latin (published as an article in 2005), I paid more attention to Polish studies on WF, and I then realized how strongly the approach of Polish linguists, both in theoretical and terminological aspects, was influenced by the ideas of Dokulil. In this way, I hit for the first time Dokulil's pattern of derivatives classification, which I then hoped to apply to Latin.

For me, particularly interesting in his concept – let us remember, based on onomasiological presumptions and structuring lexical items according to the degree of meaning's abstractness and generality[2] – is the level at which formal (morphological) and semantic criteria meet. Dokulil (1962), as the most general term combining these criteria, proposes *word-formation category*. Such a category within nouns, for example, is constituted by names of professions. Lexemes such as *decretista* 'expert in canon law' (← *decretum*), *forestarius* 'forest ranger' (← *foresta*), *mensator* 'carpenter' (← *mensa*) may be enclosed into this category in medieval Latin. They represent the identity of semantics and WF bases (all are desubstantive) but are coined with different suffixes. However, when derivatives

[1] I wish to thank the anonymous reviewer for proofreading my paper and for his/her precious remarks, as well as bibliographical suggestions, which have significantly widened my insight into the discussed issue.

[2] We leave aside presentation of Dokulil's onomasiological theory since it has been recently broadly discussed, e.g. Stekauer (2005), Panocová (2015).

characterized by semantics and WF base are, in addition, coined with the same suffix, e.g. Lat. -*tor*; they are formed according to terminology used by Dokulil: WFT, as illustrated in the Latin deverbal names of professions: *braxator* 'brewer' (← *braxare* 'to brew'), *falcastrator* 'mower' (← *falcastrare* 'to mow'), *impressator* 'printer' (← *impressare* 'to print'). Moreover, to mark any slight semantic differentiation between the lexemes belonging to the same type, it is possible to divide into further subtypes. For example, Latin nouns using -*ista* to designate musicians of very strict specializations (*clavichordista* 'clavichordist', *lutnista* 'lutenist', *organista* 'organist') can be interpreted as a *subtype* of the WFT of desubstantive names of professions using -*ista*. What is striking in this concept is certainly its hierarchical order, which goes from the most general to the most specific term.

2. Groupements associatifs, microstructure (lexicale), semantische Nische

Dokulil (1962: 75, 1986: 220) believes that WFT represents the most important WF characteristics and constitutes a nodal point in the nest of WF relations. A similar conclusion seems to emerge when looking at studies on WF in Latin. It is symptomatic that, especially from the beginning of the 1960s, an increasing number of publications pertaining to the classes of Latin derivatives with one specific affix have begun appearing, focused, from both synchronic and diachronic points of view, on both WF and semantic examinations of individual formations (Perrot 1962; Quellet 1969; Serbat 1975; Martínková 1980; Kircher-Durand 1983, 2002[3]; Fruyt 1986, 1989; Gaide 1988). It is worth noting, as well, that if the terminology concerning morphological structure of derivatives is usually, regardless of the linguistic current or national language tradition, rather well established – terms like *base*, *affix*, *suffix*, *prefix*, *formative* are widespread and well understood – the terminology concerning classes of words kept together by some common semantic and formal criteria is, in contrast, far from being unified. Linguists agree that the derivatives might be grouped together under these criteria, but, at the same time, different terms to name the distinguished items are used.

I was confronted with such terminological heterogeneity when I wanted to write in French about the classes of derivatives formed by specific suffixes in medieval Latin. Naturally, I had to familiarise myself with French terminology, first, which was, however, not always consistent with that of Dokulil (1962). The most systematic overview of the terms of WF as applied to Latin was given by Fruyt (1986). To define the items delimited by Dokulil as WFT, Fruyt uses the phrase "«groupements associatifs» ou «séries» des mots",[4] which she repeats after Marouzeau (1949). Basically, its contents are consistent with Dokulil's definition, although it does not explicitly claim to be a unifying criterion of the words forming an associative group of their WF bases. In contrast, Fruyt emphasizes morphological, as well as phonetic, character of the bases in the context of the measurement of suffix

[3] We mind tome IX of the series *Grammaire fondamentale du Latin*, *Création lexicale: la formation des noms par dérivation suffixale*, elaborated and edited by Chantal Kircher-Duran, which partially unifies selected chapters of the studies of different authors previously published elsewhere and delivers some completely new studies. In the Introduction, Kircher-Duran sketches methodological presumptions shared by the authors of the contributions, such as Saussure and Benveniste structuralism, Coseriu's structural semantics, as well as more modern currents pointing out user's linguistic sentiments.

[4] "Il existe des «groupements associatifs» ou «séries» des mots ayant en commun un suffixe au niveau du signifiant, des traits sémantiques au niveau du signifié et des caractéristiques au niveau des référents" (Fruyt 1986: 23).

productivity, arguing that the more varying the bases to which a suffix may be attached, the more productive a suffix is and the wider its strength of extension should be viewed (1986: 23).

Fruyt (1986) lists, in addition, the terms used by other scholars to describe the groups of derivatives with common affixes being distinguished as somehow specific, which one can also consider as corresponding, more or less, to WFT in Dokulil's (1962) sense, namely: *microstructure*, *micro-structure lexical*, and *semantische Nische*. Dubois (2002), the author of the first term, interprets *microstructure*[5] similarly to Dokulil's WFT, the class constituting subsystems within some larger structure, which, thanks to specific regularities, can function autonomously to some degree. Although Dubois does not at first explain what he means by specific regularities, his further explanations and especially the examples he provides – the same type of WF morphemes added before the same type of bases in French names of relatives – make it clear he thinks about the lexical items representing common semantics and, at the morphological level, common affixes.

The next phrase, *semantische Nische*, has been firmly accepted in German tradition. It was used for the first time by Baldinger (1950).[6] Its definition has been recently developed by Hüning (2009: 183) who interprets *semantic niche* as the result of semantic fragmentation becoming apparent with the passage of time in word formation processes. Hüning defines it more precisely as "the group of words (subset of morphological category) kept together by formal and semantic criteria and extensible through analogy" (ibid.).[7] Since Hüning employed it for contrastive examination (of the morphological and semantic differences between German and Dutch), he also outlines its usefulness for this kind of study. In his opinion, it allows for better generalizations about the systematic differences between two languages than the term *category*.

In turn, the phrase *micro-système lexical* has been consequently, and almost uniquely, used by Kircher-Durand (2002: 88, 149, 190, 198, 200). The context in which Kircher-Durand uses it shows that she means rather strict specific classes with this phrase. If one risked terminological comparison with Dokulil (1962, 1986), these should be placed somewhere between WFTs and WF subtypes. For example, in the case of the Latin suffix *-ensis*, primarily used to coin the names of inhabitants, Kircher-Durand delimits yet seven semantic fields in which it had been, starting from the Latin of the republic period, somehow productive. One of them is constituted by the epithets of deities derived from place names, and, as those establish a relatively extensive series, they are defined by her as *micro-système lexical* (2002: 191).

Incidentally, one yet encounters the phrase *sous-ensembles formeles*, also in the *sous-série* variant, introduced by Deaude (Kircher-Durand 2002: 290) concerning formations such as coined with *-ntia*, *-mōnia,* and *-itia* suffixes, which constitute a subset of the

---

[5] *[O]n donne le nom microstructures à certains sous-systèmes qui à l'intérieur d'une structure plus large, présentent des regularités spécifiques et une organisation qui leur assure une relative autonomie de fonctionnement. Ainsi, les noms des parenté constituent une micro-structure formée, en français, d'unitées linguistiques en nombre fini déterminées sémantiquement par les rapports qu'elles entretiennent entre elles et par rapport à un moi (ego) imaginaire, et morphologiquement par un système particulier des morphèmes (grand et petit dans grand-mère, petit-fils, etc.* (Dubois 2002: 304)

[6] *Metzler Lexicon Sprache* (2010: 260) defines it as follows: "Ableitungen mit demselben Suffix bilden oft Rheinen mit ident. WB [Wortbildungs]-Bedeutung, die K. Baldinger s. N. [semantische Nische] nannte"; "Derivatives with the same suffix often establish series with identical word-formation meaning which by K. Baldinger are named *semantic niches*".

[7] Hüning also briefly sketches the history of the term's interpretation (as by Erben 2000; Rainer 2003).

derivatives with the suffix *-ia*. Those terms seem to correspond to Dokulil's WF subtype more unequivocally.

When we compare the terms and phrases discussed, none of them so clearly emphasize the uniformity of the WF base like Dokulil. For Dokulil (1962: 75, 1986: 221), it is the foundation; for the derivatives forming common WFTs, besides the identity of general onomasiological and semantic structure, the identity of the affix must also represent the identity of the base (of the part of speech from which they are coined). Authors of other definitions speak about common morphological characteristics as well, but if any strict unity is explicitly postulated by them in this regard, it is that of the affix. However, they almost all underline that the discussed items constitute part of some larger structure ("sous-systèmes à l'intérieur d'une structure plus large" (Dubois 2002), "subset of morphological category" (Hüning 2009)). But again, none of them organize these classes in a hierarchically ordered framework to the same extent as Dokulil, which is reflected in the relevant terminology by Dokulil: *WF category* → *WFT* → *WF subtype*.[8]

However, returning to WFT and Latin, the phrase WFT as originally applied by Dokulil to synchronic analysis may also represent advantages for diachronic studies. In this perspective, it has been employed by Pultrová (2011), who accepted it as a classificatory framework when discussing the rules governing the rise, development, and differentiation of Latin formations established by the deverbative nouns and adjectives. It is interesting to note that Pultrová nowhere (in her monograph published in English) proposes any explicit definition of the term WFT[9], instead, using it as if it could be understood by non-Czech readers. Pultrová apparently believes that it may be well understood through the context. That is perhaps the first time that the term WFT has been used on such a scale in an English linguistic text, i.e. used to repeatedly recall the principle of classification of lexical items, where, in addition, no initial extra explanation was given. Even if such a circumstance was caused by the fact that the discussed monograph is a translation from Czech, and for Czech readers such an explanation was not necessary, it shows, at the same, the extent to which the term WFT has been locally (in Middle Europe) accepted in the studies of WF. One can admit as well that such an approach reflects the author's conviction about the universal applicability of the term WFT, and that is, perhaps, the reason why Pultrová decided not to introduce it in a more explicit way.[10]

Pultrová points out advantages connected to the examination of historical linguistic processes within the structure of WFT. First, as it allows (2011: 12–14) proceeding from the function and not the form of the suffix, it helps to avoid interpretative problems resulting from functional and formal differentiation of the lexemes coined with the same suffix, in other words, from suffixal polysemy and a lack of uniformity in the base words of the distinguished lexemes. It is possible thanks to the perspective imposed by WFT that combines rigorously semantic characteristics of the items grouped together with their formal uniformity. Otherwise, no wonder that the author, who decided in her monograph to classify material according to WFTs, already in its title and once more in the introduction (2011: 15),

---

[8] It is worth outlining that Dokulil's terminology had evolved with the passage of time. In Dokulil & Čechova (2011: 105 ff.), the term *category* is exclusively reserved for classifications at onomasiological level, while the previous *word-formation category* is defined as *word-formation class* (Czech *slovotvorná třída*). In addition, the phrase *word-formation subcategory* used in Dokulil (1962) is not encountered in Dokulil & Čechova (2011).

[9] By Pultrová (2011) WFT = *word-formative type.*

[10] In reality, the author, at the beginning, discusses generally Dokulil's concept of word formation and it is against this background that Pultrová introduces the term WFT (2011: 10–12).

precisely defines the categorical status of the base of the lexical items which Pultrová intends to discuss, namely that these will be deverbative nouns and adjectives.

Second, since the perspective of WFT is biased towards analysing the structure of the whole word, not only the suffix, it delivers the ground for examining the sound structure of lexical units within individual classes. Solving phonological problems from the point of view of WFT may significantly help the separation of inherited and analogical items, both in what concerns the distinguished items as a given WFT they enter and the individual formations (2011: 10).

The distinction of the inherited and analogical formations at more specific (within WFT) and more general (as WFT) levels, as can be observed in the study of Pultrová (2011), can be viewed as the third advantage owed to the application of WFT to diachronic studies.

Additionally, in the introductory chapter Pultrová (2011: 11–12) mentions the first, in effect, systematic classification of Latin derivatives, proposed by Leumann (1944); however, as she outlines, it failed to be put into practice. Leumann ordered his classification by the parts of speech, both of the word bases and the derivatives, and accordingly, thus delimited the main syntactic categories distinguished furthermore, as Pultrová names them, *traditional word-formative classes*, i.e. also taking into account semantic aspects, such as *nomina agentis* and *actionis*. Pultrová rightly points out the similarity of this concept to the classification put forward by Dokulil. As for terminological aspects, Leumann speaks about the kernel or main function (*Kernfunktion* or *Grundfunktion*) of the suffix, which should be distinguished, and stresses that this overlaps mostly with its historically primitive function.

3. Štichá's reinterpretation of Dokulil's framework

Dokulil's terminological apparatus has been recently newly approached by another Czech linguist, Štichá (2012, 2013, 2018), who tries to evaluate its usefulness especially from the point of view of WF analysis made with electronic corpora. Štichá has paid particular attention to the phrase WFT. On one hand, he points out (Štichá 2013: 341–344) its improper use (not consistent with Dokulil's definition), encountered by some scholars, but, on the other hand (2013: 342–344), its insufficiency for classifying data obtained from electronic corpora, insofar it is rigidly interpreted according to this definition. As a manifestation of improper use of the phrase, Štichá quotes relatively new studies, by Chýlová (2010) and Tůmová (2007), in which the WFTs were delimited derivatives not uniform with the part of speech of their bases, for instance, those derived from adjectives and those derived from substantives. Nevertheless, it should be emphasized that a less rigid interpretation of Dokulil's phrase has occurred before. For instance, Grzegorczykowa & Puzynina (1998: 384) claimed that within WFT one abstracts from the shape of the WF base (but, surprisingly, examples of the derivatives provided to illustrate their definition all represent uniformity of the base). Horecký (1999: 11, 2007: 26), in turn, interpreted the class of place names with the suffix -*áreň*, in Slovak, as one WFT that can make use both of a substantival and actional motivant.[11]

[11] Horecký provides examples of desubstantive *lekáreň* 'pharmacy', *vináreň* 'wine cellar', as well as deverbal *sušiareň* 'drying room' and *plaváreň* 'swimming pool'. According to the terminology used by Dokulil (1962: 75), the classes characterized by the identity of the affix and semantics but not the base should be defined as WF subcategories.

Although Štichá believes (2018: 165) that the relation between the part of speech of the WF base and the affix of the derivative is crucial for WFT delimitation, he outlines, at the same time, that it would not be suitable to reduce such delimitations only to this relation. Even if it constitutes some general derivational model it, as he observes, does not allow considering, for example, the phenomenon of affixal polysemy. As he concludes, for the users of language, when choosing a suffix, the semantic category of a word being coined with it is often apparently not relevant. That is the reason why one, when delimiting WFT, should also consider categorical meanings of the derivatives. This consideration allows, then, distinguishing the categories in which a given suffix is preferably chosen by the users. The observation that polysemic affixes are in individual categories differentiated by their productivity is foundational for Štichá's interpretation of the term WFT.

It is also worth noting that the explicit definition of WFT that he proposes (Štichá 2018: 164) is rather general and similarly not only oriented on the relation between the base and affix. He states, "by word-formation type we understand, in general sense, the sum of characteristics of a given class of derivatives, insofar this class is defined by a given set of word-formation relations".[12] In addition, he recommends paying more attention to some of these relations, especially to the following: Is a derivation base constituted of a single or multi-word item? What operation was used to coin a new word (e.g. affixation, compounding)? Is the part of speech of the base word and the derivative the same or different? Are derivatives with the same affix formed from common base words (belonging to one given category) differentiated from the category they represent? As long as such semantic differentiation occurs and, in addition, a different degree of productivity between the derivatives in question can be observed, it may be, according to Štichá, an indication to distinguish different WFTs.

He points out, at the same time, the advantage delivered by big electronic corpora of easily and quickly discerning between frequent and not frequent classes (2013: 341). However, he doesn't postulate (2013: 342) some hazardous query but suggests identifying first, at least preliminarily, the typology of WF structures we want to examine in corpus. In fact, the purpose of such a search should be to objectify our presumptions, or, perchance, to discover some other regularities which we couldn't notice without examining a huge amount of texts. He even declares to be willing to consider WF categories and WFTs to be any category or type of structure being somehow specific (Štichá 2012: 97).

As one can see, in the delimiting of WFT, the central role is assigned to the productivity of word formation structures by Štichá. One can even risk the thesis that for him it becomes the main criterion of such delimitation. Consequently, Štichá rejects (2013: 344, 2018: 165) the phrase WF subtype as entailing the delimitation of too many specific structures, reserved by Dokulil for any more specific items within WFT. Štichá (2012: 98–99) accounts for Dokulil's classification framework (emphasizing the importance of WF subtypes) by the fact that he did not dispose of big text collections and while constructing his theory, mainly used data taken from dictionaries. So, he effectively could not have had the possibility to examine the productivity of the items he considered worth distinguishing as specific ones. However, it is worth underlining that Štichá appreciates Dokulil's intuition that for an exact WF analysis it is necessary to calculate productivity at different levels and degrees. As Štichá outlines (2012: 100–101), especially Dokulil's distinction between

[12] "Slovotvorným typem v obecném smyslu rozumíme souhrn vlastností určité třídy odvozených slov definované určitým souborem slovotvorných vztahů" (Štichá 2018: 164).

absolute and relative (depending on specific styles) frequency of WF methods, types and means can significantly contribute to the studies based on electronic corpora. Štichá (2018: 169 ff.), pursuing Dokulil's distinction of systematic and empiric, absolute and relative productivity proposes his own distinction, more appropriate as he believes, of systematic (potential or at the level of langue) and real (at the level of parole), as well as hidden productivity.

The real productivity can be established insofar as in linguistic and extra-linguistic circumstances it will be considered all together. Štichá stresses the importance that statistical examination of individual groups of derivatives aiming to distinguish the most common and rarest items has in this context. Štichá also postulates that, based on statistical examination, the degrees of productivity of the affixes operating within a given WFT should be determined (2018: 169–170). In fact, he not only determines such degrees (six in total), but also strictly defines the percentage of the derivatives coined with a given affix that should be decisive about enclosing this affix to one related degree or another.

Finally, according to Štichá (2018: 166), as an important criterion of the delimitation of WFT, insofar as one strictly combines it with productivity, the morphological structure of the base word and its derivational history (from which word and with what method it was coined) should be regarded. Štichá (2013: 344–346) also provides examples of WFTs as he delimited them, based on the data extracted from the Czech National Corpus, formed namely by the deadjective names of persons with the suffix *-ec*. He places here, indeed, particular emphasis on the influence which the morphological and phonological structure of the base words might have on the real productivity of the WFTs.

In contrast, a real polemic approach of Štichá towards Dokulil can be observed when the former states his opinion concerning onomasiological theory, which for the latter is essential in interpreting WF processes. Štichá redefines the position of WFT in Dokulil's terminological framework, subjecting two fundamental phrases in this framework to critical analysis, namely *onomasiological category* and *word-formation category*. He argues (2012: 97–98) that the very general character of onomasiological categories constitutes their strengths and, at the same time, their limitations and believes that those as superfluous might be well replaced just by *word-formation categories* – the term equally extensible in abstractness and generality.[13] Insofar as WFT is concerned, he underlines that Dokulil's postulate of triple uniformity (of onomasiological category, WF base, and formative) entails difficulty when delimiting some WFTs because of the restraint number of basic conceptual categories admitted by Dokulil to represent the onomasiological structure. Since there are only four categories parallel to the main parts of speech (substance, action, quality, and concomitant circumstance) consequently, for example, the notion of *person* is not included into the set of onomasiological categories. And since for Dokulil the most important characteristic of WF base is the part of speech, consequently, for example, the Czech nouns *vyslanec* 'messenger' and *makovec* 'poppy-seed cake', both representing deadjective derivatives with the suffix *-ec*, are included in the same WFT by him. Štichá (2013: 344–345) strongly opposes such delimitations because, as he argues, the need to coin new names of persons usually results in language of a given period from other social conditions than the need to coin the names of non-persons. Thus, as he argues, the accuracy of the measurement of WFT productivity might be affected.

[13] Štichá, in his most recent systematic presentation of the issue of WF and related terminology, in academic Czech grammar (2018: 163), declares he will not intentionally use either the phrase *onomasiological category* nor the adjective *onomasiological*.

4. WFT in Štekauer's onomasiological model

The phrase WFT is also employed by Štekauer as a part of the terminological system in his onomasiological model of WF. To better understand what place it occupies there, it is suitable to be reminded of some principles of Štekauer's concept. First of all, his interest is outside of the semasiological approach, which is effectively focused on analysing the results of the process of coining new words, on "the already existing word-stock" (Štekauer 2005: 207). The onomasiological method, having an interest in the naming act itself, focuses instead on the individual steps of the naming process (Panocová 2015: 45). According to Štekauer (2005: 226), the approach he proposes aims to interrelate the cognitive abilities of the speech community with both extra-linguistic and linguistic phenomena. Consequently, the level of traditional structuralist analysis focused on the examination of relations between formal and semantic characteristics of derivatives is in it degraded and subordinated to this conceptual one (cf. Štekauer 2005: 213, figure 1). Štekauer (2005: 226) wants to see his model as a reaction to the formalism of generative morphologists. For us, however, the more interesting question is how it is positioned with respect to the model of Dokulil. Štekauer (2005: 226) clearly exposes his objective, which is to pass over the borders between traditionally distinguished WF processes by accounting for them, no matter which WF method intervenes, "by means of the same word-formation principles." Such an approach manifests in combining the two above-mentioned points of view, namely the onomasiological and semasiological, in one framework. To be more accurate, Štekauer (2005: 215–217) distinguishes and includes into one framework onomasiological and onomatological levels of analysis, as the former is concept-grounded while the latter morpheme-grounded.

In contrast, when one reads Dokulil, one can get the impression that his presentation of onomasiological theory serves much more as a preparation for the subsequent exposition of WF rules.[14] From this point of view, that he assigns such a high place in the nest of WF relations to the WFT should not be a surprise.

In turn, in Štekauer's model, the phrase WFT appears only when the computation of productivity is in question.[15] In addition, it is rather not distinguished as an autonomous term, but as one of the terms of the four terms complex in relation, denoting four levels at which productivity is to be calculated (Štekauer 2005: 221 ff.). These are namely: onomasiological types, WF types, morphological types, and WF rules. At the first level, one examines the degree to which individual productive ways of forming naming units are involved. These ways may sometimes overlap with traditionally distinguished WF methods (like prefixation and affixation), but since the concept of onomasiological types is based on cognitive premises and starts from an extra-linguistic reality which is to be named, one abstracts from the WF (formal) perspective here. At this level, it is possible to evaluate the preference of language

[14] Even if I am wrong in interpreting Dokulil's intentions, it is a fact that a similar approach, a virtual separation of the onomasiological level from the WF description, has been accepted as practice in some (if not in the majority of) studies referencing Dokulil. For example, in Polish grammar (Grzegorczykowa & Puzynina 1998: 390 ff.; Nagórko 2003: 214 ff.) authors do not discuss the principles of his onomasiological theory, but once noted that he departed from it and had delimited three categories of derivatives, namely mutational, modificational, and transpositional, they pass to their own classifications, based on the mentioned categories, without making further links, unless incidental, to the onomasiological theory. It should be added, however, that Grzegorczykowa devoted some lines to Dokulil's onomasiological concept in her previous, separate, volume (1973: 27) dealing with Polish word formation.

[15] It should be underlined that in Štekauer's model, a foundational principle is that all WF and morphological types and rules are 100% productive and calculation is restricted only to actual words (Štekauer 2005: 222–223).

users for one of two contradictory tendencies: economy of speech and explicitness of expression.

At the second level, which is related to the conceptual categories (such as agent, location, and action), naming units (words) of different structure, defined just as different WFTs, are taken into account when computing productivity. Different structure means here a set of possible combinations of conceptual categories reflected in the structure of a given naming unit. If one took examples of these combinations provided by Štekauer (2005: 222) and applied traditional WF terminology to them, one would conclude that WFTs are delimited units regardless of the type of their base and the shape of their formatives; in other words, one abstracts here, once again, from a WF viewpoint. So, some substantial differences between Štekauer's and Dokulil's concept of WFT emerge here.

Štekauer proposes, in addition, the phrase *word-formation type cluster* as unifying various WFTs regarding the particular conceptual category, i.e. encompassing WFTs used "to coin new naming units falling within one and the same conceptual category". He admits, at the same time, the possibility that the productivity of an individual WFT within a given WFT cluster may be calculated. Due to the variety of representations of the WFTs at the morphological level, he delimits as the third level of productivity the calculation of morphological types. Similarly, the particular conceptual category distinguishes the morphological types cluster encompassing all morphological types used to coin new naming units within it. The individual morphological type may be calculated in similar terms to WFT clusters.

The fourth level, WF rules, is constituted by the unity of two precedents: WFTs and morphological types. Just this last phrase, *WF rules*, corresponds most closely to the Dokulil's WFT as it represents "the unity of the onomasiological and onomatological structures" (Štekauer 2005: 223). The examples of lexical units, provided by Štekauer, which might be put at this level, additionally strengthen this impression. Of course, the question remains whether this term has a chance to be employed as frequently as Dokulil's WFT. Although it has been regularly employed after 2005 (e.g. Körtvélyessy & Štekauer 2014: 413, 415, 419), the phrase *onomasiological type* seems to be of greater importance to the presentation of the onomasiological principles on Štekauer's concept.

To summarize our comparison, Dokulil and Štekauer use the term WFT in quite different senses; the latter entirely abstracts from a WF perspective. A more accurate equivalent to Dokulil's WFT is established by Štekauer's WF rules. These two phrases are not, however, of equal importance in the two respective models. While in Dokulil's model WFT represents a crucial point in the nest of WF relations, in Štekauer's WF rules intervene only at the end of the naming process, after the conceptual level is structured. An equal rank equivalent to WFT should rather be indicated by onomasiological types since as much in Dokulil's model lexical units are structured around WFTs as in Štekauer's one around onomasiological types.[16]

[16] In practice, it can be observed when comparing the internal organization of the books or articles presenting respective points of view (as for Štekauer's model inspired studies, e.g. Panocová 2015: Chapter Five).

5. Lexico-semantic class vs. WF analysis and vs. automatic extraction of lexical data

The criterion of delimiting WFTs, as put forward by Štichá, based on productivity and emphasizing the role of big electronic corpora encourages, if not makes inevitable, methodological, as well as terminological, confrontation with the practice of classifying lexical items in the domain of computational linguistics. If we look for a term commonly used nowadays to classify sets of lexical items being extracted from big electronic corpora, it is, as it seems, the phrase *lexico-semantic class* (LSC), which merits particular attention. Obviously, it had been used before computational linguistics developed, and also in WF related studies.[17] Among these studies, an important place is occupied by the monograph of Pounder dealing with the question of paradigms operating in WF morphology, since the framework of the derivatives classification by her adopted is actually based on LSCs, defined generally as "set of bases united semantically" (Pounder 2000: 663). However, it must be underlined that the distinction of individual LSCs is not the objective of this study, which is, indeed, the identification of class-specific paradigms productive in the particular LSC. Among those, Pounder understands small domains of systematic paradigms revealed by LSC as active (Pounder 2000: 662), which are, in turn, interpreted by Štekauer as a "subparadigm of systematic paradigm," where the former differs from the latter only in size (Štekauer 2014b: 367–368). Since Pounder (2000: 663–664) wants to view LSC as "the main organizational principle of the lexicon for word-formation, within which recurring paradigm types play a role" one may define it, in a larger sense, as a semantic platform allowing the identification of further regularities, also morphological, of lexical items. One can, then, risk the hypothesis that from this point of view the function of LSC approaches somewhat like the WFT.

Pounder specifically examines German denominal and deadjectival adjective formations. An important conclusion Pounder comes to is that LSC allows the users of a language to make predictions relating to the set of the phenomena, mainly semantic, but also concerning form-meaning relations within the paradigm (2000: 662). Pounder wonders, furthermore, if the LSC may play a similar role in WF systems of other languages and if it has, in what concerns linguistic prediction, universal importance. Pounder provides a few examples from Russian and Labrador Inuktikut which make such presumption plausible (2000: 665). It is worth noting that, similarly, the concept of WFT stimulates Štichá to reflect on the possibilities of predicting lexical meaning based on the analysis of the WF structure of the derivatives (2018: 172–179). Of course, by Štichá the emphasis is more clearly laid at the starting point of the investigation, on morphology.

As stated, LSC as a phrase grouping lexical items has also occupied a particular place in the domain of computational lexical semantics. However, in this domain, the methods of its delimitation are usually abstracted from WF principles, emphasizing, by contrast, the role of context in which the words appear. One assumes (Jurafsky & Martin 2009: 671 ff.) that the examination of context similarity will enable one to detect semantic similarity. Retrieval of the LSCs is essentially based on the distributional properties of the word. The method of extracting LSCs may be either supervised or unsupervised – depending on whether we (or rather, the machine) dispose of hand-labelled (with correct sense) data or not when making a word sense disambiguation. In the latter case, the senses of a word are identified

[17] We enumerate here some works in which LSC has been adopted as the framework for classification of lexical items, regardless of the degree to which the WF viewpoint is taken into account: Pounder 2000; Korhonen & Briscoe 2004; Gamallo et al. 2007; Romeo et al. 2012; Michow 2013; Malatowska-Statkiewicz 2014.

automatically, thanks to adequate machine learning procedure. When dealing with large corpora, unsupervised methods are preferred. They consist of building clusters both of words and local syntactic contexts. As a consequence, LSCs are generated through a clustering process. Each cluster, as the result of merging its most prototypical constituents (word and contexts), represents an LSC and is used as a so-called centroid to word classification. (Gamallo et al. 2007: 37 ff.)

Partington (1998: 138) defines LSC as "a set of lexical items which have a high probability of being found together in text or collection of texts of the same genre" and argues that statistically established co-occurrence of lexical items might reflect the way in which language users organize the world. Such a cognitively inspired presumption that the results obtained from the query of corpora are not random and might reveal certain regularities of which we are sometimes not aware is essential in computational linguistics. On the other hand, the methods of establishing LSCs based on automatic extraction of lexical data sometimes give the impression of having recourse to the criteria being too vague and abstracting too much from the morphological shape of the words. Partington enumerates an example of the lexemes forming one LSC with nouns such as *table*, *chair*, and *furniture* (1998: 138). I put aside, at this point, further discussion since it would digress from the paper's central topic. My intention was here to show that the approach of computational linguistics may as much objectify the procedure of determining lexico-semantic regularities, as separate it from WF perspective.

6. Conclusion

The overview of the terms relating to the classification of lexical items being distinguished or emerging as somehow specific has shown that these terms, according to different methodological approaches, have served to bring out more or less different sets of common characteristics of concerned items. Special attention was given to the phrase WFT which was introduced by its author, Dokulil (1962), to individuate specific meeting points of morphological and semantic characteristics of the groups of derivatives. The phrase WFT appears, as well, in Štekauer's (2005) terminological apparatus, where, nevertheless, it is involved only when the onomatological (morphologically oriented) level of the process of coining new lexical units is considered, according to the onomasiological interest of the author. Moreover, WFT does not occupy an equally important place as in Dokulil's framework, in which it gets the status of the central point of WF relations; under Štekauer, WFT is reduced to the role of one of the components constituting WF rules, the phrase used by the author as a platform for unifying onomasiological (cognitively oriented) and onomatological levels. This last phrase is, effectively, that which corresponds most closely, insofar as such a comparison is justified, to WFT as understood by Dokulil.

In turn, what seems common to Dokulil and Štekauer is their ambition to propose a possibly holistic framework for describing WF relations. Their ambition is reflected in the hierarchically ordered notional systems they propose. However, while Dokulil, parting from onomasiological principles, focuses furthermore on the rules unifying morphological and semantic levels, Štekauer tends to liberate the interpretation of WF process from any formal restrictions, to examine it regardless of the WF methods and means actually chosen by the users of a language.

One has argued in this study that it is possible to indicate equivalents to the term WFT, as conceived by Dokulil, used by the scholars from other structuralist schools outside of Prague. However, the terms evoked in this context (*groupements associatifs*, *microstructure*, *semantische Nische*) do not seem to function, contrarily to Dokulil's work, as part of a wider, specially coined to describe WF relations and, in addition, hierarchically ordered apparatus. On the other hand, it is true that they almost all entail interpretations of lexical items being characterized by specific morphological and semantic identity as a subset of some more general structure.

One also paid attention to the possibilities which the concept of WFT may offer to strictly diachronic analysis of WF processes (Pultrová 2011). In the context of a diachronic approach, a possible utility of the phrase WF subtype was emphasized too.

In what concerns the development of the phrase WFT and, in a more general sense, of Dokulil's concept of WF, one gets the impression that they have undergone constant interpretation, not even as re-interpretation as much as deeper and more insightful interpretation, in the sectors already recognized by Dokulil as relevant for WF. The question of productivity stands out here as crucial. Dokulil's distinction of its different types and degrees has been resumed both by Štekauer and Štichá in the way to accept productivity as a kind of absolute criterion in their own concepts. Though Štekauer identifies WF with productivity and declares considering, in his model, only 100% productive WF rules, Štichá views productivity as reflecting virtual WF dynamics, creating the most objective criterion of delimitation of WFTs.

Since Štichá (2013) emphasizes the importance of big electronic corpora for the exact measurement of productivity, and postulates to admit them at large scale to the studies of word formation (what effectively already happens), a methodological and terminological confrontation between traditional and computational linguistics, will be, especially for the former, it seems, still a growing challenge. For example, the challenge of how the phrase LSC might be interpreted when being used for traditional WF analysis and how the automatic extraction of lexical items, as briefly shown throughout this study, remain. I am convinced that the phrase WFT imposes a certain, more systematic way of describing WF; in terms of considering regularities combining different linguistic levels and placing WF relations within some hierarchical framework, it may also successfully serve as the device of lexical classification when automatic lexical data extraction is in question. The first steps in this direction (Štichá 2013) are promising.

Abbreviations

LSC = Lexico-Semantic Class
WF = Word-Formation
WFT = Word-Formation Type

References

Baldinger, Kurt. 1950. *Kollektivsuffixe und Kollektivbegriff.* Berlin: Akademie-Verlag.

Chýlová, Helena. 2011. Odmítané slovotvorné typy v brusech jazyka českého. In *Minulost, přítomnost a budoucnost v jazyce a v literatuře* (Sborník z mezinárodní konference pořádané katedrou bohemistiky PF UJEP v Ústí nad Labem. II. díl.), 342–346. Ústí nad Labem: Univerzita J. E. Purkyně.

Daude, Jean. 2002. Les substantifs abstrait de qualité. In Kircher-Durand, Chantal (ed.), *Grammaire fondamentale du Latin IX, Création lexicale: La formation des noms par dérivation suffixale*, 225–305. Louvain: Éditions Peeters.

Dokulil, Miloš. 1962. *Tvoření slov v češtině 1. Teorie odvozování slov.* Praha: Nakladatelství Československé Akademie Věd.

Dokulil, Miloš. 1986. Obecné pojetí tvoření slov. In Horálek, Karel & Knappová, Miloslava & Dokulil, Miloš & Petr, Jan (eds.), *Mluvnice češtiny 1: Fonetika, Fonologie, Morfonologie a morfemika, Tvoření slov*, 193–230. Praha: Academia.

Dokulil, Miloš & Čechová, Marie. 2011. Tvoření slov. In Čechová, Marie & Dokulil, Miloš & Hlavsa, Zdeněk & Hrbáček, Josef & Hrušková, Zdeňka (eds.), *Čeština – řeč a jazyk*, 96–260. Praha: SPN - Státní pedagogické nakladatelství.

Dubois, Jean & Giacomo, Mathée & Guespin, Louis & Marcellesi, Christiane & Marcellesi, Jean-Baptiste & Mével, Jean-Pierre. 2002. *Dictionnaire de linguistique.* Paris: Larousse Bordas VUEF.

Erben, Johannes. 2000. *Einführung in die deutsche Wortbildungslehre* (Grundlagen der Germanistik (GrG), Band 17). Berlin: Erich Schmidt Verlag.

Fruyt, Michèlle. 1986. *Problèmes méthodologiques de dérivation à propos ds suffixes latin en -cus*. Paris: Klincksieck.

Fruyt, Michèlle. 1989. Etude sémantique des "diminutifs" latins: les suffixes -vlus, -culus, -ellus, -illus ...de-substantivaux et de-adjectivaux. Actes du Ve Colloque de Linguistique Latine (Cahiers de l'Institut de Linguistique de Louvain 15), 127–138. Louvain-la-Neuve: Peeters.

Gaide, Françoise. 1988. Les «formes élargies» du «latin vulgaire»: un cas très particulier de la dérivation. *Latomus* 47(3). 584–592.

Gaise, Francoise. 1988. *Les Substantifs masculins latins en -(i)ō, -(i)ōnis*. Louvain: Éditions Peeters.

Gamallo, Pablo & Lopes, Gabriel P. & Agustini, Alexandre. 2007. Extraction of Lexico-Semantic Classes from Text. *PICS - Publication Series of the Institute of Cognitive Science* 1. 39–47.

Glück, Helmut (ed.). 2010. *Metzler Lexikon Sprache*, 4th edition. Stuttgart: J. B. Metzler.

Grzegorczykowa, Renata. 1973. Problemy kwantyfikacji w grupie werbalnej. In Topolińska, Zuzanna & Grochowski, Maciej (eds.), *Liczba, ilość, miara. Materiały Konferencji Naukowej w Jadwisinie, 11-13 maja 1972*, 83–99. Wrocław: Zakład Narodowy im. Ossolińskich, Wydawnictwo PAN.

Grzegorczykowa, Renata & Puzynina, Jadwiga, 1998. Słowotwórstwo. In Grzegorczykowa, Renata & Laskowski, Roman & Wróbel, Henryk (eds.), *Gramatyka współczesnego języka polskiego. Moforlogia.* Warszawa: Wydawnictwo Naukowe PWN.

Horecký, Ján. 1999. Onomaziologická interpretácia tvorenia slov. *Slovo a slovesnost* 60(1). 6–12.

Horecký, Ján. 2007. An onomasiological interpretation of word-formation. *SKASE Journal of Theoretical Linguistics* 4(1). 21–27.

Hüning, Matthias. 2009. Semantic niches and analogy in word formation. *Languages in Contrast*, 9(2). 183–201.

Jurafsky, Daniel & Martin, James H. 2009. *Speech and Language Processing: An introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2nd edition. Upper Saddle River, NJ: Pearson Prentice Hall.

Kircher-Durand, Chantal. 1983. Les noms latins en -nus, -na, -num. Étude morpho-sémantique et historique d'une catégorie dérivationnelle du latin classique. *L'Information Grammaticale* 16, 40–52.

Kircher-Durand, Chantal (ed.). 2002. *Grammaire fondamentale du Latin IX, Création lexicale: la formation des noms par dérivation suffixale*. Louvain: Éditions Peeters.

Korhonen, Anna & Briscoe, Ted. 2004. Extended Lexical-Semantic Classification of English Verbs. *Proceedings of the HLT/NAACL Workshop on Computational Lexical Semantics*. Boston: Massachusetts.

Körtvélyessy, Livia & Štekauer, Pavol. 2014. Derivation in a social context. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford Handbook of Derivational Morphology*, 407–423. Oxford: Oxford University Press

Leumann, Manu. 1944. Gruppierung und Funktionen der Wortbildungssuffixe des Lateins. *Museum Helveticum* 1(3). 129–151.

Malanowska-Statkiewicz, Maria. 2014. Nazwy pokrewieństwa i powinowactwa w języku czeskim i polskim. Paradygmat słowotwórczy. *Slavica Wratislaviensia* 159. 245–254.

Martínková, Dana. 1980. Slova tvořená sufixem -ista ve slovní zásobě středověké latiny v českých zemích. *Listy filologické* 103. 144–147.

Michow, Elżbieta. 2013. *Studia nad frazeologią somatyczną języka polskiego i bułgarskiego*. Poland: Kielce.

Marouzeau, Jules. 1949. *Quelques aspects de la formation du latin littéraire*. Collection linguistique (Société de Linguistique de Paris) 53. Paris: Klincksieck.

Nagórko, Alicja. 2003. *Zarys gramatyki polskiej.* Warszawa: Wydawnictwo Naukowe PWN.

Panocová, Renáta. 2015. *Categories of Word Formation and borrowing. An Onomasiological Account of Neoclassical Formations.* Newcastle upon Tyne: Cambridge Scholars Publishing.

Partington, Alan. 1998. *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam: John Benjamins Publishing.

Perrot, Jean. 1962. *Les dérivés latin en -men et -mentum*. Paris: Klincksieck.

Pounder, Amanda. 2000. *Processes and Paradigms in Word-Formation Morphology*. Berlin: Mouton de Gruyter.

Pultrová, Lucie. 2011. *The Latin Deverbative Nouns and Adjectives*. Prague: Charles University, Karolinum Press.

Quellet, Henri. 1969. *Les dérivés latin en -or*. Paris: Klincksieck.

Rainer, Franz. 2003. Semantic fragmentation in word-formation: The case of Spanish -azo. In Sinch, Rajendra & Starosta, Stanley (eds.), *Explorations in seamless morphology*, 197–211. New Delhi: Sage Publications.

Romeo, Lauren & Mendes, Sara & Bel, Núria. 2012. Using qualia information to identify lexical semantic classes in an unsupervised clustering task. *Proceedings of the 24th International Conference on Computational Linguistics COLING, 8–15 December, 2012; Posters*, 1029–1038. Mumbai: The COLING 2012 Organizing Committee.

Rzepiela Michał. 2005. Les néologismes créés par suffixation dans le latin polonais. *Archivum Latinitatis Medii Aevii* 63. 35–44.

Serbat, Guy. 1975. *Les Dérivés nominaux latins à suffixe médiatif*. Paris: Klincksieck.

Štekauer, Pavol. 2005. Onomasiological Approach to Word-formation. In Štekauer, Pavol & Lieber, Rochelle (eds.), *Handbook of Word-formation*, 207–232. Dordrecht: Springer.

Štekauer, Pavol. 2014a. Derivational Paradigms. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford handbook of Derivational Morphology*, 354–369. Oxford: Oxford University Press.

Štekauer, Pavol. 2014b. Derivation in a social context. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford Handbook of Derivational Morphology*, 407–423. Oxford: Oxford University Press.

Štichá, František. 2012. Jak v epoše elektronických korpusů následovat Miloše Dokulila [How to follow Miloš Dokulil in the era of corpora] (Miloši Dokulilovi ke stému výročí narození). *Jazykovědné aktuality* 49(3/4). 95–107.

Štichá, František. 2013. K pojmu "slovotvorný typ" [Concerning the concept "derivational type".]. *Philologica* 72. 341–347.

Štichá, František. 2018. *Velká akademická gramatika spisovné češtiny. I. Morfologie: Druhy slov / Tvoření slov. Část* 1, 162–179. Praha: Academia.

Tůmová, Šárka. 2007. Jeden slovotvorný typ v současné němčině (Ze sborníku Zur Rolle der Sprache im Wandel der Gesellschaft/The Role of Language in Changes of Society, Turku 2002). *Časopis pro moderní filologii* 89(1). 55–59.

Michał Rzepiela
Instytut Języka Polskiego Polskiej Akademii Nauk
al. Mickiewicza 31, 31-120 Kraków
michal.rzepiela@ijp.pan.pl

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/10.pdf. ISSN 1336-782X.

Metaphor and metonymy as a means of economy of expression

Petr Kos, University of South Bohemia in České Budějovice

This paper deals with metaphor and metonymy in word formation from an onomasiological perspective. As the treatment of both metaphor and metonymy in word formation has previously been rather neglected, it aims to show that they play an indispensable role in the formation of new naming units, both morphologically simple and complex, and that they are instrumental in achieving the economy of expression and effability in instances where literal description might fail to do so. The discussion of metaphor and metonymy in word formation is based on an onomasiological model originally based on that of Štekauer's, with some major modifications, and the examples used for illustration are taken from the realm of natural organisms.

Keywords: *onomasiology*, *word formation*, *metaphor*, *metonymy*, *economy of expression*

1. Introduction

The discussion of the place of metaphor and metonymy in word formation has been rather neglected. Metaphor and metonymy have been either completely excluded from word formation, or the discussion of their role has been restricted to specific forms. An example of the former approach to metaphor and metonymy is Štekauer's onomasiological model, in which the processes of semantic shift (including metaphor and metonymy) are referred to as idiosyncratic coinages that are part of the lexical (but not the word-formation) component (cf. Štekauer 2011: 22). Examples of studies that focus on specific forms only are Benczes (2006), which deals with metaphor and metonymy in NN compounds, or Barcelona (2008), which analyses the interaction of metaphor and metonymy in bahuvrihi compounds.

In this paper, I aim to offer an onomasiological viewpoint on the role that metaphor and metonymy have in the formation of new naming units. I aim to show that metaphor and metonymy are an inherent part of word-formation and that they are instrumental in achieving an economical expression of the resulting form.

In §2, I will provide a theoretical basis for the approach to metaphor in naming. In §3, I will describe the advantages of names for natural organisms in the study of onomasiology. In §4, I will describe the onomasiological model which serves as the background for the description of the role of metaphor and metonymy in word formation. In §5, I will discuss the ways in which metaphor and metonymy contribute the economy of form, namely their effect on the complexity of the onomasiological structure, their contribution to a simpler expression of salient features, their capacity to express features that otherwise resist literal expression, and their capacity to condense multiple local features.

2. Image/resemblance metaphor

The approach taken to metaphor in this paper is that of image or resemblance metaphor. Lakoff defined image metaphors as "one-shot" metaphors, which "map only one image onto one other image" (Lakoff 1993: 229). Thus, unlike conceptual metaphors, which are based on conceptualization and map concepts, image metaphors are based on sensory perception and map mental images. A mental image in Lakoff's approach is understood as a static visual image, e.g. shape and colour.

Grady (1997) expanded the notion of image metaphor by a behavioural element, introducing the notion of resemblance metaphor, and described it as an overlap of perceived features from both source and target domains. In literature, the discussion of image/resemblance metaphors is based solely on visual perception, such as Lakoff's (1993) image of an hourglass, Grady's (1997) behaviour-based metaphor *Achilles is a lion*, and Ureña & Faber's (2010) dynamic and static images in marine biology.

However, in compliance with Ureña & Faber's claim that "a mental image need not refer to a 'mental picture' but can also refer to sensory images or image simulations in different sensory modes" (Ureña & Faber 2010: 125), I understand image/resemblance metaphors as metaphors that can also be triggered by other types of sensory perception but vision, such as smell, taste, and sound.

Another difference from the above mentioned studies on image/resemblance metaphor is that their treatment of metaphor is semasiological, i.e. they treat metaphors from the decoder's perspective, cf. the terms source and target domains, where the source domain is the name lending entity and the target domain is the concept to be named. In this paper, I approach metaphor from the coiner's point of view, i.e. how metaphoric naming units come into existence.

In my view, onomasiology in general, not only metaphor and metonymy, is in principle anchored in cognitive linguistics. As will be apparent below, other main themes of cognitive linguistics, such as perception and conceptualization of extra-linguistic reality, levels of categorization, prototypes, and saliency, are also inherent in onomasiological investigations. It is then only natural that the onomasiological models on which my approach is based, those of Štekauer and Grzega, are viewed as cognitive (cf. Štekauer 2005: 8 and Grzega 2007: 5, respectively).

3. Naming units under investigation

All the naming units used as examples in this paper come from the realm of natural organisms. They were chosen for a number of qualities that well suit onomasiological research.

Firstly, an identical referent, e.g. a bird species, is distributed, at least part of the year, over large geographical areas, as most European birds inhabit the whole of the Eurasian space. This means that terms for the same bird species are found in all languages spoken on this territory.

The identity of the concepts is achieved for study by the fact that natural organisms are discrete natural concepts and not constructs of social or cultural reality. We are thus not studying

human inventions, cf. Blank's (2003) onomasiological analysis of the terms for the match in a number of languages, which may culturally vary, but identical concepts stable in time and space. Also, natural organisms are not as prototypical to our everyday experience as, for example, the eye (cf. Tagliavini 1949; Blank 2003), so the names denoting them do not belong to the central core of our vocabulary. This fact allows for greater variability of names within individual languages, which manifest a high level of lexical creativity, leading to a larger sample of names for investigation.

Another important quality of names for natural organisms is a high level of polygenesis, i.e. they were coined in the individual languages without the interference from other languages. This was achieved by the fact that the bulk of names were formed by the country folk with little geographical mobility.

The high level of polygenesis is also apparent within various local names, leading to a number of synonyms within one language. Most natural organisms, very often the non-prototypical ones, have a high number of local names with varied motivation. For example, the names for the long-tailed tit are motivated by the colour of patches on its body, the quality of its plumage, the shape of the body with the conspicuous tail, its overall size, its vocalization, its habitat, the shape of the nest, and the material of the nest. This high level of synonymy is especially important as we may contrast different naming strategies applied to one specific species within one language.

4. Onomasiological model

As has been already mentioned, the onomasiological model used in this paper as a background for the discussion of metaphor and metonymy in word formation has its starting point in Štekauer (1998) and was partly influenced by Grzega (2007). It was originally elaborated to fully account for the creation of names for natural organisms as when the original models were applied several shortcomings became apparent.

A full description of the model and its differences from the original goes beyond the purpose of this paper, thus I will focus on those aspects which are most relevant for the current study. My description will focus on the three main levels of the model, namely the perceptual (Štekauer's conceptual level), the onomasiological, and the onomatological.

4.1 *Perceptual level*[1]

At the perceptual level, the extralinguistic referent is analysed and "both the more general, 'global' features and the more specific, 'local' features of a concept are processed" (Grzega 2005: 77).

[1] The terms *perceptual* and *conceptual* (cf. Štekauer's model) in the name for this level appear to be complementary, as the analysis of an extra-linguistic referent may be based either on perception or conceptualization. As the analysis of all naming units in this paper is based on perception, the term *perceptual level* will be used.

The global features classify the referent into an existing cognitive category, and the local features serve as the distinguishing marker within this category. So, for example, if I see a new organism, thanks to its ability to fly and its characteristic shape (global features) I classify the referent into the already existing mental category BIRD "by comparing the overall image of the referent with other images already in the mind" (Grzega 2007: 6), in other words by comparing it to the prototype of the category (cf. Rosch 1978), and its characteristic colour of wings (a local feature) may be the distinguishing marker within this category.

The ability of language users to categorize the extralinguistic reality, i.e. how specific their cognitive categories are, is determined by their experience. Generally speaking, where one sees a blue tit, another sees a tit, and someone else a bird, only. The highest level of categorization in natural organisms seems to be Berlin's (1992) life-forms, "ranging from five to ten, and among them they include the majority of all named taxa of lesser rank. These life-form taxa are named by linguistic expressions that are lexically analysed as primary lexemes, for example, *tree, vine, bird, grass, mammal*" (Berlin 1992: 15). As is apparent, this folk taxonomy does not correspond to the scientific one.

The local feature can either be static (a salient physical feature) or dynamic (an activity or relation to another entity), and in either case the local feature has a complex internal structure.

The conceptual structure of the former is

ASPECT / PART (QUALITY) FOR THE WHOLE

This, in fact, is double metonymy. The salient feature refers to one of the possible aspects of the referent – ASPECT FOR THE WHOLE, e.g. shape, colour, size, and at the same time this aspect refers to a part only or the referent as a whole, – PART FOR THE WHOLE. The third part of this structure is the quality itself, e.g. what shape, what colour, or what size. At the onomasiological level, however, not all the members of the structure will eventually find their linguistic expression.

Thus, for example, in *black-headed gull*, both QUALITY and PART are expressed, with ASPECT being left out, and in *soap-scented toadstool,* QUALITY and ASPECT are expressed, with PART being left out. In *blue tit, tailor*[2]*,* and *variegated toadstool,* QUALITY, PART, and ASPECT are expressed only, respectively.

The conceptual structure of the dynamic local feature comprises, as suggested by Štekauer (1998), the determining and determined constituents – the determining constituent is an entity in a metonymical relation to the referent, and the determined constituent expresses the type of the relation or merely an activity.

Thus, for example, in *ant-eater*, the referent is in a metonymical relation to the ant (the determining constituent), the relation being expressed by the determined constituent *eat*. In *screecher*, a local name for the swift, the conceptual structure includes the determined constituent only, as no determining constituent is expressible.

[2] The naming unit is motivated by the bird's conspicuous tail.

4.2 *Onomasiological level*

At the onomasiological level, the global features become the onomasiological base and the local features the onomasiological mark. As suggested above, at this level the language user determines how much of the onomasiological structure becomes linguistically expressed, with some of the constituents being left unexpressed.[3]

Importantly, unlike in Štekauer's approach, it is also the onomasiological base that may remain unexpressed. Consider the following examples of naming units denoting three different birds:

(1) a. *devil swallow*
b. *devil bird*
c. *devilling*
d. *devil*

(2) a. *emmet hunter*[4]
b. *eten bird*
c. *emmet*

(3) a. *black-headed gull*
b. *blackhead*

Example (1) consists of local names for the swift, a bird whose salient feature BLACK is expressed metaphorically. The first three names (1a-c) have the onomasiological base expressed, though at a different level of specificity, by the constituents *swallow, bird, -ling*, respectively, but in the last one, *devil*, the onomasiological base remains unexpressed.

In (2) are local names for the wryneck, a bird typically feeding on ants. In all these names, the referent is in metonymical relation to ants, expressed by the determining constituent, and the type of relation is given by the determined constituent. Thus, *emmet hunter* has all the constituents of the onomasiological structure expressed, *eten bird* has the determining constituent and the onomasiological base expressed, whereas in *emmet* the determining constituent is expressed only.

Traditionally, the formation of the morphologically simple naming units *devil* and *emmet* is not considered to be part of word formation, the names being instances of semantic shift, metaphor and metonymy, respectively. However, their morphologically complex counterparts suggest that such naming units undergo the same onomasiological process. We can consider them to be an extreme case of economy of expression, where one constituent of the whole onomasiological structure is expressed only, in which, however, the basic onomasiological structure of the onomasiological mark and base is still retained.

The distinction between the two naming units in (3) lies again in the (non-) expression of the onomasiological base. Nevertheless, as *blackhead* is still morphologically complex, the non-

[3] Cf. Štekauer's onomasiological types.
[4] *Emmet* and *eten* are dialectal terms for the ant.

expression of the base does not lead to its eviction from word formation processes; traditionally, it causes its shift to the category of exocentric compounds.

We can also find examples of the onomasiological base being unexpressed in other languages, e.g. Czech names for various families of dragonflies, *šídlo* 'dragonfly' (lit. 'awl') next to *šídl-atka* 'dragonfly' (lit. 'awl+suf.'), or local names for the sorrel, a plant known for its sour juice, *šťáva* 'sorrel' (lit. 'juice') next to *šťav-lík* and *šťav-el* both 'sorrel' (lit. 'juice+suf.').

The non-expression of the onomasiological base may also be systematic, as in English names for butterflies, in which the onomasiological base *butterfly*, apparently for its three-syllable structure, is never expressed, as in the butterfly names *blue, white, yellow* based on the colour of their wings; these terms would traditionally be classified as instances of conversion.

4.3 *Onomatological level*

At the onomatological level, the features from the perceptual level find their linguistic expression. The underlying principle in choosing the expression is *in search of*. We scan our known world *in search of* the linguistic material that will express the specific feature. This can be a literal expression, if the feature is expressible literally; or we can search in other conceptual domains (or even the same one) to find the same salient feature, this time yielding a metaphoric or metonymical expression, respectively.

(4) a. EN[5] *black martin*
b. CS *vlaštovka černá* (lit. 'black swallow')
c. FR *martinet noir* (lit. 'black martinet')

(5) a. EN *collier*
b. EN *devil bird*
c. CS *papežník* (lit. 'pope+suf.')
d. CS *uhlíř* (lit. 'collier')
e. SK *kominár* (lit. 'chimney sweep')
f. FI *tervapääsky* (lit. 'tar swallow')

Example (4) consists of names for the swift in which the salient feature BLACK is expressed literally, whereas in (5) the same salient feature in the same bird is found in other domains, namely collier, devil, pope[6], chimney sweep, and tar. The search for the salient feature(s) from the perceptual level in other domains thus significantly broadens the choice of naming possibilities at the onomatological level.

[5] The two-letter codes for languages are in compliance with the international norm ISO 639-1.
[6] Apparently from times when people judged the colour of the pope's garment from that of the local priest.

5. Metaphor and metonymy and the economy of form

The choice between the literal and non-literal expression, however, is not always free. A lot of salient features would simply be too long to be expressed literally or would resist literal expression altogether. Even in those cases where there seems to be a choice, the use of metaphor (and metonymy) enables us to form naming units with greater economy; of course, at the cost of their transparency. Together with Körtvélyessy et. al., I see the concept of semantic transparency to be "interrelated to that of meaning predictability" (Körtvélyessy et. al. 2015: 85). Metaphor in principle lowers the degree of meaning predictability in naming as it is not clear which salient feature(s) triggered the metaphoric expression, in other words which salient features were mapped.

Current studies that deal with "the fundamental conflict in word formation (and language in general), that between the explicitness of expression and the economy of expression" (Štekauer et al. 2005: 2) in onomasiological research focus their attention on onomasiological types (OTs), specifically on "the degree and the nature of completeness of morphematic representation of the onomasiological structure" (Körtvélyessy et. al. 2015: 92).

It will be shown below that it is also metaphor that has the capacity to affect the morphematic representation of the onomasiological structure towards a more economical expression (see §5.1); both metaphor and metonymy also allow for a morphologically simpler expression of salient features that would require a lengthy literal description (see §5.2) or would be literally inexpressible (see §5.3). Finally, I aim to show that metaphor also enables us to condense more salient features from the perceptual level into one linguistically simple expression (see §5.4).

5.1 *Affecting the complexity of the onomasiological structure*

The effect of metaphor on the onomasiological structure is twofold – it affects the expressibility of the onomasiological base as well as the complexity of the structure of the onomasiological mark.

5.1.1 *Affecting the expressibility of the base*

If we go back to (4), in which the salient feature BLACK in the various names for the swift is expressed literally, we can see that, as the colour is expressed by an adjective, such naming units always need to have at least one more constituent of the onomasiological structure expressed; be it the base, as in all instances in (4), or another constituent from the structure of the mark, as in *blackhead*, in this case PART. In naming units in which the feature BLACK is expressed metaphorically, this condition disappears. Thus, we may find *devil* next to *devil bird* or naming units, such as *collier*, the Czech *uhlíř* 'collier', and the Slovak *kominár* 'chimney sweep'. Although the latter mentioned are morphologically complex, in the naming process they are used as one unit, so from the onomasiological perspective we should see them as monemes.

The need for the economy of expression becomes more apparent when the adjectival onomasiological mark is a multi-word expression, as in the English names for *Tricholomopsis rutilans*, a mushroom with a purple cap and yellow gills.

(6) a. *purple-and-yellow agaric*
b. *plums and custard*

While the two metaphors, *plums* and *custard*, express the same salient features as the literal terms, they also enable the onomasiological base to remain unexpressed, achieving a more economical expression, of course at the cost of the transparency of the ontological category the referent belongs to.

5.1.2 *Affecting the onomasiological structure of the mark*
Metaphor also enables the compression of the constituents of the onomasiological structure of the mark into a formally less complex expression. Example (7) gives different English names for the treecreeper, a bird known for climbing up trees, and example (8) gives different French names for *Chironex fleckeri*, a jellyfish which can kill people with its venom.

(7) a. *treecreeper*
b. *tree mouse*
c. *squirrel bird*

(8) a. *piqueur marin* (lit. 'marine stinger')
b. *guêpe de mer* (lit. 'wasp of sea')

In (7) the semantic constituents TREE-CREEP-AGENT can have literal realization as in *tree-creep-er*, or the agent and the determined constituent are compressed into metaphoric *mouse,* as in *tree mouse*, or the agent and both determined and determining constituents are compressed into *squirrel*, as in *squirrel bird*.

The French examples in (8) exhibit the same pattern; the onomasiological structure $\text{STING}_{\text{VERB}}$-AGENT is either expressed literally in the morphologically complex *piqu-eur* 'stinger', or the same structure can find its expression in the metaphorical *guêpe* 'wasp', again compressing the agent and the determined, verbal, constituent into one monomorphemic realization.

The same type of compression may be found in names in which the onomasiological mark is of static nature, as in the names for the black-headed gull:

(9) a. *black-headed gull*
b. *blackhead*
c. *masked gull*
d. *hooded maw*

In (9) both metaphoric expressions, *masked* and *hooded,* compress the two constituents QUALITY (black) and PART (head) from the perceptual level into one, as both mask and hood can be understood as dark coverings of the head, allowing for a shorter, binominal, structure of the respective naming units. The same binominal structure can be achieved by omitting the onomasiological base, as in *blackhead*; the metaphors, however, this time retain the

onomasiological base, achieving a higher level of transparency of the ontological category of the referent.

The economy of expression at the onomasiological level may thus be achieved not only by not expressing some of the constituents from the onomasiological structure, which leads to different onomasiological types in Štekauer's approach, but also by compressing some of these constituents into one formal expression with the use of metaphor.

5.2 *Achieving a simpler expression of a salient feature*

Metaphor not only affects the onomasiological structure of the naming unit but also enables the expression of salient features that would otherwise require a multi-word literal description. Example (10) consists of different names for *Hygrocybe chlorophana*, a mushroom whose salient feature is the bright yellow colour.

(10) a. EN *golden waxcap*
b. EN *sulphur-colored hygrophorus*
c. CS *voskovka citronová* (lit. 'lemon waxcap')

The literal expression of the salient feature BRIGHT YELLOW (or any other shade of the colour) would most likely prove to be uneconomical for naming purposes, so by searching in other conceptual domains at the onomatological level, we succeed in capturing the complexity of the feature while confining ourselves to a simple expression by the use of metaphor, as in the metaphorical *golden*, *sulphur*, and *citronová* 'lemon$_{ADJ}$', which all express a bright yellow colour.

This ability to supply a simple term for a salient feature that would otherwise require a multi-word expression is more apparent in names in which the onomasiological mark is expressed metonymically; in this case, the salient feature has been searched for within the same conceptual domain. The following example comprises names for the sparrow hawk, a bird of prey typically feeding on (any) small birds:

(11) a. EN *sparrow hawk*
b. EN *chicken hawk*
c. SW *speckhök* (lit. 'finch hawk')
d. CS *vrabčák* (lit. 'sparrow+suf.')
e. ES *arrapapájaros* (lit. 'catch small-birds')

In (11) the only language that in a simple term distinguishes small birds and larger birds is Spanish, with its *pájaro* and *ave*, respectively, so the Spanish *arrapapájaros* can be considered to be literal. In the remaining languages, we can see various small birds that represent the salient feature SMALL BIRD, namely *sparrow*, *chicken*, and *finch*, which is an instance of A MEMBER OF A CATEGORY FOR CATEGORY metonymy.

A more illustrative example of this type of metonymy is given in (12), which are names for the swift, a bird that nests on vertical surfaces only.

(12) a. DE *steinschwalbe* (lit. ‘rock swallow’)
b. DE *mauerschwalbe* (lit. ‘wall swallow’)
c. DE *turmschwalbe* (lit. ‘tower swallow’)
d. DE *kirchschwalbe* (lit. ‘church swallow’)
e. CS *skalníček* (lit. ‘rock+suf.$_{\text{DIM}}$’)
f. ZH 楼燕 (lit. ‘tower swallow’)
g. MN [illegible] (lit. ‘tower swift’)

A salient feature, such as VERTICAL SURFACE, would hardly find its way into the name for the bird, so various instances of vertical surface on which the bird nests are chosen instead, as e.g. in the four local German names that include the metonymy A MEMBER OF A CATEGORY FOR CATEGORY in *stein* ‘rock’, *mauer* ‘wall’, *turm* ‘tower’, and *kirch* ‘church’, which are all morphologically simple terms.

5.3 *Beyond literal expression*

So far, we have dealt with salient features for which we have at least some choice between a literal and non-literal expression. Nevertheless, metaphor is also instrumental in expressing features either for which the literal description would be too long for naming purposes or for which there is no possibility of literal description whatsoever.

Ineffability, or the difficulty or impossibility to express percepts and other experiences in words (i.e. literally), is of a different degree for different perceptual modes. “It is arguable for example that, in English at least, it seems generally easier to linguistically code colors than (non-musical) sounds, sounds than tastes, tastes than smells” (Levinson & Majid 2014: 415). Levinson & Majid (2014) make a distinction between linguistic codability (a literal expression) and indirect conveyability (a metaphoric expression). The level of codability may differ in different languages, as for example the language “Yéli Dnye lacks clear color terms (or any word for ‘color’), certainly none beyond white, red and black: so to convey ‘It is blue’ in Yéli Dnye you would have to say, for example, ‘It has the surface appearance similar to the shallow sea over sand’” (Levinson & Majid 2014: 410). In other words, where codability is impossible it is replaced by conveyability. Metaphor thus conveys the feature within formal limits suitable for naming needs.

An example of a feature which to a high degree resists linguistic codability is SHAPE. In (13) are names for the swift, a bird often seen as a silhouette flying above our heads.

(13) a. EN *anchor bird*
b. HU *sarlósfecske* (lit. ‘sickle+swallow’)
c. CS *kosak* (lit. ‘scythe+suf.’)
d. CS *nůžky* (lit. ‘scissors’)
e. CS *vidlák* (lit. ‘pitchfork+suf.’)

In general, it is thinkable that we describe SHAPE literally, but such description would go far beyond the suitable length for a name. The metaphor thus allows for an economical expression where the literal one would fail to do so.

In compliance with the conceptual structure of the static feature at the perceptual level, any ASPECT is also subject to PART FOR WHOLE metonymy. Thus, as is apparent from Figure 1, (13a) applies to the silhouette of the whole bird in flight, whereas in (13b-c) it is only the shape of the spread wings and in (13d-e) it is the shape of the tail that is salient at the perceptual level. In none of these names, the constituent PART is expressed; however, as has been noted above, it is only a matter of the naming strategy at the onomasiological level what constituents of the onomasiological structure are finally expressed; consider naming units, such as *spoonbill*, *pintail*, and *sabretooth*, metaphorically expressing the quality of the shape as well as expressing the part it refers to.

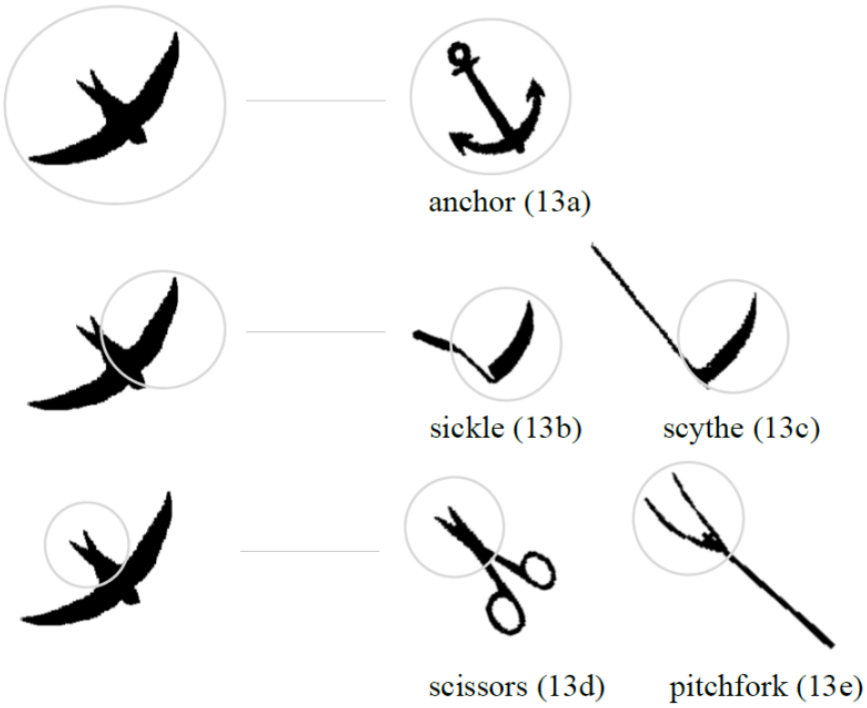


Figure 1: Metaphorical mappings in the names for the swift expressing SHAPE

In order to find a linguistic expression for the salient feature SOUND, we can either *search for* similar characteristics of the perceived vocalization in the realm of the phonemic system of our language, which results in onomatopoeia (for a detailed treatment of the correspondences between the natural sounds and a language's phonemic system, see Tsur 2001), use a verb which generally characterizes the sound (usually lexicalized onomatopoeia), or find the characteristic features of the sound in other domains, which leads to a metaphoric expression. Example (14) consists of names for the long-tailed tit, a bird whose salient feature is, among other things, its vocalization.

(14) a. EN *churr*
b. EN *churn*
c. EN *creak mouse*
d. CS *pilař* (lit. 'saw+suf.')
e. EN *bellringer*
f. CS *sýkora psí* (lit. 'dog tit')
g. EN *dog-tail*
h. SK *pánský psík* (lit. 'lord's dog')
i. DE *hundsmeise* (lit. 'dog tit')
j. DE *querrelmeise* (lit. 'quarrel tit')

Examples (14a-b) represent pure onomatopoeia, (14c)[7] is an instance of the salient feature SOUND expressed by a verb denoting sound, and the remaining examples are metaphoric, in which the characteristic qualities of the vocalization were found in the domains of sound producing entities – the saw, bell, dog, and the act of quarrelling. It is important to note that individual domains may reflect a different part of the bird's vocalization (cf. Kos 2014: 81).

The quality of sound can no longer be expressed *literally,* as all three approaches mentioned above are to a large extent approximative. None of the approaches seems to be more economical than the others; nevertheless, it appears that onomatopoeia and metaphor attempt to grasp the qualities of the perceived sound more faithfully than a general verb denoting sound.

Salient features for which there is no possibility of non-metaphoric expression altogether are, for example, SMELL or TASTE. "The specific qualities of smells […] are not lexically codable in English since there are arguably no words that identify the precise properties of smells" (Levinson & Majid 2014: 411). As an illustration for the impossibility of expressing the quality of SMELL and TASTE any other way but metaphorically may serve the wine aroma wheel[8], in which all wine aromas are described metaphorically.

In (15) we see names which were motivated by their smell or taste. *The miller* is a mushroom whose smell is described as that of freshly ground flour, and *oyster plant*'s leaves are said to taste of oysters.

(15) a. *the miller*
b. *oyster plant*

Metaphor in such cases does not only allow a more economical expression; rather, it is the only possible way to express such salient features.

5.4 *Condensing multiple local features*

[7] The constituent *mouse* in *creak mouse* is not metaphoric, as it represents the onomasiological base with an older term for the titmouse.
[8] https://www.winearomawheel.com/

At the perceptual level, some referents may also be analysed by more than one local feature. As an example consider the two English names for the ghost orchid, whose flower is white in colour, has a characteristic shape, due to the lack of leaves seems to be floating in the air, and grows in marshes, an environment which is associated with supra-natural entities.

(16) a. *ghost orchid*
b. *white frog orchid*

All these salient features are condensed in the metaphoric expression *ghost*, as all the colour, the particular shape, the ability to float in the air, and marshes are characteristic of ghosts. Alternatively, in *white frog orchid* there is no compression of the features into one simple expression – two of the features are expressed separately, the colour and the shape (by a different metaphor), and the two remaining features are missing altogether. By using the metaphor *ghost*, we thus succeed in expressing more local features from the perceptual level with less linguistic material than the literal *white* and the metaphor for the shape only *frog*; again, though, at the cost of transparency.

6. Conclusion

From an onomasiological perspective, metaphor and metonymy in word formation do not represent a creative choice only, as suggested by the title of Benczes (2006), but one of the two basic principles of the expression of salient features, analysed at the perceptual level. At the onomatological level, these features may thus be expressed literally or may be searched for in a different, or even the same, conceptual domain. Finding the salient feature in a different conceptual domain yields a metaphoric expression and finding it in the same one yields A MEMBER OF A CATEGORY FOR CATEGORY metonymy.

The structural possibility at the onomasiological level not to express the onomasiological base leads to the fact that the resulting naming unit may be morphologically simple. I come to the conclusion that such naming units, though being morphologically simple, are coined on the same onomasiological principle as the morphologically complex ones, as both are formed with the conceptual structure of an onomasiological base and a mark. The formation of these simple units, traditionally classified as semantic shift, should thus be treated on a par with the traditional categories of compounding, affixation, etc.

Although the employment of metaphor and metonymy does not necessarily lead to a more economical form of expression, the potential to do so is apparent. The economy may be achieved by affecting the complexity of the onomasiological structure, either by allowing the non-expression of the onomasiological base or by compressing the constituents of the onomasiological structure of the mark into a formally less complex expression; it may also be achieved by the ability to provide a simple expression for features that would necessarily need to be complex, if expressed literally, and metaphor also enables the expression of features that would otherwise resist a literal expression altogether. Last but not least, metaphor allows for the

expression of multiple features, analysed at the perceptual level, into one word, a phenomenon, which can be understood as an implosion of several mental images into one word.

The potential of metaphor and metonymy to form a more economical expression may play its role in the choice of the linguistic form at the onomatological level if both, literal and non-literal, options are possible. As proved by Štekauer et al. (2005), certain sociolinguistic factors influence the choice between economy of expression and its transparency as, for example, the level of education appears to play a role: "while native speakers with university education prefer more precise names, lower educated speakers are more frequently driven by the principle of economy of expression" (Štekauer et al. 2005: 46). The comparison of local bird names, coined by country folk, and standard names, often coined by scientists, suggests a higher frequency of metaphoric names coined by country folk (cf. Kos 2011). However, the exact correlation between the use of metaphor (and metonymy) and the sociolinguistic factors needs to be further studied in more detail.

Sources

Desfayes, Michel. 1998. *A thesaurus of bird names: Etymology of European lexis through paradigms*. Sion, Switzerland: Musée cantonal dhistoire naturelle et La Murithienne.

Hladký, Josef. 1996. *The Czech and the English names of mushrooms*. Brno: Masarykova Univerzita.

Klůz, Zdeněk. 1977. *Naše ptactvo v lidovém názvosloví a vědecké terminologii*. Praha: Státní zemědělské nakladatelství.

Lockwood, William Burley. 1993. *The Oxford dictionary of British bird names*. Oxford: Oxford University Press.

Sibagu – Bird names in East and Southeast Asian languages. (http://www.sibagu.com) (Accessed 2018-07-28.)

References

Barcelona, Antonio. 2008. The interaction of metonymy and metaphor in the meaning and form of "bahuvrihi" compounds. *Annual Review of Cognitive Linguistics* 6(1). 208–281.

Benczes, Réka. 2006. *Creative compounding in English: the semantics of metaphorical and metonymical noun-noun combinations*. Amsterdam: John Benjamins.

Berlin, Brent. 1992. *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies*. Princeton: Princeton University Press.

Blank, Andreas. 2003. Words and concepts in time: Towards diachronic cognitive onomasiology. In Eckardt, Regine & von Heusinger, Klaus & Schwarze, Christoph (eds.), *Words in Time: Diachronic Semantics from Different Points of View*, 37–65. Berlin: Mouton de Gruyter.

Grady, Joseph. 1997. *Foundations of meaning: Primary metaphors and primary stress*. Berkley: University of California. (Dissertation thesis.)

Grzega, Joachim. 2005. Onomasiologial approach to word-formation. A comment on Pavol Štekauer: Onomasiological approach to word-formation. *SKASE Journal of Theoretical Linguistics* 2(2). 76–81.

Grzega, Joachim. 2007. Summary, supplement and index for Grzega, Bezeichnungswandel, 2004. *Onomasiology Online* 8. 18–196.

Körtvélyessy, Lívia & Štekauer, Pavol & Zimmermann, Július. 2015. Word-Formation strategies: Semantic transparency vs. formal economy. In Bauer, Laurie & Körtvélyessy, Lívia & Štekauer, Pavol (eds.), *Semantics of complex words*, 85–113. Dordrecht: Springer.

Kos, Petr. 2011. *Bird names in English: A linguistic analysis*. Praha: Univerzita Karlova v Praze. (Dissertation thesis.)

Kos, Petr. 2014. Zvukomalebná pojmenování ptáků v angličtině. *Časopis pro moderní filologii* 96. 74–81.

Lakoff, George. 1993. The Contemporary theory of metaphor. In Ortony, Andrew (ed.), *Metaphor and thought*, 202–251. Cambridge: Cambridge University Press.

Levinson, Stephen & Majid, Asifa. 2014. Differential ineffability and the senses. *Mind & Language* 29(4). 407–427.

Rosch, Eleanor. 1978. Principles of categorization. In Rosch, Eleanor & Lloyd, Barbara (eds.), *Cognition and Categorization*, 27–48. Hillsdale: Lawrence Erlbaum Associates.

Štekauer, Pavol. 1998. *An onomasiological theory of English word-formation*. Amsterdam: John Benjamins.

Štekauer, Pavol. 2005. *Meaning predictability in word formation: Novel, context-free naming units*. Amsterdam: John Benjamins.

Štekauer, Pavol & Chapman, Don & Tomaščíková, Slávka & Franko, Štefan. 2005. Word-formation as creativity within productivity constraints: Sociolinguistic evidence. *Onomasiology Online* 6. 1–55.

Štekauer, Pavol. 2011. Fundamental principles of an onomasiological theory of English word-formation. In Grzega, Joachim (ed.), *A Recollection of 11 Years of Onomasiology Online (2000–2010): All Articles Re-Collected*, 15–54. (http://www1.ku-eichstaett.de/SLF/EngluVglSW/OnOn-Total.pdf) (Accessed 2018-07-28.)

Tagliavini, Carlo. 1982 [1949]. *Di alcuni denominazioni della "pupilla" (Studio di onomasiologia, con speciale riguardo alle lingue camito-semitiche e negro-africane)*. In Tagliviani, Carlo (ed.), *Scritti minori*, 529–568. Bologna: Pàtron.

Tsur, Reuven. 2001. *Onomatopoeia: Cuckoo-language and tick-tocking: The constraints of semiotic systems*. (https://www.tau.ac.il/~tsurxx/Cuckoo_onomatopoeia.html) (Accessed 2018-07-28.)

Ureña Gómez, José Manuel & Faber, Pamela. 2010. Reviewing imagery in resemblance and non-resemblance metaphors. *Cognitive Linguistics* 21(1). 123–149.

Petr Kos
Department of English Studies
Faculty of Arts
University of South Bohemia in České Budějovice
Branišovská 31a, 370 05 České Budějovice
Czech Republic
kos@ff.jcu.cz

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/11.pdf. ISSN 1336-782X.

Word-formation preferences of non-natives

Richard Skultéty Madsen, Section of English Business Communication, Department of Culture and Global Studies, Aalborg University, Denmark

This paper presents the preliminary results of a study of the word-formation preferences in their second/foreign languages of non-native speakers. The study attempts to bring together linguistic typology and second language acquisition. Several pairs of native – non-native languages, for example Danish (NL) – German (NNL), Slovene (NL) – English (NNL), were tested. The informants are university students of the respective foreign languages. It was investigated which word-formation processes (premodification, postmodification, or combinations hereof) are preferred by the informants by measuring which word-formation processes come to the informants' minds first when asked to produce new words from certain keywords. The preferences of word-formation processes were correlated both with the informants' native languages and the target language. The purpose of the correlation was to examine whether the word-formation choices of the informants are influenced more by the typological nature of their mother tongues or by the typological nature of the target languages. The differences between the informant groups could not be attributed to their mother tongues unequivocally. Instead, general intergroup tendencies were found with respect to preferences in word formation.

Keywords: *word formation*, *cross-linguistic influence*, *second language acquisition*, *linguistic typology*, *statistical analysis*

1. Introduction

The purpose of this paper is to investigate whether there are differences between various groups of learners of English as a second language with respect to their preferences of forming words in English as a function of their first languages (L1). This study is thus on the crossroads between linguistic typology and second language acquisition although the main focus is on the production of the acquired second language, not on its acquisition per se. The use of the second language is tested in laboratory in a psycholinguistic experiment, not investigated in corpora. Hence, the informants' preferences are measured in terms of what new word forms they provide first in the experiment, not in terms of how frequently and in what contexts they use certain word forms.

At the time of writing this article, usable data were only available from three groups of informants, Danish, Israeli and Slovene university students, all partaking in the experiment with English as L2. Some data were also collected from informants with Danish and Slovene L1 and German L2. However, the amount of these data is unfortunately miniscule and does not make possible a comparison. It is planned to pursue the collection of further data in the future both with respect to L1's and L2's.

2. Theory and hypotheses

2.1 *Theory*

The main theory of this study is the theory of cross-linguistic influence, i.e. one's mother tongue (L1) – and likely other languages one has acquired – influences the acquisition of a new language (Lado 1957; Johnson 2008; Jarvis 2011). It has also been demonstrated that languages acquired more recently can retroactively influence the languages one has acquired earlier (Pavlenko & Jarvis 2002; Jarvis & Pavlenko 2008; Madsen 2015), but this possibility is not explored in this study.

The main hypothesis is thus a fairly general one: There is a statistically detectable difference between how groups of learners with different mother tongues prefer to make up new words in their target language. To test this hypothesis, informants with typologically different mother tongues, currently Danish, Hebrew and Slovene, have been enlisted to provide data on their use of English as their L2.

A hypothesis based on concrete differences between the informants' mother tongues is that the Danish informants will use compounding more than the other groups. This hypothesis is based on the nature of Danish grammar, or perhaps rather Danish pragmatics (Togeby 2003), preferring compounds to adjectival or prepositional paraphrases. For example, *universitetsprofessor* (university professor) in Danish is *univerzitetni professor* (university-y professor) or *professor na univerzi* (professor at university) in Slovene (Svane 1958; deBray 1980; Melita Koletnik, personal communication).

Apart from the above-mentioned hypotheses based on the theory of cross-linguistic influence, the opportunity was taken to investigate a couple of other hypotheses as well because the data collected for this project are also suitable for testing hypotheses below. One of these hypotheses is that the informants will prefer suffixes to prefixes. This has been posed as a general cross-linguistic tendency and has as such nothing to do with cross-linguistic influence in language acquisition (St. Clair et al. 2009; Ramscar 2013).

Another hypothesis concerns the assumption that the informants, when asked to make new words, will prefer making new lexemes instead of *just* providing different forms of the same lexeme expressed by the keyword. In other words, the informants will prefer using derivation and compounding to using inflection. The justification for this hypothesis is an informal observation that laymen tend to equate words with lexemes, and thus if their task is to make new words, they will tend to provide new lexemes based on the keyword (i.e. use derivation and/or compounding) instead of giving various inflectional forms of the keyword.

Here is an overview of the hypotheses to be tested:

1) There is a statistically significant difference between the informant groups' preferences with respect to word formation.
2) The Danish informants use compounding more than the Israeli and Slovene informants.
3) All the informant groups use postmodification more than premodification.
4) All the informant groups use derivation/compounding more than inflection when producing new "words".

In the rest of this section, major theoretical challenges are discussed briefly.

2.2 *Affix or root?*

The original idea was to measure the differences between the informant groups in terms of their use of derivation by way of affixation and compounding, i.e. the combining of roots. However, it was soon to become apparent that the demarcation of roots from affixes is not at all clear-cut (Štekauer et al. 2012). Especially, deciding the status of a number of morphemes as being prefixes or roots proved to be difficult. These morphemes include among others *out*, *over*, *under*, *geo*, *hyper*, *super*, *uni*.

Looking at the matter superficially, it may be preferred to consider the elements of Germanic origin (*out*, *over*, *under*, etc.) to be roots inasmuch they can appear as orthographic words. In contradistinction, the elements of Latin/Greek origin may be preferred as prefixes since they do not ordinarily constitute orthographic words. Lexicographers do not seem to agree. For instance, American Heritage Dictionary (2018) defines *geo* a prefix whereas Random House Kernerman Webster's College Dictionary (2018) calls it a combining form, i.e. a root. Some of the morphemes above can in fact appear as orthographic words as abbreviations of words in which they would otherwise be prefixes, for instance *hyper(active)*, *homo(sexual)*, *trans(sexual)*. Additionally, Bauer (2017) notes that these morphemes are sometimes written as orthographic words as though they were treated as adjectives by the language users.

Since it is beyond the scope of this paper to try to settle the matter of affix vs root once and for all, a more practical and hopefully less controversial approach has been adopted. Instead of counting the occurrence of prefixes, suffixes and roots, modification has been used as the metric for measuring the differences between the informant groups. Three categories of modification have been distinguished: (1) the keywords of the questionnaire being premodified, (2) the keywords being postmodified, (3) the keywords being used as premodifiers. In the first category fall prefixes (e.g. ***re**volve*) and roots preceding the keyword (e.g. ***arm**chair*). In the second category fall suffixes (e.g. *vision**ary***) and verbal particles (e.g. *get **away***). The third category comprises keywords modifying a following element in a compound (e.g. ***chair**man*).

In this way, the distinction between prefixes and roots becomes irrelevant since both prefixes and roots placed before other roots are premodifiers. Thus, it does not matter whether *over* is considered a prefix or a root because it is a premodifier in say *overworked* regardless of its morphological status. However, it is a different matter for the distinction between suffixes and roots because a suffix postmodifies the element that precedes it, and a root is premodified by the element that precedes it. This was a matter to consider for, among other elements, *log(y)* (stemming from Greek *logos* and frequently appearing in names of branches of science). On etymological grounds, it should be considered a root; however, one might also argue that it is a suffix in modern English on a par with *-ism*. Ultimately, it was decided to count as a root following the reasoning in Merriam-Webster (2018), considering it a combining form. The same applies to the element *phobe*.

A special case is that of the keyword *trans*. When it was used in its canonical prefixal (American Heritage Dictionary 2018) sense (e.g. in *transaction*), it was counted as a premodifier, not as a premodifying keyword. When it was used in its sense of 'transgender' (e.g. *transphobic*) it was counted as a root, and thus as a premodifying keyword.

Finally, there were few responses that could not be analyzed in terms of modifications, either. These are dealt with separately in the analysis.

2.3 *To be or not to be analyzed?*

Testing the hypotheses posited requires a morphological analysis of the responses so that it can be ascertained whether the informants use prefixes, suffixes or roots, or as discussed above premodification or postmodification. In certain cases, however, it is not clear whether a part of a response should indeed be analyzed or not. It is for instance the case when words such as *nature*, *nation* or *aggressive* appear in compounds provided by the informants. These words are clearly complex morphologically; however, it is not given that the informants are or can be expected to be aware of the complexity since these words have been borrowed from Latin and formed according to the morphotactic properties and rules of Latin, which people without classical training are not likely to know. There is no reason to assume that a significant number of the informants know Latin.

A decision as to whether to decompose members of compounds was made on word by word basis. Words such as *nature* and *nation* were not decomposed because although their endings should be recognizable by the informants, their root is not likely familiar since it does not appear in modern English freely, and for laymen it is, therefore, not conspicuous that these words are actually related. On the other hand, *aggressive* and similar words were decomposed because even though its root is not used freely in modern English, it should be recognizable since it appears in a number of other words, e.g. *progressive*, *aggressor*, *regression*, which even laymen with a little fantasy can diagnose as related (having to do with movement).

A similar situation arises when the informants use suffixes such as -*ation*, which are originally a combination of several suffixes (four in the case of -*a*-*t*-*io*-*n*). Should the use of -*ation* and similar complex suffixes count as adding one or several suffixes? Since, as discussed above, it cannot be expected of the informants to know Latin, it was decided that such suffixes count as one.

2.4 *Inflection vs derivation*

As described above, one of the hypotheses concerns the expectation that the informants will use derivation and compounding more than inflection to create new words from the keywords given. However, it has transpired that it is not an easy matter to decide whether a response contains inflectional or derivational elements. Beyond the theoretically challenging question as to whether participles should be considered inflectional or derivational forms, homography causes problems too. For instance, without context, it is impossible to know whether the informants meant *thought* as a noun (derivation) or a past tense form (inflection), not to mention the possible participial reading as well. Unfortunately, the design of the investigation, namely the lack of context in the responses, does not make it possible to solve this issue. Therefore, only forms unambiguously inflectional were counted as such, that is the plural of nouns, 3sg present of verbs and the superlative of adjectives. In the case of keywords ambiguous between an adjective and a verb reading (e.g. *clean*), the -*er* suffix was counted as the comparative, thus inflectional, giving inflection the benefit of the doubt.

2.5 *How not to bias the informants with respect to part of speech?*

There was a challenge with the German version of the questionnaire. It was – just as the English version (see below in §3) – to represent all declinable parts of speech. The English

version intentionally contains keywords which are ambiguous between parts of speech, for instance *work*. The German version was to contain similarly ambiguous words as well. However, as is well known, German standard orthography requires that nouns be capitalized; consequently, potential ambiguity is reduced considerably. The solution was to write all keywords without capitalization and write a note for the informants that it was done on purpose.

3. Method

3.1 *Questionnaire design*

As the method of data collection, this study has used questionnaires administered electronically either through Moodle (for the Danish informants) or through Google Documents (for the Israeli and Slovene informants). The informants were asked to make as many words as they could based on the 22 keywords given. The informants were given no instruction as to what method of word formation to use; they were free to use whatever method they saw fit. However, they were not allowed to use aids such as dictionaries, and there was a time limit of 20 minutes for making new words of all the keywords. These limitations were strictly enforced for the Danish informants as they were under surveillance during the filling in of the questionnaires. However, it is not possible to ascertain whether the other informant groups kept the rules of the game or not.

Naturally, all the informants were given the same set of keywords to build on. However, whereas the Israeli and Slovene informants were given all the keywords at the same time in the same questionnaire, the Danish informants were given the keywords in two mutually exclusive half-subsets. The reason for this was that the Danish informants were given the questionnaire in class in order to secure their proper participation in the project; the time used on the questionnaire was, however, to be limited to 2 x 10 minutes so that the filling in the questionnaire would not be detrimental to the curriculum. On the other hand, the Israeli and Slovene informants filled in the questionnaire outside the classroom, which can explain their relatively low participation.

The keywords given in the questionnaire are the following: ***vision***, *persuade*, ***passive***, *pose*, *human*, ***evolve***, *chair*, *girl*, *world*, *high*, *graph*, *use*, *form*, *think*, *green*, *scribe*, *help*, *work*, *get*, ***trans***, *clean*, *nerve.* Both Moodle and Google Documents were programmed to present the keywords to the informants in a random order (Oppenheim 1992; Dörnyei 2014).

An attempt was made to represent all the inflectional parts of speech (nouns, verbs and adjectives) equally and both roots of Germanic and roots of Latin/Greek origin. Except for the highlighted items, all the keywords are monomorphemic roots. The polymorphemic keywords were included in order to test whether the informants were prepared to separate the morphemes from one another in their formation of new words. *Trans* was included because it can be both a prefix (or premodifier cf. §2.2) and a root (as the abbreviation of *transsexual*), and it was thus tested which use the informants would prefer.

3.2 *Preparation of the responses for analysis*

The preparation of the responses for the analysis was an unexpectedly lengthy process, involving several steps. As it turned out, several informants did not use any conventional

separators (e.g. comma) to separate their responses to the same keyword from each other. Nor were the informants as a group consistent in using UK vs US spelling, not to mention plain misspellings.

Consequently, the responses had to be read through by a human in order to detect whether a sequence of words in the response of the informants to a certain keyword was meant as a compound or separate responses. When compounds were detected, hyphens and spaces between the elements of the compounds were ignored because since the informants did not write texts, it was impossible to know whether their orthographical practice had any significance meaning-wise.

At the same time as detecting compounds, all responses were converted to lowercase, and clearly incorrect responses, e.g. the providing of synonyms of the keyword instead of derivations/compounds based on the keyword, were discarded. Misspelled responses were marked for later reanalysis. All responses were standardized into US spelling in order to ease automated analysis later on.

Additionally, unconventional but in principle possible responses, e.g. *subvision* or *impersuade*, were annotated as such; however, later on in the analysis, they were treated just as ordinary responses, i.e. words which are attested in a dictionary. On the other hand, responses that violate morphotactic rules of English derivation were eliminated from the database, e.g. *worldness* because *-ness* is not ordinarily added to nouns, but to adjectives, and *world* is not attested as an adjective (Huddleston & Pullum 2002).[1]

In the next step, misspelled responses were normalized into their standard orthography if and only if the intended word could be recovered beyond reasonable doubt. For instance, *usefull* was rejected as it could equally be a misspelling of both *useful* and *usefully* whereas say *visionery* was transformed into *visionary* and later analyzed as that word form.

In the final step of raw data preparation, the remaining responses were partitioned into their constituting morphemes. This step was done largely automatically with the use of a custom-made morphological analyzer in Microsoft Excel (Madsen unpublished). Manual intervention was only performed when needed. The morphological analyzer produced output consisting of prefixes, roots and suffixes in accordance with the original idea of using affixes and roots, i.e. derivation and compounding as metrics. This output was then converted into elements of modification in accordance with the description in §2.2 above. Also, responses not containing modifying elements were tagged as such.

Even though many of the keywords can belong to several parts of speech, a response was only counted as conversion if the informant marked it as such unequivocally, e.g. in the case of the keyword *form* as *to form* or *a form*. If the keyword was merely repeated in a response, as was the case with many responses of the Danish informants, it was considered an invalid response.

Apart from dividing the responses into their constituent morphemes, it was noted whether deletion was employed. It is relevant for the polymorphemic keywords, for which

[1] Admittedly, a clear-cut distinction between the accepted and not accepted unconventional responses may be difficult to draw. Since the informants' intention with the responses, i.e. what they thought the responses meant, was unavailable, the decision whether to accept one response and reject another one was based mainly on morphotactic rules presented in grammar books and not in dictionaries and corpus searches. Indeed, one of the reasons for conducting this research was to analyze students' creativity. In fact, it turned out in some cases that unconventional responses judged acceptable had been used by people as for instance proper names: Subvision, a music band.

deletion entails the removal of one of the elements and then adding new elements, for instance when producing *visually* from *vision*.

4. Analysis

4.1 *General comparison*

To start with, Table 1 presents some basic statistics of the data used.

Table 1: Basic statistics of the database

	DNK	**SLO**	**ISR**
number of informants	69	15	15
raw responses in orthographic words	4447	1220	1309
responses including compounds	4326	1197	1248
responses/informant	62.7	79.8	83.2
deletions	27	*11*	*21*
typos	167	41	35
recovered typos	108	28	26
well-formed but unattested words	68	8	10
unrelated words	127	60	10
incorrect derivations	41	11	6
valid responses	3898	1096	1225
valid responses/informant	56.5	73.1	81.7
valid responses/informant/keyword	2.7	3.3	3.7

As can be seen, the Israeli informants were the most productive in making new words. The following two figures show the differences between the informant groups for each of the keywords in terms of both tokens and types in the responses.

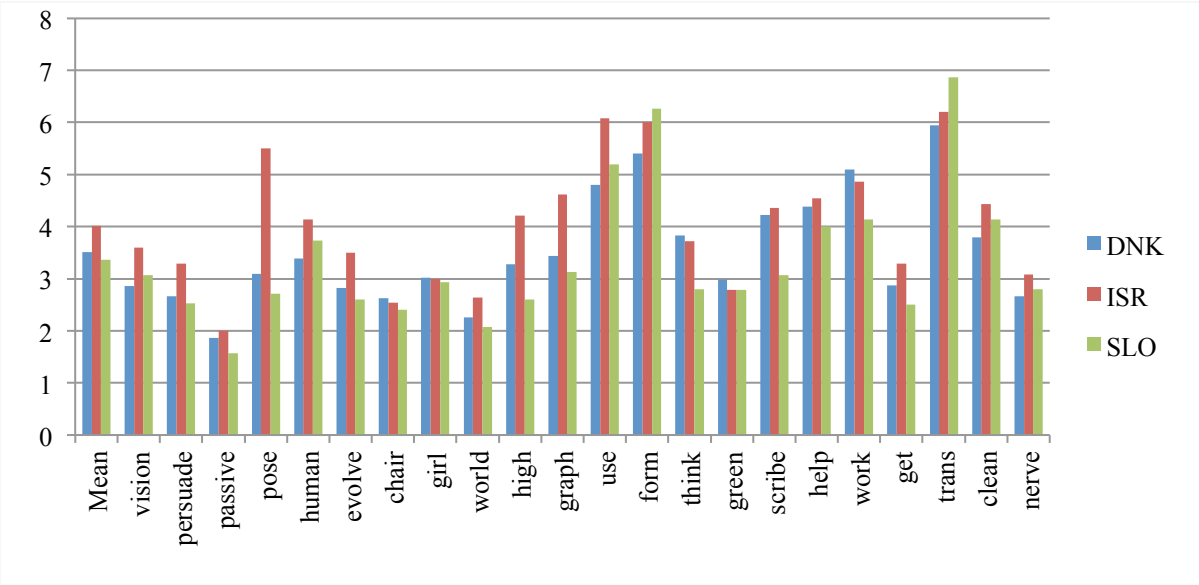


Figure 1: The informants' responses by token

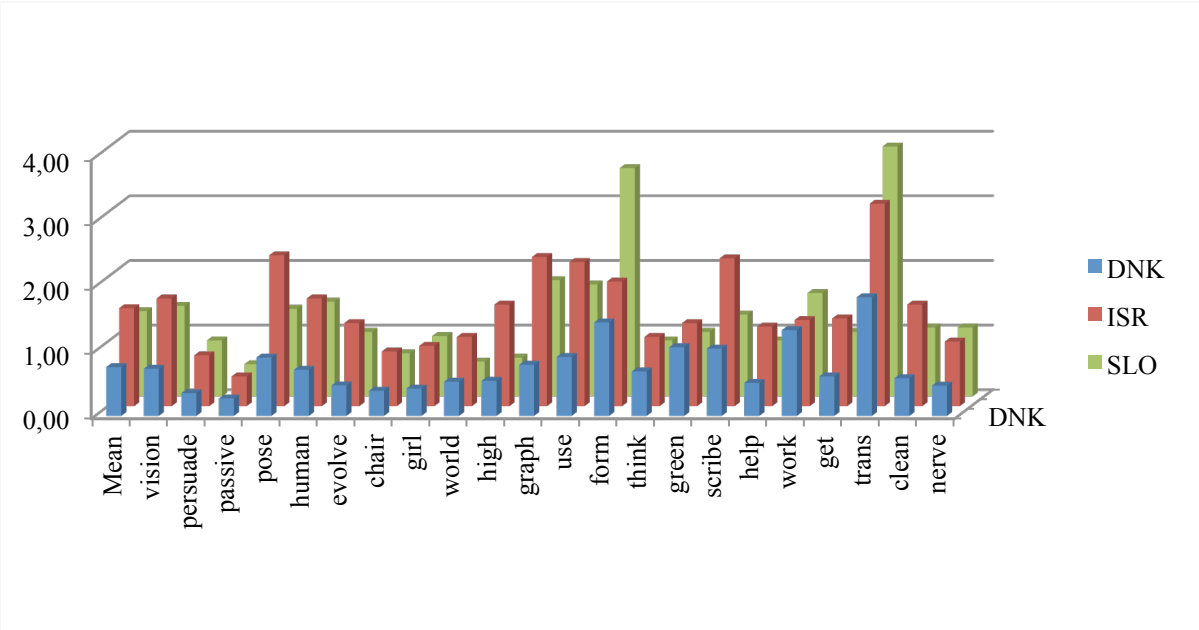


Figure 2: The informants' responses by type

The difference between the informant groups is significant when it comes to type. The Danish informants show significantly lower variety in their responses than the Israeli and Slovene informants. Table 2 shows the differences and their statistical significance as estimated by the two-tailed heteroscedastic t-test (Harmon 2014). Since multiple t-tests are used instead of ANOVA, the significance level of alpha is adjusted down to 0.017 from the usual 0.050 (Abdi 2007). This alpha is used in all the statistical estimations.

Table 2: The differences between the groups in terms of number of responses

	By token			By type		
	Means		p	Means		p
DNK vs ISR	3.52	4.02	0.149	0.75	1.52	3.29E-5
DNK vs SLO	3.52	3.36	0.668	0.75	1.32	8.70E-3
ISR vs SLO	4.02	3.36	0.090	1.52	1.32	0.389

The reason why the mean values by token by keyword in Table 1 and Table 2 differ from each other is that in Table 1, the mean was calculated by the number of all responses per group divided by the number of informants. However, not all informants attempted to make new words from all the given keywords. Hence, the averages in Table 2 have been calculated by only taking those informants into account who actually wrote a response to a given keyword. Thus, the divisor differs from keyword to keyword. The two different ways of calculating the mean for a group of informants are as follows:

In Table 1, the number of all responses (tokens) for all keywords combined is divided by the number of all the informants in the group. In Table 2, the number of responses (tokens and types, respectively) to an individual keyword is first divided by the number of informants who responded to that keyword. Then, these means are summed, and their sum is subsequently divided by the number of keywords. The same method is used in all the calculations below, except that premodifying elements, postmodifying elements and keywords used as premodifiers, respectively, are substituted for responses in the formula.

Despite the differences, which in two cases are statistically significant, it is interesting to note that the distribution of the number of responses per keyword is very similar for all three informant groups as can be seen from the graphs in Figure 1 and Figure 2. The informants seem to have found the same keywords more or less challenging. This is also shown in Table 3 below. Four of the five most inspiring keywords are the same for all three informant groups, and three of the five most tedious keywords are also the same.

Table 3: Most and least inspiring keywords

	Top 5						Bottom 5				
DNK	trans	form	work	use	help	…	persuade	nerve	chair	world	passive
ISR	trans	use	form	pose	work	…	girl	green	world	chair	passive
SLO	trans	form	use	work	clean	…	persuade	get	chair	world	passive

Another rough measure of the difference between the groups is the complexity of their responses. Complexity is measured in terms of the number of morphemes added to the keyword in a response. In this calculation, the removal of a morpheme from a polymorphemic keyword counts as adding a morpheme. Thus, the morpheme complexity of *vis-(u)al-ly* based on *vision* is 3. Inflectional and non-modifying morphemes are included in the calculation. Table 4 tabulates the mean complexity of the responses.

Table 4: Average complexity of the responses in terms of added morphemes

	By token			By type		
	Means		p	Means		p
DNK vs ISR	1.210	1.270	0.289	1.481	1.501	0.790
DNK vs SLO	1.210	1.246	0.483	1.481	1.427	0.420
ISR vs SLO	1.270	1.246	0.730	1.501	1.427	0.318

As can be seen, the informant groups are fairly close to one another, and the differences are not significant statistically. The fact that the per-token means are lower than the corresponding per-type means shows that the majority of responses contained only one added morpheme.

4.2 *Analysis of modification types*

Table 5 summarizes the findings for modification in terms of item per keyword per informant. As in Table 2, only the informants who actually gave a response to a keyword have been taken into account. It must be noted that the modification types do not exclude each other in that the informants might use all three methods in one single response. Moreover, one single response may well contain several instances of the same modification type except – quite naturally – when the keyword is used as a premodifier. None of the keywords used as premodifier was repeated within the same response, and such a repetition would unlikely make sense. Inflectional morphemes are excluded from the calculation. Inflection is dealt with separately in §4.4 below.

Table 5: Average frequency of the modification types

	By token				By type			
	keyword premo-dified	Key-word post-modi-fied	keyword used as premo-difier	non-modifi-cational response	keyword premo-dified	keyword postmo-dified	keyword used as premo-difier	non-modifica-tional response
DNK	0.938	2.385	0.338	0.002	0.364	0.547	0.112	0.002
ISR	1.141	2.951	0.318	0.046	0.696	1.191	0.172	0.046
SLO	1.013	2.596	0.270	0.000	0.646	1.017	0.114	0.000

It transpires clearly that all informant groups prefer postmodifying the keywords to premodifying them by a ratio of about 3:1 by token and 2:1 by type. Using the keywords as premodifiers is the least preferred method, being by up to a magnitude less frequent than postmodification. This is despite the fact that the keyword *trans* lent itself to being used as a premodifier, and it was the keyword eliciting most responses. Were it not for such an eminently premodifying keyword, the difference between the modification types would likely have been even more marked. Non-modificational responses were given in diminishingly few

instances. Yet they merit special attention and are therefore discussed in a special section below partly because they were totally unexpected and partly because virtually only the Israeli informants provided such responses.

Table 6 shows the statistical significance (p-values) between all three types of modification. No statistics were calculated for the non-modificational responses because of their small amount.

Table 6: Statistical significance of intragroup differences between modification types

	By token			**By type**		
	keyword premo-dified vs keyword postmo-dified	keyword postmodified vs keyword used as premodifier	keyword used as premodifier vs keyword premodi-fied	keyword premodi-fied vs keyword postmodi-fied	keyword postmodified vs keyword used as premodifier	keyword used as premodifier vs keyword premodified
DNK	0.001	0.000	0.066	0.115	0.000	0.014
ISR	0.000	0.000	0.014	0.024	0.000	0.004
SLO	0.001	0.000	0.029	0.152	0.000	0.007

As can be seen, the differences are significant – typically by far – in all but five cases, which are highlighted. Table 7 below enumerates the p-values for the differences between the informant groups with respect to the modification types.

Table 7: Statistical significance of intergroup differences within respect to modification types

	By token			**By type**		
	keyword premodified	keyword postmodified	keyword used as premodifier	keyword premodified	keyword postmodified	keyword used as premodifier
DNK vs ISR	0.633	0.155	0.875	0.077	0.000	0.296
DNK vs SLO	0.862	0.603	0.601	0.167	0.022	0.962
ISR vs SLO	0.767	0.430	0.684	0.834	0.449	0.295

There is only one case – highlighted – in which the informant groups perform significantly differently from each other: The Danish informants use postmodification by type significantly less than the other two groups. The Danish informants performed most uniformly of the groups by providing relatively few different types of responses. This probably explains why this difference between the informant groups is significant.

4.3 *Compounding*

Since the status of some morphemes is uncertain (see §2.2 above), the calculations concerning compounding presented here are limited to the keywords used as premodifiers. It is hoped that it is uncontroversial that the keywords are roots; thus, when they are used as premodifiers, the resulting constructions should count as compounds beyond reasonable doubt. See §2.2 for the treatment of the keyword *trans*. Table 8 presents the calculations concerning compounding.

Table 8: Frequency of compounding

	By token			**By type**		
	Means		p	Means		p
DNK vs ISR	0.338	0.318	0.875	0.112	0.172	0.296
DNK vs SLO	0.338	0.270	0.601	0.112	0.114	0.962
ISR vs SLO	0.318	0.270	0.684	0.172	0.114	0.295

As can be seen, there is very little difference between the informant groups. The Danish informants do seem to use compounding slightly more than the other two groups token-wise. However, because of their rather uniform responses, also apparent in other measures (cf. e.g. §4.2), they slightly lag behind the other informants' type-wise.

4.4 *Inflection vs derivation*

Table 9 shows the frequency of inflectional morphemes used by the informants. Only explicit inflectional morphemes (plural of nouns (including vowel change as in *man/men*), 3sg of verbs, comparative of adjectives, and superlative of adjectives) have been counted.

Table 9: Frequency of inflection compared to derivational/compounding modification

	By token				**By type**			
	keyword premodi-fied	keyword postmodi-fied	keyword used as premodi-fier	Inflec-tion	keyword premodi-fied	keyword postmodi-fied	keyword used as premodi-fier	Inflec-tion
DNK	0.938	2.385	0.338	0.687	0.364	0.547	0.112	0.121
ISR	1.141	2.951	0.318	0.642	0.696	1.191	0.172	0.190
SLO	1.013	2.596	0.270	0.416	0.646	1.017	0.114	0.190

As can be seen, inflection dwarfs compared to modification. It is type-wise roughly on the same order of magnitude as keywords used as premodifiers, which is the least frequent way of modification used by the informants. Inflection does *outperform* keywords used as

premodifiers token-wise, which suggests that the informants tended to use inflection in the more frequent responses, especially the Danish informants.

4.5 *Non-modificational responses*

At the design stage of this study, only derivation, compounding (including phrasal verbs) and inflection were expected as processes for making new words or new word forms. However, some responses from the Israeli informants and one response from a Danish informant consist of constructions that are not modificational in nature, i.e. they are not derivations, compounds, inflections, or prototypical phrasal verbs. Most of them (all from Israeli informants) are constructions in which *get* serves as an auxiliary or copula verb, or even as a full verb: *get it*, *get older*, *got married*, *got engaged*, *getting jiggy with it*, *got hurt* and *get on one's nerves*. Three responses are noun phrases: *citizen of the world*, based on the keyword *world*; *fear of heights*, based on the keyword *high;* and *on nerve*, based on the keyword *nerve.* Each of these unusual responses was only provided in one instance, though.

5. Conclusion

Below is an overview of the interpretation of the analysis with respect to the individual hypotheses.

Hypothesis 1: There is a statistically significant difference between the informant groups' preferences with respect to word formation.

It must be concluded that the differences between the groups are not statistically significant. Hence it cannot be ascertained that any of the existing differences is caused by the differing mother tongues of the informants (Jarvis 2000). However, this does not necessarily falsify the theory of cross-linguistic influence. The lack of statistically significant differences may be caused by the small sample size of the Israeli and Slovene informants, and may also be caused by the fact that the informants are semi-professional users of English, and thus any differences between the groups that may have existed at an earlier stage of their acquisition of English may have disappeared by now. However, it is remarkable that virtually only the Israeli informants provided phrases or clause-like structures as responses. This begs for explanation and therefore merits further investigation.

Hypothesis 2: The Danish informants use compounding more than the Israeli and Slovene informants.This hypothesis could not be corroborated. There were no significant differences between the informant groups.

Hypothesis 3: All the informant groups use postmodification more than premodification. This hypothesis has been confirmed. Even though the differences between the modification types do not always reach statistically significant levels, the tendencies are clearly in favor of postmodification.

Hypothesis 4: All the informant groups use derivation/compounding more than inflection when producing new "words".

This hypothesis has been confirmed. Unfortunately, there are as yet only remotely sufficient data on English as L2. Since English does not have much inflection left insofar it has many more derivational morphemes than inflectional morphemes, and since there can be only one inflectional morpheme in a word whereas there can be any number of derivatives or compounding elements, it is not surprising that derivation and compounding outweigh

inflection by a large margin. The study will have to be repeated with an L2 that has a richer inflectional system than English to see if inflection might, nevertheless, play a larger role in laymen's word-formation preferences.

The overall conclusion of this study must be that there is more work to be done, for the theory of cross-linguistic influence could be neither refuted nor corroborated on the basis of the available data. Apart from collecting more data from more informants and informant groups (both further L2's and L1's), the design of the study should also be changed so that ambiguities in the responses can be avoided.

References

Abdi, Hervé. 2007. The Bonferonni and Šidák Corrections for Multiple Comparisons. In Salkind, Neil (ed.), *Encyclopedia of Measurement and Statistics*, 103–107. Thousand Oaks, CA: Sage.

American Heritage Dictionary. 2018. (http://www.thefreedictionary.com/) (Accessed 2018-08-31.)

Bauer, Laurie. 2017. *Compounds and Compounding*. Cambridge: Cambridge University Press.

De Bray, Reginald George Arthur. 1980. *Guide to the South Slavonic languages*. Columbus, Ohio: Slavica Publishers.

Dörnyei, Zoltán. 2014. *Questionnaires in Second Language Research: Construction, Administration, and Processing*, Kindle edition. Abingdon: Taylor and Francis.

Harmon, Mark. 2014. *Practical and Clear Graduate Statistics in Excel: The Excel Statistical Master.* Mark Harmon. Kindle Edition.

Huddleston, Rodney and Pullum, Geoffrey K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Jarvis, Scott. 2000. Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning* 50(2). 245–309.

Jarvis, Scott. 2011. Conceptual transfer: Crosslinguistic effects in categorization and construal. *Bilingualism: Language and Cognition* 14(1). 1–8.

Jarvis, Scott & Pavlenko, Aneta. 2008. *Crosslinguistic influence in language and cognition*. New York: Routledge.

Johnson, Keith. 2008. *An Introduction to Foreign Language Learning and Teaching*. Upper Saddle River, N.J: Pearson. Kindle Edition.

Lado, Robert. 1957. *Linguistics across Cultures*. Ann Arbor: University of Michigan Press.

Madsen, Richard. 2015. L2 Influence on L1 with Respect to Constituent Order in Translations from English into Danish. In Veselovská, Ludmila & Parrott, Jeffrey K. & Janebová, Markéta (eds.). *Proceedings of the CECIL'S 5: 5th Central european Conference in Linguistics for Postgraduate Students*, 144–150. Olomouc: Palacký University.

Madsen, Richard. Unpublished. Morphological analyzer in Microsoft Excel VBA.

Merriam-Webster. 2018. (https://www.merriam-webster.com/dictionary/combining%20form) (Accessed 2018-08-31.)

Oppenheim, A. N. Bram. 1992. *Questionnaire design, interviewing and attitude measurement*. London: Continuum.

Pavlenko, Aneta & Jarvis, Scott. 2002. Bidirectional transfer. *Applied Linguistics* 23. 190–214.

Ramscar, Michael. 2013. Suffixing, prefixing, and the functional order of regularities in meaningful strings. *PSIHOLOGIJA* 46(4). 377–396.

Random House Kernerman Webster's College Dictionary. 2018. (http://www.thefreedictionary.com/) (Accessed 2018-08-31.)

St. Clair, Michelle C. & Monaghan, Padraic & Pamscar, Michael. 2009. Relationships Between Language Structure and Language Learning: The Suffixing Preference and Grammatical Categorization. *Cognitive Science* 33(7). 1317–1329.

Štekauer, Pavol & Valera, Salvador & Körtvélyessy, Lívia. 2012. *Word-Formation in the World's Languages: A Typological Survey*. Cambridge: Cambridge University Press.

Svane, Gunnar Olaf. 1958. *Grammatik der slowenischen Schriftsprache*. Copenhagen: Rosenkilde & Bagger.

Togeby, Ole. 2003. *Fungerer denne sætning? Funktionel dansk sproglære*. [Does this sentence work? A functional grammar of Danish]. Copenhagen: Gads Forlag.

Richard Skultéty Madsen
Section of English Business Communication, Department of Culture and Global Studies
Aalborg University, Denmark
richard@cgs.aau.dk

In SKASE Journal of Theoretical Linguistics [online]. 2019, vol. 16, no. 1 [cit. 2018-16-01]. Available on web page http://www.skase.sk/Volumes/JTL39/pdf_doc/12.pdf. ISSN 1336-782X.