INDIVIDUATING LEXEMES IN LFG

Andrew Spencer
University of Essex

Proceedings of the LFG15 Conference

Miriam Butt and Tracy Holloway King (Editors)

2015

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

Lexicalist models of morphosyntax need to have a way of individuating lexemes. I adopt Spencer's (2013) Generalized Paradigm Function Morphology (GPFM) model of lexical relatedness, in which lexemes/lexical entries are individuated by means of an arbitrary index, distinct from phonology, syntax or semantics. I sketch an LFG interpretation of the GPFM model of argument structure which crucially appeals to semantic function (SF) roles to specify morphosyntactic category. I focus on transpositions, deploying the notion of complex SF role, (which can be used to label corresponding c-structure nodes). I distinguish true transpositions and meaning-bearing transpositions from transpositional lexemes, distinct lexemes which do not involve an added semantic predicate. I illustrate the latter with deverbal adjectives such as English *interesting*. The analysis crucially appeals to GPFM's LI attribute. I show that the LI attribute is virtually coextensive with LFG's PRED attribute. GPFM and LFG are therefore compatible.

1 Introduction

Linguistic theories, especially lexicalist models, need an explicit account of the lexicon, that is, an explicit characterization of what information goes into lexical representations and how lexical entries are related to each other ('lexical relatedness').[*] This has not been a major focus of research in LFG but it is a prerequisite for a study of morphology/word structure and especially for a study of the relation between word structure and phrase/sentence structure. A case in point is the problem of 'mixed categories'. The most widespread of these are the 'transpositions', illustrated by (productive) action nominalizations (*his writing the letter so quickly*) and deverbal participles (*the girl quietly reading a book in the corner*). These typically have the 'external syntax' of a derived morphosyntactic category while retaining a good deal of the syntax (and semantics) of the base category ('internal syntax').

I summarize a model of lexical representation based on Stump's (2001) Paradigm Function Morphology (PFM), and the model of Generalized Paradigm Function Morphology (GPFM) proposed in Spencer (2013), which extends the model of lexical relatedness implicit in PFM to all types of (paradigmatic, regular) relatedness. An important set of questions for the GPFM model centres around the issue of how to distinguish one lexeme (or lexical entry) from another. This is important in lexeme-based models such as PFM, GPFM and, in fact, any model which effectively draws an architectural distinction between inflection (intra-lexemic relatedness) and derivation (inter-lexemic relatedness). However, it is well-known that there is no principled way to distinguish inflection from derivation (Spencer 2013). This means that there is no automatic way of deciding when two word forms are forms of the same lexeme or of different lexemes.

The GPFM model makes considerable appeal to a notion of Lexemic Index (LI), an arbitrary integer serving to individuate lexemes. Other models adopt sig-

[*] I am grateful to Yoko Sugioka, to an external reviewer and the editors for very helpful discussion.

nificantly different ways of individuating lexemes. I argue that the GPFM Lexemic Index essentially fulfills the same function as LFG's PRED attribute, as it is used in the current architecture. This means that it should be relatively more easy to integrate the GPFM model into LFG architecture than models of morphology (such as those discussed below in Section 4) which adopt different strategies for individuating lexical entries.

I make explicit a number of assumptions that remained implicit in Spencer (2013), specifically concerning the typology of lexical relatedness, but also regarding the details of how transpositions are to be analysed in the model. I then consider the notion of 'transpositional lexeme', briefly introduced in Spencer (2013) extending the empirical base which motivates that notion.

2 Lexical representations and lexical relatedness

2.1 Lexical representations

The GPFM model of lexical relatedness (Spencer, 2005, 2013) assumes that a lexeme is a simple four-dimensional vector, ⟨FORM, SYN(TAX), SEM(ANTICS) L(EXEMIC) I(NDEX) ⟩, which can be represented schematically as in (1):

(1) Lexical entry as four-dimensional vector

FORM:	representation of the (inflected) word form
SYN:	grammatical function array argument structure collocational/selectional restrictions other syntactic properties
SEM:	e.g. Lexical-Conceptual Structure
LI:	a unique integer (LI)

Here, I assume that FORM means:

- a specification of the 'root' of the lexeme and any non-predictable stem forms (this effectively subsumes phonological information)
- a specification of any non-predictable inflected word forms
- a pointer to an (inflectional) morphology engine (cf. PFM)

SYN means:

- GF structure: a specification of the array of grammatical functions associated with a predicate's entry
- argument structure: a list of the predicate's (external and internal) arguments (possibly with thematic role labels)
- a specification of other types of selection, for instance, whether a verb takes an infinitival complement, an *ing*-complement, or both

SEM means: Some appropriate lexical semantic representation. In Spencer (2013) I assume some paired-down version of Jackendoff's Lexical Conceptual Structure (Jackendoff, 1990; Levin and Rappaport Hovav, 2005), including Jackendoff's ontological categories Thing, Event, Property. Note: this type of conceptual/lexical representation is to be distinguished from 'denotational', 'model theoretic', 'type theoretic' semantic representation (see Bach and Chao, 2009, 2012 on this distinction). In practice, this means that two SEM representations might be 'equivalent' when the corresponding typed representations are distinct, since the latter change with morphosyntactic context. For instance, a noun such as *cat* in *catfood* is treated here as semantically identical to the *cat* of *a black cat*, even though in the noun-noun compound *cat* is an attributive modifier and so it will presumably be of type ⟨⟨e,t⟩,⟨e,t⟩⟩, not ⟨e,t⟩.

The LEXEMIC INDEX (LI) corresponds in function to what Sag (2012) refers to as the Lexical Identifier (LID). However, in GPFM it is a unique arbitrary integer identifying each distinct lexeme, somewhat in the manner of a key in a database.

A schematic example of a lexical entry is given in (2).

(2) Schematic/simplified entry for 'write'

FORM: /raɪt, roʊt, rɪtn̩/

SYN: ⟨SUBJ, OBJ⟩
⟨Ext-arg, Int-arg, e⟩
Int-arg = linguistic/musical text

SEM: λx,y[$_{\text{Event}}$ **write′**(x,y)]

LEXEMIC INDEX: 59 (or more perspicuously, WRITE)

The precise definition of 'argument structure representation' is not relevant to this discussion, and so I will just represent arguments as ⟨ARG1, ARG2, …⟩. The crucial point is that argument structure includes a set of 'semantic function roles' (SF roles): the 'R' role for nouns (Williams, 1981), the event role, 'E', for verbs (as found in neo-Davidsonian approaches to lexical semantics, and also the approaches such as that of Bierwisch, 1983, Wunderlich, 1997). For adjectives I follow typological tradition (e.g. Croft, 2003, see also Spencer, 1999, 2005) in assuming that the primary function of adjectives as that of attributive modification (of a noun). Therefore, the SF role of an adjective is 'A*', where the '*' is a notational device to indicate that the adjective's SF role is coindexed with 'R', the SF role of the noun it modifies. For instance, a representation such as *Swedish* ⟨A*$_x$⟨x⟩⟩ *man* ⟨R*$_x$⟩ will correspond roughly to the representation λx(**Swedish**(x) ∧ **man**(x)) (see Spencer, 2013, 33f for further discussion and motivation).

The SF roles largely, if not entirely, subsume the functions of c-structure labels, N, V, A. The rationale for adopting this view of SF roles is:

(i) tradition: the 'R' and 'E' roles are already widely deployed in the literature (e.g. Lieber, 2004
(ii) the problem of mixed morpholexical categories (especially transpositions).

Spencer (1999) proposes representing transpositional categories such as action nominals, relational adjectives and deverbal participles with complex or composite SF roles. Thus, an action nominal derived from a verb with a-structure ⟨E⟨…⟩⟩ is the name of an event predicate and has the composite SF role ⟨R⟨E⟨…⟩⟩⟩, while a participle is the attributive modifier representation of that predicate: ⟨A*⟨E⟨…⟩⟩⟩.

We often find that the inflectional paradigm of a lexeme includes periphrases, that is, combinations of a form of the main lexeme together with inflected forms of one or more functional (auxiliary, ancillary) lexemes. Moreover, some lexemes have a basic form which itself consists of more than one lexeme. A clear example of this would be lexicalized but nonetheless transparent compounds, such as *catfood, lightpen, men's room*, … as well as productive noun incorporation in many languages. Other cases include many light verb constructions (*render assistance to*), some verb-particle constructions (*pull apart*), and so on. This means that the paradigm function defining the word forms of a lexeme will sometimes define multiword expressions (better, multilexeme expressions, MLEs), and each of the component lexemes (main verb and auxiliary verb, for instance) will have its own lexical entry. A lexical entry for such a multiword lexeme will therefore itself be defined in terms of a *set* of lexical entries. Now, in GPFM, the basic lexical entry of a lexeme is itself the result of applying the paradigm function to a pairing of Lexemic Index and a null feature set, effectively capturing the idea that the lexical entry is the starting point for the definition of an inflectional paradigm, and the starting point for all the derivational morphology defined with that lexeme as its base. This means that a multiword lexeme will be similar to a periphrastic inflected word form, but one whose parts are the realization solely of the lexical meaning, not of any functional/inflectional features. What this means is that we can conflate the terms 'lexeme' and 'lexical entry', provided we bear in mind that lexemes can be complex and themselves consist of lexemes, and that those MLEs can be the result of productive, paradigmatic lexical processes, so that the complex lexemes themselves are not (necessarily) 'listemes'.[1]

2.2 Lexical relatedness

There is a long tradition of discussing the way that word meanings are related to each other and this is the basis of much lexicography, thesaurus construction and lexicology generally (Cruse, 1986), but words can be related in a variety of ways, depending how we understand the notion 'word' and 'related'.

First, we must distinguish lexemes from (inflected) word forms of a lexeme. For instance, the word forms {*write*, *writes*, *wrote*, *writing*, *written*} are all related

[1] But see Stump, forthcoming, for a dissenting view.

to a lexeme, WRITE, by virtue of realizing the FORM attribute of that lexeme's lexical entry. These forms are also all related to each other in purely formal terms, either by simple affixation (*write, writes, writing*), irregular (unpredictable) ablaut (*write, wrote*), or both (*write, written*). By the same token, the forms *writes, written* or *writing, wrote* are lexically related, albeit indirectly (via the root or stem of the lexeme). FORM-based relatedness cross-cuts syntactic and semantic relatedness and clearly is independent of lexemic identity: thus, even if *re-write* is a form of a distinct lexeme from *write* the two are still related by FORM (as well as by SEMANTICS). Some words are related by FORM without enjoying any other relationship. Thus, the various readings of the homonymous lexemes DRAW (as in *draw a picture* vs. *draw blood, conclusions, breath*) share exactly the same array of inflected forms. Similarly, the verb UNDERWRITE shares all the forms of WRITE together with the meaningless prefix *under-* even though they share nothing else.

At the level of syntax two words may be related in a variety of ways. They may be argument structure variants of each other (*give Fido a bone* ∼ *give a bone to Fido*, they may be semantically related in a way that has syntactic repercussions (*The cakes burned* ∼ *Alfred burned the cakes*), they may be related constructionally by addition of an extra argument (*They baked her a cake, She whistled her way out of the room*. For a given language and for a given pair of alternants it remains an open question whether we should regard the alternants as variants of the same lexeme or not.

The examples just cited also illustrate ways in which words may be semantically related. In canonical derivation we see the SEM representation of some base lexeme enriched by addition of a semantic predicate, as in WRITER, 'person/thing realizing the SUBJECT argument of the verb WRITE'. In some cases the meaning relation is not expressed formally, in which case we may speak of zero derivation or of coercion. Much of the literature on polysemy, including that inspired by the notion of the Generative Lexicon (Pustejovsky, 1995) revolves around the question of how to define relatedness at the semantic level. Where we have what Apresjan (1974) calls 'systematic polysemy' (*reguljarnaja mnogoznačnost'*) we are, presumably, dealing with distinct interpretations for a single lexeme. This contrasts with 'accidental polysemy' as in *branch (of a tree)* vs. *branch (of a bank)*, which presumably is effectively homonymy between two distinct (though formally related) lexemes, but the problem lies then in distinguishing systematic from accidental polysemy.

A preliminary summary of lexical relatedness types is provided in (3).

(3) Two words $£_1$, $£_2$ are lexically related iff
- $FORM(£_1) = \Phi(FORM(£_2))$, where Φ is some morphophonological relation); and/or
- $SYN(£_1) = \Sigma(SYN(£_2))$, where Σ is some operation over syntactic representations; and/or
- $SEM(£_1) = M(SEM(£_2))$, where M is some operation over semantic representations

The Φ operation defines systematic allomorphy (that is, allomorphy that is not suppletive). The Σ operation will include argument structure alternations and possibly other types of systematic syntactic operation. The M operation defines (presumably monotonic increasing, Koontz-Garboden, 2007) alterations which we can think of informally as 'adding a semantic predicate'. Non-trivial application of M will therefore mean that SEM($£_1$) ⊂ SEM($£_2$), where '⊂' means that SEM($£_2$) is an 'extension' of SEM($£_1$). In the most general case, this means that the set of entailments associated with SEM($£_2$) properly includes the set of entailments associated with SEM($£_1$). Two lexical items will also be lexically related if LI($£_1$) = Λ(LI($£_2$)), for some function, Λ, but it will not be necessary to include this in the definition, because any such pair will also be related in terms of FORM and/or SYN and/or SEM representations according to (3). In the special cases where (specific applications of) Φ, Σ, M, Λ instantiate the identity function we have straightforward cases of relatedness.

We may now ask what restrictions there might be on the way that two arbitrary lexical items might be related. Spencer (2013, 139) proposes the maximally simple answer to this question, (4):

(4) *Principle of Representational Independence (PRI)*:
Each of the four principal attributes or dimensions of a lexical representation can define lexical relatedness independently of the other three.

A clear example of the application of the PRI is found with a very common but relatively understudied type of lexical relatedness, the transposition. Here, the morphological and syntactic category is changed but the lexical content (SEM) remains the same. A true transposition can be shown to be a form of the base lexeme, which means that shares its LI with the base. The most studied transposition is the action nominal (Koptjevskaja-Tamm, 1993; Vendler, 1968; Zucchi, 1993; Alexiadou, 2010a,b and for LFG Nikitina, 2008). Spencer (2005) shows that we can find productive instances of all of the six logically possible ways of transposing the major lexical categories, N, V, A, into each other.

Spencer (2013) follows a number of authors in coding grammaticalized, productive transpositions in terms of a category of REPRESENTATION (REPR, cf. the Russian term *representacija*): an action nominal is the nominal representation of a verb lexeme. Despite the change in category it remains a form of a verb and does not change the lexemic status of that verb: *writing* is a form of the verb lexeme WRITE, despite being categorially a noun (in part), or an adjective (in part). It is thus a feature of a transposition that it shares the Lexemic Index of its base. Thus, in a (canonical) transposition, FORM($£_1$) = Φ(FORM($£_2$)), SYN($£_1$) = Σ(SYN($£_2$)) for non-trivial Φ, Σ, but SEM($£_1$) = SYN($£_2$) and, crucially, LI($£_1$) = LI($£_2$).

3 Lexical representations and the Lexemic Index in LFG

The transpositions (which Kuriłowicz, 1966 referred to as 'dérivation syntaxique') raise significant questions for the way that lexical representations interact with syntactic representations. The architecture of LFG is particularly well suited to exploring these interactions because of the way that it factorizes parallel aspects of syntactic structure mirroring the factorization of lexical structure in GPFM (the parallel between LFG and GPFM is not entirely accidental, of course). In Spencer (2013) I sketch a 'proof of concept' of the way that the GPFM representations might be coded using LFG feature structures and this is deployed for the analysis of participles in Spencer (this volume).

The GPFM architecture relies heavily on being able to distinguish forms of a single lexeme from (related) forms of distinct lexemes. We will see later how lexemic individuation can be used to characterize rather subtle distinctions in word behaviour across languages. The LI is a useful formal (or at least, notational) device for recording when we regard two forms as belonging to the same lexeme or not.

In LFG lexemes were not originally identified as linguistic objects: and the main interface between lexicon and syntax was mediated through word forms, not lexemes. The nearest thing to a formal specification of lexemehood was the PRED value, which, however, united a lexeme's name, its grammatical functions (where appropriate) and some (implicit) characterization of its meaning. In contemporary LFG the semantic properties of the lexeme/lexical entry are defined in terms of the semantic projection σ (Andrews, 2008, Asudeh, 2012). In practice, this leaves only the kinds of PRED functions sketched in (5).

(5) Functions of the PRED value:

For nouns	[PRED 'man'], [PRED 'Named: "Harriet"']
For adjectives:	[PRED 'Swedish⟨SUBJ⟩']
For prepositions:	[PRED 'between⟨SUBJ, OBJ⟩']
For verbs:	[PRED 'give⟨SUBJ, OBJ1, OBJ2⟩']

On this view, the PRED value therefore provides (exactly) two sorts of information:

(i) A specification of the grammatical function array associated with the predicate: GF-ARRAY = ⟨SUBJ, OBJ1, OBJ2⟩
(ii) A unique 'name' for the lexeme: 'man', 'Harriet', 'Swedish', 'between', 'give' (cf. PRED FN)

We can therefore split the PRED attribute into two attributes, GF-ARRAY and 'unique name', as shown in (6). The 'unique name' is, of course, the LI. The GF-ARRAY is a specification of the 'construction type' that the lexeme belongs to (intransitive verb, transitive adjective and so on).

(6) Factorization of PRED attribute

PRED ‘give⟨SUBJ, OBJ1, OBJ2⟩’ ⇒
LI: GIVE
GF-ARRAY: ⟨SUBJ, OBJ1, OBJ2⟩

PRED ‘seem⟨XCOMP⟩ SUBJ’ ⇒
LI: SEEM
GF-ARRAY: ⟨XCOMP⟩ SUBJ

Note: A true transpositions does not have a distinct PRED FN value (LI) from the base lexeme, even if its a-structure and GF-ARRAY are different from that of the base.

4 Individuating lexemes: various views

In this section I consider three contrasting proposals for individuating lexemes. In the first, the lexeme is identified in terms solely of its root. In the second, the lexeme’s meaning is the principal marker. I argue that both these approaches are flawed. The third alternative is conceptually the simplest: the LI is an arbitrary label (for instance, a unique integer) with no other properties, as proposed in Stump (2002) and particularly in Spencer (2013).

4.1 The simplest set of assumptions

In the first model, all inflection and derivation is realized by separate ‘functional heads’ (‘morphemes’) each with its own lexical entry. All morphology is compounding, and compounding is an operation in syntax (‘merge’). This is the essence of standard Distributed Morphology (DM) (Marantz, 1997).

In terms of the GPFM model sketched above, we can say that, in DM, each distinct FORM representation corresponds uniquely to a single SYN, SEM representation.[2] An entailment of this model is that no lexical items whatsoever, in any language whatsoever, can show (root) suppletion (see Borer, 2013, 398-99 for a very explicit statement). This appears to be untrue (Corbett, 2007, Harley, 2014, Spencer, to appear).

It is very difficult to see how such a model could account for the more articulated aspects of lexical structure discussed here without smuggling into the model some equivalent to the LI, so I will ignore such models from now on.

[2] In fact, it is generally assumed that this is true only for lexical roots, not for functional heads. However, it is very unclear how that lexical∼functional distinction is supposed to be drawn in this model.

4.2 Second possibility: take SEM value/representation as its index

Sag (2012) deploys an attribute 'Lexical Identifier' (LID) in his model of lexical structure. The LID is an attribute of the HEAD|SYN attribute. It is, however, tagged with the semantic representation in such a way that two lexical entries with the same meaning (including synonyms) will have the same LID. This in itself is only a problem if we believe that there are true synonyms. However, it does mean that the LID has to be tied to rather subtle aspects of CONTENT and usage (including collocations) if it is to be a coherent notion. There are, however, other problems with Sag's conception of the LID.[3]

The first problem is that there is morphology which changes a lexeme's meaning but without changing its LID. There are regular argument structure alternations such as productive causative forms of verbs, and below I discuss the case of meaning-bearing noun-to-adjective transpositions in Selkup. There is also a rather subtle technical problem with Sag's LID: the SBCG model inherits from standard HPSG a distinction between linguistic objects and their descriptions. For Sag a lexical entry is a description of an object, a lexeme, which is a member of the types of linguistic expression. However, Sag encounters difficulties in providing an illustration of a lexeme. This is because HPSG is essentially defined in terms of word forms. Sag therefore has to take an arbitrarily chosen inflected form of a lexeme in order to illustrate it as a linguistic object (the preterite form *laughed* to illustrate the lexeme LAUGH, Sag 2012, 99–101). But that means that Sag is talking about word forms, not the lexemes of which those words are forms. But it is now not possible to see how the LID is individuating lexemes as opposed to word forms of lexemes. I can see no way around this conceptual difficulty.

Bonami's (2015) analysis of periphrasis deploys the LID to express the idea that an auxiliary verb is part of the realization of an inflectional property in a (true) periphrasis. For instance, in the Latin perfect passive periphrasis *amatus est* '(he) is/has been loved' we would say that SUM ('be') had INFL attribute with LID SUM, but also has its HEAD|LID feature tagged with the HEAD|LID of AMO ('love'). This expresses the intuition that the auxiliary lexeme LID SUM is used to realize part of the inflectional paradigm of a lexical verb such as the lexeme 'love', LID AMO. However, as Bonami points out, this device will not work if the LID is tied to the SEM representation, because, in LFG terms, it would give rise to a clash of (old-style) PRED values.

4.3 LI = arbitrary stipulated label (e.g. integer)

To avoid potential confusion I shall from now on refer to the arbitrary indexing marker for a lexeme as the Lexemic Index (LI), following Spencer (2013). It is the LI that serves to record particular decisions about how lexemes are individuated. The LI on this account is an attribute of a lexeme, distinct from the FORM, SYN,

[3] Expletive elements such as *it, there* presumably require LIs (as a reviewer reminds me) but it is difficult to see how that can be achieved if the LI is tied to a semantic representation.

SEM attributes.[4]

5 Transpositions and transpositional lexemes

Assuming the feature system sketched in Spencer (2013), in which REPR defines (canonical) transpositions, along with the notational conventions used there, a participle is defined as the result of applying the generalized paradigm function to the pairing $\langle \mathcal{V}, \rho \rangle$, where $\mathcal{V}$ is the LI of a verb and ρ = {[REPR[TRANS2A[V2A]]], …}. This function defines a new set of word forms and crucially redefines the SF role from $\langle E\langle x,\ldots\rangle\rangle$ to $\langle A^*_x,\langle E\langle x,\ldots\rangle\rangle\rangle$. Being an adjective the participle now has a set of inflected forms which are entirely absent in finite verbs forms, namely, the agreements in features such as gender, number, case (and definiteness, in the case of Baltic languages; see Spencer, this volume). This means that the inflectional morphology must treat them as though they were distinct lexical entries (even though they remain in an important sense 'forms of the base verb lexeme'). Spencer (2013) argues that the morpholexical categories which are defined by the SF roles are associated by default with a morpholexical signature. This specifies precisely which properties a word is inflected for. Since all the transposed participial forms are now categorially adjectives (as well as, to some extent, still verbs) they are subject to that default specification. It is a matter of language-particular stipulation precisely how many of the verbal properties are retained and in what form.

Adjectives in a language such as Latin, Sanskrit or Lithuanian are subject to a default specification under which their SYN attribute bears an agreement attribute, AGR (Haug and Nikitina, 2012, Spencer, this volume). This default specification will now apply to the participle's SYN attribute, meaning that appropriate agreement relations can be defined in the syntax. By a further default specification, the syntactic agreement AGR attribute is copied in the FORM attribute, as a morphological AGR attribute. This ensures that the inflectional morphology will provide the actual inflected forms required by the syntactic representations. In the case of languages such as Sanskrit and Lithuanian the FORM|AGR attribute will be further specified with information specifying which inflectional and accentual class the participle belongs to (just as in the case of ordinary adjectives).

The discussion of participles in Spencer (2013) is somewhat sketchy and so I will here flesh out some of the details of how such transpositions are to be handled (the general principles will apply to all six types of true transposition). Specifically, I clarify an important detail concerning the way that inflected forms of transpositions are defined.

In Stump's (2001) PFM model the paradigm function for a lexeme defines a single set of inflectional properties for that class of lexemes. In Spencer (2013) I

[4] In GPFM, the functions which define the FORM, SYN and SEM values of a lexeme or word form are defined in terms of the LI. Thus, if we assume a LI CAT, the form /kat/, the SYN value [REF[ARG1[ ]]], and the SEM value λx.**cat**(x) are related as values of FORM$\langle$CAT,{Ø}$\rangle$, SYN$\langle$CAT,{Ø}$\rangle$, SEM$\langle$CAT,{Ø}$\rangle$ (where '{Ø}' is the null feature set). Details can be found in Spencer (2013), but they are not directly relevant to the point being made in this paper.

appeal to a morpholexical signature, a declaration of precisely those properties for which a given class inflects, notated as value, MORSIG, of the FORM attribute in the lexical entry. By the *Principle of Inflectional Specifiability* (Spencer 2013, 199) a lexeme has to be furnished with such a signature before it can be inflected. But this means that transposition will in general bear the wrong morpholexical signature. For instance, a deverbal participle will bear the signature of a verb, not that of an adjective. If we treated the participle as though it were purely derivational we could simply overwrite the verb's morpholexical signature and replaced it with that of an adjective. However, then we would then lose the fact that the participle retains crucial verb properties. In this respect the participle is a prototypical instance of a paradigmatically mixed category and this has to be reflected in its morpholexical signature.

Unfortunately, very little of cross-linguistic validity can be said about paradigmatic categorial mixing. Malchukov (2004, 2006) has very interesting observations on inflectional property mixing in action nominals. He asks to what extent a verb loses its verbal properties and to what extent it gains nominal properties. We can ask the same question about participles (cf. also Lehmann, 1984) but that work has yet to be done. Although there are strong tendencies apparent, there is, it would appear, no way to predict precisely which verb properties will be lost and which adjective properties will be acquited. The precise content of the MORSIG therefore has to be stipulated on a language-by-language basis.

There are two aspects to defining inflections: (i) what properties does the lexeme inflect for? and (i) what are the actual inflected word forms? The word forms will depend on the inflectional and morphological class the word belongs to. In Lithuanian there is a clearly defined morphological class of Adjective (with various declensional and accentual subclasses). The MORSIG attribute therefore includes an attribute MORCLASS which defines the word's [MCLASS: Adjective] and its inflectional (declensional) and accentual) class, INFLCLASS, ACCENTCLASS. At the same time the MORSIG bears an attribute INFL which specifies the agreement properties of an adjective. This is shown in Figure 1, for the adjective *báltas* 'white'.

A verb will have a MORSIG attribute which specifies a MORCLASS|MCAT value Verb, and INFL properties of subject agreement and tense-aspect-mood-voice inflection, as seen in Figure 2, for the verb *dìrbti* 'work' (with the most common feature values included for illustration).

The representation of the present active participle is shown in Figure 3 from the verb *dìrbti*.

The morphology of Lithuanian defines a participial stem form (or forms) for each class of verbs and these stems are then inflected like adjectives. Thus, to define the present active participle (PPA) of a verb $\mathcal{V}$ we need the Generalized Paradigm Function (GPF) shown in (7). The f_{form} component of the GPF defines the FORM template representation shown in Figure 3. The f_{syn} component defines the compound A-STR representation. The $f_{sem,\ li}$ components induce no change, of course, and are omitted.

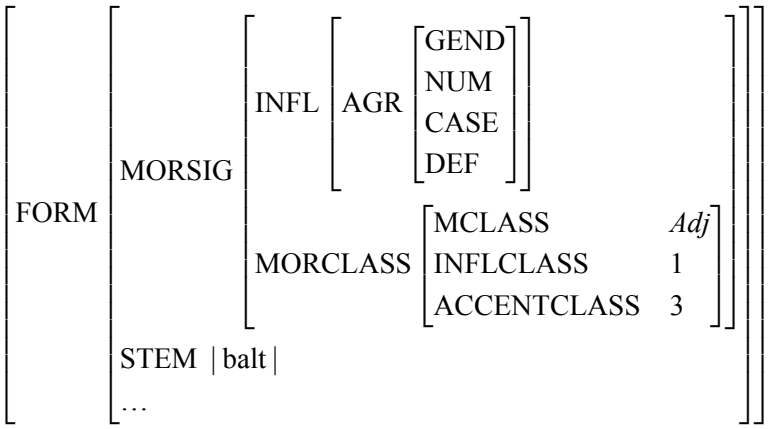

Figure 1: FORM attribute from lexical entry for Lithuanian adjective *báltas* 'white'

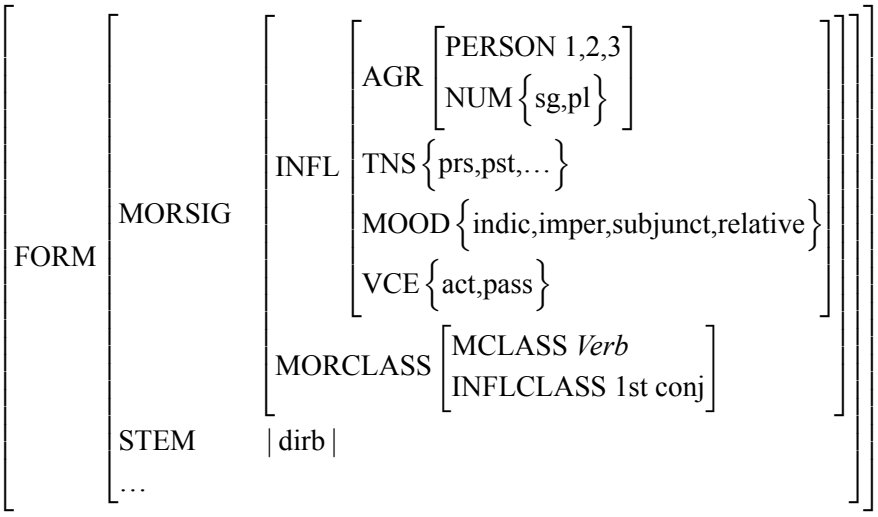

Figure 2: FORM attribute from lexical entry of Lithuanian verb *dìrbti* 'work'

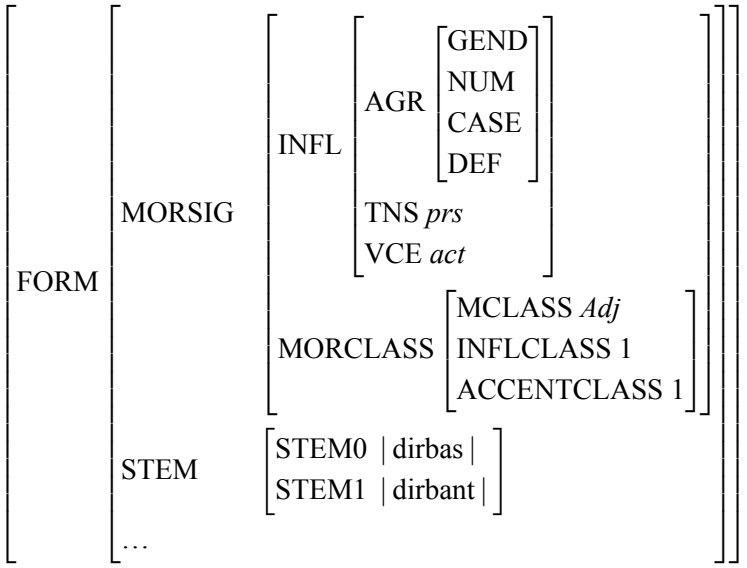

Figure 3: FORM attribute for present active participle form of Lithuanian *dìrbti*, *dirbq̃s* 'working'

(7) $\text{gpf}(\langle \mathcal{V}, \{[\text{REPR:V2A}], \text{TNS:PRS}, \text{VCE:ACT}\}\rangle) =$
$f_{form}(\langle \mathcal{V}, \{[\text{REPR:V2A}], \text{TNS:PRS}, \text{VCE:ACT}\}\rangle) \wedge$
$f_{syn}(\langle \mathcal{V}, \{[\text{REPR:V2A}], \text{TNS:PRS}, \text{VCE:ACT}\}\rangle)$

The f_{syn} function essentially adds the A-MOD semantic function role to the EV role in the verb's argument structure. The operation of the f_{form} function is more complicated. First, the function (re)defines the MORSIG attribute of the participle. Some of the verb features are copied to the participle's MORSIG (voice, tense) others are overwritten (all the MORCLASS and AGR features). The MOOD attribute is absent from the participle's MORSIG. The representation in Figure 3 shows the participle's stem set, not the fully realized paradigm. The STEM attribute houses the Stem Formation information needed to define the final forms of the participle, as described in detail by Stump (2001) for the very similar system of Sanskrit participles.

True transpositions have no effect on the SEM or LI attributes of the base lexeme. This is particularly clear in the case of the pure relational adjectives of the Samoyedic language Selkup. These are noun-to-adjective (N2A) transpositions (Spencer, 2013, Chapter 10). In Selkup relational adjectives are derived from nouns marked for possessor agreement inflection, that is, clearly inflected forms of the noun, yet they still change the word's morpholexical class from noun to adjective. Moreover, the base noun is still 'visible' to the syntax, in that it can still be modified by an attributive modifier (adjective), even though it bears the relational adjective suffix and is morphosyntactically now an adjective. This is shown schematically in (8), where (8a) is the morphological structure, while (8b) is the structure implied

by the attribute-modifier syntax of the construction.

(8) a. *old* *canoe*-RELADJ *oar*
b. [[*old* *canoe*]-RELADJ] *oar*
'oar from the old canoe'

Examples such as (8) clearly show that the relational adjective form is a 'mixed category', behaving like a noun with respect to inflection and modification by attribute, but itself functioning as an adjectival attributive modifier.

Selkup has two additional types of N2A transposition, a locative relational adjective ('situated in/at my canoe'), and a similitudinal relational adjective ('similar to my canoe'). These are also derived from possessor marked nouns, so they are also true transpositions, but they involve addition of a semantic predicate, hence, they are not canonical transpositions (canonically, it is only derivational morphology which alters the SEM attribute of an entry). The existence of such transpositions demonstrates (again) that lexemes can't be individuated solely in terms of semantic differentiation: the similitudinal/locative relational adjectives are semantically enriched compared to the base noun lexeme but they are still forms (adjectival representations) of that noun lexeme and not distinct lexemes.

Relational adjectives can also be found in languages such as English, Russian, and many others. However, unlike the relational adjectives of Selkup, these largely lack the original noun properties either inflectional or syntactically. In fact, they behave like distinct lexemes and respect Lexical Integrity. This means that their base noun is not visible to syntactic modification, for instance. Thus, the relational adjective *prepositional*, does not allow its base noun, *preposition*, to be modified by an adjective such as *spatial* (9a), though the noun *preposition* can be modified by *spatial* in the compound noun *preposition phrase*, (9b).

(9) a. *preposition* ⇒ *prepositional* ⇒ *prepositional phrase*
b. *preposition* ⇒ *spatial preposition*
c. *[[*spatial preposition*]*al phrase*
= 'spatial phrase headed by a preposition'
≠ 'phrase headed by a spatial preposition'
d. *spatial preposition* ⇒ *spatial preposition phrase*]
= 'spatial phrase headed by preposition' *spatial* [*preposition phrase*]

Example (9c) can only mean 'spatial phrase headed by a preposition'.

What this suggests is that the English relational adjective *prepositional* and all those like it are now lexemes in their own right, not transposed forms of nouns (after all). And yet they are semantically identical to their base nouns. For this reason the expression *prepositional phrase* is entirely synonymous with the noun-noun compound preposition phrase. Spencer (2103) labels this type of derived lexeme a

'transpositional lexeme'.[5]

The existence of transpositional lexemes is unexpected given standard assumptions about inflection and derivation, because derivational morphology is supposed to create new lexemes with enrichment of the semantic representation, yet the transpositional lexeme has the same SEM respresentation as its base. However, transpositional lexemes are exprected if we accept the Principle of Representational Independence, and, indeed, we would have to invoke some additional principle to exclude them.

It is reasonable to ask whether other types of transpositional relation give rise to transpositional lexemes. Arguably, there is a host of such examples in English. Obvious examples are provided by action nominals. If *shooting* in *Lee Harvey Oswald's shooting Kennedy (was appalling)* is a V2N transposition, then *shooting(s)* in *There has been a shooting/have been a lot of shootings (in Dallas)* is presumably a distinct lexeme, but one whose meaning, 'act(s) of shooting', is non-distinct from that of the verb.

Spencer (2013) discusses the property nominalization (A2N transposition) found in many languages. While in some languages such nominalizations may be true transpositions,[6] showing the same apparent violations of Lexical Integrity found with Selkup relational adjectives, in a language like English we appear to be dealing with a transpositional lexeme. Again, the lexical semantics of a nominal such as *happiness, popularity, ...* may be the same as that of the base adjective but that adjective can no longer be modified (*very popular* vs **very popularity*), nor can inflected forms be nominalized (*happiest* vs **happiestness, *most happiness*).

English has (true) adjectives derived from participles, both active: *interesting, boring, charming, flattering, challenging, ...*, and passive: *broken, disappointed, wounded, enlightened, interested, bored, flattered, challenged, ...* These adjectives sometimes acquire entirely idiosyncratic meanings, but where the basic verb denotes a psychological predicate their basic meaning can be directly related to that of the verb they are (etymologically) derived from, as indicated in (10).

(10) a. *This book is very interesting/This is a very interesting book* $\approxeq$ *this book interests me a great deal*

b. *Her reaction was in some ways surprising/That was an in some ways surprising reaction* $\approxeq$ *Her reaction surprised us in some ways*

The example illustrated in (10a) raises the question of the role of a degree modifier such as *very* in such constructions. At first sight, *interesting* appears to be an ordinary gradable, like *tall*, because it is modified by *very*. Degree modifiers such as *very* cannot modify genuine participles: **the book very interesting me (is about*

[5]Of course, we find considerable variation here. Very often the relational adjective construction will have distinct nuances from compounds or prepositional phrase structures: *mathematics course* $\neq$ *course of mathematics* $\neq$ *mathematical course*. In other cases, one or other expression will be simply unavailable: *religious history* $\sim$ *??religion history, telephone system* $\sim$ *??telephonic system*. This variation is in stark contrast to the productivity of genuine N2A transpositions.

[6]The Japanese nominalizer *-sa* may be a case in point (Sugioka, 2011).

morphology). On the other hand, participial adjectives such as *interesting* differ from ordinary property adjectives because they can be modified by words which modify the verb element of the participle: *a physically/mentally challenging task* = a task which challenges someone physically/mentally, *a partially broken chair* = a chair which has been broken but not completely, and soon. Even with ordinary gradable adjectives we find degree modifiers which do not modify a straightforward degree component but which entail a more complex kind of semantic modification: *a deceptively tall building, a suspiciously enthusiastic response*. Given this we can conclude that *very* in (10a) modifies the degree to which an entity interests another entity.

If this reasoning is correct then most of the participial lexicalized adjectives turn out to be transpositional in the sense that they inherit their basic meaning from (one) meaning of the base verb and do not add a semantic predicate. However, they are distinct (adjectival) lexemes so they do not inherit the full array of verbal properties from their base, and in particular cannot be modified in the manner of verbs (though they can take typical adjectival modifiers, with atypical meanings).[7]

6 Conclusions

I have argued that the model of lexical relatedness presented in Spencer (2013), Generalized Paradigm Function Morphology, is readily compatible with the LFG architecture. A central feature of that model is the deployment of a Lexemic Index (LI), which serves to individuate lexical entries/lexemes. That LI turns out to have essentially the same function as the naming component of LFG's PRED attribute.

One of the reasons why it has been so difficult for morphologists to distinguish inflection from derivation is the existence of transpositions such as deverbal participles, which appear to be forms of the base lexeme and fail to alter the semantic content of the base, just like canonical inflection, but which do change the morphological and syntactic category of that base, just like derivation. In the GPFM model relatedness is factorized so that it is not actually important to decide whether a type of relatedness is strictly inflectional, derivational, or some intermediate type such as a transposition. I also argue that the LI has to be an arbitrary label, such as a unique integer, rather than being identified with phonological or semantic representations. One of the reasons why the LI can't be identified with the lexeme's semantics is that there are transpositions such as the relational adjectives of Selkup, two of which do change the semantic representation, adding similitudinal and locational meanings,

[7] Spencer (forthcoming) proposes that the Principle of Representational Independence be restricted to just those cases in which the LI is changed. The proposal there is that where two words are related but have distinct LIs then they must also be distinct along some dimension of 'interpretation', i.e. either their SEM value has to be distinct or, failing that, their a-structure. This would then rule out a situation in which a language with very rich inflection has lexemes with thousands of different forms, each one of which could in principle be given its own LI. However, the revised statement of the PRI in Spencer (forthcoming) would also rule out pure synonymy, which may or may not be an advantage. I leave this to future research.

but which can be formed from nouns inflected for possessor agreement and which are thus themselves effectively inflected forms of the base noun lexeme, and not the product of derivational morphology.

On the other hand, the LI also allows us to distinguish the truly transpositional type of adjective from superficially similar cases such as English participial adjectives of the type *interesting*. These are like true (participial) transpositions in that they do not add a predicate to the SEM representation of the base verb, and yet they can be shown to behave like distinct, derived lexemes. They are thus instances of a 'transpositional lexeme', adding to the types identified in Spencer (2013). The LI, now PRED, attribute plays a crucial role in distinguishing this type from true transpositions, thus providing further motivation for retaining that attribute in the LFG architecture.

References

Alexiadou, Artemis. 2010a. Nominalizations: a probe into the architecture of grammar. Part I: the nominalization puzzle. *Language and Linguistics Compass* 4, 496–511.

Alexiadou, Artemis. 2010b. Nominalizations: a probe into the architecture of grammar. Part II: the aspectual properties of nominalizations, and the lexicon vs. syntax debate. *Language and Linguistics Compass* 4, 512–523.

Andrews, Avery. 2008. The role of PRED in LFG+glue. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG08 Conference*, pages 46–67, Stanford, CA: CSLI Publications.

Apresjan, Jurij Derenikovič. 1974. *Leksičeskaja semantika. Sinonimičeskie sredstva jazyka*. Moscow: Nauka.

Asudeh, Ash. 2012. *The Logic of Pronominal Resumption*. Oxford: Oxford University Press.

Bach, Emmon and Chao, Wynn. 2009. On Semantic Universals and Typology. In Morten H. Christiansen, Chris Collins and Shimon Edelman (eds.), *Language Universals*, pages 152–173, Oxford: Oxford University Press.

Bach, Emmon and Chao, Wynn. 2012. Semantic types across languages. In Claudia Maienborn, Klaus von Heusinger and Paul Portner (eds.), *Semantics. An International Handbook of Natural Language Meaning*, volume X, pages 2537–2558, Berlin: De Gruyter Mouton.

Bierwisch, Manfred. 1983. Semantische und konzeptuelle Repräsentationen lexikalischer Einheiten. In Rudolf Ružičká and Wolfgang Motsch (eds.), *Untersuchungen zur Semantik, Studia Grammatika 22*, pages 61–99, Berlin: Akademie Verlag.

Bonami, Olivier. 2015. Periphrasis as collocation. *Morphology* 25, 63–110.

Borer, Hagit. 2013. *Structuring Sense: Taking Form*. Oxford: Oxford University Press.

Corbett, Greville G. 2007. Canonical Typology, Suppletion, and Possible Words. *Language* 83(1), 8–42.

Croft, William. 2003. *Typology and Universals*. Cambridge: Cambridge University Press, second edition.

Cruse, Alan D. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.

Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40, 225–275.

Haug, Dan and Nikitina, Tanya. 2012. The many cases of non-finite subjects: the challenge of "dominant" participles. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG12 Conference*, pages 292–311, Stanford, CA: CSLI Publications.

Jackendoff, Ray S. 1990. *Semantic Structures*. Cambridge, MA: The MIT Press.

Koontz-Garboden, Andrew. 2007. *States, changes of state, and the Monotonicity Hypothesis*. Ph.D. thesis, Stanford University Department of Linguistics, Stanford, CA.

Koptjevskaja-Tamm, Maria. 1993. *Nominalizations*. London: Routledge.

Kuriłowicz, Jerzy. 1966. Dérivation lexical et dérivation syntaxique (contribution à la théorie des parties du discours). In Eric Hamp, Fred Householder and Robert Austerlitz (eds.), *Readings in Linguistics volume 2*, pages 42–50, Chicago: University of Chicago Press.

Lehmann, Christian. 1984. *Der Relativsatz: Typologie seiner Strukturen, Theorie seiner Funktionen, Kompendium seiner Grammatik*. Tübingen: Gunter Narr Verlag.

Levin, Beth and Rappaport Hovav, Malka. 2005. *Argument Realization*. Cambridge: Cambridge University Press.

Lieber, Rochelle. 2004. *Morphology and lexical semantics*. Cambridge: Cambridge University Press.

Malchukov, Andrej L. 2004. *Nominalization/verbalizations: constraining a typology of transcategorial operations*. Munich: LINCOM EUROPA.

Malchukov, Andrej L. 2006. Constraining nominalization: function-form competition. *Linguistics* 44, 973–1008.

Marantz, Alec. 1997. No escape from syntax: don't try morphological analysis in the privacy of your own lexicon. In A. Dimitriadis, L. Siegel, C. Surek-Clark and A. Williams (eds.), *University of Pennsylvania Working Papers in Linguistics*, volume 4, pages 201–225, Department of Linguistics, University of Pennsylvania.

Nikitina, Tatiana. 2008. *The Mixing of Syntactic Properties and Language Change*. Ph. D.thesis, Stanford University, Stanford, CA.

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge, MA: The MIT Press.

Sag, Ivan A. 2012. Sign-Based Construction Grammar: an informal synopsis. In Hans C. Boas and Ivan A. Sag (eds.), *Sign-Based Construction Grammar*, pages 69–202, Stanford, CA: CSLI Publications.

Spencer, Andrew. 1999. Transpositions and argument structure. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1998*, pages 73–102, Dordrecht: Kluwer Academic Publishers.

Spencer, Andrew. 2005. Towards a typology of 'mixed categories'. In C. Orhan Orgun and Peter Sells (eds.), *Morphology and the Web of Grammar. Essays in Memory of Steven G. Lapointe.*, pages 95–138, Stanford, CA: CSLI Publications.

Spencer, Andrew. 2013. *Lexical Relatedness: A Paradigm-based Model*. Oxford: Oxford University Press.

Spencer, Andrew. This volume. Participial relatives in LFG.

Spencer, Andrew. Forthcoming. How to individuate lexemes. In Daniel Siddiqi and Heidi Harley (eds.), *Morphological Metatheory*, Amsterdam: John Benjamins Publishing Co.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2002. Morphological and syntactic paradigms: arguments for a theory of paradigm linkage. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 2001*, pages 147–180, Dordrecht: Kluwer Academic.

Stump, Gregory T. Forthcoming. *Inflectional Paradigms*. Cambridge: Cambridge University Press.

Sugioka, Yoko. 2011. Nominalization affixes and multi-modularity of word formation. In Etsuyo Yuasa, Tista Bagchi and Katharine Beals (eds.), *Pragmatics and Autolexical Grammar. In Honor of Jerry Sadock*, pages 143–162, John Benjamins Publishing Co.

Vendler, Zeno. 1968. *Adjectives and Nominalizations*. The Hague: Mouton de Gruyter.

Williams, Edwin. 1981. Argument structure and morphology. *The Linguistic Review* 12, 18–114.

Wunderlich, Dieter. 1997. Cause and the structure of verbs. *Linguistic Inquiry* 28, 27–68.

Zucchi, Alessandro. 1993. *The Language of Propositions and Events*. Dordrecht: Kluwer Academic Publishers.