Chapter 12

On lexical entries and lexical representations

Andrew Spencer
University of Essex

Lexicalist models of syntax share with lexeme-and-paradigm models of morphology the assumption that the primary unit of the lexicon is the lexeme, an abstract representation of properties unifying a set of inflected word forms. Lexicalist syntactic models (such as Head-driven Phrase Structure Grammar, henceforth HPSG, and Sign-Based Construction Grammar, henceforth SBCG) distinguish modelled linguistic objects from descriptions of objects. A description, but not an object, can be a partial (underspecified) representation. However, a lexeme is by definition only partially specified, being underspecified for all those morphosyntactic properties that its word forms realize (the lexeme DOG realizes neither singular nor plural, unlike the word forms *dog, dogs*). This implies that lexemes are descriptions, not objects, which is incompatible with assumptions about the type hierarchy for signs and the lexicon in HPSG/SBCG. If we relax the definition of full specification to admit lexemes as objects then the question arises as to how many properties can be left unspecified. I argue for a maximally underspecified model. Even the declaration of properties for which the given class of lexemes inflects (the 'morpholexical signature', MORSIG) is underspecified to the extent that its contents are predictable. This entails that an inflected word form of a lexeme can be defined only after the MORSIG attribute is specified. Derivation, a lexeme-to-lexeme mapping, can therefore be defined over the same maximally underspecified lexical representations, whose inflection is then typically governed by a different morpholexical signature (e.g. when the derivation changes word class). All such specifications are given by default statements, which are overridden for irregular items. Verb-to-adjective transpositions (participles) are members of the verb's paradigm yet inflect according to the adjectival paradigm (the 'adjectival representation' of a verb). This gives the effect of a 'lexeme-within-a-lexeme', posing a challenge for lexeme-and-paradigm models. I present an analysis in which the definition of the participle is driven by a feature REPRESENTATION. This (re-)defines the MORSIG attribute, creating a representation which is identical to that of an adjective, while remaining part of the verb's paradigm. I discuss some of the implications of this analysis for lexical relatedness, the lexical type hierarchy of SBCG and the morphology-syntax interface.

Andrew Spencer. On lexical entries and lexical representations. In Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer (eds.), *The lexeme in descriptive and theoretical morphology*, 277–301. Berlin: Language Science Press.
DOI:10.5281/zenodo.1407009

1 Introduction

The notion of word is by definition central to lexicalist models of syntax, so one would imagine that morphology, too, would occupy a central place in the construction of such models. However, there is as yet surprisingly little consensus between morphologists and syntacticians over fundamental aspects of word structure and the relations between words and syntax or semantics. In addition, I will argue that there is a systematic unclarity in conceptualizations of wordhood even amongst those of us who accept the primacy of the lexeme notion and its role in morphosyntax ('lexeme-and-paradigm' models). One central ontological question is 'what kind of a thing is a word?' The problem is that, whereas inflected word forms can be regarded as 'concrete' linguistic objects which combine with each other to form phrases (another type of object), lexemes are by their nature more abstract: they are ultimately representations which unite a set of related inflected word forms without themselves being a form. They are therefore underspecified representations, in the sense that they are not specified for the various morphosyntactic properties their word forms realize. The dictionary is a set of lexemes, so it, too, is an abstract construct.

The question of what lexemes are is made more acute when we examine a somewhat neglected, but theoretically and conceptually important, type of lexical relatedness, the (true) transposition, illustrated in this paper by the Russian deverbal participle. A participle is the adjectival 'representation' (Haspelmath 1996) of a verb. As such, it is part of the paradigm of a verb and yet it inflects exactly like an adjective and demonstrates much of the external syntax of an adjective (a true participle is used principally as an attributive modifier to a noun). Shifting morphosyntactic category in this fashion is characteristic of derivation, i.e. lexeme formation, yet a true participle (that is, a participle that has not undergone lexicalization, or some other process of grammaticalization) is not an autonomous lexeme, independent of its verb base, any more than the past tense or the infinitive form of a verb is an autonomous lexeme. The participle thus gives the appearance of being a 'lexeme-within-a-lexeme', posing obvious difficulties for any simple characterization of lexeme-and-paradigm inflectional morphology, and especially to the inferential-realizational (I-R) class of models in Stump's (2001) typology.

In this paper I investigate some of these questions against the backdrop of the class of I-R models called Paradigm Function Morphology (PFM: Stump 2001, Bonami & Stump 2016). Specifically, I will assume the overall architecture of a model of lexical relatedness proposed in Spencer (2013), Generalized Paradigm Function Morphology (GPFM). I confront the proposals about lexical representations and lexical relatedness made in GPFM with influential proposals put forward within the variant of HPSG developed by Sag (2012), Sign-Based Construction Grammar (SBCG). I argue that the HPSG/SBCG conception of the lexeme conceals important conceptual inconsistencies. In particular, a lexeme can only be described by a feature structure (FS) that is partially specified. However, this means that technically a lexeme is just a description and not an object. Yet the architecture of the HPSG lexicon demands that lexemes be bona fide linguistic objects, not descriptions of objects.

If we simply declare the lexemes as objects then the question arises as to how much the lexeme can be underspecified. Building on the defaults-based GPFM model I argue that a lexeme is best regarded as a *maximally* underspecified object, bearing all and only those properties which are not predictable from default specifications.[1] I show how the maximally underspecified lexemic representation can help solve the question of the status of transpositions such as participles.

I make a number of background assumptions.

- A dictionary is a list of lexemes.
- Inflectional morphology operates according to I-R principles and defines a paradigm for each class of lexemes, each cell of which is occupied by a pair ⟨ω,σ⟩ for the set σ of morphosyntactic properties realized by the word form ω.
- A fully specified representation of a lexeme includes a specification of a set of syntactic properties, a semantic representation (which for convenience I take to be a simplified form of Lexical Conceptual Structure, Jackendoff 1990) and a unique identifier, variously called the Lexical Identifier (LID), the Lexical Index, or the Lexemic Index (LI). (This is comparable in function to the lexicographer's lemma.)
- The syntactic properties of a lexeme include a specification of its argument structure (ARG-ST).
- The ARG-ST attribute of a lexeme includes a semantic function role (SF role, Spencer 2013), canonically R for nouns, E for verbs and A for adjectives

The chapter is structured as follows. I open by outlining four possible ways of representing lexemes, the fourth of which relies heavily on the device of defaults and overrides operating over a maximally underspecified entry. The next section addresses the question of whether a lexeme can be regarded as an object or not, and how many of its properties can be underspecified.

In §4 I turn briefly to the model of lexical representation proposed in Spencer (2013), and specifically to the way in which an inflectional feature declaration (MORSIG, 'morpholexical signature') can be defined and deployed in a defaults-based model of lexical representation. Against this background §5 addresses the architecturally important question of the place of transpositions such as deverbal participles. These are an important test case because they raise questions of lexemic identity and category membership: the participle behaves as a 'quasi-lexeme', without being the output of derivational lexeme formation proper. I deploy an attribute REPRESENTATION to define transpositions. I discuss the way that the adjectival inflectional paradigm can be incorporated into the paradigm of a verb by appropriate use of the MORSIG attribute. I illustrate with a description of the Russian participial system. I contrast the behaviour of true participles

[1] This corresponds to Sag's 2012 notion of listeme. The listeme has a somewhat unclear status in SBCG, but Sag explicitly describes it as a description and not an object, so it is not a perfect correspondent to the conception of lexeme proposed here.

with that of transpositional lexemes (Spencer 2013, 2016), which are derived autonomous lexemes formed from transpositions such as participles.

In §6 I ask how transpositions might be incorporated into a multiple inheritance hierarchy but note two problems. First, multiple inheritance hierarchies are not straightforwardly capable of distinguishing, say, the adjectival representation of a verb (participle) from the verbal representation of an adjective (inflecting predicative adjective). Second, there is in any case virtually no discussion in the morphological literature of transpositions and hence no consensus on how their morphological properties should be accounted for. I conclude with a tentative list of questions which arise from the discussion.

I will close this introduction with a terminological note. I shall simplify discussion wherever possible by assuming the correctness of my approach and taking the lexeme to be effectively identical to its description. That is, a lexeme is a dictionary entry, an abstract underspecified representation, which we can think of as a meta-representation, unifying the concrete representations in the complete set of its word forms. The obvious synonym for 'dictionary entry' is 'lexical entry'. However, in constraints-based syntactic models the notion of 'lexeme' is rather poorly developed, and the term 'lexical entry' is often (though not invariably!) used to refer not to the abstract object listed in a dictionary but rather to a concretely instantiated inflected word form of a lexeme. This terminological ploy is confusing, but is now ingrained. Following Dalrymple et al. (2015), I shall therefore adopt the term 'lexemic entry' for the standard lexicographic notion of dictionary entry. I will avoid the term 'lexical entry' and refer to the representation (fully or partially specified) of an inflected form as the lexical representation of that word form. This is more than a question of mere terminology, especially in HPSG, but proper evaluation of the issues would require a separate study.

2 The nature of the lexeme

In principle there are a good many ways in which dictionary entries can be represented. It will be useful to consider four of these. The first possibility is to list every inflected form separately with a complete specification of all its properties, whether idiosyncratic or predictable. This will include (i) all the morphological properties, such as inflection class, (ii) syntactic properties such as argument structure, including the SF (semantic function) roles, valence, selection, collocation, lexicosyntactic class features and others, together with (iii) contextual properties or properties relating to usage such as register, connotations, and other, not strictly linguistic, properties that a competent user would be expected to know about the word (what is sometimes called 'encyclopaedic information', though this term is difficult to pin down). I shall call this mode of representation the **unindexed full word form listing** model. Some psycholinguistic models of the mental lexicon appear to have essentially this structure. It does not define a dictionary entry in any direct sense because every word form of every lexeme has the same representational status as any other: *dog* and *dogs* are only marginally more related to each other on this model than are *dog* and *dig* or *dogs* and *geese*.

The unindexed full word form listing model effectively excludes any standard understanding of the notion of dictionary entry, therefore. However, it would be possible to reconstruct the traditional notion of dictionary entry by providing all the forms that unite under a given lexeme with a unique **lexemic index**. Thus, *dog, dogs* would both have the index DOG, distinct from that of *dig* (DIG) or *geese* (GOOSE). This would then define our second model of lexical representation, which I will call the **indexed full word form listing** model. The LI would have to be a secondary property associated with each component of a lexemic entry, FORM, SYN, SEM. At the level of FORM this would mean indexing the lexeme's root, its various stem forms and all its inflected forms (unless these were able to inherit the LI of their stems). At the level of SYN, SEM each individual subattribute (syntactic class, argument structure or whatever, depending on one's syntactic assumptions) would be furnished with the same LI, as would the basic meaning or lexical conceptual structure and any other aspects of meaning. This use of a lexemic index is very similar to that proposed by Jackendoff (1997) and integrated into the Simpler Syntax model (Culicover & Jackendoff 2005), though their model makes rather different assumptions about the structure of inflected words because it retains the morphemic concept and therefore is not strictly speaking lexeme-based.

These first two models share the property that all inflected word forms are fully listed. In such models there is effectively no morphology defining the lexical relatedness that holds between word forms of the same lexeme. In order to capture formal similarity/identity between word forms it would therefore be necessary to postulate lexical redundancy rules (Jackendoff 1975, Bochner 1993) or inflectional templates (Ackerman et al. 2009).

The third model I shall call the **fully specified lexemic entry** model. The term 'fully specified' refers to the fact that on this model (along with the previous two models) the lexemic entry includes fully predictable information about the FORM, SYN, SEM representations as well as unpredictable, idiosyncratic information. For instance, if all syntactic nouns in the language are also morphological nouns (i.e. if the language lacks category mixing with respect to the noun class) then the property of inflecting as a noun, that is, being a morphological noun, can be deduced from the SYNCAT label. However, under the fully specified lexemic entry model such a word would still be given the attribute [MORCAT *noun*] or the equivalent as part of its FORM representation. Where this third model differs from the previous two is in the important assumption that (regularly) inflected word forms are not included as part of the lexicon as such. Rather, such a model follows lexicographic tradition in abstracting away from inflected word forms, instead, defining them by means of a separate 'inflectional engine', such as PFM. On the fully specified lexemic entry model, the lexeme-as-dictionary-entry is accorded a special ontological status, that of a linguistic object. Depending on how such a model is implemented formally it may or may not be necessary to individuate dictionary entries by means of the arbitrary LI attribute. However, traditional lexicography certainly makes use of something very close to an LI in the form of a lemma or headword. An arbitrary label of this sort appears to be the most natural way of individuating entries.

The fourth model of lexical representation is the **underspecified lexemic entry** model, argued for in Spencer (2013). This model deploys the logic of default inheritance to ab-

stract away fully predictable lexical information. The lexemic representation in this case includes just the information that cannot be inferred by default from other aspects of the representation or from other facts in the grammar of the language. Thus, in our previous example, if the specification [MORCAT *noun*] is fully predictable from the specification [SYNCAT *noun*] then the MORCAT specification need not be stated in the lexemic entry itself (indeed, there need be no mention of the attribute MORCAT at all).

To see how the underspecified lexical entry model might define dictionary entries, consider a word such as TREE. This minimally has to specify a phonological form for the basic stem form (root), STEM_0 = /tri:/, as well as minimal information about the kind of meaning the word has. As far as morphosyntax and especially inflection is concerned it hardly matters, of course, what kind of a thing a tree is (much less where to draw the line between trees and bushes). Also, the difference between abstract and concrete denotations seems to have little grammatical import, in English. However, it is important to know that TREE denotes some type of *Thing* and that it is countable, in contrast to words such as VEGETATION, or WOOD (in the sense of 'material coming from a tree'). Informally, we can distinguish count *Things* and mass *Things* with a subscript: $\textit{Thing}_c$/$\textit{Thing}_m$. However, for English we should also have some way of representing the fact that TREE (and IDEA) denotes something which is not a sexed higher animal, such as a person or a horse and which therefore can only be referred to as *it*, not as *s/he*. In languages which distinguish a 'vegetable' gender (e.g. Bininj-Gunwok) we might need to indicate the fact that TREE (and perhaps VEGETATION but not IDEA) denotes a kind of plant. In other languages with semantically-driven gender other distinctions would have to be made. These observations hold for the determination of inflectional properties. However, for a specification of derivational morphology it is often necessary to appeal to very subtle nuances of meaning (Fradin & Kerleroux 2003).

The point of this discussion of lexical semantics is that once the right semantic properties are fixed much of the rest of the lexemic representation can be deduced by default. Thus, if an English lexeme belongs to the *Thing* ontological category (as opposed to the category *Event* or *Property*) then by default it will be a noun, with an argument structure that includes the SF role R. A syntactic noun will also be a noun morphologically, and if it is of subcategory $\textit{Thing}_c$ it will have a singular and plural form. This is more than just a modern version of the notional parts-of-speech theory, however. Being defaults, all these inferences can, of course, be overridden by more specific lexical stipulations. Thus, a noun such as JOURNEY is ontologically an *Event* but grammatically it is a noun, so that the inference from *Event* to SF role E to [SYNCAT/MORCAT *verb*] is overridden in the lexemic entry (for instance, by stipulating that its SF role is a simplex R). Moreover, in many languages there will be non-default morphological information to stipulate in addition to the phonology of the root. For instance, the Russian noun STOLOVAJA 'canteen; dining room' is a noun syntactically, but it has the morphology of a (feminine gender) adjective, thus its [MORCAT *adjective*] value cannot be inferred from its [SYNCAT *noun*] value and has to be stipulated in the lexemic entry in some way. In some cases, not all argument structure or complementation properties can be deduced from the semantic representation so those would need to be specified lexically. Some of the contextual properties of a

lexeme such as special register, connotations, or other details of usage may also diverge from the default and will therefore have to be recorded in the lexeme's entry. But the limiting case of a lexical representation in the underspecified lexemic entry model is a pure pairing of basic meaning with the form of the root (what Sag 2012 refers to as a 'listeme'; see §3) .

3 Lexemes as objects or descriptions

The principal question to be addressed in this paper is: what kind of a representation is a dictionary (lexemic) entry? Specifically, is it a linguistic object in its own right? In this section I discuss the answers proposed in Sag's (2012) summary of SBCG.

In SBCG, as in HPSG generally, a distinction is drawn between linguistic objects and the representational technology used to describe those objects, notably feature structures (FSs) or attribute-value matrices (AVMs). An inflected word form, for example, is a linguistic object, but it can be described in various ways, including partial feature descriptions which underspecify certain aspects of the representation. A linguistic object proper, however, cannot be thus underspecified. This means, for instance, that Sag's listeme, the barest possible representation of a lexemic entry, must be a description, not an object in its own right.

Sag (p. 98) introduces the notion of the lexeme into the model, giving it a special place in the type hierarchy of signs shown in Figure 1. This hierarchy defines the lexeme as a lexical sign, just like a word form. However, word forms appear as parts of syntactic phrases which can ultimately be pronounced, and so they count as linguistic expressions. A lexeme cannot be pronounced. This is not because it is some kind of 'covert expression', however (like *gap* and *pro* in Sag's type hierarchy). A lexeme is an altogether different kind of sign, in fact, a unique type given the hierarchy in Figure 1.

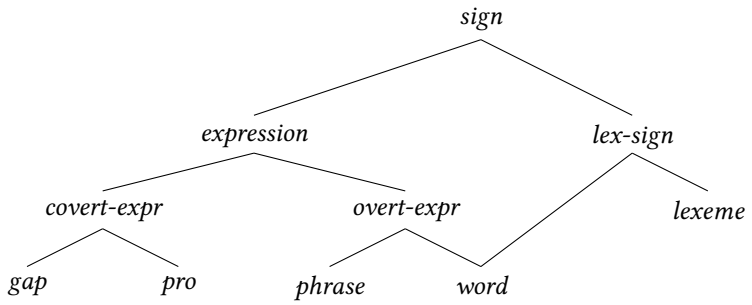

Figure 1: Sag's (2012) type hierarchy

Sag provides examples of representations of word forms from English (plurals, past tense forms) and in his Fig. 6 (p. 101), here reproduced as Figure 2, he gives the example of the lexeme LAUGH. Notice that this representation actually seems to specify the word

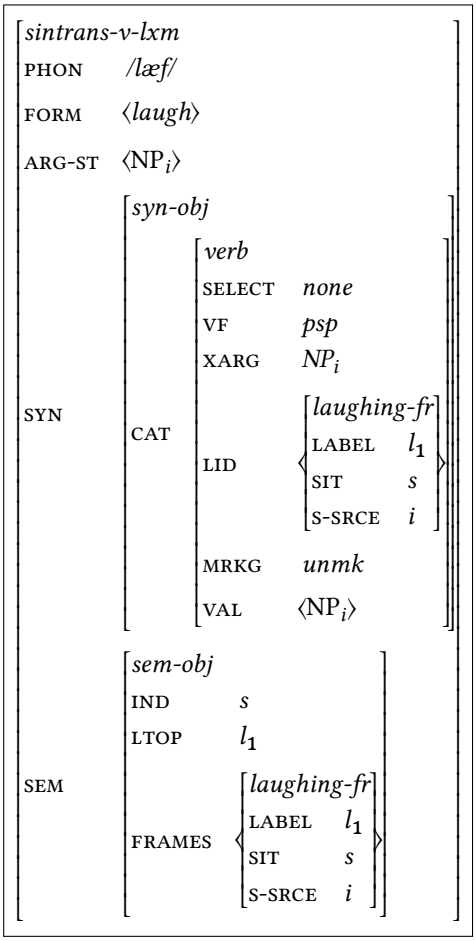

Figure 2: Sag's (2012: 111) representation of the lexeme LAUGH

form *laughed*, in that it bears the feature [VFORM *psp*]. It is worth citing Sag's justification for this choice of representation:

> [T]he value *psp* illustrated here […] represents an arbitrary expositional choice — any value of VFORM would satisfy the requirements imposed by the *laugh* listeme. And each such choice gives rise to a family of well-formed FSs licensed by that listeme. (Sag 2012: 99)

Sag here appeals to the LAUGH listeme. In SBCG a listeme *licenses* modelled linguistic objects. This means that it places restrictions on what properties a modelled object or sign may have (p. 105). Another way of characterizing the listeme is as "a lexeme description in the lexicon" (p. 107).

The type *lexeme* plays a central role in SBCG, in that it is the starting point for all morphology (Sag is here following PFM and related models). Inflection and derivation are modelled by means of morphological functions. An inflectional rule such as the English preterite (past tense) is modelled by a *preterite-cxt*, whose mother is the past tense form and whose daughter is the lexeme whose past tense form is being defined. A derivational rule is given by a construction whose mother is the derived lexeme and whose daughter is the base lexeme.

Sag summarizes the morphological functions by saying (p. 113) that they express "<…> the relation between the forms of two lexemes or the relation between the form of a lexeme and the form of a word that realizes that lexeme." This sounds like an expression of conventional wisdom in lexeme-based morphology, but it hides a serious conceptual flaw. This centres around the way that Sag's formulation uses the term 'form'. The problem is apparent from Sag's description of the lexeme LAUGH. He is obliged to provide this representation with an arbitrary inflectional feature specification, in effect defining not the lexeme as such but one of its inflected forms. This is because a lexeme is meant to be a modelled object, a subtype of *sign*, and a linguistic object must be fully specified. But the whole point of defining a lexemic level of representation is to abstract away from actual (concrete) word forms. This means that the lexeme is effectively a description, in fact a partial description, of the full set of word forms. But that is completely incompatible with Sag's type hierarchy and, indeed, with any coherent interpretation of the HPSG lexicon.

Given this reasoning we seem to have two logical courses of action. Either we can reconstruct the HPSG lexicon without recourse to the type *lexeme*, or we can redefine the notion of linguistic object in such a way as to make a dictionary entry a kind of modelled object, even though it appears to be underspecified. I shall adopt the second approach.

I propose to treat the lexicon as more than just a convenient descriptive fiction, as would be implied by a strict application of the object~description distinction. Rather, I take the lexicon to be a network of mentally represented (or representable) objects which can be defined and described by FSs just like (utterable and unutterable) linguistic expressions.

By simply declaring a dictionary (lexemic) entry to be a kind of object we solve the immediate problem: the lexeme can remain a type of sign, and can be a supertype of other

signs. Its unusual position in being partially underspecified is now reflected in the type hierarchy: only the *expression* type has to be fully specified, a lexical sign may be only partially specified (*lexeme*), though when a lexical sign is also a subtype of *expression* (*word*) it, too, can, and must, be fully specified.

Now, once we admit the possibility of an underspecified entity as an object in the linguistic ontology we are immediately faced with two sets of questions. The most general of these is 'are there other linguistic objects which can be less than fully specified? Can *any* partially specified representation be interpreted as a modelled object? If so, then what is the content of the original object~description distinction?" It seems that we should not be allowed to postulate such objects except in very special circumstances. But if we admit lexemes as less than fully specified objects what prevents us from postulating entirely arbitrary types? The simplest answer is to say that it is an architectural (i.e. stipulated) property of linguistic expressions that they be fully specified. However, whether this is really true may depend on how we perceive linguistic specification. Presumably, an object of type *word* such as *dogs* is to be regarded as a fully specified object and not a description, even when, for instance, its intonation and other prosodic characteristics are not specified. But in the strictest sense a word form remains partially underspecified until its full phonetic realization is given. Indeed, the same is true of sentences, which can be uttered with a very wide variety of affective intonation contours even when realizing one and the same set of discourse or information-structure functions.

The second question is more immediately relevant: if we are to admit as an object a lexeme underspecified for its inflection properties, how much further can we go with the underspecification? For instance, we might want to say that our lexeme LAUGH is underspecified for its inflectional properties by virtue of bearing the attribute values [TENSE u, VFORM u, SUBJAGR u, …] or whatever, where 'u' means 'not yet specified value', or we may wish to make the more radical proposal that LAUGH lacks the actual attributes [TENSE, VFORM, SUBJAGR, …]. This may turn out to be little more than a matter of notational convention, but in a more radical vein we can ask why we can't regard Sag's maximally underspecified listeme as a default lexeme object. In other words, can we not adopt the underspecified lexemic entry model for dictionary entries, as proposed in Spencer (2013)? We will see that the question assumes particular importance in defaults-based models of morphology such as PFM, where the lexeme concept finds its most elaborated implementation, and especially GPFM, where defaults define all aspects of lexical representation. Before turning to a consideration of the lexeme concept in such models I first discuss an important but generally neglected aspect of lexical representation and its relation to inflectional morphosyntax.

4 The morpholexical signature (MORSIG)

A lexeme of a given morpholexical class, such as 'noun', will (typically!) inflect for properties particular to that class (say, NUMBER, CASE, DEFINITENESS, POSSESSOR AGREEMENT) and may have intrinsic properties which determine its morphosyntax, such as GENDER. The actual set of properties is stipulated for each language, so a grammar has to include a

declaration of that set. In the Generalized Paradigm Function Morphology(GPFM) model of Spencer (2013) I refer to this declaration as the **morpholexical signature** (MORSIG). In GPFM the MORSIG attribute is itself treated as a default property with respect to lexemic entries/representations. By this I mean that the properties which make up the MORSIG are true of every regular lexeme of the given class, so it would be redundant to specify that information in the lexemic entry itself.

In Spencer (2013) I treat the MORSIG as a value of the FORM attribute, i.e. as a morphological property of a lexeme, but this is an oversimplification. It is well-known that the set of features needed to define a lexeme's syntactic distribution, and the set of grammatical meanings expressed by inflected word forms, are often at variance with the set of features needed to define the inflected morphological forms themselves. The most obvious mismatches are found in periphrases. We often find that the morphological form of one of the elements of the construction bears properties which contradict the feature content expressed by the periphrasis as a whole. Elsewhere, the morphological element may be morphomic and therefore not associated with any meaning, or the periphrasis may express a meaning in the manner of an idiom, so that no part of it can sensibly be associated with the meaning of the periphrasis as a whole (Brown et al. 2012). Periphrasis therefore motivates a distinction between m-features and s-features (mnemonically, morphological/syntactic features, Sadler & Spencer 2001). Similarly, Stump has argued for a modification of the original Paradigm Function Morphology (PFM) model, 'PFM1' (Stump 2001), in favour of a model, 'PFM2', which draws a distinction between FORM and CONTENT paradigms, on the basis of mismatches such as syncretisms, deponency and a variety of others (Stump 2002, 2006, 2016a,b). The obvious way to capture such distinctions in lexical representations is to assume that there is a SYN|MORSIG attribute which is mapped to a FORM|MORSIG attribute by means of a function, Stump's 'paradigm linkage'. By default, paradigm linkage is the identity function, in the sense that the FORM paradigm or m-feature set is identical to the CONTENT paradigm/s-feature set.

In GPFM the relation between the most highly underspecified lexical representation and a fully specified word form is mediated by two sets of functions. The second of these is effectively identical to the paradigm function of PFM2. It maps a pairing of $\langle L,\sigma\rangle$, for LI *L*, feature set σ, to a pair $\langle\omega,\sigma\rangle$, where ω is the corresponding inflected word form. This function is, however, only defined for a complete and coherent feature set. In other words the function cannot be defined for a representation which lacks a specification of those features for which the lexeme inflects, that is, the MORSIG. Therefore, to be inflectable the lexeme's MORSIG attribute needs first to be specified (*Inflectional Specifiability Principle*, Spencer 2013: 199). This is achieved by the first of the two functions, the default specification of MORSIG for a given morphosyntactic lexical category.

An illustration of how this works can be given by (a simplified version of) the Turkish noun (following the discussion in Stump 2016a: 175–179). The minimal lexical information needed for, say, the word EV 'house' is shown in Figure 3 (using English as a metalanguage). Turkish grammar stipulates that a count noun inflects for the properties shown in Figure 4. The FORM|MORSIG attribute is almost identical except for a well-known syncretism between the 3sg possessed form of 'houses', and the 3pl possessed forms of

'house/houses' and the ordinary unpossessed plural. We would expect these to take the forms *evler, evlerler, evler* respectively, but the form *evlerler* is reduced by haplology to *evler*. Clearly, the FORM paradigm makes fewer distinctions than the CONTENT paradigm.

$$\begin{bmatrix} \text{FORM} & \left[\text{STEM}_0 \left[\text{PHON } /\text{ev}/\right]\right] \\ \text{SEM} & \left[\text{Thing}_c \ \lambda x.\text{house}(x)\right] \\ \text{LI} & \textit{HOUSE} \end{bmatrix}$$

Figure 3: Lexemic entry for Turkish EV 'house'

$$\begin{bmatrix} \text{SYN}|\text{MORSIG} & \begin{bmatrix} \text{NUMBER} & \{ sg,pl \} \\ \text{CASE} & \{ nom,acc,gen,dat,loc \} \\ \text{POSS} & \begin{bmatrix} \text{PERSON} & \{ 1,2,3 \} \\ \text{NUMBER} & \{ sg,pl \} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 4: MORSIG for Turkish count noun lexeme

In PFM2 this mismatch is defined via a Correspondence function, ***Corr***, which specifies the distinct FORM features and CONTENT features and which defines the mismatches giving rise to syncretism, deponency and so on. The details are not relevant here so I simply assume the existence of the ***Corr*** mapping.

5 Lexical relatedness and the role of the Lexemic Index

The notion of lexemic representation (lexeme, lexical entry) plays an important role in the I-R class of models. This is especially true of GPFM, because that model attempts to unify inflection with (regular, productive, paradigmatic) derivational morphology. If we say, for the sake of argument, that English Subject Nominal (SubjNom) formation is paradigmatic then we can define it by recourse to a derivational feature (cf. Stump 2001: 257) SN, such that the generalized paradigm function, GPF, will map a verb lexeme to its subject nominal: GPF($\langle L$, SN$\rangle$) = $\langle L'$, SN$\rangle$, where L' is the LI of the subject nominal of the verb L. However, the GPF cannot apply in exactly the way that the PF applies in PFM2. In PFM2 the PF maps a pairing of ⟨LI,features⟩ to a word form (via the ***Corr*** function). But the output of a derivational function has to be some representation of an independent lexeme. This means that when a derivational feature is in the domain of the GPF it must map to a representation of that derived lexeme, not to a word form. But the standard architecture of PFM2 (including the ***Corr*** function) does not permit this. The problem is at heart very familiar: while inflectional morphology specifies word forms that realize the particular morphosyntactic property set of a lexeme, derivational morphology effects

wholesale changes in syntactic and semantic representations, undermining the basic I-R assumptions under which morphology simply serves to realize property sets.

In the GPFM model of Spencer (2013), derivational morphology requires the GPF to perform a kind of 'deletion' of the base lexeme's properties, followed by respecification by means of defaults driven by the enriched SEM representation of the derived lexeme. However, a more parsimonious way to represent derivational morphology is to map the maximally underspecified base lexeme's entry to a maximally underspecified derived entry. This obviates the need to delete most of an entry's specifications, in that they are lacking in any case. Thus, for the lexeme DRIVE and its SubjNom DRIVER a schematic application of the GPF would be as in Figure 5 (where SN(DRIVE) is a function from LIS to LIS governed by the derivational feature, defining the LI of the derived lexeme, DRIVER). This type of application can be thought of as an elaborated, feature-driven word formation rule (*wfr*) in the sense of Aronoff 1976.

$$\begin{bmatrix} \text{FORM} & \begin{bmatrix} \text{STEM}_0|\text{PHON} & /\text{draɪv}/ \\ \text{STEM}_{\text{PST}}|\text{PHON} & /\text{droʊv}/ \\ \text{STEM}_{\text{PSTPTCP}}|\text{PHON} & /\text{drɪv}/ \end{bmatrix} \\ \text{SEM} & \left[\text{Event } \lambda x, y.\text{drive}(x, y)\right] \\ \text{LI} & \text{DRIVE} \end{bmatrix}$$

$$\Downarrow$$

$$\begin{bmatrix} \text{FORM} & \left[\text{STEM}_0|\text{PHON} \quad /\text{draɪv}/\oplus/\text{ə}/\right] \\ \text{SEM} & \left[\text{Thing } \lambda x[\text{person}(x) \wedge \exists y.\text{drive}(x, y)]\right] \\ \text{LI} & \text{SN}(\text{ DRIVE}) \end{bmatrix}$$

Figure 5: Derivation of DRIVER from DRIVE

Now, the output of the GPF is the representation of a *Thing*, so by default it will have all the morphosyntactic properties of a noun.[2] In languages with nominal inflectional classes the GPF may additionally have to specify which inflectional class the derived noun belongs to, as a FORM property overriding whatever the default specification for noun inflection class is, just as would be the case with a simplex (underived) lexemic entry belonging to a non-default inflectional class. The function in Figure 5 fails to transfer the non-default (stipulated) specification of the past tense and past participle stems from the base verb to the subject nominal, giving rise to a kind of despecification. There is an important rationale behind the despecification of lexemic entries in Spencer (2013): derivation, unlike inflection, leads to lexical opacity. Thus, the derived lexeme DRIVER lacks any specification which would identify it as having a base with past tense or past participle forms, irregular or otherwise, or, indeed, any of the morphosyntactic proper-

[2] DRIVER is a count noun, of course. I assume that this can be made to follow from the fact that a driver is a subtype of person.

ties associated with a finite verb. In this case the failure of the past and past participle forms to be inherited by the derived noun is the consequence of the definition of the MORSIG attribute for nouns as opposed to that for verbs. The GPF for SubjNom specifies exactly one STEM_0 form (for regular lexemes). This can be unified with the default MORSIG specification associated with *Thing* lexemes. Since the *Thing* ontological category does not license inflectional (s-feature) paradigm properties other than NUMBER in English, there would be no way for any tense or participle features to unify with the MORSIG attribute once it is specified. The only additional assumption that we need to make here is that SubjNom derivation is the kind of lexical relatedness which defines an entirely new MORSIG (i.e. one which 'deletes' the MORSIG of the base entry). I return later in this section to the question of how we characterize the class of relatedness functions which fail to preserve the base lexeme's MORSIG attribute in this way.

In true derivational morphology the LI of the output lexeme is always distinct from that of the base. This reflects the most significant difference between derivational types of lexical relatedness, on the one hand, and types of lexical relatedness broadly thought of as inflectional, on the other hand: derivation defines new lexemes while inflection defines forms of lexemes. However, in GPFM, preservation or alteration of the LI is just one parameter of relatedness, almost entirely independent of other parameters (this is the *Principle of Representational Independence*, Spencer 2013: 139). In particular, we systematically encounter two types of situation in which the crucial feature of the relatedness is the preservation or change of the base lexeme's LI.

The first of these is the class of relatedness types called **transpositions**, in which the morphosyntactic class of a word changes, as in typical derivation, but in which there is no creation of a novel lexeme with a distinct LI. In a canonical transposition the SEM value, that is, the conceptual content of the representation, does not change either.

The second type of case is very similar. Here the lexical relation defines a distinct lexeme but does not alter the conceptual content of the base. These are what I have called **transpositional lexemes** (Spencer 2013: 275; 359–60; Spencer 2016). Simple examples are adjectives derived from participles such as *interesting, bored* or so-called relational adjectives (in English and other European languages) such as *prepositional, ferrous*. These contrast with superficially similar cases in which the derived adjective differs semantically from its (etymological) base: *budding (linguist), harrowing (experience), gaping (hole); outspoken, unspoken, incensed, poised; popular (= 'well-liked'), spectacular*. Distinguishing true transpositions from transpositional lexemes and transpositional lexemes from other, often homophonous, adjectives is important for understanding the nature of lexical representations and types of lexical relatedness. In some cases, the only difference between the lexical representation of a true transposition and that of the homophonous transpositional lexeme is the difference in LI. However, in many cases the transpositional lexeme has different syntactic privileges from the homophonous transposition by virtue of being an independent lexeme. For instance, the adjective *interesting* has the complementation properties of an adjective, not of a verb or a true participle, as seen by comparing the true participle in (1) with the true adjective in (2).

(1) *interesting* = participle

a. the book (*very) interesting the children

b. * The book seems interesting the children.

(2) *interesting* = adjective

a. the book most interesting to the children

b. The book seems interesting to the children.

Comparable examples can be found with Russian participles and participial lexemes.

A clear instance of a true transposition is the (deverbal) participle, familiar from many languages, including almost all Indo-European languages. In Russian, for instance, we find four participles, realizing the properties [VOICE {act, pass}], [ASPECT {*pfv*, *ipfv*}] (Spencer 2017). These inflect exactly like adjectives and their principal function is that of attributive modifier to a noun. However, in addition to expressing the verbal properties of voice and aspect the participles also retain the argument structure/complementation of the base verb, including quirky case assignment. They are thus prototypical examples of mixed categories.

In Spencer (2013, 2017) I argue that participles belong to the base verb's paradigm in the broadest sense, and that this means their LI is that of the base verb. In an I-R model this means that the participles are defined by a ⟨feature, value⟩ pair, just like tense or number forms, and I propose the feature REPR(ESENTATION), following Russian descriptive tradition (see, for instance, Kuznecova et al. 1980, Helimski 1998 for the Samoyedic language Selkup, which is particularly rich in transpositions; see also Haspelmath 1996).

Following Spencer (2017) I notate the feature as REPR⟨K,Λ⟩, denoting a transposition from category K to category Λ. For example, a participle would be defined by the feature REPR⟨V,A⟩.[3] The GPF(⟨*V*,{REPR⟨V,A⟩,σ}⟩) applies to a verb lexeme *V* and defines a participle realizing features σ. For instance, the Russian perfective passive participle *udar´onn-* from UDARIT´ 'hit, strike' is defined by (3).

(3) GPF(⟨UDARIT´,{REPR⟨V,A⟩,{[ASPECT *pfv*],[VOICE *pass*]}}⟩).

The GPF (3), however, only defines the stem of the participle. In order to inflect it as an adjective it must be given an appropriate MORSIG, inheriting CONCORD (agreement) features from the adjective class, permitting the participle to agree with the head noun. This addition to the MORSIG is an automatic consequence of redefining the morphosyntactic class as *adjective*. The technical details of exactly how this is achieved are provided in Spencer (2017). The GPF which defines the stem of the participle defines a lexical representation which is thus very similar to that of a (maximally underspecified) simplex adjective before it receives the default MORSIG specification. In this way the participle resembles an automonous adjectival lexeme, whilst remaining a form (better, the adjectival representation) of the verb, what we could call a 'quasi-lexeme'.

[3] The labels 'V, A' are for convenience. In fact, it is likely that all 'capital letter' lexical/phrasal ('c-structure') category labels (N, V, A, P) can be dispensed with, in favour of appeal to more fine-grained properties, especially the SF roles (Spencer 1998, 1999, 2013: 322–23; see also Chaves 2014 for similar remarks) .

Here is, in broad outline, how the GPF would deliver the quasi-lexeme form *udar´onn-*. A (partial) FS for the MORSIG of a typical transitive verb is shown in Figure 6. The FS in Figure 6 shows those morphosyntactic properties that are reflected in the grammatical system of Russian. It does not, however, tell us what the inflected forms are. This is because that FS defines the CONTENT paradigm feature set, not the FORM paradigm set. For instance, [TENSE *fut*] is only expressed morphologically in [ASPECT *pfv*] verb forms; in imperfective verb forms future tense is expressed periphrastically. Similarly, [VOICE *pass*] is only expressed synthetically in imperfective verb forms (where it actually borrows forms marked [REFLEXIVE *yes*]); in perfective verb forms it is expressed again periphrastically.

$$\left[\begin{array}{ll} \text{SYN} \mid \text{MORSIG} & \left[\begin{array}{ll} \text{ASPECT} & \{\, pfv,ipfv \,\} \\ \text{VOICE} & \{\, act,pass \,\} \\ \text{TENSE} & \{\, prs,pst,fut \,\} \\ \ldots & \\ \text{REPR}\langle \text{V,A} \rangle & \left[\begin{array}{ll} \text{ASPECT} & \{\, pfv,ipfv \,\} \\ \text{VOICE} & \{\, act,pass \,\} \end{array}\right] \end{array}\right] \end{array}\right]$$

Figure 6: Partial MORSIG for a Russian transitive verb

The somewhat complex mapping between CONTENT and FORM paradigms in Russian verbs is explored in greater detail in Spencer (2017). The precise characterization of the FORM or m-features for Russian verbs is controversial (as it is for most languages, including English). In Spencer (2017), for instance, I argue that the FORM paradigm has a single-valued m-[TENSE *prs-fut*] feature, accounting for both the present tense inflections of imperfective verb forms and the (identical) future tense inflections of perfective verb forms. Likewise, the CONTENT paradigm feature s-[TENSE *pst*] is expressed by a morphomic l-participle form ([VFORM *lptcp*]), which has no semantic interpretation of its own but which co-realizes s-[MOOD *conditional*] in conjunction with the particle *by*. Elsewhere, by default the l-participle realizes the CONTENT paradigm s-[TENSE *pst*] feature value. The specification [TENSE *pst*] has no FORM/m-feature counterpart.

The partial specification in Figure 6 also shows us that a transitive verb in Russian has four participial forms, listed in Table 1, where the parenthesized suffixes (-ij, …), (-yj, …) indicate the agreement inflections.

Table 1: Participles of Russian UDAR´IT´ 'hit'

udar´-aju-šč(-ij, …)	imperfective active
udar´-aje-m(-yj, …)	imperfective passive
udar´-i-vš(-ij, …)	perfective active
udar´-on-n(-yj, …)	perfective passive

Given the MORSIG in Figure 6 the GPF can apply to a pairing ⟨*U*,π⟩, where *U* is the LI of UDAR´IT´ and π is a mnemonic shorthand for the set of participial features {[REPR⟨V,A⟩],{[ASPECT *pfv*],[VOICE *pass*]}}. In the original PFM models (PFM1 and PFM2) the paradigm function serves solely to define inflected forms (and periphrastic realizations of certain inflectional features). In terms of the lexical representational schemas discussed so far this means that the PF operates solely at the level of the FORM attribute. In GPFM the PF is generalized to four functions, operating over the FORM, SYN, SEM, LI attributes. The first of these, f_{form}, is the classical PF. For ordinary inflectional morphology the f_{syn}, f_{sem}, f_{li} functions have no material effect and behave like identity functions. Thus, the GPF for pure inflection collapses with the classical PF. However, for paradigmatic derivational morphology all four functions can introduce non-trivial changes as we saw earlier in the case of the derivation of DRIVER from DRIVE.

The case of transpositions such as participles is midway between that of pure or canonical inflection and derivation. The LI and SEM attributes remain unchanged but both FORM and SYN attributes have to be (re-)specified. Following Spencer (1999, 2013), in Spencer (2017) I assume that the category of a transposition is defined in terms of a complex SF role. A simplex verb has the SF role [ARG-ST|SF E] and an adjective the SF role [ARG-ST|SF A]. A participle is the adjectival representation of a lexeme with SF role E. The notion 'adjectival representation' is captured by defining a complex SF role ⟨A⟨E⟩⟩. To simplify the exposition I shall assume that the complex SF role is cashed out as a complex category label, [A [V]] (at the SYN level SYNCAT|[A [V]], at the FORM level MORCAT|[A [V]]).[4] The GPF for a participle, as defined by the attribute REPR⟨V,A⟩ will define a form with this new category, as shown in (4).

(4) $f_{syn}(\langle U,\pi\rangle)$ = …
[SYN|SYNCAT V] ⟹ [SYN|SYNCAT [A [V]]]

The transpositional feature specification π will also define a restatement of the MORSIG attribute for the participle, as shown in (5).

(5) [ASPECT], [VOICE] ⊂ [SYN|MORSIG]

The statement in (5) is more specific than the default specification and hence it will override that default. However, the participles in Russian (unlike some languages) are actually adjectival forms. Therefore, their lexical representations must include a feature defining their agreement properties, which for convenience I will label CONCORD. This feature must be included there, in the participle's MORSIG. However, that fact, together with the definition of [CONCORD], is inherited from elsewhere in the grammar in the definition of adjectival inflection, as shown in (6).

(6) a. SF ⟨⟨A … ⟹ [CONCORD] ⊂ [SYN|MORSIG]
b. [NUMBER], [GENDER], [CASE] ⊂ [CONCORD]

[4] In fact, it seems that the device of complex SF roles allows us to dispense entirely with traditional syntactic category labels (see also footnote 3).

Declaration (6) is so formulated that it applies to any word type whose 'outermost' category label is defined by the complex SF ⟨⟨A …. This will trivially include simplex adjectives, of course, but it also includes (true transpositional) participles (SF ⟨⟨A⟨E⟩⟩) and true relational adjectives (SF ⟨⟨A⟨R⟩⟩). Russian participles are well-behaved morphologically and so they will inherit very nearly all the FORM|MORSIG properties implied by the SYN|MORSIG specification.[5]

We are now in a position to state the full GPF defining the perfective passive participle, an extension of the GPF shown schematically in (3). This is shown in (7). It defines the object represented by the FS given in Figure 7.

(7) GPF for the perfective passive participle of UDARIT′ 'hit'
Where U is the Lexemic Index of the lexeme UDARIT′ 'hit' and π is the feature set [REPR⟨V,A⟩ [ASPECT *pfv*, VOICE *pass*]], the passive perfective participle stem form is defined by a generalized paradigm function, GPF(⟨U,π⟩) =

(i) f_{form}(⟨U,π⟩) =
[FORM STEM_{ppp} = PHON STEM_0(U)⊕onn = /udar′onn/]

(ii) f_{syn}(⟨U,π⟩) =

$$\left[\begin{array}{ll} \text{SYN} & \left[\begin{array}{ll} \text{SYNCAT} & \left[\text{A}\left[\text{V}\right]\right] \\ \text{ARG-ST} & \langle(x), y\rangle \\ \text{MORSIG} & \left[\begin{array}{ll} \text{ASPECT} & \textit{pfv} \\ \text{VOICE} & \textit{pass} \end{array}\right] \end{array}\right] \end{array}\right]$$

where (x) denotes the suppressed external argument of the passive.

(iii) f_{sem}(⟨U,π⟩), f_{li}(⟨U,π⟩) are the 'identity function' (no change in representation).

The redefinition of the MORSIG attribute to include two attributes inherited from the verb base together with the new CONCORD attribute is part of the morphosyntactic definition of 'participle' in Russian. However, the subsequent inflection of the participle as an adjective follows entirely from the more general characterization of adjectives, independently of their origin. For instance, it is equally applicable to a purely derivational adjective such as *svet-l-yj* 'bright, light' from *svet* 'light', or *krov-av-yj (režim)* 'bloody (regime)' from *krov′* 'blood'. This means that the participle feature ensemble π defines an underspecified lexical representation which has exactly the same type of structure as an

[5]The main caveats here concern participles used as predicates, where there are a number of restrictions. The participle also retains crucial verb properties such as complementation and even quirky case assignment, so we need to ensure that those properties are inherited by the participle when the GPF is applied to π. This would require a much more detailed discussion of the lexical representation of verbs, so I refer the reader to Spencer (2017) where some of those details are worked out.

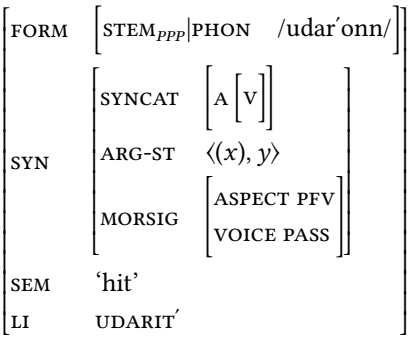

Figure 7: "Quasi-lexemic" feature structure for Russian passive perfective participle *udarˊonn*

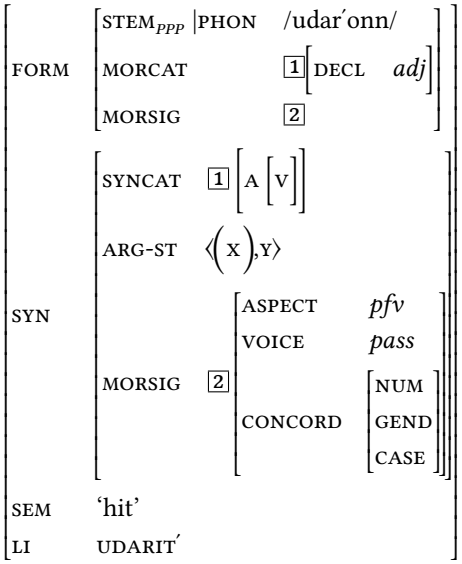

Figure 8: Feature structure for passive perfective participle *udarˊonn* after default specification of MORSIG

independent simplex or derived adjectival lexeme. It is in this respect that the participle behaves as a quasi-lexeme, having the inflectional and morphosyntactic potential of an adjective but remaining a 'form' (more precisely, representation) of the base verb.

The analysis now brings us back to one of the questions posed earlier — is the representation in Figure 7 an object or a description?

If we regard Figure 7 as a description (vs. object) then it would presumably have to describe an object of type word. But this would entail that it describes some particular inflected form, say, the feminine instrumental plural. But the participle is not specified for those or any other CONCORD features, just as Sag's FS for LAUGH is underspecified for any inflectional feature set. This makes the participle FS look exactly like a lexemic entry, which *ex hypothesi* is an object not a description. It is this object that I have informally referred to as a quasi-lexeme. However, from the perspective of the grammatical system, it *is* a lexeme, albeit not one which is independent of its verb base.

The participle shares its Lexemic Index with the base verb in all its inflected forms. However, it is easy to imagine such a representation undergoing the simplest type of lexicalization, namely, to acquire its own unique LI. This would happen if the participle were recategorized as a simplex adjective, that is a member of the morphosyntactic category [A] rather than [A [V]]. This is then the representation of a transpositional lexeme of the type *interesting*. Russian, too, has such converted participial lexemes, though they often do not correspond to English transpositional lexemes. Examples are *potr´asájuščij* 'amazing' from *potr´asát´* 'to amaze', *izmúčonnyj* 'exhausted' from *izmúčit´* 'to exhaust' and many others (see Spencer 2017 for further discussion) . The crucial point is that these derived adjectival lexemes do not seem to differ from their verb bases in their semantics, just like true transpositions, yet they behave syntactically like independent lexemes.

6 Lexemes and types

We have arrived at the conclusion that the lexical representation of a participle is non-distinct in crucial ways from the representation of a lexeme, and for this reason the grammar will treat it as a linguistic object, akin to a lexeme. This invites the conclusion that the participle is, in fact, a subtype of the type *lexeme* in the hierarchy proposed by Sag (2012), say, *ptcp-lxm*. The problem would then be to define where *ptcp-lxm* fits in the type hierarchy. A participle inherits from both adjectives and verbs, as illustrated in Figure 9, adapting Sag's hierarchy for English (with obvious modifications for Russian).

This would be in keeping with Malouf's (2000) approach to deverbal nominalizations. However, there are a number of problems with this solution. One of these relates to the 'directionality' or 'headedness' of transpositions: a transposition is a representation of its base lexeme. In that respect a participial quasi-lexeme bears the same relationship to a verb that, say, the past tense form bears. But this is not captured in a hierarchy such as that sketched in Figure 9, where the relation between *verb-lxm, adj-lxm*, the two mothers of the participle *ptcp-lxm*, is equal. As a result, there will be no way of distinguishing between the adjectival representation of a verb and the verbal representation of an adjective (that is, a transpositional predicative adjective heading a finite clause and bearing inflections for verb features such as tense-mood-aspect-polarity or subject agreement).

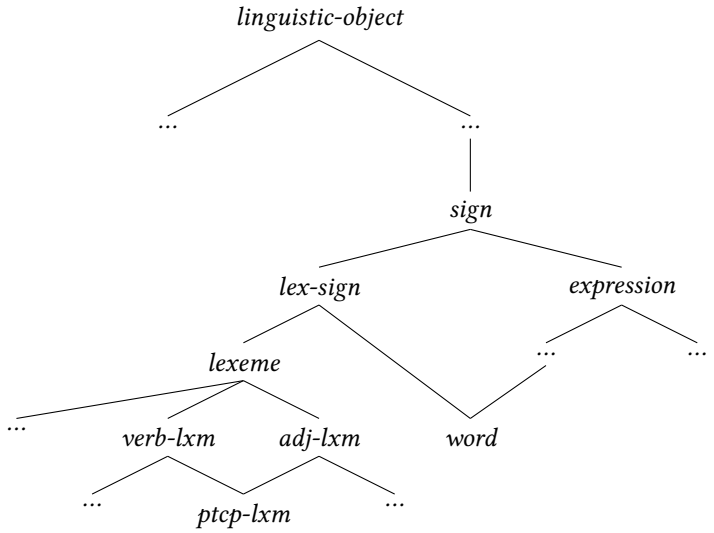

Figure 9: Revised partial type hierarchy

Perhaps, then we should adopt a different approach. Since participles are morphologically derived we can set up a construction type in SBCG (or a lexical rule in standard HPSG) which would perform the same role as the GPF applied to the REPR feature in GPFM. Sag defines two sorts of morphological construction relevant to us in this context, the *infl-cxt* and the *deriv-cxt*.

(8) $\textit{infl-cxt}: \begin{bmatrix} \textsc{mtr} & \textit{word} \\ \textsc{dtrs} & \textit{list(lexeme)} \end{bmatrix}$ (Sag 2012: 115)

(9) $\textit{deriv-cxt}: \begin{bmatrix} \textsc{mtr} & \textit{lexeme} \\ \textsc{dtrs} & \textit{list(lex-sign)} \end{bmatrix}$ (Sag 2012: 119)

The formulation in (9) additionally permits derivation from word forms, but in general derivation is defined over lexemes and to simplify the discussion I will assume that this is always the case. If we take a participle to be a subtype of *lexeme*, then participle formation will be a subtype of the derivational construction shown in (9).

One issue that has to be resolved when incorporating morphological models into lexicalist syntactic models arises from the fact that I-R models of morphology are generally based on default inheritance logic, while the syntactic models generally avoid the use of defaults and overrides. An important proposal for marrying the two systems is given by Bonami & Samvelian (2015) in the context of analysing periphrastic constructions in Persian (see also Bonami & Webelhuth 2012). The details depend on the specifics of their

analysis, but the overall import of their proposal is a 'meta-constraint' on signs of type *word*, such that a word is licenced in the (HPSG) syntax only if a corresponding representation of it is also licensed in the (PFM) morphology (Bonami & Samvelian 2015: 32). In effect, they treat the PFM morphology as a 'black box' whose outputs bear properties that can be recognized by the syntax.

The interface for canonical inflection works well. However, the proposals do not touch directly on other types of morphology, notably derivation and transpositions. Presumably, the interface principle could be extended so as to apply between a morphological engine and the HPSG lexicon. A major problem here is the lack of consensus over how to handle derivational morphology in I-R models. In PFM there has been very little discussion of derivation and no discussion of transpositions.[6] Concrete proposals for derivation and transpositions can be found in the Network Morphology model of Brown & Hippisley (2012) but it is not clear how that model would interface with syntax. Moreover, it is not clear how the Network Morphology model distinguishes between transpositions and canonical derivation, and between these and the (non-canonical) phenomenon of transpositional lexemes.

A detailed set of proposals for defining lexical relatedness is given in Spencer (2013), where I show that there are many other types of relatedness between words in addition to canonical inflection, canonical derivation and true (canonical) transposition. Any model of the lexicon has to be able to account for all these types. They include meaning-changing inflection, meaning-changing transposition, derivation which involves no change at all in FORM properties (morphologically inert derivation) and others. The conceptual problem here is that any of these types of relatedness might be part of the paradigmatic grammatical system in a given language, in which case the morphological means by which they are all expressed cannot be distinguished. Therefore, the same kind of machinery has to be deployed for paradigmatic derivation as for inflection. Given our current assumptions this means some form of paradigm function, defined in terms of defaults and overrides, and the challenge is therefore to ensure that the lexical representations so defined are compatible with the kinds of representations deployed in the syntax.

7 An agenda for lexical representation

The foregoing discussion raises more question than it answers, but the questions are important for lexicalist, constraints-based models generally, and for theories of lexical representation and morphology generally. Here, by way of a conclusion I summarize the main issues that have emerged.

- Are lexemes partially specified linguistic objects?
- What is the relationship between transpositional quasi-lexemes and canonical lexemes?

[6] This includes Stump (2016a,b), which are concerned exclusively with FORM/CONTENT mismatches.

- How do we ensure that I-R morphological models can interface with constraints-based syntactic models, including all aspects of paradigmatically organized morphology?
- To what extent can the morphological functions/constructions proposed in Sag (2012) be retained in their current form? To what extent can such constructional types, or their more traditional incarnations in standard HPSG, be made compatible with I-R models?

Finally, the most difficult question of all is the oldest and the one with the widest significance: what kind of a thing is a dictionary entry? Is it a real, mentally represented linguistic construction or is it merely the convenient fiction of the lexicographer? We cannot address this question without providing very explicit answers to the representational and ontological questions raised in this paper, and so I present my discussion of those questions as a modest contribution towards answering the much bigger question.

Abbreviations

ARG-ST	Argument Structure (attribute)
FS	feature structure
GPFM	Generalized Paradigm Function Morphology
HPSG	Head-driven Phrase Structure Grammar
I-R	inferential-realizational (model)
LI	Lexical/Lexemic Index
LID	Lexical Identifier
PFM	Paradigm Function Morphology
SBCG	Sign-Based Construction Grammar
SF	semantic function (role)
wfr	Word Formation Rule

References

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar: Form and acquisition*, 54–82. Oxford: Oxford University Press.

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge: MIT Press.

Bochner, Harry. 1993. *Simplicity in generative morphology*. Berlin: Mouton de Gruyter.

Bonami, Olivier & Pollet Samvelian. 2015. The diversity of inflectional periphrasis in Persian. *Journal of Linguistics* 51(2). 327–382.

Bonami, Olivier & Gregory T. Stump. 2016. Paradigm Function Morphology. In Andrew Hippisley & Gregory T. Stump (eds.), *The Cambridge Handbook of Morphology*, 449–481. Cambridge: Cambridge University Press.

Bonami, Olivier & Gert Webelhuth. 2012. The phrase-structural diversity of periphrasis: A lexicalist account. In Marina Chumakina & Greville G. Corbett (eds.), *Periphrasis: The role of syntax and morphology in paradigms* (Proceedings of the British Academy 180), 141–167. Oxford: Oxford University Press/British Academy.

Brown, Dunstan, Marina Chumakina, Greville G. Corbett, Gergana Popova & Andrew Spencer. 2012. Defining 'periphrasis': Key notions. *Morphology* 22. 233–275.

Brown, Dunstan & Andrew Hippisley. 2012. *Network Morphology: A defaults based theory of word structure*. Cambridge: Cambridge University Press.

Chaves, Rui P. 2014. Grammatical alignments and the gradience of lexical categories. In Philip Hoffmeister & Elisabeth Norcliffe (eds.), *The Core and the Periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*, 167–220. Stanford: CSLI publications.

Culicover, Peter & Ray S. Jackendoff. 2005. *Simpler Syntax*. Oxford: Blackwell Publishers.

Dalrymple, Mary, Louise Mycock & John Lowe. 2015. Morphology in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG15 conference*, 43–62. Stanford: CSLI Publications.

Fradin, Bernard & Françoise Kerleroux. 2003. Troubles with lexemes. In Geert Booij, Janet DeCesaris, Angela Ralli & Sergio Scalise (eds.), *Selected papers from the third Mediterranean Morphology Meeting*, 177–196. Barcelona: IULA – Universitat Pompeu Fabra.

Haspelmath, Martin. 1996. Word-class-changing inflection and morphological theory. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1995*, 43–66. Dordrecht: Kluwer Academic Publishers.

Helimski, Eugene. 1998. Selkup. In Daniel Abondolo (ed.), *The Uralic Languages*, 548–579. London: Routledge.

Jackendoff, Ray S. 1975. Morphological and semantic regularities in the lexicon. *Language* 51(3). 639–671.

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

Jackendoff, Ray S. 1997. *The architecture of the language faculty*. Cambridge: The MIT Press.

Kuznecova, Ariadna Ivanovna, Evgenij Arnol'dovič Xelimskij & Elena Vladislavovna Gruškina. 1980. *Očerki po sel'kupskomu jazyku*. Moscow: Izdatel'stvo moskovskogo universiteta.

Malouf, Robert. 2000. *Mixed categories in the hierarchical lexicon*. Stanford: CSLI Publications.

Sadler, Louisa & Andrew Spencer. 2001. Syntax as an exponent of morphological features. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2000*, 71–96. Dordrecht: Kluwer Academic Publishers.

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

Spencer, Andrew. 1998. Relational adjectives and the redundancy of lexical categories. In Geert Booij, Angela Ralli & Sergio Scalise (eds.), *Proceedings of the First Mediterranean Conference of Morphology*, 19–29.

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

Spencer, Andrew. 2013. *Lexical relatedness: A paradigm-based model*. Oxford: Oxford University Press.

Spencer, Andrew. 2016. How are words related? In Daniel Siddiqi & Heidi Harley (eds.), *Morphological metatheory*, 1–26. Amsterdam: John Benjamins Publishing Co.

Spencer, Andrew. 2017. Split-morphology and lexicalist morphosyntax: The case of transpositions. In Claire Bowern, Laurence Horn & Raffaella Zanuttini (eds.), *On looking into words (and beyond)*, 385–421. Berlin: Language Science Press.

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 & Jaap van Marle (eds.), *Yearbook of morphology 2001*, 147–180. Dordrecht: Kluwer Academic Publishers.

Stump, Gregory T. 2006. A theory of heteroclite inflectional paradigms. *Language* 82. 279–322.

Stump, Gregory T. 2016a. *Inflectional paradigms: Content and form at the syntax-morphology interface*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2016b. Paradigms at the interface of inflectional morphology. In Daniel Siddiqi & Heidi Harley (eds.), *Morphological metatheory*, 27–58. Amsterdam: John Benjamins Publishing Co.