# Chapter 8

# Lexeme and flexeme in a formal theory of grammar

Olivier Bonami
Laboratoire de linguistique formelle, Université Paris Diderot

Berthold Crysmann
Laboratoire de linguistique formelle, CNRS

This paper deals with the role played by the notion of a lexeme in a constraint-based lexicalist theory of grammar such as Head-driven Phrase Structure Grammar. Adopting a Word and Paradigm view of inflection, we show how the distinction between lexemes, individuated by their lexical semantics, and flexemes, individuated by their inflectional paradigm, can fruitfully be integrated in such a framework. This allows us to present an integrated analysis of stem spaces, inflection classes, heteroclisis and overabundance.

It is often observed by morphologists that contemporary work in theoretical morphology has little impact on formal theories of grammar, which on average are content with a view of morphology quite close to that of offered by the post-Bloomfieldian morphemic toolkit. A notable exception to this situation is the pervasive use in Head-driven Phrase Structure Grammar (henceforth HPSG) of the distinction between WORDS and LEXEMES familiar from Word and Paradigm approaches to morphology (see among many others Robins 1959, Hockett 1967, Matthews 1972, Zwicky 1985, Anderson 1992, Aronoff 1994, Stump 2001, Blevins 2016). In this paper we reevaluate the role of the lexeme in HPSG in the light of 20 years of research, and in particular of recent attempts to integrate a truly realisational theory of inflection within the HPSG framework (Crysmann & Bonami 2016). We conclude that current theorizing conflates two distinct notions of an abstract lexical object: lexemes, which are characterised in terms of their syntax and semantics, and FLEXEMES (Fradin & Kerleroux 2003), which are characterised in terms of their inflectional paradigm. We propose distinct formal representations for lexemes and flexemes, and explore the benefits of the distinction for a formally explicit theory of morphology and the morphology-syntax interface.

The structure of the paper is as follows. In Section 1, we present the standard view of the lexeme in contemporary HPSG, and show that lexemes are given a dual representation, as a distinct type of signs and as the value of the feature LID. In Section 2, we present

Olivier Bonami & Berthold Crysmann. Lexeme and flexeme in a formal theory of grammar. In Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer (eds.), *The lexeme in descriptive and theoretical morphology*, 175–202. Berlin: Language Science Press. DOI:10.5281/zenodo.1407001

Information-based Morphology (IbM),[1] an HPSG-compatible realisational approach to inflection, and show that lexemes-as-signs have no role to play in an HPSG using IbM as its inflectional component. In Section 3 we discuss Fradin and Kerleroux's distinction between lexemes and flexemes, and argue that this should be encoded by distinguishing a feature LID and the values it can take from *pid* objects: while the former reside in syntactic/semantic representations, the latter are found in inflection proper. Finally in section 4 we discuss the consequences of the distinction between LID and *pid* for the modelling of heteroclisis and overabundance.

# 1 The lexeme in standard HPSG

## 1.1 Lexemes as a distinct type of lexical signs

Most current work in Head-driven Phrase Structure Grammar (henceforth HPSG; Pollard & Sag 1994, Ginzburg & Sag 2000, Sag et al. 2003) and its variant Sign-Based Construction Grammar (henceforth SBCG; Boas & Sag 2012) embraces the notion of a lexeme, familiar from Word-and-Paradigm approaches to morphology. Under this view, a lexeme is an abstract lexical object encapsulating what is common to the collection of words belonging to the same inflectional paradigm. Although the details are complex and disputed, it is uncontroversial enough to assume that a lexeme may be comprised of some amount of phonological information (in the form of a stem, a collection of stem alternants, a consonantal pattern, etc.), morphological information (e.g. inflection class information), syntactic information (at the very least part of speech and valence information), and semantic information corresponding to a notion of 'lexical meaning' (plus linking of semantic roles to syntactic dependents). Inflection is then concerned with the relation between (abstract) lexemes and (concrete) words,[2] while 'word formation', more adequately called LEXEME FORMATION (Aronoff 1994), is concerned with morphological relations between lexemes.

Since the late 1990s a growing consensus has emerged within HPSG that lexemes should be treated as signs on a par with words.[3] That is, the hierarchy of linguistic objects includes the subhierarchy in Figure 1. Syntactic rules may form phrases by combining signs of type *syn-sign*, while rules of morphology manipulate only signs of type *lex-sign*.

This is intended to implement the notion of strong lexicalism. First, words constitute the interface of morphology and syntax, since they belong to both types. Second, morphology and syntax are discrete components of grammar inasmuch as some aspects of

[1] The framework is presented and elaborated in Bonami & Crysmann (2013, 2016), Crysmann (2017), Crysmann & Bonami (2016). The name is intended as a reference to Pollard & Sag's (1987) *Information-based Syntax and Semantics*. In IbM, the notion of information in the sense of feature logic plays a central role in determining morphological wellformedness, defined in terms of exhaustive expression of morphosyntactic properties. Furthermore, IbM implements Paninian competition on the basis of subsumption, a measure of informativity in feature logic.

[2] Alternatively, within an *abstractive* conceptualisation of morphology (Blevins 2006), where words are seen as primitives rather than derived objects, inflection is concerned with the relation between words in a paradigm, and the abstract notion of a lexeme captures what is common between these words.

[3] See Bonami & Crysmann (2016) for a thorough overview of work on morphology in HPSG.

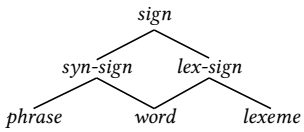


Figure 1: A standard HPSG subhierarchy of signs

the feature geometry of signs will be specific to phrases or lexemes; likewise, this architecture allows for the possibility that the kind of combinatory rules relating phrases to their component parts be very different from the kind of combinatory rules relating words to their component parts.

Although this is by no means an obligation, as we will see below, standard practice in HPSG and SBCG in the past two decades has been to assume an Item and Process view of morphology (Orgun 1996, Riehemann 1998, Koenig 1999, Müller 2002, Sag et al. 2003, Sag 2012), where the word-lexeme opposition captures the difference between inflection and lexeme formation. Rules of inflection map a lexeme to a word, rules of derivation map a lexeme to a lexeme, rules of composition map two lexemes to a lexeme. The three toy rules in Figures 2, 3 and 4 illustrate the basic architecture.

$$\left[\begin{array}{ll} \textit{word} & \\ \textsc{phon} & \boxed{1}\text{+/z/} \\ \textsc{synsem} & \boxed{2}\left[\begin{array}{ll} \textsc{head} & \textit{noun} \\ \textsc{num} & \textit{pl} \end{array}\right] \\ \textsc{m-dtrs} & \left\langle\left[\begin{array}{lll} \textit{lexeme} & & \\ \textsc{phon} & \boxed{1} & \\ \textsc{synsem} & \boxed{2} & \end{array}\right]\right\rangle \end{array}\right]$$

Figure 2: Simplified rule of regular English plural formation.

$$\left[\begin{array}{lll} \textit{lexeme} & & \\ \textsc{phon} & \boxed{1}\text{+/ɚ/} & \\ \textsc{synsem} & \left[\textsc{head} \quad \textit{noun}\right] & \\ \textsc{m-dtrs} & \left\langle\left[\begin{array}{lll} \textit{lexeme} & & \\ \textsc{phon} & \boxed{1} & \\ \textsc{ss|hd} & \textit{verb} & \end{array}\right]\right\rangle & \end{array}\right]$$

Figure 3: Simplified rule of English Agent noun formation.

Formally, morphological rules are modeled on a par with phrase-structure rules, except for the fact that, in inflectional and derivational rules, the relation between the phonology of the mother (the output lexical sign) and the phonology of the daughter (the input lexical sign) is specified syncategorematically: affixes are not signs, but bits

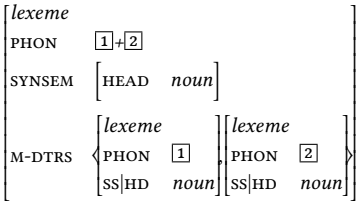


Figure 4: Simplified rule of English noun-noun compound formation

of phonology added by rule.[4] The main difference between inflection and lexeme formation rules lies in the fact that inflection does not modify the SYNSEM value, but merely expresses some of its aspects. The main specificity of composition is that the input (the daughter signs) consists of two lexemes rather than one. Figures 5 and 6 illustrate typical morphological analyses within such a framework.

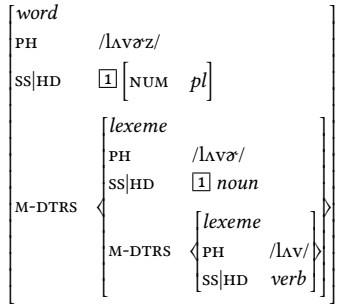


Figure 5: Analysis of the noun *lovers* under an Item-and-Process view of morphology

## 1.2 Lexeme identifiers

It is sometimes necessary for a lexical entry or syntactic construction to be able to select a particular lexical item in its environment. One clear case of this is that of flexible idioms. Consider the idiom *pull strings* 'try something'. As the examples in (1) make clear, while the idiomatic meaning is present only when the object of *pull* is headed by the lexeme *string*, the noun may occur in either singular or plural form, and combine with a variety of determiners and modifiers (Bargmann forthcoming).

(1) a. There I learned whom [*sic*] my secret advocate was, the man who had pulled strings to get me the teaching job in the midst of a terrible economy, and who

[4]For a dissenting view see Emerson & Copestake (2015).

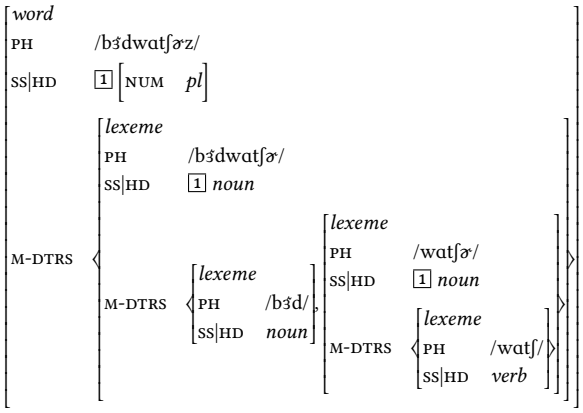


Figure 6: Analysis of the noun *birdwatchers* under an Item-and-Process view of morphology

had pulled more strings to allow me to keep it, and who had then pulled even more strings to have my commission assigned to the Abwehr.[5]

b. You'll never know the trouble I had, and the strings I had to pull to get you back from Berlin.[6]

c. We have to remember that Jacob was at their wedding. Just how many strings did he pull?[7]

d. So I didn't pull any string. Didn't need to.[8]

e. When I got the job, I thought to myself, "Someone upstairs finally pulled a string for me".[9]

f. No string was pulled, it was based on merit.[10]

This type of situation motivated the introduction of the feature LID (or LEXEME IDENTIFIER) as a head feature projecting to phrasal level information as to which lexeme heads a phrase (Sag 2007, 2012).[11] Simplifying matters considerably, one can see the constructions above as licensed by the two idiomatic lexical entries in Figure 8, which contrast with the two ordinary entries in Figure 7: a special lexical entry of *pull* with idiomatic meaning selects specifically for an object headed by a form of *strings* with idiomatic meaning. The postulation of a specific LID value for idiomatic *string* allows idiomatic *pull*

[5] K. Ryan, *The Somnambulist*, New York: iUniverse, 2006.
[6] K. McDermott, *The time of the corncrake*, Victoria: Trafford, 2004.
[7] http://www.losttv-forum.com/forum/showthread.php?t=65542. Accessed on November 26, 2016.
[8] http://www.justusboys.com/forum/archive/index.php/t-437037.html. Accessed on November 26, 2016.
[9] http://ultraphrenia.com/2016/10/02/a-cigarette-break-behind-heavens-gate. Accessed on November 13, 2016.
[10] http://obafemayor02.blogspot.fr/2013_03_24_archive.html. Accessed on November 26, 2016.
[11] Note that a very similar role is played by the feature LISTEME in Soehn (2006) and Richter & Sailer (2010).

to select for a specific combination of an inflectional paradigm with an idiomatic meaning, while abstracting away from inflectional and syntactic variability in the makeup of the object of *pull*.

$$\begin{bmatrix} \textit{lexeme} \\ \textsc{phon} & /\text{p}\upsilon\text{l}/ \\ \textsc{head} & \begin{bmatrix} \textsc{lid} & \textit{pull-lid} \end{bmatrix} \\ \textsc{val} & \langle \text{NP}_i, \text{NP}_j \rangle \\ \textsc{cont} & \begin{bmatrix} \textit{pull-rel} \\ \textsc{act} & i \\ \textsc{und} & j \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \textit{lexeme} \\ \textsc{phon} & /\text{st}\imath\text{ɪŋ}/ \\ \textsc{head} & \begin{bmatrix} \textsc{lid} & \textit{string-lid} \end{bmatrix} \\ \textsc{cont} & \textit{string-rel} \end{bmatrix}$$

Figure 7: Ordinary lexical entries for *pull* and *strings*

$$\begin{bmatrix} \textit{lexeme} \\ \textsc{phon} & /\text{p}\upsilon\text{l}/ \\ \textsc{head} & \begin{bmatrix} \textsc{lid} & \textit{idiomatic-pull-lid} \end{bmatrix} \\ \textsc{val} & \langle \text{NP}_i, \text{NP}_j [\textsc{lid}\ \textit{idiomatic-string-lid}] \rangle \\ \textsc{cont} & \begin{bmatrix} \textit{use-rel} \\ \textsc{act} & i \\ \textsc{und} & j \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \textit{lexeme} \\ \textsc{phon} & /\text{st}\imath\text{ɪŋ}/ \\ \textsc{head} & \begin{bmatrix} \textsc{lid} & \textit{idiomatic-string-lid} \end{bmatrix} \\ \textsc{cont} & \textit{influence-rel} \end{bmatrix}$$

Figure 8: Idiomatic lexical entries for *pull* and *strings*

The feature LID provides a useful mechanism for spreading lexical information in syntactic structures that has been used since in the analysis of complex predicates (Müller 2010) and periphrastic inflection (Bonami & Webelhuth 2012, Bonami & Samvelian 2015, Bonami 2015, Bonami et al. 2016). It also provides a direct encoding of lexemic identity. Since LID is a head feature, and inflected words share the HEAD value of the lexeme they are derived from, all inflected forms of a lexeme will have the same LID. Under the natural assumption that all lexemes have a distinct LID value, whether two words instantiate the same lexeme can thus be deduced by inspection of their LID values, without examining their derivation history.

# 2 The lexeme in a Word and Paradigm version of HPSG

## 2.1 Going Word and Paradigm

While an Item and Process view of morphology has been dominant in the HPSG literature, over the last 20 years a number of authors have become more vocal in advocating the incorporation into HPSG of a Word and Paradigm view of inflection (see among others Erjavec 1994, Miller & Sag 1997, Ackerman & Webelhuth 1998, Crysmann 2002,

Bonami & Boyé 2006, Bonami & Webelhuth 2012, Bonami 2015, Bonami & Samvelian 2015, Crysmann & Bonami 2016). Under such a view, rules of inflection do not incrementally specify how a basic sign is augmented with morphosyntactic information and phonological exponents; rather, a full morphosyntactic specification of the word is given as input to a system of rules of exponence indicating how such a specification is partially realised by exponents in various positions with respect to the basic stem. The arguments in favour of such a move are the usual ones (Matthews 1974, Zwicky 1985, Anderson 1992, Stump 2001, Brown & Evans 2012): systems of exponence depart too strongly from a one-to-one correspondence between form and content for the Item and Process view to make sense in the general case. We will not rehearse these arguments here, and simply make the sociological observation that Word and Paradigm approaches have over the last two decades become the *de facto* standard for theoretical and typological reasoning on inflection systems.

Recent attempts at implementing Word and Paradigm inflection in HPSG come in two flavors. One the one hand, Bonami & Webelhuth (2012), Bonami (2015), Bonami & Samvelian (2015) explicitly interface Paradigm Function Morphology (Stump 2001, 2016) with HPSG through a set of relational constraints. On the other hand, Crysmann & Bonami (2016) design a realisational framework for inflection native to the HPSG architecture, *Information-based Morphology* (IbM), making heavy use of the underspecification techniques provided by a typed feature structure formalism.

Figure 9 illustrates the main features of IbM by way of the analysis of a rather simple inflected word, the French verb *buvions* 'we drank'. IbM specifies the inflectional system of a language as a set of constraints relating a word's SYNSEM value to its PHONology. In the present example, a word realising the past imperfective of the verb BOIRE in the context of a 1PL subject is constrained to have the string /byvjɔ̃/ as its phonological realisation. The specification of these constraints makes use of three intermediate, strictly morphological, representations. The feature MS (standing for 'morphosyntactic properties') encodes those syntactic and semantic properties of the word that are relevant to inflection, in a format suitable for the expression of constraints on exponence. The feature MPH (standing for 'morphs') indicates the set of morphs making up the word, indexed for their position within the word (PC, standing for 'position class'). Finally, the feature RR (standing for 'realisation rules') indicates which generalisations on the relationship between morphosyntactic properties and morphs license the particular association between form and content instantiated in that word. Importantly, realisation rules relate a *set* of morphosyntactic properties (listed under MUD, standing for 'morphology under discussion') to a *set* of morphs (listed under MPH). Thus, while in this simple example, there is a one-to-one mapping between properties and morphs, IbM realisation rules can just as easily accommodate cumulative exponence (*m* properties : 1 morph), extended exponence (1 property : *n* morphs), overlapping exponence (*m* properties : *n* morphs), and zero exponence (*m* properties : 0 morphs).

The relationship between the various features is regulated by a set of general principles that we will only state in prose here; we refer the reader to Bonami & Crysmann (2013) or Crysmann & Bonami (2016) for a more explicit formulation. Let us start with the relation-

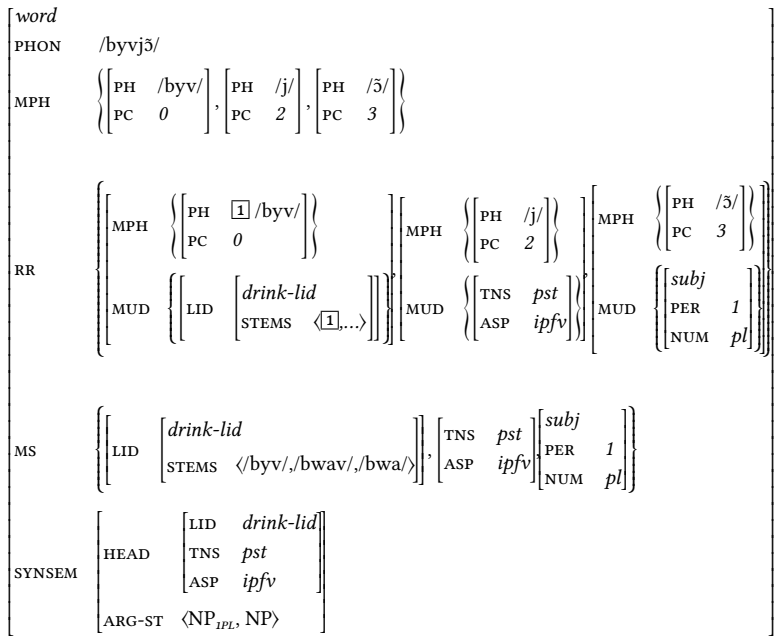


Figure 9: A sample IbM analysis: the French IPFV.2PL word *buvions* 'we drank'

ship between the SYNSEM and MS values. This is regulated by a set of language-specific constraints, since which aspects of syntax and semantics are realised by inflection is a highly parochial matter. Two features of this interface are worth noting. First, lexeme-specific information on inflection class and stem alternants is included in MS inside the *lid* value. In particular, a list-valued feature STEMS provides an indexed set of stem alternants, also known as a stem space (Bonami & Boyé 2006).[12] The choice of a particular stem is then effected by a realisation rule of *stem selection* (Stump 2001), picking out the appropriate value in this list, depending on the morphosyntactic context; In the present instance, the default of picking the first stem applies. In other words, in IbM, even the stem is taken to be the realisation of some word-level information, namely lexical identity. Second, MS values are relatively flat in comparison to SYNSEM values, consisting of a set of small feature structures, rather than one large, deeply recursive feature structure. This is necessitated by the different demands of morphological and syntactic combination.[13]

[12] Bonami & Boyé (2006) argue that the French stem space has 12 coordinates. for simplicity we show only 3 in the example in Figure 9.

[13] The distinction between SYNSEM and MORSYN may also be used to account for mismatches between content and form at the morphology-syntax interface, as variously captured in the literature by distinguishing syntactic and morphological features (Sadler & Spencer 2001, Corbett & Baerman 2006, Bonami 2015) or content and form paradigms (Stump 2006, 2016).

We may now turn to the relationship between MS and RR. This is regulated by a principle of morphological wellformedness: the MS value of a word must be identical to the disjoint union of the MUDs of the realisation rules. In other words, each morphosyntactic property must be realised by exactly one rule, although a single rule may realise multiple properties at once.[14]

Finally, the relation between RR, MPH and PHON is rather straightforward. First, the MPH value of a word is the union of the MPH values of its realisation rules: in other words, every morph must be licensed by at least one realisation rule, although a realisation rule may license more than one morph (extended or overlapping exponence), or even no morph at all (zero exponence). Second, a word's phonology is determined by appending the phonology of its morphs in accordance with the linear sequence of position class indices. Note that, although the system of position class indices encodes the notion of a morphotactic template, it does so with appropriate flexibility. There is no notion of an 'empty position' in the template: position class indices regulate the relative order of morphs, but morph ordering is not effected by putting bits of phonology in slots, just by appending bits of phonology in order. More importantly, realisation rules may partially underspecify the position they assign morphs to, allowing one to capture an unprecedented set of situations of variable morphotactics. Note also that, although a realisation rule may encode zero exponence, it is not equivalent to a zero morpheme: having no morph as one's exponent is not the same thing as having a morph with no phonological realisation. In particular, since no empty morphs are postulated, no sybilline decisions need to be taken as to the positioning of inaudible elements.

## 2.2 The role of the lexeme in IbM

Now that we have outlined the main features of IbM, let us consider the role of the lexeme in such a framework. Remember that in classical HPSG, inflection rules take the form of unary rules relating an abstract sign, the lexeme, to a surface sign, the inflected word. IbM has no use for such a notion of inflection rule, since inflection is stated directly as a relation between content and form at the word level. On the other hand, IbM makes crucial use of the notion of a lexeme identifier to state lexeme-specific phonological and morphological information; and the word/lexeme opposition is still a useful way of capturing the relationship between lexical entries and inflected words, and making a clear distinction between lexeme formation and inflection.

We thus assume that, while there are no inflectional lexical rules, there is a general constraint on objects of type *word* to the effect that they are the realisation of a lexeme, as indicated in (2). This constraint enforces the monotonic character of inflection: unlike derivation, inflection does not modify syntax or semantics but merely realises whatever features are made available by paradigm structure and compatible with the syntactic context. This is enforced by the identity of SYNSEM values at the *lexeme* and *word* levels.

[14] Implicit here are two assumptions familiar from Paradigm Function Morphology: (i) if two realisation rules are appropriate in some context, only the rule realising more content may apply (Panini's Principle); and (ii) there exists a universal rule of default non-realisation, ensuring that a property set remains unrealised if and only if the inflection system provides no other rule for its realisation.

$$(2)\quad word \rightarrow \begin{bmatrix} \textsc{synsem} & \boxed{1} \\ \textsc{m-dtrs} & \langle \begin{bmatrix} lexeme \\ \textsc{synsem} & \boxed{1} \end{bmatrix} \rangle \end{bmatrix}$$

As a consequence of (2), an inflected word will inherit any constraint imposed by the lexeme's lexical entry within SYNSEM, including, crucially, lexical identity and stem alternants as specified through the LID feature. Note that we assume the PHON attribute to be appropriate only for *syn-sign* objects (that is, words and phrases): lexemes constrain the phonology of their inflected forms through the STEMS feature instead (Bonami & Boyé 2006). The inflection-specific features MPH, RR and SC are appropriate for *word*s only. The format of lexical entries and lexeme formation rules is essentially unchanged.

# 3 Lexemes and flexemes

In this section we build on the general architecture just presented and argue that a distinction between two notions of lexical identity needs to be made.

## 3.1 Introducing the flexeme

Up to now, we have assumed a simple relationship between lexemes and inflectional paradigms: the value of the same feature LID is used for purposes of lexeme selection and for purposes of individuating inflectional paradigms. In doing so we have been following standard practice in realisational morphology, where paradigm functions take 'lexemes' (Stump 2001, 2016) or equivalently a 'lexemic index' (Spencer 2013) as an argument.

In an important but rarely cited paper, Fradin & Kerleroux (2003) note that matters are not so simple, for reasons having to do with lexical ambiguity and the division of labour between inflection and lexeme formation.[15] Rules of inflection are not generally concerned with matters of lexical ambiguity: from the point of view of inflection, the two French verbs DEVOIR$_1$ 'must' and DEVOIR$_2$ 'owe' are indistiguishable, as they have the same (highly irregular) inflectional paradigm. From the point of view of derivation, however, things are different. Derived lexemes normally relate to one sense of their base: for instance, while the French noun FILLE is ambiguous between two readings FILLE$_1$ 'girl' and FILLE$_2$ 'daughter', the diminutive FILLETTE 'small girl' only relates to the first.[16] Fradin & Kerleroux (2003) argue that this warrants a distinction between two kinds of abstract lexical objects: *lexemes* and *flexemes*. Inflection is about flexemes, while derivation is about lexemes. Because of the pervasive nature of lexical ambiguity, a single flexeme often corresponds to multiple lexemes.

[15] We purposefully use the general term 'lexical ambiguity' because whether the relevant examples are instances of polysemy or homonymy does not affect the argument.

[16] This very short summary does not do justice to Fradin and Kerleroux's insights, which build on an examination of the compatibility of various lexeme formation rules in French (Fradin & Kerleroux 2003) with various families of meanings. See also Fradin & Kerleroux (2009) for more discussion.

In the remainder we follow Walther (2013) in assuming that inflection is strictly concerned with flexemes, and propose an implementation of the lexeme-flexeme distinction in IbM.

## 3.2 LID and PID

Within an HPSG view of the world, it is tempting to capture the relationship between lexemes and flexemes in terms of underspecification in an inheritance hierarchy: flexemes would then be abstract groupings of lexemes. Suppose for concreteness a hierarchical organisation of LID values such as that indicated in Figure 10. Rules of inflection can then be stated in terms of the supertype *fille*, while lexemes are properly individuated in terms of the subtypes; and hence FILLETTE can be uniquely related to the lexeme whose LID value is $fille_1$.

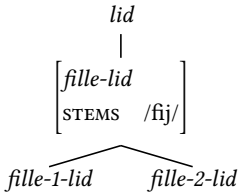


Figure 10: A first pass at flexemes in HPSG: flexemes as underspecified LID values

While this is technically feasible, such an approach only obscures the orthogonal roles played by the two notions. As illustrated above, IbM LID values are structured objects, which encompass all lexically-specified information relevant to inflection, including most notably stem alternants and inflection class. Such information is clearly irrelevant to syntax, although it is an indispensable component of inflection. On the other hand, studies that use LID for purposes of syntactic selection presuppose a tight correspondence between LID values and lexical semantic identity, and have no use for purely morphological information on stem alternants or inflection classes. In particular, Sag 2012 argues that LID values are to be identified with the main semantic predicate associated with a lexeme. One clear advantage of this convention is avoidance of redundancy in lexical entries: it is not necessary to postulate a new symbol as the LID value of each lexeme, since such a symbol is already present in the lexical entry as the constant designating the lexeme's main semantic predicate.

We now propose to clarify the situation by adopting Sag's view of LID. This entails that, for purposes of inflection, a separate index must be posited that individuates words according to which flexeme they instantiate. We call this index PID, standing for 'paradigm identifier'. While LID resides in HEAD and is thus available for selection in idioms, complex predicate constructions, or periphrastic constructions, PID is a top-level feature carried by signs of type *lexeme* only. As such it can be specified by lexical entries or manipulated by lexeme formation rules. In addition, it is universally constrained to be present among the features realised by inflection through inclusion in MS, as indicated in (3). This

is crucial to ensuring that inflection is always concerned with the realisation of lexical identity.

(3) $$word \rightarrow \begin{bmatrix} \textsc{ms} & \left\{ \boxed{1}, \ldots \right\} \\ \textsc{m-dtrs} & \left\langle \begin{bmatrix} \textit{lexeme} \\ \textsc{pid} \quad \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$

In this architecture then, lexical entries need to specify both an LID and a PID value. To elaborate on the same example, an appropriate analysis of FILLE would posit two lexical entries sharing the same PID object while having different LID values, as indicated in Figure 11.

$$\begin{bmatrix} \textit{lexeme} \\ \textsc{ss} \quad \begin{bmatrix} \textsc{cat|hd|lid} & \boxed{1}\ \textit{girl-rel} \\ \textsc{sem|restr} & \left\{ \boxed{1} \right\} \end{bmatrix} \\ \textsc{pid} \quad \textit{fille-pid} \end{bmatrix} \quad \begin{bmatrix} \textit{lexeme} \\ \textsc{ss} \quad \begin{bmatrix} \textsc{cat|hd|lid} & \boxed{1}\ \textit{daughter-rel} \\ \textsc{sem|restr} & \left\{ \boxed{1} \right\} \end{bmatrix} \\ \textsc{pid} \quad \textit{fille-pid} \end{bmatrix}$$

Figure 11: Proposed lexical entries for the two lexemes FILLE.

Under this analysis, the two lexemes FILLE are related by virtue of having indistinguishable PIDs, but they are still distinguishable in terms of LID. Hence, as indicated in the lexical entry in Figure 12, the derived noun FILLETTE adds diminutive semantics (*dim-rel*) to the semantics of its base which is constrained to be that lexeme with LID *girl-rel*, i.e., the left-hand lexeme in Figure 11. This captures the notion of formal lexical identity at the level of PID while implementing Fradin and Kerleroux's insight that derivational morphology operates on fully specific rather than underspecified lexemes.

$$\begin{bmatrix} \textit{lexeme} \\ \textsc{ss} \quad \begin{bmatrix} \textsc{cat|hd|lid} & \boxed{1} \begin{bmatrix} \textit{dim-rel} \\ \textsc{inst} \quad \boxed{x} \end{bmatrix} \\ \textsc{sem|restr} & \left\{ \boxed{1}, \boxed{2} \right\} \end{bmatrix} \\ \textsc{pid} \quad \textit{fillette-pid} \\ \textsc{m-dtrs} \quad \left\langle \begin{bmatrix} \textit{lexeme} \\ \textsc{ss} \quad \begin{bmatrix} \textsc{cat|hd|lid} & \boxed{2} \begin{bmatrix} \textit{girl-rel} \\ \textsc{inst} \quad \boxed{x} \end{bmatrix} \\ \textsc{sem|restr} & \left\{ \boxed{2} \right\} \end{bmatrix} \\ \textsc{pid} \quad \textit{fille-pid} \end{bmatrix} \right\rangle \end{bmatrix}$$

Figure 12: Proposed lexical entry for the lexeme FILLETTE 'small girl'.

## 3.3 Individuating flexemes: stem spaces

We now turn to the nature of *pid* objects. Evidently, there should be enough distinct PID values to be able to distinguish each flexeme from one another; that is necessary and sufficient to capture the notion of a flexeme. In the context of a typed-feature structure ontology, however, it is very natural to use PID to capture all aspects of inflectional identity. We thus take PIDs to be structured objects providing enough phonological and inflectional information to deduce a whole paradigm with minimal redundancy: Hence, at the very least, for the simplest inflectional systems, a basic stem. For systems of any complexity, this basic information needs to be supplemented with inflection class information (if there is more than one inflectional strategy) and information on stem alternants (if there are unpredictable stem alternations).

We illustrate a simple approach to the encoding of stem alternations by adapting the HPSG analysis of French conjugation presented in Bonami & Boyé (2006). French verbs exhibit pervasive stem alternations, illustrated in Table 1 in the indicative present subparadigms. Regular verbs from the first conjugation use a uniform stem in the present, and regular verbs from the second conjugation use an augmented stem in /-s/ in the plural. In addition to these two patterns, however, there are hundreds of irregular verbs instantiating others, which can be grouped into three types: either there is one stem for the singular and one for the plural, or the same stem is used for the singular and for the third plural, or three different stems are used following the pattern illustrated by BOIRE.[17]

Table 1: Sample French present indicative paradigms illustrating recurrent stem alternation patterns

| | 1SG | 2SG | 3SG | 1PL | 2PL | 3PL | |
|---|---|---|---|---|---|---|---|
| LAVER 'wash' | lav | lav | lav | lav-ɔ̃ | lav-e | lav | (1st conjugation) |
| FINIR 'finish' | fini | fini | fini | finis-ɔ̃ | finis-e | finis | (2nd conjugation) |
| ENVOYER 'send' | ɑ̃vwa | ɑ̃vwa | ɑ̃vwa | ɑ̃vwaj-ɔ̃ | ɑ̃vwaj-e | ɑ̃vwa | |
| JOINDRE 'join' | ʒwɛ̃ | ʒwɛ̃ | ʒwɛ̃ | ʒwaɲ-ɔ̃ | ʒwaɲ-e | ʒwaɲ | (other patterns) |
| BOIRE 'drink' | bwa | bwa | bwa | byv-ɔ̃ | byv-e | bwav | |

Given the pervasive nature of these alternations and the general unpredictability of the shapes of the alternants, Bonami & Boyé (2003a) build on previous work by Aronoff (1994), Brown (1998), Hippisley (1998), and Stump (2001), and posit that each lexeme is associated with a STEM SPACE, a vector of phonological shapes indicating the shape of the stem used in some zone of the paradigm. Limiting attention again to the stems found in the indicative present, the stem space of the verbs under consideration is indicated in Table 2: Stem 1 the default stem, Stem 2 is used in the 3PL, and Stem 3 is used in the singular.

[17] Bonami & Boyé (2006) deliberately set apart a handful of highly irregular and very frequent verbs instantiating an unpredictable form in the 1SG, 1PL or 2PL.

Table 2: Stem spaces for a sample of French verbs in the present indicative

| | Stem 1 | Stem 2 | Stem 3 |
|---|---|---|---|
| LAVER 'wash' | lav | lav | lav |
| FINIR 'finish' | finis | finis | fini |
| ENVOYER 'send' | ɑ̃vwaj | ɑ̃vwa | ɑ̃vwa |
| JOINDRE 'join' | ʒwaɲ | ʒwaɲ | ʒwɛ̃ |
| BOIRE 'drink' | byv | bwav | bwa |

In the context of an Item-and-Process view of inflection, Bonami & Boyé (2006) propose to encode stem spaces as the value of a feature carried by lexemes, and posit a hierarchy of stem space types capturing different patterns of identity among coordinates in the stem space. This analysis can be readily adapted to the current framework by assuming that stem spaces are represented inside *pid* objects using a list-valued feature STEMS. Let us first consider the lexical entry of BOIRE 'drink'. This needs to list three unpredictable stems, as indicated in Figure 13.

$$\left[\begin{array}{ll} \textit{lexeme} & \\ \textsc{ss}|\textsc{cat}|\textsc{hd} & \left[\begin{array}{ll} \textit{verb} & \\ \textsc{lid} & \textit{drink-rel} \end{array}\right] \\ \textsc{pid} & \left[\begin{array}{ll} \textit{boire-pid} & \\ \textsc{stems} & \langle /\text{byv}/, /\text{bwav}/, /\text{bwa}/ \rangle \end{array}\right] \end{array}\right]$$

Figure 13: Lexical entry for BOIRE 'drink'

The grammar then needs to specify in which context each element in STEMS is to be used. Following insights from Stump (2001: chap. 6), we assume that this is effected by STEM SELECTION RULES, a special kind of realisation rule that selects a stem alternant for insertion. The relevant rules are presented in Figure 14.[18]

The first rule states that, by default, lexical identity (i.e. PID) is realised by inserting the first element on the STEMS list as a morph in position 0.[19] The two other rules add some allomorphic conditioning: the second element is only used if the morphosyntactic context is that of a 3PL subject, while the third is used when it is that of a SG subject.

Note that the stem selection rules are in no way sensitive to inflection class. This is in keeping with Bonami and Boyé's (2003b, 2006) analysis, which starts from the assumption that all variation in French conjugation originates in differential distributions of

[18] We use the em dash ('—') to denote an unconstrained string of segments. '—' in a STEMS value thus indicates that the shape of that stem is not constrained by the rule, type, or lexical entry under consideration.

[19] This rule can be thought of as capturing an inflectional universal, as it simply states that some stem must be provided for every word. In systems without unpredictable stem allomorphy, this will be the sole element on the STEMS list. In systems with stem allomorphy, by convention, we place the default stem alternant first.

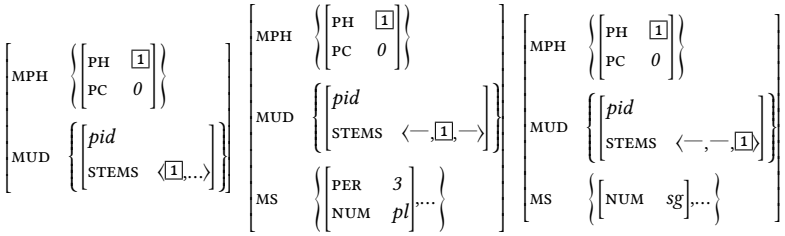


Figure 14: Stem selection rules for French present indicative

alternants in the stem space. That being said, it is useful to characterise classes of flexemes in terms of the patterns of identity they instantiate. In the present context, such a classification can be stated in the form of a type hierarchy of *pid* objects, as indicated in Figure 15.

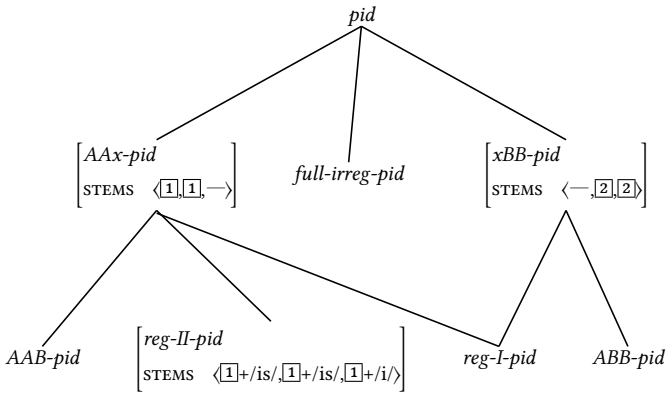


Figure 15: Hierarchy of *pid* subtypes capturing aspects of the French verbal stem space

The hierarchy of *pid* objects highlights the structure of the system, and allows the grammar writer to minimise redundancy in the stamement of lexical entries. In particular, all regular verbs can be described with mention of the first stem only, while different types of irregulars necessitate information on two or more stems in different coordinates of the stem space. More sample lexical entries are provided in Figure 16 for illustration. Note that the lexical entry for BOIRE of Figure 13 does not need to mention a subtype of *pid* explicitly, since *full-irreg-pid* is the only subtype compatible with the listing of three distinct stems.

Finally, the distinction between PID types and stem inventories provides a simple account of situations where two verbs belonging to different stem alternation types have the same basic stem, as is the case e.g. with TAPIR 'hide' and TAPISSER 'paper', wich have both have a basic stem /tapis/, witness the ambiguous PRS.1PL /tapisɔ̃/ 'we hide'/'we paper'. Figure 17 shows the relevant lexical entries.

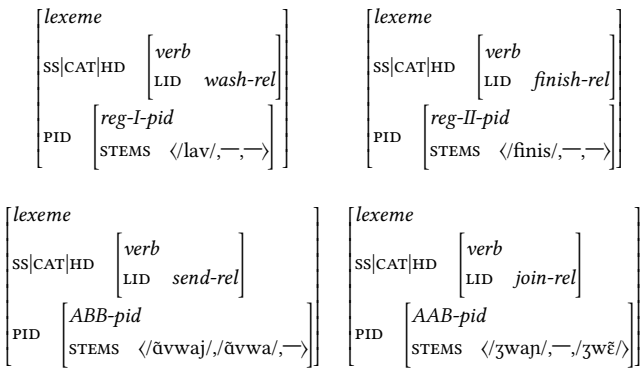


Figure 16: Lexical entries for a sample of French verbs

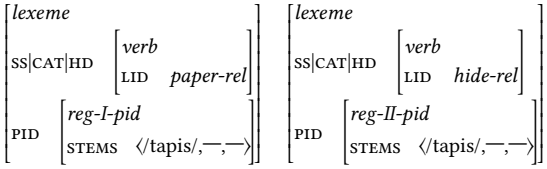


Figure 17: Lexical entries for two French verbs with homophonous basic stems

To sum up then, PID provides a natural locus for the representation of lexical information on stem alternations, and allows for a natural encoding of Bonami and Boyé's notion of a stem space. In addition, in a system where (by hypothesis) all variation in inflection is located in the stems, the indication of a specific vector of stem alternants is sufficient to fully individuate flexemes. In such a system, the hierarchy of *pid* values is merely used to limit the statement of redundant information in lexical entries.

## 3.4 Individuating flexemes: affixal inflection classes

We now turn to the role of PID in a system with nontrivial affixal inflection classes. As an illustration, let us examine a subset of the Czech nominal declension system. Table 3 provides partial paradigms for four nouns belonging to four of the major inflection classes of masculine inanimate and neuter nouns.

The distinction between hard and soft declension is correlated with the phonological properties of the stem-final consonant; however, it is not in general possible to categorically predict whether a noun will belong to a hard or soft declension on the basis of the phonological shape of its stem. Groups of declensions do share characteristics of exponence; in particular, it is evident from the table that some exponent strategies are common to the soft declensions (e.g. *-e* marking the GEN.SG), to the masculine declensions (e.g. *-ů* in the GEN.PL), or to larger groups of declensions (e.g. *-ům* is used in the DAT.PL accross the declensions shown here, except in the soft neuter). These observations

Table 3: Partial declension for the four inflection classes of Czech inanimate nouns

| | | MASCULINE | | NEUTER | |
|---|---|---|---|---|---|
| | | hard | soft | hard | soft |
| SG | NOM | most | pokoj | měst-o | moř-e |
| | GEN | most-u | pokoj-e | měst-a | moř-e |
| | DAT | most-u | pokoj-i | měst-u | moř-i |
| | ACC | most | pokoj | měst-o | moř-e |
| | VOC | most-e | pokoj-i | měst-o | moř-e |
| | LOC | most-ě | pokoj-i | měst-ě | moř-i |
| | INS | most-em | pokoj-em | měst-em | moř-em |
| PL | NOM | most-y | pokoj-e | měst-a | moř-e |
| | GEN | most-ů | pokoj-ů | měst | moř-í |
| | DAT | most-ům | pokoj-ům | měst-ům | moř-ím |
| | ACC | most-y | pokoj-e | měst-a | moř-e |
| | VOC | most-y | pokoj-e | měst-a | moř-e |
| | LOC | most-ech | pokoj-ích | měst-ech | moř-ích |
| | INS | most-y | pokoji | měst-y | moř-i |
| | | ‘bridge’ | ‘room’ | ‘town’ | ‘sea’ |

motivate arranging flexemes in a hierarchy of classes, so that the application of rules of exponence can be restricted to arbitrary collections of declension classes. We thus propose a simpler hierarchy of *pid* objects reflecting the distinction between hard and soft declensions, as indicated in Figure 18.

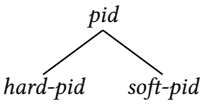


Figure 18: Premiminary hierarchy of *pid* subtypes for Czech declension

In addition, we propose that, since gender is inherent for nouns (in contrast to agreement gender) yet still conditions inflectional realisation, it should be represented as part of PID. Hence the lexical entries of the 4 nouns under consideration are as indicated in Figure 19. Note that traditional declensions correspond to a combination of a *pid* subtype and a gender value.[20]

[20]This bidimensional representation of declension classes is possible because gender is a strict predictor of inflection class in Czech: all members of each declension class belong to the same gender. Some declension classes corresponding to different genders are very similar, but always differ in at least one paradigm cell: e.g. masculine TÁTA ‘dad’ inflects like a feminine hard noun in only about half of its paradigm cells. Also note that a full description of the system would require more subtypes of *pid*, as there are more than two classes per gender, and hence organizing the *pid* hierarchy as a dense semi-lattice of inflection class groupings (Beniamine & Bonami 2016-09).

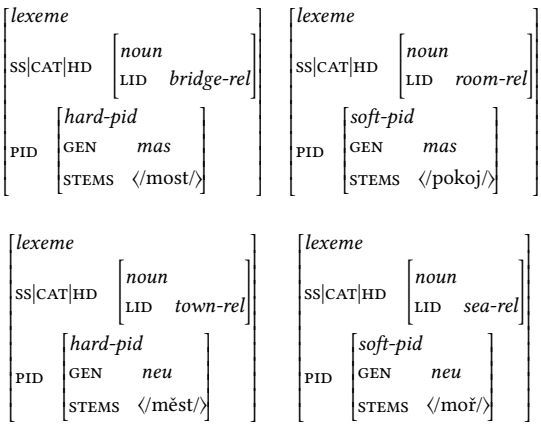


Figure 19: Preliminary lexical entries for a sample of Czech nouns

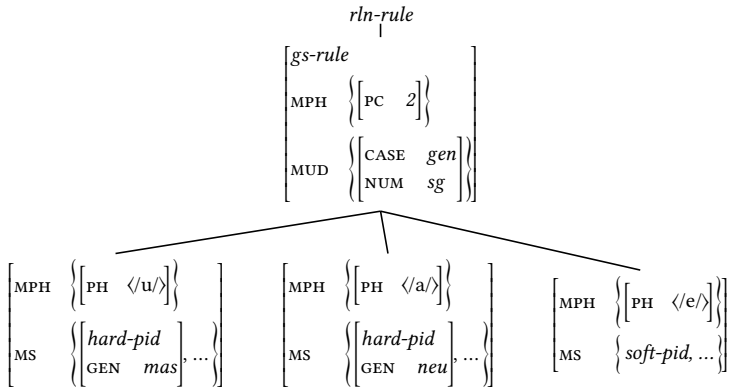


Figure 20: Preliminary realisation rules for Czech GEN.SG

To see how this hierarchy helps in capturing the distribution of exponents in Czech, consider the partial hierarchy of rules of exponence for the expression of GEN.SG in Figure 20. The three rules have the same general structure: they associate a specific phonological shape with the expression (through the MUD value) of the GEN.SG, but place a condition on that expression by restricting the MS value to contain specific information in its *pid* value. That is, they limit the use of an exponent to flexemes belonging to a particular inflection class or group of inflection classes. The first two rules express the conditioning in terms of both a type in the *pid* hierarchy and a gender value. The third one, however, does not mention gender, and hence can apply both in the case of masculine and neuter soft nouns.

This simple example illustrates how the typed feature structure architecture allows for a straightforward statement of generalisations on exponence across declension types by

locating inherent inflectional information in PID values and conditioning the application of rules of exponence to families of possible PID values.

We conclude this section by noting that the use of stem spaces, inherent features such as gender, and type of PID does not necessarily exhaust the inventory of relevant information that should be coded inside PID for the languages of the world. For instance, Bonami & Lacroix (2011) proposed that lexical information on thematic suffixes in the conjugation of the Kartvelian language Laz should be stored as the value of a dedicated feature inside the PID, since information on the shape of the thematic affix needs to be lexically stipulated but the affix is neither always present nor always contiguous to the root; and Crysmann & Bonami (2017) propose a concrete implementation of that idea in the context of Estonian declension. Our general claim is that PID should be the sole locus of lexically stipulated information on inflection.

## 4 Flexemes and overabundance

In previous sections we have justified the distinction between lexemes and flexemes by arguing that a single flexeme (characterised by a single inflectional paradigm) may correspond to multiple lexemes (characterised by different lexical semantic and/or syntactic properties). In this final section we explore situations where one may want to argue the opposite: multiple flexemes corresponding to a single lexeme.

Although we have not made use of it yet, the analytic scheme defined in the previous section certainly leaves room for such a possibility. Both for French verbs and Czech nouns, we have proposed that *pid* objects be organised in a hierarchy, capturing families of inflectional behavior. The lexical entries used thus far all introduce a PID value corresponding to a specific leaf type in the hierarchy: hence one flexeme for each lexeme. However, if some lexical entries were to refer to some *pid* supertype, this would authorise multiple inflectional behaviours for the same lexeme – hence, in a sense, multiple flexemes for one lexeme.

As a matter of fact, both French conjugation and Czech declension provide examples of phenomena that are insightfully analysed in this fashion. The phenomena at hand fall under the general heading of OVERABUNDANCE (Thornton 2011, 2012, to appear), that is, of situations where a single lexeme has multiple realisations for the same set of morphosyntactic properties.

First consider the French verb ASSEOIR. There is considerable variation in the realisation of different paradigm cells of this verb, leading to free variation at least for some paradigm cells in some varieties (Bonami & Boyé 2010). Limiting ourselves again to the indicative present, there seem to be two equally felicitous forms for each person-number combination in Standard French, as indicated in Table 4.

Although this situation could be described in terms of overabundance in individual paradigm cells, such an approach would not capture the fact that the forms seem to be organised in two distinct paradigms, each with two stem alternants, and each instantiating a different but familiar pattern of stem allomorphy: the /aswa/ /aswaj/ contrast follows an ABB pattern similar to that of ENVOYER (see Table 1), while the /asje/ /asɛj/

Table 4: The two main indicative present subparadigm of ASSEOIR 'sit'

| 1SG | 2SG | 3SG | 1PL | 2PL | 3PL |
|---|---|---|---|---|---|
| aswa<br>asje | aswa<br>asje | aswa<br>asje | aswaj-ɔ̃<br>asɛj-ɔ̃ | aswaj-e<br>asɛj-e | aswa<br>asɛj |

contrast follows an AAB pattern similar to that of JOINDRE. It is thus more perspicuous to describe this case of overabundance as involving two different stem spaces, and hence two different PID values, rather than variation in individual paradigm cells. Figure 21 shows two appropriate lexical entries corresponding to the two paradigms of ASSEOIR that readily integrate with the analysis presented in Section 3 and account for overabundance directly.

$$\left[\begin{array}{l} \textit{lexeme} \\ \textsc{ss|cat|hd}\ \left[\begin{array}{ll} \textit{verb} & \\ \textsc{lid} & \textit{sit-rel} \end{array}\right] \\ \textsc{pid}\ \left[\begin{array}{ll} \textit{AAB-pid} & \\ \textsc{stems} & \langle /\text{asej}/, —, /\text{asje}/ \rangle \end{array}\right] \end{array}\right] \quad \left[\begin{array}{l} \textit{lexeme} \\ \textsc{ss|cat|hd}\ \left[\begin{array}{ll} \textit{verb} & \\ \textsc{lid} & \textit{sit-rel} \end{array}\right] \\ \textsc{pid}\ \left[\begin{array}{ll} \textit{ABB-pid} & \\ \textsc{stems} & \langle /\text{aswaj}/, /\text{aswa}/, — \rangle \end{array}\right] \end{array}\right]$$

Figure 21: Lexical entries for two variants of the verb ASSEOIR 'sit'

The French verb ASSEOIR exemplifies a case of stem-based overabundance, which is readily accommodated by having two stem spaces for a single lexeme. Let us now turn to Czech and discuss a situation of exponent-based overabundance.

In Section 3.4 we discussed the fact that the Czech inflection system distinguishes 'hard' and 'soft' declensions. As it happens, some lexemes follow a hybrid or 'mixed' pattern that does not clearly fall into one type or the other, but rather makes use of both hard and soft exponents. However, this has different manifestations for neuter and masculine inanimate nouns, as evidenced by the examples in Table 5.

The paradigm of the mixed neuter noun KUŘE 'chicken' exhibits HETEROCLISIS (Stump 2006): KUŘE inflects like a soft noun in the singular, but like a hard noun in the plural. By contrast, the paradigm of the mixed masculine noun PRAMEN 'spring' exhibits a combination of heteroclisis and partial overabundance. In the plural, PRAMEN inflects like a hard noun; in the singular, it may inflect either like a hard noun or like a soft noun. Correctly capturing the difference between these two types of mixed inflectional behaviour is a serious challenge for any theory of inflection.

Both behaviours are readily accomodated in the present framework, using a more refined hierarchy of PID values. The crucial insight is that overabundance amounts to ambiguity, i.e. disjunctive membership of two inflection classes, whereas heteroclisis involves simultaneous membership of two classes: while the former is modelled straightforwardly by means of underspecification, corresponding to the JOIN in the semi-lattice

Table 5: Overabundance and Heteroclisis in Czech declension

| | | MASCULINE | | | NEUTER | | |
|---|---|---|---|---|---|---|---|
| | | hard | mixed | soft | hard | mixed | soft |
| SG | NOM | most | pramen | pokoj | měst-o | kuř-e | moř-e |
| | GEN | most-u | pramen-u~pramen-e | pokoj-e | měst-a | kuř-et-e | moř-e |
| | DAT | most-u | pramen-u~pramen-i | pokoj-i | měst-u | kuř-et-i | moř-i |
| | ACC | most | pramen | pokoj | měst-o | kuř-e | moř-e |
| | VOC | most-e | pramen-e~pramen-i | pokoj-i | měst-o | kuř-e | moř-e |
| | LOC | most-ě | pramen-u~pramen-i | pokoj-i | měst-ě | kuř-et-i | moř-i |
| | INS | most-em | pramen-em | pokoj-em | měst-em | kuř-et-em | moř-em |
| PL | NOM | most-y | pramen-y | pokoj-e | měst-a | kuř-at-a | moř-e |
| | GEN | most-ů | pramen-ů | pokoj-ů | měst | kuř-at | moř-í |
| | DAT | most-ům | pramen-ům | pokoj-ům | měst-ům | kuř-at-ům | moř-ím |
| | ACC | most-y | pramen-y | pokoj-e | měst-a | kuř-at-a | moř-e |
| | VOC | most-y | pramen-y | pokoj-e | měst-a | kuř-at-a | moř-e |
| | LOC | most-ech | pramen-ech | pokoj-ích | měst-ech | kuř-at-ech | moř-ích |
| | INS | most-y | pramen-y | pokoji | měst-y | kuř-at-y | moř-i |
| | | ‘bridge’ | ‘spring’ | ‘room’ | ‘town’ | ‘chicken’ | ‘sea’ |

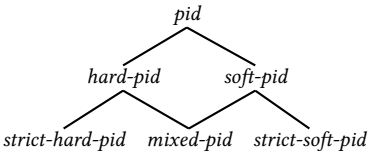


Figure 22: Improved hierarchy of *pid* subtypes capturing heteroclite Czech declension classes

of *pid* types, the latter can be captured by overspecification, i.e. the MEET, as shown by the type hierarchy in Figure 22.

Figure 23 shows schematically to which PID value each noun is assigned, and Figure 24 which PID value rules of exponence for the GEN.SG (left hand side) and NOM.PL (right hand side) are restricted to. More detailed lexical entries and rules of exponence are presented below in Figures 25 and 26. Any noun can be inflected using a realisation rule declared with a compatible PID value. That is, any point in the hierarchy that is identical to that of the noun, dominates it, or is dominated by it.

As shown in Figure 23, nouns belonging to non-mixed declensions are assigned to either of the two simple leaf types *strict-hard-pid* (MOST, MĚSTO) and *strict-soft-pid* (POKOJ, MOŘE). The heteroclite noun KUŘE is assigned to *mixed-pid*, and hence may inflect using either *hard* or *soft* exponents, but not *strict-hard* or *strict-soft* ones. The assignment of exponents to *pid* values (shown in Figure 24) ensures that it must use soft exponents in the singular, yet hard exponents in the plural. By contrast, the overabundant noun PRAMEN is assigned to an underspecified inflection class, namely *hard-pid*. As such it may

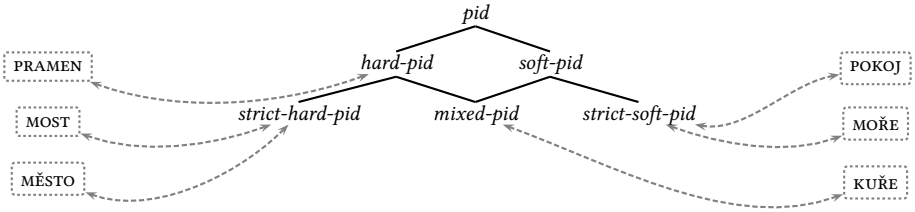


Figure 23: Schematic representation of inflection class assignment for Czech nouns

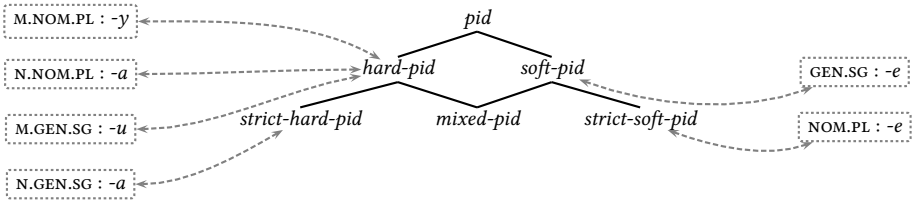


Figure 24: Schematic representation of the scope of rules of exponence for Czech nouns

$$\begin{bmatrix} \textit{lexeme} \\ \textsc{ss|cat|hd} & \begin{bmatrix} \textit{noun} \\ \textsc{lid} & \textit{bridge-rel} \end{bmatrix} \\ \textsc{pid} & \begin{bmatrix} \textit{strict-hard-pid} \\ \textsc{gen} & \textit{mas} \\ \textsc{stems} & \langle /\text{most}/ \rangle \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \textit{lexeme} \\ \textsc{ss|cat|hd} & \begin{bmatrix} \textit{noun} \\ \textsc{lid} & \textit{spring-rel} \end{bmatrix} \\ \textsc{pid} & \begin{bmatrix} \textit{hard-pid} \\ \textsc{gen} & \textit{mas} \\ \textsc{stems} & \langle /\text{pramen}/ \rangle \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \textit{lexeme} \\ \textsc{ss|cat|hd} & \begin{bmatrix} \textit{noun} \\ \textsc{lid} & \textit{room-rel} \end{bmatrix} \\ \textsc{pid} & \begin{bmatrix} \textit{strict-soft-pid} \\ \textsc{gen} & \textit{mas} \\ \textsc{stems} & \langle /\text{pokoj}/ \rangle \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} \textit{lexeme} \\ \textsc{ss|cat|hd} & \begin{bmatrix} \textit{noun} \\ \textsc{lid} & \textit{city-rel} \end{bmatrix} \\ \textsc{pid} & \begin{bmatrix} \textit{strict-hard-pid} \\ \textsc{gen} & \textit{neu} \\ \textsc{stems} & \langle /\text{město}/ \rangle \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \textit{lexeme} \\ \textsc{ss|cat|hd} & \begin{bmatrix} \textit{noun} \\ \textsc{lid} & \textit{chicken-rel} \end{bmatrix} \\ \textsc{pid} & \begin{bmatrix} \textit{mixed-pid} \\ \textsc{gen} & \textit{neu} \\ \textsc{stems} & \langle /\text{kuře}/ \rangle \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \textit{lexeme} \\ \textsc{ss|cat|hd} & \begin{bmatrix} \textit{noun} \\ \textsc{lid} & \textit{sea-rel} \end{bmatrix} \\ \textsc{pid} & \begin{bmatrix} \textit{soft-pid} \\ \textsc{gen} & \textit{mas} \\ \textsc{stems} & \langle /\text{moře}/ \rangle \end{bmatrix} \end{bmatrix}$$

Figure 25: Lexical entries for six Czech nouns

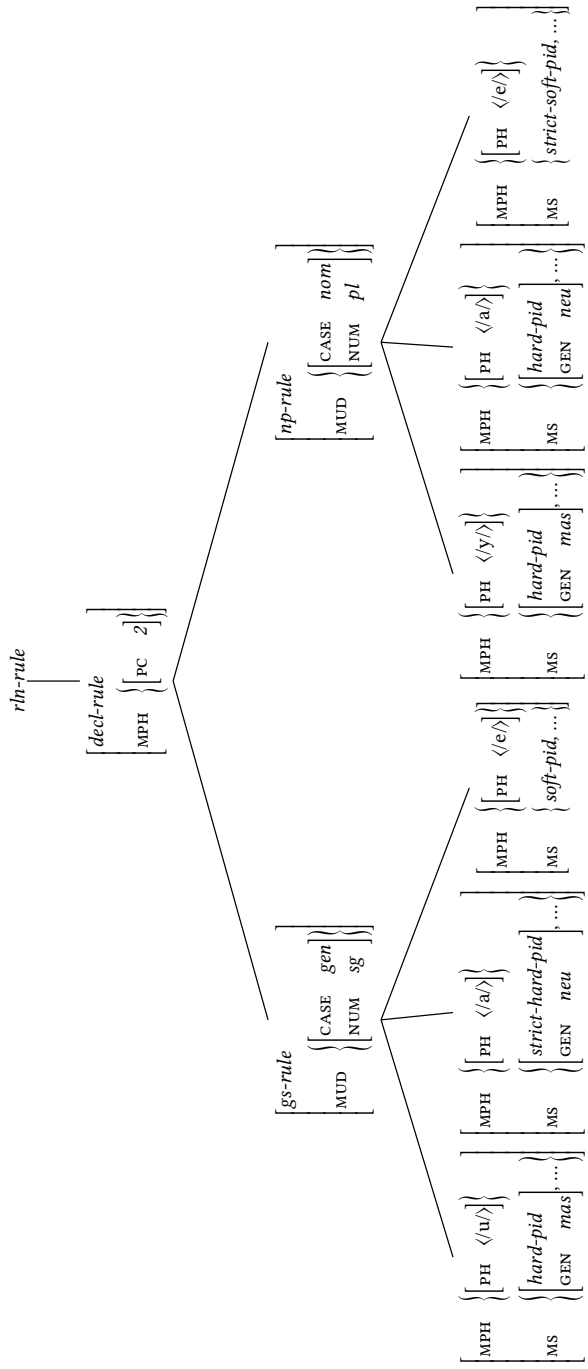


Figure 26: Realisation rules for Czech GEN.SG and NOM.PL nouns

use any one of *hard-pid*, *strict-hard-pid*, *mixed-pid*, or *soft-pid* exponents, but, crucially, not *strict-soft* exponents. This accounts pretty concisely for its contrasting behaviour in the singular and the plural: since the GEN.SG exponent *-e* is only *soft-pid*, there are two GEN.SG exponents available for PRAMEN, which is thus overabundant: inflection with *-e* by resolving *soft-pid* demanded by the rule and *hard-pid* demanded by PRAMEN to the hetoroclite type *mixed-pid*, or else with *-u*, by the sheer fact that this is the exponent available for all *hard-pid* words, whether strict or heteroclite. By contrast, the NOM.PL exponent *-e* is constrained to *strict-soft*. As such, it is inaccessible to PRAMEN, which hence behaves like a simple hard masculine noun in the plural.

We have thus established that mixed overabundant declensions can be accommodated by assigning a lexeme to a supertype in the *pid* hierarchy, while mixed heteroclite declensions can be accommodated by introducing a subtype intermediate between the hard and soft declensions.

The discussion in this section has exhibited the benefits of associating multiple *pid* objects with a single LID value to address some situations of overabundance; which amounts to positing that a single lexeme may correspond to multiple flexemes. We by no means claim that all overabundance phenomena are best thought of in such terms; See Thornton (this volume) for relevant discussion. Rather, we suggest that, where overabundance results from a lexeme being ambiguous between two classes of paradigms, lexically underspecified *pid*s make good sense of the situation.

## 5 Conclusions

In this paper we have addressed the representation of lexical identity in morphology. Following Fradin & Kerleroux (2003), we have argued that a distinction should be made between lexemes, individuated in terms of lexical semantics, and FLEXEMES, individuated in terms of inflectional paradigms. We have then shown that lexemes and flexemes stand in a many-to-many relation: in cases of lexical ambiguity, one flexeme realises multiple lexemes; in at least some situations of overabundance, multiple flexemes realise the same lexeme. We have shown how this distinction can be integrated into Information-based Morphology by providing words with two independent indices: LID and PID.

The distinction between LID and PID clarifies the role of lexical identity at the interface between inflectional morphology and syntax: syntax cares about lexemes, but not flexemes; inflectional morphology cares about flexemes, but not about lexemes. In the present framework, this is captured by the fact that LID is not represented in MS, the input to rules of inflection. Arguably, the distinction is also useful to clarify the role of lexical identity in lexeme formation. Recent work on French lexeme formation has highlighted the many-to-many nature of lexeme formation rules (see Bonami & Crysmann 2016: §3.1 and references cited therein): typically, a single formal process may be associated with multiple meanings, and the same type of meaning may be realised by multiple processes. Bonami & Tribout (2012) and Tribout & Bonami (2014-07) explore how the LID/*pid* can be used to make sense of that distinction. In their analytic scheme, lexeme formation rules are organised in a bidimensional multiple inheritance hierarchy, with

one dimension laying out formal strategies, and the other dimension describing a syntactic/semantic operation. Formal strategies determine a new *pid* from that of the base, while syntactic/semantic operations amount to constructing a new LID from that of the base.

More work is needed to integrate Bonami and Tribout's insights into IbM, but this integration paves the way towards a general, underspecification-based framework for morphological analysis.

## Acknowledgments

We thank Gilles Boyé and Jana Strnadová for their comments. This work was partially supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-LABX-0083).

## References

Ackerman, Farrell & Gert Webelhuth. 1998. *A theory of predicates*. Stanford: CSLI Publications.

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge: MIT Press.

Bargmann, Sascha. forthcoming. *A semantic representation account of English VP-idioms*. Goethe-Universität Frankfurt am Main dissertation.

Beniamine, Sacha & Olivier Bonami. 2016-09. A comprehensive view on inflectional classification. In *Annual meeting of the linguistic association of Great Britain*. Paris.

Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42(3). 531–573.

Blevins, James P. 2016. *Word and paradigm morphology*. Oxford: Oxford University Press.

Boas, Hans & Ivan A. Sag (eds.). 2012. *Sign-based construction grammar*. Stanford: CSLI Publications.

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

Bonami, Olivier, Robert D. Borsley & Maggie Tallerman. 2016. On pseudo-non-finite clauses in Welsh. In Doug Arnold, Miriam Butt, Berthold Crysamnn, Tracy Holloway King & Stefan Muller (eds.), *Proceedings of the joint 2016 conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, 104–124.

Bonami, Olivier & Gilles Boyé. 2003a. Supplétion et classes flexionnelles dans la conjugaison du français. *Langages* 152. 102–126.

Bonami, Olivier & Gilles Boyé. 2003b. Supplétion et classes flexionnelles dans la conjugaison du français. *Langages* 152. 102–126.

Bonami, Olivier & Gilles Boyé. 2006. Deriving inflectional irregularity. In *Proceedings of the 13th international conference on HPSG*, 39–59. Stanford: CSLI Publications.

Bonami, Olivier & Gilles Boyé. 2010. La morphologie flexionnelle est-elle une fonction ? In Injoo Choi-Jonin, Marc Duval & Olivier Soutet (eds.), *Typologie et comparatisme, hommage offert à Alain Lemaréchal*, 21–35. Leuven: Peeters.

Bonami, Olivier & Berthold Crysmann. 2013. Morphotactics in an information-based model of realisational morphology. In Stefan Müller (ed.), *Proceedings of HPSG 2013*, 27–47. Stanford: CSLI Publications.

Bonami, Olivier & Berthold Crysmann. 2016. The role of morphology in constraint-based lexicalist grammars. In Andrew Hippisley & Gregory T. Stump (eds.), *The Cambridge handbook of morphology*, 609–656. Cambridge: Cambridge University Press.

Bonami, Olivier & René Lacroix. 2011. Unblurring the inflection/derivation divide in Laz. Paper presented at the 8th Mediterranean Morphology Meeting, Cagliari.

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

Bonami, Olivier & Delphine Tribout. 2012. Underspecification and the semantics of lexeme formation. Vienna. http://www.llf.cnrs.fr/sites/llf.cnrs.fr/files/biblio//diapos-vienne-v5.pdf, accessed 2016-11-06. Paper presented at the 15th International Morphology Meeting.

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. 1998. Stem lndexing and morphonological selection in the Russian verb: A network morphology account. In R. Fabri, A. Ortmann & T. Parodi (eds.), *Models of inflection*, 196–224. Tübingen: Niemeyer.

Brown, Dunstan & Roger Evans. 2012. Morphological complexity and unsupervised learning: Validating Russian inflectional classes using high frequency data. In F. Kiefer, M. Ladányi & P. Siptár (eds.), *Current issues in morphological theory: (Ir)regularity, analogy and frequency*, 135–162. Amsterdam: John Benjamins.

Corbett, Greville G. & Matthew Baerman. 2006. Prolegomena to a typology of morphological features. *Morphology* 16. 231–246.

Crysmann, Berthold. 2002. *Constraint-based coanalysis*. Universität des Saarlandes dissertation.

Crysmann, Berthold. 2017. Inferential-realisational morphology without rule blocks: An information-based approach. In Andrew Hippisley & Nik Gisborne (eds.), *Defaults in morphological theory*, 182–213. Oxford: Oxford University Press.

Crysmann, Berthold & Olivier Bonami. 2016. Variable morphotactics in Information-based Morphology. *Journal of Linguistics* 52(2). 311–374.

Crysmann, Berthold & Olivier Bonami. 2017. Atomistic and holistic exponence in Underspecified Realisational Morphology. In *Proceedings of the 24th international conference on HPSG*.

Emerson, Guy & Ann Copestake. 2015. Lacking integrity: HPSG as a morphosyntactic theory. In Stefan Müller (ed.), *Proceedings of the 22nd international conference on Head-driven Phrase Structure Grammar*, 75–95. Stanford: CSLI Publications.

Erjavec, Tomaž. 1994. Formalizing realizational morphology in typed feature structures. In Gosse Bouma & Gertjan van Noord (eds.), *Proceedings of the fourth CLIN meeting*, 47–58. Groningen.

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.

Fradin, Bernard & Françoise Kerleroux. 2009. L'identité lexémique. In Bernard Fradin, Françoise Kerleroux & Marc Plénat (eds.), *Aperçus de morphologie du français*, 83–102. Saint-Denis: Presses Universitaires de Vincennes.

Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative investigations. The form, meaning, and use of English interrogatives*. Stanford: CSLI Publications.

Hippisley, Andrew. 1998. Indexed stems and Russian word formation: A Network Morphology account of Russian personal nouns. *Linguistics* 36. 1039–1124.

Hockett, Charles F. 1967. The Yawelmani basic verb. *Language* 43. 208–222.

Koenig, Jean-Pierre. 1999. *Lexical relations*. Stanford: CSLI Publications.

Matthews, P. H. 1972. *Inflectional morphology. A theoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.

Matthews, P. H. 1974. *Morphology*. Cambridge: Cambridge University Press.

Miller, Philip & Ivan A. Sag. 1997. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory* 15. 573–639.

Müller, Stefan. 2002. *Complex predicates: Verbal complexes, resultative constructions, and particle verbs in German*. Stanford: CSLI Publications.

Müller, Stefan. 2010. Persian complex predicates and the limits of inheritance-based analyses. *Journal of Linguistics* 46. 601–655.

Orgun, Cemil Orham. 1996. *Sign-based morphology and phonology*. University of California, Berkeley dissertation.

Pollard, Carl & Ivan A. Sag. 1987. *Information-based syntax and semantics*. Stanford: CSLI Publications.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Richter, Frank & Manfred Sailer. 2010. Phraseological clauses in constructional HPSG. In *Proceedings of HPSG 2009*, 297–317. Stanford: CSLI Publications.

Riehemann, Susanne Z. 1998. Type-based derivational morphology. *Journal of Comparative Germanic Linguistics* 2. 49–77.

Robins, R. H. 1959. In defense of WP. *Transactions of the Philological Society*. 116–144.

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. 2007. Remarks on locality. In Stefan Müller (ed.), *Proceedings of HPSG 2007*, 394–414. Stanford: CSLI Publications.

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.

Sag, Ivan A., Thomas Wasow & Emily Bender. 2003. *Syntactic theory: A formal introduction*. 2nd edition. Stanford: CSLI Publications.

Soehn, Jan-Philipp. 2006. *Über Bärendienste und erstaunte Bauklötze – Idiome ohne freie Lesart in der HPSG*. Frankfurt: Peter Lang.

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

Stump, Gregory T. 2001. *Inflectional morphology. A theory of paradigm structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2006. Heteroclisis and paradigm linkage. *Language* 82. 279–322.

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

Thornton, Anna M. 2011. Overabundance (multiple forms realizing the same cell): A non-canonical phenomenon in Italian verb morphology. In Martin Maiden, John Charles Smith, Maria Goldbach & Marc-Olivier Hinzelin (eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology*, 358–381. Oxford: Oxford University Press.

Thornton, Anna M. 2012. Reduction and maintenance of overabundance. A case study on Italian verb paradigms. *Word Structure* 5(2). 183–207.

Thornton, Anna M. to appear. Overabundance: A canonical typology. In Francesco Gardani, Hans Christian Luschützky & Franz Rainer (eds.), *Competition in morphology*. Berlin: Springer. U. L'Acquila.

Tribout, Delphine & Olivier Bonami. 2014-07. The polysemy of lexeme formation rules. In *Semantics of derivational morphology: Empirical evidence and theoretical modeling*. Düsseldorf.

Walther, Géraldine. 2013. *De la canonicité en morphologie: perspective empirique, théorique et computationnelle*. Université Paris Diderot dissertation.

Zwicky, Arnold M. 1985. How to describe inflection. In *Proceedings of the 11th meeting of the Berkeley Linguistics Society*, 372–386.