Deriving Inflectional Irregularity

Olivier Bonami

Université Paris-Sorbonne & Laboratoire de Linguistique Formelle (CNRS UMR 7110)

Gilles Boyé

Université Bordeaux 3 & ERSS (CNRS UMR 5610)

Proceedings of the HPSG06 Conference

Linguistic Modelling Laboratory
Institute for Parallel Processing
Bulgarian Academy of Sciences
Sofia
Held In Varna

Stefan Müller (Editor)

2006

CSLI Publications

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

Abstract

Conventional wisdom holds that productive morphology is regular morphology. Drawing evidence from French, we argue that the description of many lexeme formation processes is simplified if we hold that a productive rule may give rise to inflectionally irregular lexemes. We argue that the notion of a *stem space* allows for a straightforward description of this phenomenon: each lexeme comes equipped with a vector of possibly distinct stems, which serve as bases for inflectional form construction. The stem space is structured by default relations which encode the regular pattern of inflection; (partial) irregularities occur when a lexeme specifies a stem space violating the default relations. Derived irregularity is then the effect of a productive lexeme formation rule which specifies an irregular stem space for its output.

1 Productive irregular inflection

1.1 Background

A central issue in the modeling of inflection is how the notion of (ir)regular inflection is taken into account. A distinct possibility is to give the notion no theoretical status (see e.g. Stump, 2001). For morphologists that seek to preserve the intuition that irregular inflection necessitate specific modeling, there are two options. Either we take regularity to be a design property of morphological systems, and thus try to model every possible process as regular, limiting irregularity to the description of blatant suppletion phenomena; this position is the default for most morphophonological work in the tradition of Chomsky and Halle (1968), and is strongly defended by proponents of distributed morphology (Halle and Marantz, 1993). Or we take regularity to be an empirical property, that is manifest in performance: speakers are able to inflect an unknown lexeme according to a regular pattern, but will not be able to inflect a lexeme according to an irregular pattern. This position is assumed in much of the psycholinguistic litterature on inflection, and defended forcefully, if somewhat partially, by Pinker (1999) and work cited therein.

In this paper we assume the second position. Note that we do not commit ourselves to any particular view of the processing of inflection, but simply assume that (ir)regularity is a real grammatical phenomenon, that is manifest not only in psycholinguistic behavior but also in language change and in synchronic grammar. Our main empirical argument concerns the status of lexeme formation rules: we show that despite conventional wisdom, the output of lexeme formation rules should not always be considered inflectionally regular.

#	lexeme	trans.	ms. sg.	fem. sg.
(i)	RAPIDE	'fast'	/ʁapid/	/ʁapid/
(ii)	VIEUX	'old'	/vjø/	/vjɛj/
(iii)	PETIT	'small'	/pəti/	/pətit/
(iv)	BREF	'brief'	/bʁɛf/	/bʁɛv/
(v)	GITAN	'gipsy'	/ʒitɑ̃/	/ʒitan/
(vi)	RAGEUR	'rageful'	/ʁaʒœʁ/	/ʁaʒøz/
(vii)	DIRECTEUR	'directorial'	/diʁɛktœʁ/	/diʁɛktʁis/

Table 1: Inflection of a few adjectives in the singular

1.2 French adjective inflection

French adjectives inflect for both gender and number. Here we concentrate on singular forms of the adjectives. Table (1) gives a sample of inflectional forms for a number of typical adjectives. Case (i) clearly corresponds to a regular pattern: identical forms in the masculine and the feminine is what happens for the majority of existing French adjectives, as well as for borrowed adjectives and for adjectives formed by nonconcatenative morphological processes such as clipping (e.g. *sensas*, clipped form of *sensationnel* 'sensational' is /sɑ̃sas/ in the masculine and the feminine). It is also the pattern used by speakers facing a novel adjective not resembling anything known. Case (ii) clearly corresponds to an irregular pattern, since it holds for exactly one lexeme, and is usually treated as a case of suppletion.

Cases (iii) through (v) are the object of some debate in studies on French adjectives. In generative descriptions of French morphophonology, starting with Schane (1968), these are usually considered to be regular cases exhibiting phonologically governed alternations; but one may doubt that this is the whole story, since there are numerous nonalternating adjectives that meet the description of the relevant rule. For instance, case (iii) is described by Dell (1985) by postulating an underlying form /pətit/, and a rule deleting word final obtruants. This rule does not apply in the feminine, because the feminine morpheme is a suffix /ə/ which will be deleted later in the derivation. Yet there are non-alternating obstruent-final adjectives, such as *mat* 'matte', *net* 'clean', *bath* 'hip', *out* 'out', etc. Similar rules postulated to account for (iv) and (v) face problems with nonalternating *paf* 'drunk', *ouf* 'crazy', *gnangnan* 'soppy', *marron* 'brown'. Thus a more realistic analysis would take the inflectional alternations to be the manifestation of a variety of inflectional classes of adjectives. Concretely, we assume four different inflectional classes, specifying the functions in table 2 as exponents for masculine and feminine singular.[1] In such a setting we end up with two distinct notions of (ir)regularity: a regular lexeme belongs to the default, first inflectional class. Being irregular may either mean be-

[†]We thank the reviewers and the audience of the HPSG06 conference, and in particular Berthold Crysmann and Ivan A. Sag, for their comments and suggestion. The analysis presented in section 3 benefited considerably from discussions with Aurélien Giraud.

[1]We assume throughout an inferential-realizational approach to inflection (Stump, 2001).

class	example	proposed stem	ms. sg.	fem. sg.
A	RAPIDE	/ʁapid/	id	id
B	PETIT	/pətit/	delete final C.	id
C	BREF	/bʁɛ/	⊕/f/	⊕/v/
D	GITAN	/ʒita/	nasalize final V.	⊕/n/

Table 2: Inflectional classes for adjectives (first version)

longing to a nondefault inflectional class, like *petit*, or specifying suppletive forms, like *vieux*.

1.3 The problem: adjectives in *-eur*

The most interesting cases in table 1, cases (vi) and (vii), are not usually discussed in the context of adjective inflection. What is interesting is that adjectives in these classes have a uniform formation: class (vi) adjectives are all the output of a rule forming adjectives from the basic stem of the verb (the stem occurring in the present indicative 1pl and 2pl); class (vii) adjectives are the output of a rule forming adjectives from a 'Latinate' stem of the verb, which is formed by suffixing /at/ to the basic stem in most cases, but may take other forms. Note that both formation rules have the same categorial and semantic effects, to the point that many descriptions of French do not recognize them as distinct rules; yet their morphophonology is clearly distinct.[2]

Now, there is little hope of treating adjectives in classes (vi) and (vii) as cases of regular inflection. First, the relation between the masculine and the feminine cannot be seen as the effect of a regular phonological alternation: starting from the masculine, we have three options for forming the feminine of an adjective ending in /œʁ/, either /øz/ (as in *rageuse* /ʁaʒøz/), /ʁis/ (as in *directrice* /diʁɛktʁis/)

[2]Most studies of French derivational morphology do not explicitly discuss adjectives in *-eur*. This is certainly due to the fact that many adjectives in these two classes are homophonous with an agent noun, so that it is usually assumed without discussion that the noun is derived from the verb and the adjective a converted noun. Two arguments show that this is not correct. First, Corbin and Corbin (1991) shows that while it is easy to derive the nominal semantics from the adjective, the opposite route is problematic. Second, there are good reasons to think that gender is not an inflectional category for nouns in French: most nouns, including quite a number of human-denoting nouns (e.g. *personne* 'person') are found in only one gender, and apparent cases of gender-opposed pairs are best analyzed as pairs of independent lexemes related by mere semantic closeness (e.g. *bouc* 'male goat' vs. *chèvre* 'female goat'), derivational rules (e.g. *dinde* 'female turkey' vs. *dindon* 'male turkey') or parallel derivation from adjectives (e.g. *italien* 'male Italian' vs. *italienne* 'female Italian'). But if nouns have just one gender, then there is no single noun lexeme that could serve as the base for conversion for the masculine and feminine forms of the adjective *directeur*. On the other hand, if the adjective is the base, then the nouns *directeur* and *directrice* are the result of two parallel conversion processes from the adjective.

or /œʁ/ (as in *inférieure* /ɛ̃feʁjœʁ/ ‘inferior’). If we start from the feminine, we also have two options for the masculine of an adjective in /øz/: either /œʁ/ (as with adjectives in class (vi)) or /ø/ (as in denominal adjectives such as *respectueux* ‘respectful’, etc.).

Second, we might assume that cases (vi) and (vii) correspond to two further inflectional classes of adjectives, specifying respectively $\langle \oplus/\text{œʁ}/, \oplus/\text{øz}/ \rangle$ and $\langle \oplus/\text{œʁ}/, \oplus/\text{ʁis}/ \rangle$ as exponents in the singular. Yet these inflectional classes would have the very peculiar feature of each containing only lexemes derived from a single formation process. This contrasts strongly with the classes discussed in table 2, which all contain both derived and root lexemes, as exemplified in (1).

(1) Class A:
   i. *rapide* ‘fast’, *joli* ‘pretty’, *gai* ‘joyful’, etc.
   ii. *bancaire* ‘(of a) bank’, *mortel* ‘mortal’, *algébrique* ‘algebraic’, etc.

   Class B:
   i. *petit* ‘small’, *grand* ‘large’, *gros* ‘big’, etc.
   ii. *venteux* ‘windy’, *grossier* ‘crude’, *lyonnais* ‘from Lyon’, etc.

   Class C:
   i. *bref* ‘brief’, *naïf* ‘naïve’, etc.
   ii. *pensif* ‘thoughtful’, *alternatif* ‘alternative’, etc.

   Class D:
   i. *bon* ‘good’, *fin* ‘thin’, *plan* ‘flat’, etc.
   ii. *alpin* ‘alpine’, *euclidien* ‘Euclidian’, *pâlichon* ‘pale-ish’, etc.

We conclude that no satisfying analysis of adjectives in *-eur* is forthcoming in a traditional morphological setting. The following sections show that introducing the notion of a *stem space* offers a third, more satisfying possibility.

2 Motivating the stem space

2.1 The stem space of French verbs

Starting with (Aronoff, 1994), a number of recent studies challenge the idea that lexemes are associated with a single phonological representation, the lexeme’s stem. Lexemes should rather be associated with a vector of possibly different phonological representations, what Bonami and Boyé (2002) call a *stem space*; each inflectional or derivational rule specifies which coordinate in the vector it uses as its input. Such analyses have been proposed, among others, by Aronoff (1994) for Latin conjugation, Sadler et al. (1997) for Russian nominalizations, Brown (1998) for Russian conjugation, Pirelli and Battista (2000) for Italian conjugation, Stump (2001) for Sanskrit declension, Bonami and Boyé (2002) for French conjugation, Boyé and Cabredo Hofherr (2006) for Spanish conjugation. We illustrate with evidence from French for uniformity.

lexeme	1sg	2sg	3sg	1pl	2pl	3pl
LAVER ‘wash’	/lav/	/lav/	/lav/	/lav-ɔ̃/	/lav-e/	/lav/
TORDRE ‘bend’	/tɔʁ/	/tɔʁ/	/tɔʁ/	/tɔʁd-ɔ̃/	/tɔʁd-e/	/tɔʁd/
MOURIR ‘die’	/mœʁ/	/mœʁ/	/mœʁ/	/muʁ-ɔ̃/	/muʁ-e/	/mœʁ/
BOIRE ‘drink’	/bwa/	/bwa/	/bwa/	/byv-ɔ̃/	/byv-e/	/bwav/

Table 3: Present indicative conjugation

Inflectional systems often exhibit alternations which have no synchronic phonological motivation and concern arbitrary subparts of the paradigm. For instance, in French, in the present indicative, there is a partition between (i) the three singular forms, (ii) the plural 1 and 2 forms, and (iii) the plural 3 form. While there is a systematic similarity between members of each cell in the partition,[3] the content of the different cells may differ in arbitrary ways, as illustrated in table 3.

One can account for this pattern by assuming that French verbal lexemes come equipped with a stem space with at least three slots.[4] Each inflectional rule specifies which slot it uses as a base, and what phonological modification is made on this base. Slot 1 serves as the base for present 1pl and 2pl inflection, slot 2 serves for 3pl, and slot 3 for singular forms.

A direct advantage of the stem space is that it allows for an account of the diversity of patterns of irregular conjugation. In French, fully regular (so-called ‘first group’) verbs have identical stems in slots 1, 2, and 3. Irregular verbs may need either two or three distinct stems, but an exhaustive examination of the French lexicon shows that no verb has identical stems in slots 1 and 3 but a different stem in slot 2. Bonami and Boyé (2002) proposes to account for this by stating that the slots are related by default relations, which may be overruled by irregular lexemes. Slot 1 is identical to slot 2 by default, and slot 2 is indentical to slot 3, but there is no default relation between slot 1 and 3, which accounts for the observed pattern.

Further evidence for the stem space comes from the fact that lexeme formation rules are also sensitive to different slots. For instance, as illustrated in table 4, the rule constructing deverbal adjectives in *-eur/-euse* uses slot 1 as its base, while the rule constructing nominal V-N compounds relies on slot 3.

2.2 The stem space of French adjectives

A different type of argument in favor of the notion of a stem space comes from the inflection of French adjectives. In section 1, we suggested an analysis of French

[3] With the exception of a handful of maximally irregular verbs; see Bonami and Boyé (2002) for discussion.

[4] In a full analysis of French conjugation, 12 distinct slots are necessary.

base	stem 1	stem 3	*eur*/*euse* Adj.	V-N compound
laver ‘wash’	/lav/	/lav/	laveur /lavœʁ/ ‘washer’	lave-mains /lavmɛ̃/ ‘washbowl’
tordre ‘bend’	/tɔʁd/	/tɔʁ/	tordeur /tɔʁdœʁ/ ‘bender’	tord-boyaux /tɔʁbwajo/ ‘rotgut’
boire ‘drink’	/byv/	/bwa/	buveur /byvœʁ/ ‘drinker’	boitout /bwatu/ ‘stemless glass’
soutenir ‘support’	/sutən/	/sutjɛ̃/	souteneur /sutənœʁ/ ‘pimp’	soutien-gorge /sutjɛ̃gɔʁʒ/ ‘bra’

Table 4: Two lexeme formation processes

adjectives in terms of inflectional classes specifying the relationship between a single stem and two inflectional forms. The following data from Bonami and Boyé (2005) shows that this analysis is inadequate.

First, French adjectives take a special form in the masculine singular when preceding a vowel-initial noun, which we call the Masculine Singular Liaison Form (MSLF).[5] That this is a distinct inflectional form of the adjective is shown by the fact that it can be suppletive or defective (Morin, 2003). But when it is not, the form is either identical to the ‘ordinary’ masculine singular or to the feminine singular (table 5), in accordance with the generalization in (2). This situation is problematic, because for some adjectives there is a discrepancy between the morphosyntactic features manifested in syntax (masculine singular) and the morphosyntactic features expressed by the form (feminine singular).[6]

(2) If the masculine singular form ends in a consonant, then the MSLF is identical to the masculine singular. Otherwise it is identical to the feminine singular.

Second, French deadjectival adverbs in *-ment* are systematically formed on the feminine form, as illustrated in table 5. This is so despite the fact that adverbs do not inflect for gender in French, so that there is no sense in which the adverb can be said to be feminine.

Both observations argue in favor of a morphomic account (Aronoff, 1994): adjectives have two distinct stems, which express no morphosyntactic features by

[5] See Bonami et al. (2004) for an HPSG analysis of French liaison.

[6] Perlmutter (1998) and Tranel (1996) attempts to account for this data in an optimality-theoretic setting, by ranking phonological markedness constraints higher than syntactic agreement constraints. See Bonami and Boyé (2003, 2005) for a detailed criticism.

adjectival lexeme	MAS.SG form	FEM.SG form	MSLF	derived adverb
RAPIDE	ʁapid	ʁapid	ʁapid	ʁapidmɑ̃
PETIT	pəti	pətit	pətit	pətitmɑ̃
BEAU	bo	bɛl	bɛl	bɛlmɑ̃
VIF	vif	viv	vif	vivmɑ̃
FORT	fɔʁ	fɔʁt	fɔʁ	fɔʁtəmɑ̃
RÊVEUR	ʁɛvœʁ	ʁɛvøz	ʁɛvœʁ	ʁɛvøzmɑ̃

Table 5: Distribution of adjective stems

class	example		slot 2 $\rightsquigarrow$ slot 1
A	RAPIDE	/ʁapid/	identity
B	PETIT	/pətit/	delete final C
C	BREF	/bʁɛv/	devoice final C
D	GITAN	/ʒitan/	delete final C, and nasalize preceding V

Table 6: Inflectional classes for adjectives (final version)

themselves. (Bonami and Boyé, 2005) implement this idea by stating that adjectives have a two slot stem space, with different morphological processes selecting the appropriate slot as stated in (3).

(3) a. The masculine singular form is identical to stem 1.
b. The feminine singular form is identical to stem 2.
c. If stem 1 is consonant final, then the MSLF is identical to stem 1; otherwise it is identical to stem 2.
d. The lexeme formation rule for adverbs in *-ment* selects stem 2 as its input.

Notice that in the context of this analysis, both masculine singular and feminine singular have a null exponent in French; all the action occurs in stem selection rather than in exponence.[7] In this context, the inflectional class partition proposed in table 2 must be recast, not as a series of distinct ways of relating inflectional forms, but as a series of ways of relating the slots in the stem space, as shown in table 6.

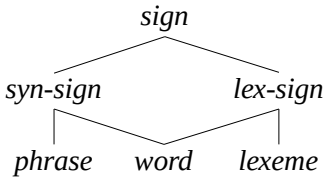
[7]By contrast, plural number has a uniform exponent, the latent consonant /z/.

3 Modeling stem spaces in HPSG

3.1 Stem spaces for verbs

There are two important issues when modeling the stem space in an HPSG grammar. First, one has to decide what status the stem space has. Bonami and Boyé (2002) treat each stem as a distinct member of the lexical hierarchy, typed for the slot it occupies in the stem space and the lexeme it belongs to. A simpler alternative is to assume that the stem space is a data structure internal to the lexical entry of a lexeme.[8] Thus we assume that lexemes carry a feature STEMS with features corresponding to each slot in the stem space (5), and that inflectional rules such as (6) take this as their input.[9]

(4)



a. *syn-sign* $\rightarrow \left[\text{PHON} \quad \textit{phon}\right]$

b. *lex-sign* $\rightarrow \left[\text{M-DTRS} \quad \textit{list}(\textit{lexeme})\right]$

c. *phrase* $\rightarrow \left[\text{DTRS} \quad \textit{list}(\textit{syn-sign})\right]$

d. *word* $\rightarrow \left[\text{M-DTRS} \quad \langle \textit{lexeme} \rangle\right]$

e. *lexeme* $\rightarrow \left[\text{STEMS} \quad \textit{stem-space}\right]$

(5) a. *v-lexeme* $\rightarrow \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{STEMS} & \textit{v-stem-space} \end{bmatrix}$

b. *v-stem-space* $\rightarrow \begin{bmatrix} \text{SLOT1} & \textit{phon} \\ \text{SLOT2} & \textit{phon} \\ \text{SLOT3} & \textit{phon} \\ \ldots & \end{bmatrix}$

[8] This type of analysis can be traced back to Pollard and Sag's (1987, p. 213) suggestion that lexical entries of irregulars contain a specification of their principal parts. It is also similar to the analysis of irregular inflection defended by (Spencer, 2004) as part of Generalized Paradigm Function Morphology.

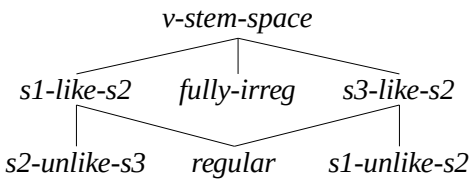
[9] We take phonological representations to be (at least) lists of objects of type *segment*; the hierarchy of segment types allows for an equivalent of phonetic feature decomposition. To improve readability, where possible, we note lists of segments as sequences of sans-serif IPA symbols, rather than using the standard HPSG notation for lists and types. Thus e.g. t is a shorthand for $\langle$*t-seg*$\rangle$.

(6) *prst-indic-1pl*→
$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \boxed{1}\oplus\tilde{\text{ɔ}} \\ \text{SYNSEM} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{TENSE} & \textit{prst} \\ \text{MOOD} & \textit{indicative} \end{bmatrix} \\ \text{SUBJ} & \left\langle \text{NP}[\textit{1pl}] \right\rangle \end{bmatrix} \\ \text{M-DTRS} & \left\langle \begin{bmatrix} \textit{v-lexeme} & \\ \text{STEMS}\vert\text{SLOT1} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$

3.2 Relations within the stem space: Giraud (2005)

The second issue is to decide on a way of encoding the default relations structuring the stem space. Bonami and Boyé (2002) rely on an ontologically quite promiscuous system, using a combination of online type construction (Koenig, 1999) and default specifications (Lascarides and Copestake, 1999). At the other end of the spectrum, Giraud (2005) proposes a much more conventional implementation, where stem spaces are typed for the morphophonological relations they verify, and the regular case just corresponds to the stem space type verifying the maximal number of relations. (7) is basically a simplified version of Giraud's proposal specifying only the part of the hierarchy of verbal stem spaces needed to account for the present indicative. Leaf types correspond to particular conjugation patterns, and inherit from intermediate types stating a morphophonological relation between stem slots.

(7)



a. *s1-like-s2*→
$$\begin{bmatrix} \text{SLOT1} & \boxed{1} \\ \text{SLOT2} & \boxed{1} \end{bmatrix}$$

b. *s3-like-s2*→
$$\begin{bmatrix} \text{SLOT2} & \boxed{1} \\ \text{SLOT3} & \boxed{1} \end{bmatrix}$$

In Giraud's system, individual lexemes need to specify (i) a stem space type, and (ii) enough stems to fill up the stem space, as illustrated by the following lexical entries.

(8) a. *laver*:
$$\begin{bmatrix} \text{STEMS} & \begin{bmatrix} \textit{regular} & \\ \text{SLOT1} & \textsf{lav} \end{bmatrix} \end{bmatrix}$$

b. *valoir*: $\begin{bmatrix} \text{STEMS} & \begin{bmatrix} \textit{s2-unlike-s3} \\ \text{SLOT2} & \text{val} \\ \text{SLOT3} & \text{vo} \end{bmatrix} \end{bmatrix}$

3.3 Regularity as default

While its formal clarity speaks in favor of Giraud's system, an obvious disadvantage is that (ir)regularity is not modeled directly. There is no formal difference between the type *regular* and the other leaf types of the stem space corresponding to the fact that regular verbs have a special status. Moreover, the lexical entry of every regular verb needs to include an explicit specification of the fact that this verb is regular, which goes against all evidence that speakers assume verbs to be regular in the absence of contradictory information. Clearly, a system where only irregulars would need explicit specification is more desirable.

We conclude that while Giraud's proposal succeeded in eliminating the need for online type construction to model the stem space, it did not eliminate the need for defaults. We thus propose to introduce a single modification to Giraud's system, the default specification in (9). By default lexemes are assumed to have a regular stem space. This means that every lexeme will inherit all morphophonological relations that are not incompatible with the morphophonological information in its lexical entry. Thus if a lexeme lists only the content of one of its slots, it will be of type *regular*. If it lists two distinct phonologies for slot 1 and slot 2, the stem space's type cannot be a subtype of *s1-like-s2*, and thus it will be of type *s1-unlike-s2*. The only way for a lexeme to be fully irregular is for it to list three distinct phonologies in slots 1, 2, and 3. (10) lists appropriate lexical entries for verbs with the four distinct patterns.

(9) *verb-lexeme*$\rightarrow \begin{bmatrix} \text{STEMS} & / \textit{regular} \end{bmatrix}$

(10) a. *laver*: $\begin{bmatrix} \text{STEMS} & \begin{bmatrix} \text{SLOT1} & \text{lav} \end{bmatrix} \end{bmatrix}$

b. *valoir*: $\begin{bmatrix} \text{STEMS} & \begin{bmatrix} \text{SLOT2} & \text{val} \\ \text{SLOT3} & \text{vo} \end{bmatrix} \end{bmatrix}$

c. *mourir*: $\begin{bmatrix} \text{STEMS} & \begin{bmatrix} \text{SLOT1} & \text{muʁ} \\ \text{SLOT2} & \text{mœʁ} \end{bmatrix} \end{bmatrix}$

d. *boire*: $\begin{bmatrix} \text{STEMS} & \begin{bmatrix} \text{SLOT1} & \text{byv} \\ \text{SLOT2} & \text{bwav} \\ \text{SLOT3} & \text{bwa} \end{bmatrix} \end{bmatrix}$

3.4 Discussion

The analysis proposed above crucially relies on the use of a default specification. Moreover this default cannot be considered to be simply of an abbreviatory nature, as e.g. the default specifications in Ginzburg and Sag (2000) can. In Ginzburg and Sag (2000)'s grammar, defaults are only used to avoid stating intuitively redundant constraints on types that are listed in the hierarchy anyway. In the current proposal, however, the default is used to constrain the members of an open lexicon: what we are attempting to model is the fact that speakers treat unknown verbs (that is, verbs that are just entering their lexicon) as regulars. Thus specifying in individual lexical entries the information represented by the default is not an option, and would amount to not model the relevant property in the grammar.[10]

Since the use of default specifications is controversial in HPSG, it is worth asking whether another way of accounting for regularity can be found. As an anonymous reviewer suggests, one possibility would be to use attempts to use online type construction, which is explicitly introduced by Koenig (1999) as a way of modeling productive morphological processes (of which regular inflection is arguably an instance). However all our attempts have failed. Here we consider two possible routes that illustrate why online type construction is not adequate.

One possible analysis, which is closest to the present proposal, is to cross-classify lexemes for stem-space type and some other dimension, say, the type of content they have. Figure 1 illustrates such an approach. Here we state at the level of lexemes the classification that was stated at the level of stem spaces in Giraud's approach. Irregular lexemes are explicitly listed as belonging to a particular stem space type, whereas regulars are not, but *can* inherit from the *regular* type. This is indicated by the dotted line from *regular* to *laver-lxm* in figure 1. Such a system predicts only one stem (irregular) stem space for irregular lexemes, and predicts a regular stem space as one possibility for regular lexemes. However the problem is that it overgenerates, since nothing precludes e.g. *laver-lxm* from having a common subtype with *s1-unlike-s2*. As far as we can see, the only way to avoid such overgeneration is to augment *laver-lxm*'s lexical entry with some information incompatible with *s1-unlike-s2*—in other word, to state explicitly in the lexical entry that *laver* is a regular verb, which is precisely what we set out not to do.

Another option is to modify the form of the lexicon so that the blocking issue does not arise. Suppose that we follow Bonami and Boyé (2002) and treat stems as objects in the lexical hierarchy, rather than simply phonological objects within a lexeme's lexical entry. Regular relations between slots are modeled as lexical rules such as those in (11) relating two stems. Within such a system, the issue is not to block irregular patterns for regular verbs, but to block the application of relevant

[10] Of course, an alternative is to take it that the default character of regular inflection is a psycholinguistic issue that needs not be modeled in the grammar *per se* but can be left to a model of performance. However we do not know of any model of inflectional performance that both recognizes a status for regularity and does not presuppose that the competence grammar provides a characterization of regularity.

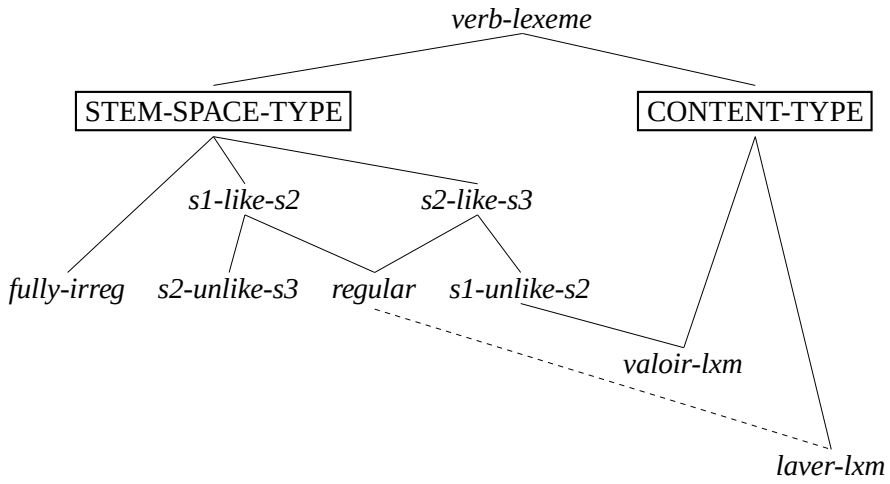


Figure 1: A failed analysis based on online type contruction

lexical rules to irregulars. Since stems are signs, one could use morphosyntactic information to this effect. For instance, the lexical entry for the slot 1 stem of *valoir* would include a feature specification ensuring that it cannot serve as the basis for a present singular form (12). The construction of an overregular slot 3 stem val for *valoir* is not blocked as such, but this stem will never be used as the base for an inflected form of the verb.

(11) a. *slot-2-stm*→
$$\begin{bmatrix} \textsc{synsem} & \boxed{1} \\ \textsc{m-dtrs} & \left\langle \begin{bmatrix} \textit{slot-1-stm} \\ \textsc{synsem} & \boxed{1} \\ \textsc{phon} & \boxed{2} \end{bmatrix} \right\rangle \\ \textsc{phon} & \boxed{2} \end{bmatrix}$$

b. *slot-3-stm*→
$$\begin{bmatrix} \textsc{synsem} & \boxed{1} \\ \textsc{m-dtrs} & \left\langle \begin{bmatrix} \textit{slot-2-stm} \\ \textsc{synsem} & \boxed{1} \\ \textsc{phon} & \boxed{2} \end{bmatrix} \right\rangle \\ \textsc{phon} & \boxed{2} \end{bmatrix}$$

(12) *valoir*:
$$\begin{bmatrix} \textsc{phon} & \textsf{val} \\ \textsc{synsem}|\textsc{cat} & \begin{bmatrix} \textsc{head}|\textsc{tense} & \textit{imperfective} \end{bmatrix} \vee \\ & \begin{bmatrix} \textsc{head} & \begin{bmatrix} \textsc{tense} & \textit{present} \\ \textsc{mood} & \textit{indicative} \end{bmatrix} \\ \textsc{subj} & \left\langle \begin{bmatrix} \textsc{ind} & \begin{bmatrix} \textsc{nb} & \textit{pl} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

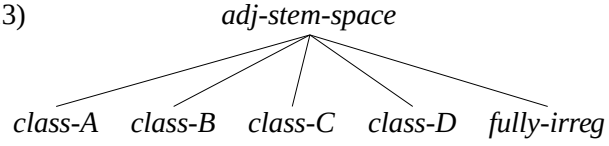
While such an approach arguably models irregularity as such, and is formally more conservative than the default-based approach defended here, it has a number of conceptual and empirical drawbacks. First, the feature specifications one needs to include in the description of stems of irregular verbs have a strong ad-hoc flavor. Second, the use of morphosyntactic features to ensure blocking is contradictory with the morphomic nature of the stem space; this is problematic for the modeling of derivation: the rule for V-N compounds can no longer state that its base is a slot-3 stem, because /val/ is not blocked as a slot-3 stem as such—rather the use of /val/ in the present singular is blocked. Finally, such a model forces one to structure the stem space of regular verbs as a directed tree, which is problematic for the modeling of morphophonological opacities in paradigms: as Bonami and Boyé (2006a) argues, although the full inflection of a regular verb can always be deduced from the knowledge of one stem, it is not always the same slot that must be known.

To conclude this section, our attempts to avoid the use of defaults in the modeling of regularity have failed. In the absence of an explicit alternative, we take it that the use of defaults is the only known way to model regularity in an HPSG implementation of the stem space.

3.5 Extending the analysis to adjectives

The general approach to stem spaces just outlined can be adapted to the analysis of adjectives with just a few modifications. Remember that we assume adjectives to have a two-slot stem space, where slot 1 is used for ordinary masculine forms and slot 2 for all feminine forms. The inflectional classes postulated in table 2 can be recast as types of stem spaces, as shown in (13);[11] note that contrary to what happens with verbs, types of adjectival stem spaces are mutually exclusive, because the constraints they impose on the structure of the stem space are incompatible. Class A is the default type; thus the vast majority of adjective lexical entries need not mention a stem space type, but can just specify the content of a slot (15a). Lexemes belonging to a different class must specify the stem space type (15b-d), and true irregulars such as *vieux* need to specify the content of both slots.

(13)



a. *class-A*→ $\begin{bmatrix} \text{SLOT}1 & \boxed{1} \\ \text{SLOT}2 & \boxed{1} \end{bmatrix}$

[11] In fact it is more satisfactory to treat the alternating endings as parts of the stems, rather than inflectional exponents, since they show up in derived lexemes: e.g. *petitesse* /pətitɛs/ 'smallness', not */petiɛs/; *pensivement* /pɑ̃sivmɑ̃/ 'thoughtfully', not */pɑ̃simɑ̃/.

b. *class-B*→$\begin{bmatrix}\text{SLOT1} & \boxed{1} \\ \text{SLOT2} & \boxed{1}\oplus\langle \textit{cons}\rangle\end{bmatrix}$

c. *class-C*→$\begin{bmatrix}\text{SLOT1} & \boxed{1}\oplus\text{f} \\ \text{SLOT2} & \boxed{1}\oplus\text{v}\end{bmatrix}$

d. *class-D*→$\begin{bmatrix}\text{SLOT1} & \boxed{1}\oplus\langle\text{nasal}(\boxed{2})\rangle \\ \text{SLOT2} & \boxed{1}\oplus\langle\boxed{2}\textit{oral-vow}\rangle\oplus\text{n}\end{bmatrix}$

where:

i. $\text{nasal}(\text{a}) = \tilde{\text{a}}$

ii. $\text{nasal}(\varepsilon) = \text{nasal}(\text{i}) = \text{nasal}(\text{y}) = \tilde{\varepsilon}$

iii. $\text{nasal}(\text{o}) = \text{nasal}(\text{ɔ}) = \tilde{\text{ɔ}}$

(14) *adj-lexeme*→$\left[\text{STEMS } \textit{adj-stem-space/class-A}\right]$

(15) a. *rapide*:$\left[\text{STEMS} \quad \left[\text{SLOT2} \quad \text{ʁapid}\right]\right]$

b. *petit*:$\left[\text{STEMS} \quad \begin{bmatrix}\textit{class-B} \\ \text{SLOT2} \quad \text{pətit}\end{bmatrix}\right]$

c. *bref*:$\left[\text{STEMS} \quad \begin{bmatrix}\textit{class-C} \\ \text{SLOT2} \quad \text{bʁɛv}\end{bmatrix}\right]$

d. *gitan*:$\left[\text{STEMS} \quad \begin{bmatrix}\textit{class-D} \\ \text{SLOT2} \quad \text{ʒitan}\end{bmatrix}\right]$

e. *vieux*:$\left[\text{STEMS} \quad \begin{bmatrix}\text{SLOT1} \quad \text{vjø} \\ \text{SLOT2} \quad \text{vjɛj}\end{bmatrix}\right]$

With these stem space specifications, we can now state appropriate inflectional rules on the basis of (Bonami et al., 2004)'s analysis of liaison. The ordinary masculine is specified as [LFORM −], which means that it can be used in contexts where liaison cannot occur, e.g. before a consonant-initial noun or post-nominally.[12] For the MSLF, we implement the stem selection rule described in (3d) using a function that inspects the phonology of the slot1-stem.

(16) a. *masc-sg-adj*→$\begin{bmatrix}\textit{word} & \\ \text{PHON} & \boxed{1} \\ \text{SYNSEM} & \begin{bmatrix}\text{HEAD} & \textit{adj}[\textit{mas},\textit{sg}] \\ \text{LFORM} & -\end{bmatrix} \\ \text{M-DTRS} & \left\langle\begin{bmatrix}\textit{adj-lexeme} & \\ \text{STEMS}|\text{SLOT1} & \boxed{1}\end{bmatrix}\right\rangle\end{bmatrix}$

[12] However it *can* occur before a vowel-initial noun, because liaison is not obligatory for prenominal adjectives; see (Bonami et al., 2004) for extended discussion and analysis.

b. *MSLF-adj*→
$$\begin{bmatrix} \textit{word} & \\ \textsc{phon} & \textsf{select-stem}(\boxed{1},\boxed{2}) \\ \textsc{synsem} & \begin{bmatrix} \textsc{head} & \textit{adj}[\textit{mas,sg}] \\ \textsc{lform} & + \end{bmatrix} \\ \textsc{m-dtrs} & \left\langle \begin{bmatrix} \textit{adj-lexeme} & \\ \textsc{stems} & \begin{bmatrix} \textsc{slot}1 & \boxed{1} \\ \textsc{slot}2 & \boxed{2} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

where

i. $\textsf{select-stem}(\boxed{1}\langle \ldots, \textit{cons}\rangle, \boxed{2}) = \boxed{1}$

ii. $\textsf{select-stem}(\boxed{1}\langle \ldots, \textit{vow}\rangle, \boxed{2}) = \boxed{2}$

c. *fem-sg-adj* →
$$\begin{bmatrix} \textit{word} & \\ \textsc{phon} & \boxed{1} \\ \textsc{synsem} & \left[\textsc{head} \quad \textit{adj}[\textit{fem,sg}]\right] \\ \textsc{m-dtrs} & \left\langle \begin{bmatrix} \textit{adj-lexeme} & \\ \textsc{stems}|\textsc{slot}2 & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$

4 Modeling derived irregularity

We can now turn to our account of derived irregularity. Notice that in the current setup, a lexeme formation rule does not derive a single stem from a single stem, but it derives a stem space from another stem space. Thus every specification of the stem space that is open to lexical entries is also open to lexeme formation rules. A rule may just specify a single slot of the stem space, in which case the output of the rule will fall in the default inflection pattern. Or it may specify extra information that is incompatible with the default pattern, in which case one ends up with an output that is inflectionally irregular despite being derived productively.

Now let us turn to a few examples of adjectival lexeme formation rules. These have the exact same typology as root adjectives. Denominal adjectives in *-aire* fall into the default class A (17). The rule for adjectives *-eux* must specify that its output falls in class B (18). Finally, we come to the crucial case: rule (19) for adjectives in *-eur* with a feminine in *-euse* directly specifies two stems for its output. Since no inflectional class can accommodate two stems with such a morphophonological relation, the output of the rule necessarily ends up with a type *fully-irreg* stem space.

(17) *_aire-adj-lxm*→
$$\left[\begin{array}{ll}\textsc{stems} & \left[\textsc{slot}2 \quad \boxed{1}\oplus\text{ɛʁ}\right] \\ \textsc{synsem} & \left[\textsc{head} \quad \textit{adj}\right] \\ \textsc{m-dtrs} & \left\langle\left[\begin{array}{ll}\textsc{head} & \textit{noun} \\ \textsc{stems}|\textsc{slot}1 & \boxed{1}\end{array}\right]\right\rangle\end{array}\right]$$

(18) *_eux-adj-lxm*→
$$\left[\begin{array}{ll}\textsc{stems} & \left[\begin{array}{ll}\textit{class-B} & \\ \textsc{slot}2 & \boxed{1}\oplus\text{øz}\end{array}\right] \\ \textsc{synsem} & \left[\textsc{head} \quad \textit{adj}\right] \\ \textsc{m-dtrs} & \left\langle\left[\begin{array}{ll}\textsc{head} & \textit{noun} \\ \textsc{stems}|\textsc{slot}1 & \boxed{1}\end{array}\right]\right\rangle\end{array}\right]$$

(19) *_eur/_euse-adj-lxm*→
$$\left[\begin{array}{ll}\textsc{stems} & \left[\begin{array}{ll}\textsc{slot}1 & \boxed{1}\oplus\text{œʁ} \\ \textsc{slot}2 & \boxed{1}\oplus\text{øz}\end{array}\right] \\ \textsc{synsem} & \left[\textsc{head} \quad \textit{adj}\right] \\ \textsc{m-dtrs} & \left\langle\left[\begin{array}{ll}\textsc{head} & \textit{verb} \\ \textsc{stems}|\textsc{slot}1 & \boxed{1}\end{array}\right]\right\rangle\end{array}\right]$$

The case of deverbal adjectives in *-eur* with a feminine in *-rice* is entirely parallel, but with two complications. First, we must account for the special form of the verbal stem these adjectives are based on. Following Bonami et al. (to appear), we assume that French verbs have an extra slot for a special stem, which never shows up in inflection, but serves as the base for at least three lexeme formation rules: the rule for nominalizations in *-ion*, the rule for adjectives in *-eur*/*-rice*, and the rule for adjectives in *-if*.In the default case, this stem is obtained by adding *at* to the end of the stem in slot 1. The corresponding stem slot is labelled SLOT13 in (20) because it comes in addition to the 12 slots necessary for a full treatment of French conjugation.

Second, we must account for the fact that some adjectives in *-eur* in this class are defective in the feminine. Specifically, all adjectives which have a (nondefault) stem 13 ending in /s/ have no feminine form; e.g. *antidépresseur* 'antidepressive' **antidépressrice*. To account for this, we assume that the stem in slot 2 is the empty list if the input's stem 13 does not end in /t/. We assume that it is a general constraint on inflectional rules that they need a phonologically nonempty input; thus no feminine form will be generated from the lexical entry of *antidépresseur*.

(20) *_eur/_rice-adj-lxm*$\rightarrow$
$$\left[\begin{array}{ll} \textsc{stems} & \left[\begin{array}{ll} \textsc{slot}1 & \boxed{1}\oplus\text{œʁ} \\ \textsc{slot}2 & \mathsf{f}_{rice}(\boxed{1}) \end{array}\right] \\ \textsc{synsem} & \left[\textsc{head} \quad \textit{adj}\right] \\ \textsc{m-dtrs} & \left\langle\left[\begin{array}{ll} \textsc{head} & \textit{verb} \\ \textsc{stems} & \left[\textsc{slot}13 \quad \boxed{1}\right] \end{array}\right]\right\rangle \end{array}\right]$$

where

a. $(\boxed{3} = \text{t}) \rightarrow \mathsf{f}_{rice}(\boxed{2} \oplus \boxed{3}) = \boxed{2} \oplus \text{tʁis}$

b. $(\boxed{3} = \langle \textit{seg} \rangle \wedge \boxed{3} \neq \text{t}) \rightarrow \mathsf{f}_{rice}(\boxed{2} \oplus \boxed{3}) = \textit{elist}$

5 Conclusions

In this paper we presented a general approach to stem allomorphy based on the notion of a *stem space*. While a previous HPSG implementation of the stem space has been presented in Bonami and Boyé (2002), the current, streamlined approach has a number of distinct advantages. It is compatible with a surface-oriented account of phonological opacities in regular inflection (Bonami and Boyé, 2006a); it is more easily embeddable in a model of morphological performance (Bonami and Boyé, 2006b); and as shown in the present paper, it interacts correctly with data from derivational morphology, accounting directly for the otherwise mysterious phenomenon of derived irregularity.

One issue we did not discuss at all is the modeling of (ir)regular exponents: all the irregularities discussed in the present paper correspond to cases of morphomic stem allomorphy. This is mainly due to the fact that, in French, irregular exponents turn out to be a sporadic phenomenon at best; for instance, there are exactly 5 verbs with irregular *forms*, whereas there are more than 350 verbs with irregular *stems*. For languages with real inflection classes though, the issue of (ir)regular exponents must of course be taken seriously. We submit that the use of hierarchies of inflection patterns should be applicable in such cases too, the difference being that patterns are characterized by the relation between stems and forms rather than the relations among the stems discussed here.

References

Aronoff, Mark. 1994. *Morphology by itself*. Cambridge: MIT Press.

Bonami, Olivier and Boyé, Gilles. 2002. Suppletion and stem dependency in inflectional morphology. In Franck Van Eynde, Lars Hellan and Dorothee Beerman (eds.), *The Proceedings of the HPSG '01 Conference*, Stanford: CSLI Publications.

Bonami, Olivier and Boyé, Gilles. 2003. La nature morphologique des allomor-

phies conditionnées : les formes de liaison des adjectifs en français. In *Actes du troisième forum de morphologie*, Lille.

Bonami, Olivier and Boyé, Gilles. 2005. Construire le paradigme d'un adjectif. *Recherches Linguistiques de Vincennes* 34, 77–98.

Bonami, Olivier and Boyé, Gilles. 2006a. Remarques sur les bases de la conjugaison, ms, Université Paris 4 & Université Bordeaux 3.

Bonami, Olivier and Boyé, Gilles. 2006b. Subregular defaults in French conjugation, paper read at the 12th International Morphology Meeting, Budapest.

Bonami, Olivier, Boyé, Gilles and Kerleroux, Françoise. to appear. L'allomorphie radicale et la relation flexion-construction. In Bernard Fradin, Françoise Kerleroux and Marc Plénat (eds.), *Aperçus de morphologie*, Saint-Denis: Presses de l'Université de Vincennes.

Bonami, Olivier, Boyé, Gilles and Tseng, Jesse. 2004. An integrated analysis of French liaison. In Gerhard Jaeger (ed.), *Formal Grammar 2004 Preproceedings*.

Boyé, Gilles and Cabredo Hofherr, Patricia. 2006. The structure of allomorphy in Spanish verbal inflection. In *Cuadernos de Lingüística*, volume 13, pages 9–24, Instituto Universitario Ortega y Gasset.

Brown, Dunstan. 1998. Stem lndexing and Morphonological Selection in the Russian Verb: a Network Morphology Account. In R. Fabri, A. Ortmann and T. Parodi (eds.), *Models of Inflection*, pages 196–224, Niemeyer.

Chomsky, Noam and Halle, Morris. 1968. *The sound pattern of English*. Harper and Row.

Corbin, Danièle and Corbin, Pierre. 1991. Un traitement unifié du suffixe *-ier(e)*. *Lexique* 10, 61–145.

Dell, François. 1985. *Les règles et les sons*. Paris: Hermann.

Ginzburg, Jonathan and Sag, Ivan A. 2000. *Interrogative Investigations. The Form, Meaning, and Use of English Interrogatives*. Stanford: CSLI Publications.

Giraud, Aurélien. 2005. A detailed treatment of suppletion in French verbal morphology, paper read at the Second International Workshop on Constraint-Based Grammar, University of Bremen.

Halle, Morris and Marantz, Alex. 1993. Distributed morphology and the pieces of inflection. In Kenneth Hale and Samuel Jay Keyser (eds.), *The view from building 20*, pages 111–176, Cambridge: MIT Press.

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

Lascarides, Alex and Copestake, Ann. 1999. Default representation in constraint-based frameworks. *Computational Linguistics* 25, 55–105.

Morin, Yves-Charles. 2003. Remarks on prenominal liaison consonants in French. In S. Ploch (ed.), *Living on the Edge: 28 Papers in Honour of Jonathan Kaye*, pages 385–400, Berlin: Mouton de Gruyter.

Perlmutter, David M. 1998. Interfaces: explanation of allomorphy and the architecture of grammars. In Steven G. Lapointe, Diane K. Brentari and Patrick M. Farrell (eds.), *Morphology and its relation to phonology and syntax*, pages 307–338, Stanford: CSLI Publications.

Pinker, Steven. 1999. *Words and Rules*. New York: Basic Books.

Pirelli, Vito and Battista, Marco. 2000. The Paradigmatic Dimension of Stem Allomorphy in Italian Verb Inflection. *Rivista di Linguistica* 12.

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

Sadler, Louisa, Spencer, Andrew and Zaretskaya, Marina. 1997. A morphomic account of a syncretism in Russian deverbal nominalizations. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1996*, pages 181–215, Dordrecht: Kluwer.

Schane, Sanford. 1968. *French Phonology and Morphology*. Cambridge: MIT Press.

Spencer, Andrew. 2004. Generalized Paradigm Function Morphology, ms., University of Essex.

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

Tranel, Bernard. 1996. French liaison and elision revisited: a unified account within Optimality Theory. In Claudia Parodi, Carlos Quicoli, Mario Saltarelli and Maria Luisa Zubizarreta (eds.), *Aspects of Romance Linguistics*, pages 53–78, Amsterdam: John Benjamins.