Stem spaces in abstractive morphology: A look at defectiveness in French conjugation

Gilles Boyé
Université Bordeaux-Montaigne & UMR5263

1 Introduction

Inflectional morphology descriptions usually adopt a top-down perspective using, for example, a partition of the lexicon into more or less fine-grained inflectional classes and describing the different classes (e.g. Network Morphology: Corbett & Fraser 1993, Brown & Hippisley 2012 or Natural Morphology: Kilani-Schoch & Dressler 2005), or a set of stems for lexemes and rules of realizations for feature bundles (e.g. A-Morphous Morphology: Anderson 1992 or Paradigm Function Morphology: Stump 2001). With Blevins (2006) and Ackerman et al. (2009), a different type of description with an *abstractive* approach has appeared built around Information Theory (Shannon, 1948), word-based and revolving around the Paradigm Cell Filling Problem (PCFP) in (1).

(1) Given exposure to an inflected wordform of a novel lexeme, what licenses reliable inferences about the other wordforms in its inflectional family?

This has lead to a new line of word-based descriptions hingeing on implicative relations between forms, for example the *dynamic principal parts* of Stump & Finkel (2013), or the *joint predictiveness* of Bonami & Beniamine (2016)).

In this paper, we revisit French conjugation and the analysis of defectiveness as suppletion of a null stem proposed by Boyé (2000) and Boyé & Cabredo Hoffherr (2010) in the light of implicative morphology.

2 Defectiveness as suppletion of a null stem

The analysis proposed by Boyé & Cabredo Hoffherr relied on a partition of the conjugation paradigm into morphomic zones systematically sharing the same stem. Figure 1 illustrates the French conjugation stem space of Bonami & Boyé (2002). The colors indicate the cells sharing the same stem, the Grace convention (Rajman et al., 1997) is used for the labels.[1]
In this case as in the case of Spanish conjugation, the shape of the stem space corresponded to the frontiers of defectiveness. French verb CLORE, for example, lacks indicative imperfective, simple past and subjunctive imperfective entirely but it also has gaps for indicative present 1PL and 2PL following the outline of the gray and the mid green stem space.

The proposal was straightforward: a null stem blocked the derivation of inflectional forms based on it. For a syntagmatic account of inflection cast, for example, in Paradigm Function Morphology (Stump, 2001), the selection of a null stem would prevent the production of an inflected form.

However, in an abstractive word-based approach to inflectional morphology, this type of analysis is not possible anymore because stems do not have a primary place in the inflectional system.

[1] For finite forms, the label is composed of three parts: one character for the mode, on character for the tense, and two the person. For example, pi1S stands for present indicative 1st person singular.

pi1S	pi2S	pi3S	pi1P	pi2P	pi3P
ii1S	ii2S	ii3S	ii1P	ii2P	ii3P
fi1S	fi2S	fi3S	fi1P	fi2P	fi3P
pc1S	pc2S	pc3S	pc1P	pc2P	pc3P
ps1S	ps2S	ps3S	ps1P	ps2P	ps3P
ai1S	ai2S	ai3S	ai1P	ai2P	ai3P
is1S	is2S	is3S	is1P	is2P	is3P
—	pI2S	—	pI1P	pI2P	—
inf	pP	ppMS	ppMP	ppFS	ppFP

Figure 1: The stem space of French conjugation according to Bonami & Boyé (2002)

3 Defectiveness as the remainder of predictability

In this new context, predictiveness cannot be used to propagate defectiveness because most lexemes appear with many gaps in any given corpus. As noted by Bonami & Beniamine (2016) and Boyé & Schalchli (to appear), the forms found even in large samples cover only a part of the grammatically defined paradigms, and the problem of defectiveness seems not to be how to predict it but rather how it is avoided by speakers even in the face of very sparsely populated paradigms with many missing forms: the PCFP for all verbs.

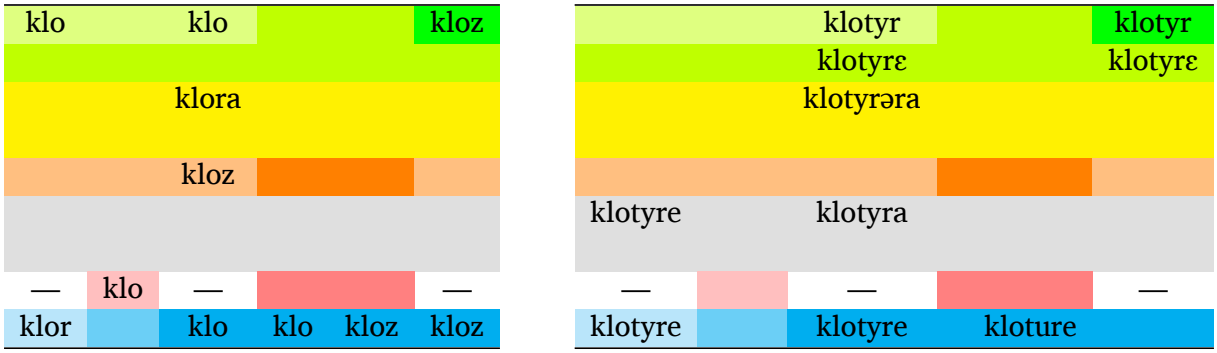

Figure 2: The forms of CLORE ('to close') and CLÔTURER ('to close') found in Lexique3

To capture both the filling strategy and the defectiveness phenomena we propose an analysis based on two steps.

First, we use SWIM (Boyé, 2017) to fill gaps by evaluating the converging predictions made by existing co-forms for every pair of cells, searching for the largest cliques of concurring predictions to fill the paradigm of each and every lexeme. This allows to generalize inflection classes without exemplary paradigms but with sufficient partial paradigms to cover the whole system. Because of its cliquing mechanism, SWIM does not allow to generalise inflectional classes that do not possess a complete cover.

For CLÔTURER, the sample does not contain an exemplary paradigm but every pair of cells in its inflection class is documented and SWIM can infer content of all the missing cells and find a clique of related forms that almost fills the whole paradigm (Fig. 3). But for CLORE, there is no support for an inflectional class that would be congruent with the known forms and the generalisations possible will be limited to partial generalisations over a very small subset of cells (Fig. 4). This leads to a low recall for this type of verbs.

To fix this caveat, we propose, in a second step, to use zero-entropy rules to predict the remaining missing forms. To avoid marginal rules, we filter zero-entropy rules based on their scopes, i.e. the number of co-pairs involved in establishing the rules in question (100 co-forms

klotyr	klotyr	klotyr	klotyrɔ̃	klotyre	klotyr
klotyrɛ	klotyrɛ	klotyrɛ	klotyrjɔ̃	klotyrje	klotyrɛ
klotyrərɛ	klotyrəra	klotyrəra	klotyrərɔ̃	klotyrəre	klotyrərɔ̃
klotyrərɛ	klotyrərɛ	klotyrərɛ	klotyrərjɔ̃	klotyrərje	klotyrərɛ
klotyr	klotyr	klotyr	klotyrjɔ̃	klotyrje	klotyr
klotyrɛ	klotyra	klotyra	klotyram	klotyrat	klotyrɛr
klotyras	klotyras	klotyra	klotyrasjɔ̃		klotyras
—	klotyr	—	klotyrɔ̃	klotyre	—
klotyre	klotyrâ	klotyre	klotyre	klotyre	klotyre

Figure 3: The forms of CLÔTURER (‘to close’) generated by SWIM

klo		klo			kloz
	klora	klora			
kloz	kloz	kloz			kloz
—	klo	—			—
klor		klo	klo	kloz	kloz

Figure 4: The forms of CLORE (‘to close’) generated by SWIM

at least). At this stage, we use the zero-entropy rules on all available co-forms to get predictions for missing forms. If all zero-entropy predictions for a cell concur, the gap is filled with the common prediction, if the zero-entropy rules diverge, the gap remains.

With this method, gaps are filled if they belong to a zero-entropy zone where at least one form is known. The emerging zones are similar as those calculated by Bonami & Boyé (2014). While these generalisations fill the gaps for “abstractible” stems, they will leave blank the gaps for “unknown” stems leading to defectiveness that will outline a space equivalent to a stem space.

4 Conclusion

The emergence of stem space like zones in abstractive morphology is natural. As Bonami & Boyé (2014) observed, zero-entropy zones are the image of what the authors used to consider stem spaces and what Stump & Finkel (2013) call distillations. In the case of defectiveness however, what we observe is the negative image of stem spaces of distillation: the lack of a base for predictions creates a dark hole in the paradigm that conforms to the shapes of neighbouring distillations leaving an empty space corresponding to a set of distillations.

Working with realistic data, a lot remains to be desired in the processing, on one hand some zero-entropy rules emerge in random places, even after filtering, and create alternate propositions preventing some correct predictions to emerge, on the other hand, some zones that should be identified as zero-entropy zones fail to appear because of a lack of information in the initial sample, but a new analysis of defectiveness is definitely possible along these lines.

In this perspective, it is natural for defectiveness to have gaps that have the same shape as stem spaces not because they are the suppletion of a null stem but rather because they are what is left after all other generalisations have been made. Defectiveness corresponds lexemes

belonging to inflectional patterns that can't be generalised because they possess some forms that block the general cliquing and gaps that correspond incidentally to a stem space.

References

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

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

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

Bonami, Olivier & S. Beniamine. 2016. Joint predictiveness in inflectional paradigms. *Word Structure* 9(2). 156–182.

Bonami, Olivier & Gilles Boyé. 2002. Suppletion and dependency in inflectional morphology. In Frank van Eynde, Lars Hellan & Dorothee Beermann (eds.), *Proceedings of the hpsg-2001 conference*, Stanford: CSLI Publications.

Bonami, Olivier & Gilles Boyé. 2014. De formes en thèmes. *Foisonnements morphologiques. Etudes en hommage à Françoise Kerleroux* 17–45.

Boyé, Gilles. 2000. *Problèmes de morpho-phonologie verbale en français, espagnol et italien*: Université Paris 7 dissertation.

Boyé, Gilles. 2017. Inflectional morphology without inflectional classes: a french conjugation reflection. Paper presented at ISMo17.

Boyé, Gilles & Patricia Cabredo Hoffherr. 2010. Defectiveness as stem suppletion in French and Spanish verbs. In Matthew Baerman, Greville G. Corbett & Dunstan Brown (eds.), *Defective paradigms: Missing forms and what they tell us* (Proceedings of the British Academy 163), 35–52. The British Academy Oxford: Oxford University Press.

Boyé, Gilles & Gauvain Schalchli. to appear. Paradigms in realistic data: The paradigm cell finding problem. *Morphology* .

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

Corbett, Greville G. & Norman M. Fraser. 1993. Network morphology: a DATR account of Russian nominal inflection. *Journal of Linguistics* 29. 113–142.

Kilani-Schoch, Marianne & Wolfgang Dressler. 2005. *Morphologie naturelle et flexion du verbe français*. Tübingen: Gunter Narr Verlag.

Rajman, Martin, Josette Lecomte & Patrick Paroubek. 1997. *Format de description lexicale pour le français. Partie 2: Description morpho-syntaxique*. Tech. Rep. GRACE, GTR 3-2.1 LIMSI.

Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27. 379–423 & 623–656.

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

Stump, Gregory T. & Raphael Finkel. 2013. *Morphological typology: From word to paradigm*. Cambridge: Cambridge University Press.