



HAL
open science

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

Gilles Boyé

► **To cite this version:**

Gilles Boyé. Stem spaces in abstractive morphology: a look at defectiveness in French conjugation. International Symposium of Morphology, Sep 2019, Paris, France. hal-04873605

HAL Id: hal-04873605

<https://hal.science/hal-04873605v1>

Submitted on 8 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

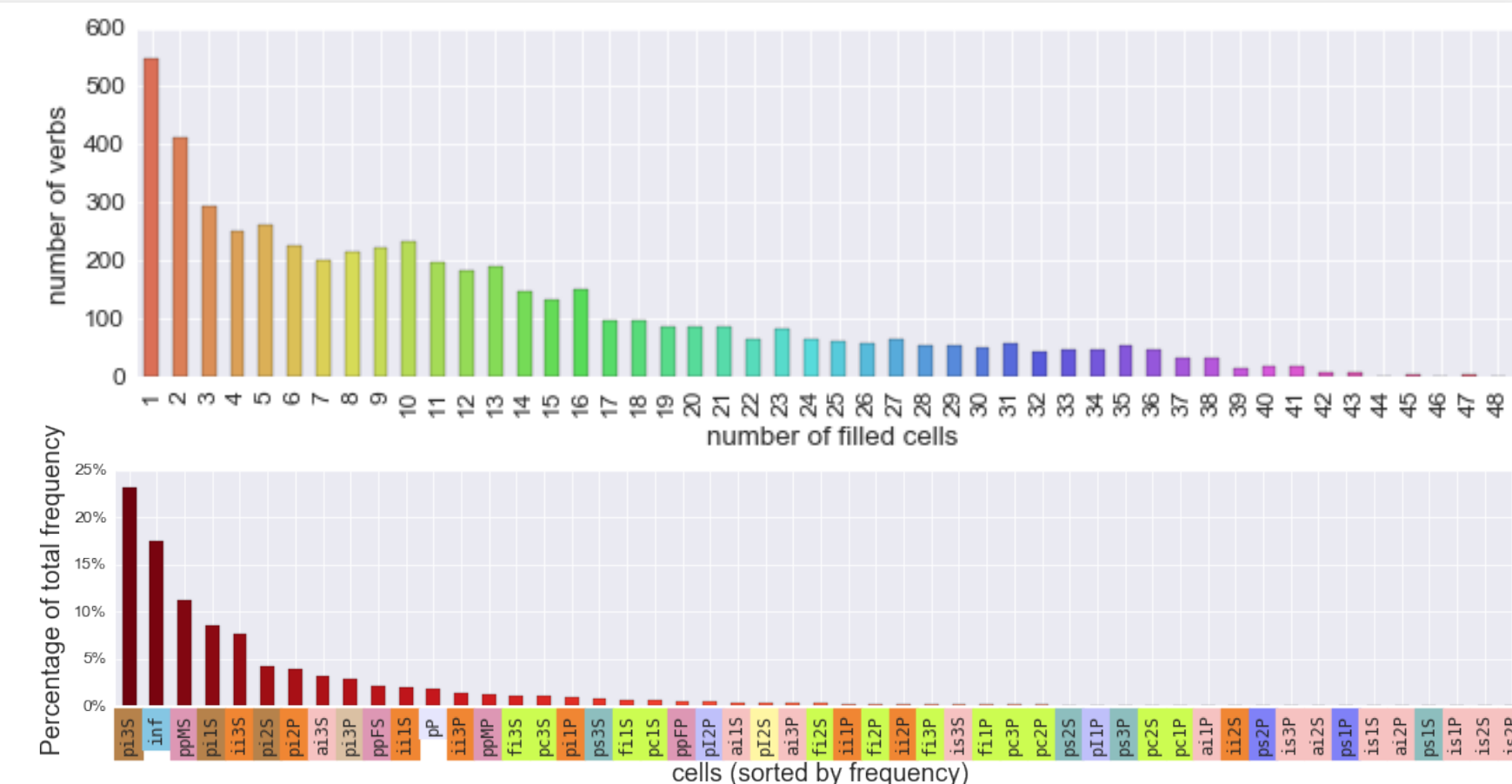
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Defective verbs in French

Find a lexicon with frequencies, Get paradigms

- Form lexicons with phonology & frequency
 - Lexique3 (New, 2006)
- Frequency-based sampling
 - 5384 verbs (73539 forms):
 - no exemplary paradigm
 - 38 verbs with more than 40 forms
 - 1593 verbs with less than 5 forms
 - 548 verbs with 1 form



Find defective paradigms, Get their distributions

- Traditional grammars consider:
 - fully inflecting verbs
 - defective verbs missing parts of the paradigm
 - clore, enclorre, éclore
 - abstraire, braire, distraire, soustraire, traire
 - vestigial verbs
 - *gésir, *apparoir,



Basic Data

TRAIRE	1sg	2sg	3sg	1pl	2pl	3pl
ind. prs	trae	trae	trae	trɛjɔ	trɛje	trɛ
ind. ipf	trɛje	trɛje	trɛje	trɛjɔ	trɛje	trɛje
ind. fut	trɛra	trɛra	trɛrɔ	trɛrɔ	trɛrɔ	trɛrɔ
cond. prs	trɛra	trɛra	trɛrɔ	trɛrɔ	trɛrɔ	trɛrɔ
subj. prs	trɛ	trɛ	trɛ	trɛjɔ	trɛje	trɛ
ind. ps	?	?	?	?	?	?
subj. ipf	?	?	?	?	?	?
imper. prs	—	trɛ	—	trɛjɔ	trɛje	—
inf.	trɛr	trɛr	trɛ	trɛ	trɛr	trɛr
part. prés.	trɛr	trɛr	trɛ	trɛ	trɛr	trɛr
part. passé	trɛr	trɛr	trɛ	trɛ	trɛr	trɛr
non-fini	trɛr	trɛr	trɛ	trɛ	trɛr	trɛr

OUTLINE

Boyé & Cabredo Hofherr (2010) defines defectiveness as the result of three different processes: i) a frequency effect, ii) an indeterminacy effect, iii) a taboo effect. They propose that all three correspond to the suppletion of a null-stem in the stem space of a lexeme blocking the realization of the corresponding part of the paradigm.

The analysis proposed here recasts Boyé & Cabredo Hofherr's insights in an abstractive approach.

The gist of the argument is that defectiveness does not rely solely on particular frequency distribution but also on partial paradigm with no PCFP solution: incompletable paradigms.

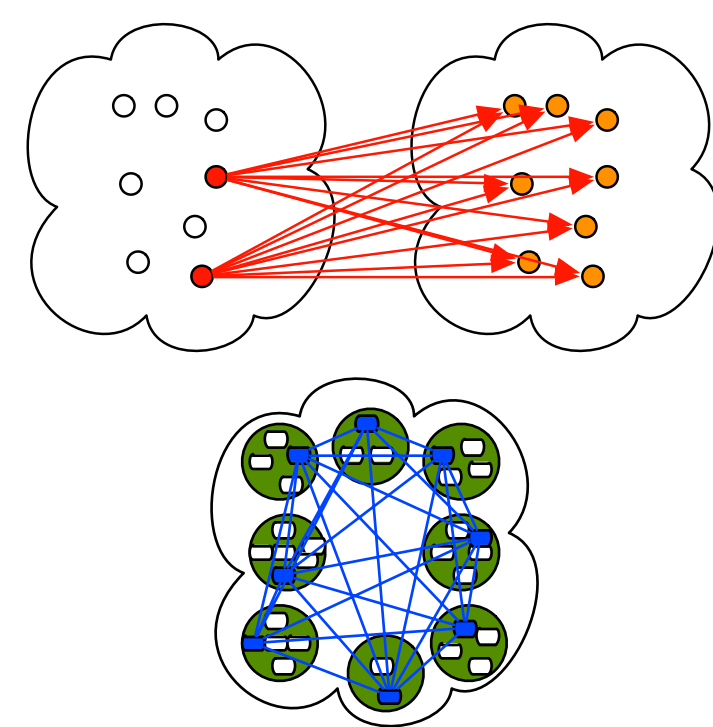
TRAIRE	1sg	2sg	3sg	1pl	2pl	3pl
ind. prs	trae	trae	trae	?	?	?
ind. ipf	?	?	trɛje	?	?	?
ind. fut	?	?	trɛra	?	?	?
cond. prs	?	?	trɛra	?	?	?
subj. prs	trɛ	?	?	?	?	?
ind. ps	?	?	?	?	?	?
subj. ipf	?	?	?	?	?	?
imper. prs	—	trɛ	—	?	?	—
inf.	trɛr	trɛr	trɛ	trɛ	?	?
part. prés.	trɛr	trɛr	trɛ	trɛ	?	?
part. passé	trɛr	trɛr	trɛ	trɛ	trɛr	trɛr
non-fini	trɛr	trɛr	trɛ	trɛ	?	?

TRAIRE	1sg	2sg	3sg	1pl	2pl	3pl
ind. prs	trae	trae	trae	trɛjɔ	trɛje	trɛ
ind. ipf	trɛje	trɛje	trɛje	trɛjɔ	?	trɛje
ind. fut	trɛra	trɛra	trɛrɔ	?	?	trɛrɔ
cond. prs	trɛra	trɛra	trɛrɔ	?	?	trɛrɔ
subj. prs	trɛ	trɛ	trɛ	?	?	trɛ
ind. ps	?	?	?	?	?	?
subj. ipf	?	?	?	?	?	?
imper. prs	—	trɛ	—	trɛjɔ	trɛje	—
inf.	trɛr	trɛr	trɛ	trɛ	trɛr	trɛr
part. prés.	trɛr	trɛr	trɛ	trɛ	trɛr	trɛr
part. passé	trɛr	trɛr	trɛ	trɛ	trɛr	trɛr
non-fini	trɛr	trɛr	trɛ	trɛ	trɛr	trɛr

An abstractive analysis for stem space

Paradigm Cell Filling with SWIM

- Preparing cells and forms:
 - Making optimal paradigms (Boyé & Schalchli 2016, 2019)
 - grouping cells with compatible content
 - Abstracting analogies (Bonami & Boyé 2014, Boyé 2016)
 - phonological transformations between pairs of cells
 - statistical distributions for output
 - Generating forms (Boyé 2015, 2016)
 - using all known forms
 - filling all cells
- Extracting paradigms:
 - Cliquing forms
 - inflectional forms are all analogically connected
 - Comparing candidates
 - the best clique is the largest and the most cohesive



Zero Entropy

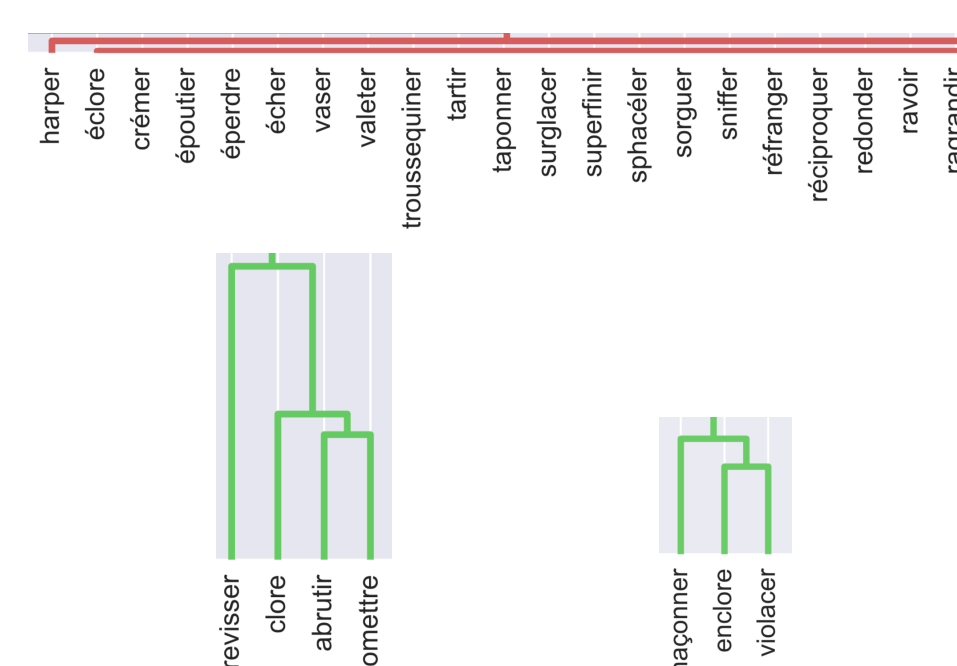
- Bonami & Boyé (2014):
 - Abstractive analysis
 - cell to cell analogies
 - whole lexicon
 - New partition
 - zero entropy zones
 - similar to stem space
 - but different

Generalizations

- SWIM:
 - relies only on cliquing analogies
 - higher precision
 - lower recall
 - Zero Entropy:
 - fills gaps with reliable analogies
 - one rule
 - 100+ co-pairs

Clusterize distributions, Compare with neighbors

- When making distribution clusters:
 - most defectives pattern with fully inflecting verbs
 - éclore, ravoir & crémer, vaser, sniffer
 - defectives that share the same paradigm separate
 - clore, éclore, enclorre



Full inflection

- incomplete paradigm
- paradigm cell filling
- zero entropy extension
- completion

Defectiveness

- incomplete paradigm
- paradigm cell filling
- zero entropy extension
- residual zero-entropy gaps

CONCLUSION

Our abstractive analysis based on Lexique3 sparse data solves partially the Paradigm Cell Filling Problem. However, for defectiveness, it runs into problems because of the interaction between overabundance and entropy. The original insights of Boyé & Cabredo Hofherr are captured without using a stem space, suppletion of a null-stem, or stems at all. It relies on the fact i) that defective verbs cannot be generalized from the known forms using any of the patterns from the fully inflecting lexemes, ii) that any zero-entropy zone will be filled if one of its member is known, leaving gaps composed of zero-entropy zones.

Selected 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.

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

Boyé, Gilles. 2015. *Small worlds inflectional morphology: A fragment for french conjugation*. Paper presented at CMDTM 2015.

Boyé, Gilles. 2016. Pour une modélisation surfaciste de la flexion, le cas de la conjugaison du français. In *SHS web of conferences*, vol. 27, EDP Sciences.

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

Boyé, Gilles & Patricia Cabredo Hofherr. 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. 2016. The status of paradigms. In Andrew Hippisley & Gregory Stump (eds.), *The cambridge handbook of morphology*. Cambridge Handbooks in Language and Linguistics, 206–234. Cambridge: Cambridge University Press.

Boyé, Gilles & Gauvain Schalchli. 2019. *Realistic data and paradigms: the paradigm cell finding problem*. *Morphology* 29(2). 199–248.

New, Boris. 2006. Lexique 3: Une nouvelle base de données lexicales. In *Actes de la Conférence Traitement Automatique des Langues Naturelles (TALN)*, 892–900.

Other defective verbs

Romani verbs

- marave
 - j'veis lui marave sa gueule à alias
 - elle lui a marave sa race à la prof
 - il lui marave la tronche quand il veut
- chourave, bicrave, bédave, ...

Verlan verbs

- pécho
 - comme si j'avais du mal à me pécho un keum
 - un coureur de la Telekom s'est fait pécho...
 - pécho les reunois, pécho les reubeus, take no prisoners
- téje, kèn, beflan, ...

and recently on Twitter

- follow
 - je te follow...
 - il m'a follow donc je l'ai follow
 - arrête de me follow !

	1sg	2sg	3sg	1pl	2pl	3pl
ind. prs	folo	folo	folo	?	?	folo
ind. ipf	?	?	?	?	?	?
ind. fut	?	?	?	?	?	?
cond. prs	?	?	?	?	?	?
subj. prs	?	?	?	?	?	?
ind. ps	?	?	?	?	?	?
subj. ipf	?	?	?	?	?	?
imper. prs	—	folo	—	?	?	—
inf.	folo	folo	folo	folo	folo	folo
part. prés.	folo	folo	folo	folo	folo	folo
part. passé	folo	folo	folo	folo	folo	folo
non-fini	folo	folo	folo	folo	folo	folo

Results

SWIM1 SWIM2

MSP		OMP		MSP		OMP	
precision	recall	precision	recall	precision	recall	precision	recall
98.6 %	72.8 %	98.6 %	80.2 %	98.4 %	82.3 %	98.2 %	89.5 %
98.4 %	88.7 %	98.5 %	92.7 %	98.2 %	93.3 %	98.2 %	96.1 %

The zero-entropy generalizations fill out all the simple past and imperfective subjunctive if one form of the zone is in the SWIM output.

Problems

Zero-entropy and overabundance

- Zones of systematic overabundance
 - with sparse input data
 - the overabundance is not systematic
 - but it can be detected by reflexive relations
 - generalization in output data
 - is made via multiple outputs
 - making the entropy non null
 - Zero-entropy zones with systematic overabundance are missed
- Zone of asymmetric zero-entropy
 - are not detected

