Quantitative evidence for paradigm structure

Olivier Bonami

Université Paris Diderot

18th International Morphology Meeting
Budapest, June 10, 2018
An inflectional morphologist’s view on derivational paradigms I

- The idea of a paradigmatic view of derivational morphology is certainly not new
  - See among many others: van Marle (1984); Becker (1993); Bochner (1993); Booij (1997); Pounder (2000); Roché et al. (2011)
- Yet this idea has been faced with skepticism by many, in particular by many inflectional morphologists.
- I see three immediate causes for this:
  1. Unclarity as to what the term ‘paradigm’ designates.
  2. Purported properties setting apart derivation from inflection
  3. The fact that our conceptualizations of inflectional paradigms and derivational families seem incompatible.
- The present talk reflects my own point of view on the issue: I will present arguments meant to convince the skeptical inflectional morphologist.
An inflectional morphologist’s view on derivational paradigms II

I will make 3 points:

1. As we learn more about inflection systems, we have fewer reasons to believe that inflection and derivation differ in the relevant ways.
2. A common conceptualization encompassing both inflectional paradigms and structured derivational families is possible.
3. Arguments for paradigmatic organization in inflection can be redeployed fruitfully in the context of derivation.

Abstractive point of view (Blevins, 2006): focus on relations between surface words, as they can be inferred from direct observations of usage.

Instrumented approach:

- Generalizations are extracted from large lexica and/or corpora
- Computational implementation provides an operational, fully explicit formulation of linguistic hypotheses.

I focus mainly on French.
Renouncing skepticism
Classical arguments against derivational paradigms

- Derivational families can not be structured into paradigms because...
  1. **Lexical gaps**: Paradigms are supposed to be exhaustive, but derivational families are full of gaps.
  2. **Variation**: Paradigms are supposed to have a unique form in each of their cells, but derivational families contain lots of doublets.
  3. **Semantic irregularity**: Paradigms are supposed to encode reliable contrasts, but derived forms differ in unpredictable ways from their bases.

- In each case, I will argue that what we have learned on inflection in the past two decades changes the picture.
1. Renouncing skepticism

1.1. Gaps
Gaps

The skeptic’s argument:
- Postulating paradigms supposes that we have words to fill the cells in these paradigms.
- In inflection this is fine, because inflection is “fully productive”.
  - This has to be so, otherwise the demands of syntax could not be met (“inflection is obligatory”).
- On the other hand, derivation is usually less than fully productive: there are lots of gaps.
  - This has to be so, because new lexemes are coined only as the need for them arises.
- So, paradigms in derivation do not make sense because they would be hollow.
Problem 1: the requirements of syntax

- Paradigm cells exhibit a Zipfian distribution (Blevins et al., 2017).

Frequency of verbs by paradigm cell in the French Treebank (Abeillé et al., 2003)
Problem 1: the requirements of syntax

- As a result, even at very large corpus sizes, inflectional paradigms do not “fill up” on average (Bonami and Beniamine, 2016).

Average number of distinct orthographic forms for verbs from the Lefff lexicon (Sagot, 2010) when progressing through the FrWac corpus (Baroni et al., 2009)
Problem 2: “full productivity”

- Although syntax may require any forms of any lexeme, most forms of most lexemes will never be required.
- Given this, it is unclear what “full productivity” means.
  - Operational measures of productivity (Baayen, 2001; O’Donnell, 2015) are inherently gradient.
  - As Gaeta (2007) shows, some inflectional processes are less productive than some derivational processes.
- This strongly suggests that, while inflectional relations may be more productive than derivational relations on average, they are not in general.
Problem 3: Defectiveness

We are used to thinking of defectiveness as an anomaly, unlike lexical gaps.

The notion of defectiveness itself is gradient (Sims, 2015):

- Defective forms are usually attested in large enough corpora.
- Note the contrast with the fact that many nondefective forms are not attested.
- Defectiveness is the failure for a form to reach an expected frequency of occurrence, given prior knowledge on the frequency distribution of inflected forms for comparable lexemes.
- Crucially, defectiveness is thus doubly gradient:
  - The frequency may be more or less close to zero
  - The unexpectedness of that frequency may be more or less large.
- No reason to think that the same does not hold for “lexical gaps”.
1. Renouncing skepticism

1.2. Variation
The skeptic’s argument:

- Postulating paradigms supposes that we can identify a unique word to fill each paradigm cell.
- In inflection this is fine, because doublets are vanishingly rare. Exceptions can and should be reduced.
  - This has to be so, because inflection is a function (Stump, 2001; Bonami and Boyé, 2007).
- In derivation, more often than not, there are multiple lexemes for the same derivational category, which may or may not contrast semantically (Fradin, to appear).
- So, paradigms in derivation do not make sense because cells would be overpopulated.
Thornton (2011, 2012, forthcoming) was instrumental in demonstrating that overabundance is a real and widespread phenomenon, directly falsifying the claim that doublets do not occur in inflection.

Hence, if there is a difference between inflection and derivation here, it is at most a difference of extent.

So, what is the extent of the difference?
Overabundance II

Guzman Naranjo and Bonami (2016) on Czech nominal declension:

<table>
<thead>
<tr>
<th>NOM</th>
<th>GEN</th>
<th>DAT</th>
<th>ACC</th>
<th>VOC</th>
<th>LOC</th>
<th>INS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG</td>
<td>1.3%</td>
<td>2.8%</td>
<td>1.2%</td>
<td>2.1%</td>
<td>0.7%</td>
<td>10.0%</td>
</tr>
<tr>
<td>PL</td>
<td>8.6%</td>
<td>2.5%</td>
<td>4.2%</td>
<td>1.6%</td>
<td>1.5%</td>
<td>4.9%</td>
</tr>
</tbody>
</table>

Proportions of lexemes attested in more than one form for each paradigm cell – SYN v4 corpus (Hnátková et al., 2014, 4.3 billion tokens), forms validated in the MorfFlex lexicon (Hajič and Hlaváčová, 2013)

Compare numbers for French derivational families documented in the *Démonette* database (Hathout and Namer, 2014).

<table>
<thead>
<tr>
<th>Morphosemantic category</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>1.6%</td>
</tr>
<tr>
<td>Action noun</td>
<td>16.5%</td>
</tr>
<tr>
<td>Agent noun</td>
<td>0.7%</td>
</tr>
</tbody>
</table>

Proportions of categories attested in the form of more than one lexeme in the FrWaC corpus (Baroni et al., 2009, 1.6 billion tokens)
Although a more principled comparison is in order, the evidence points to comparable amounts of overabundance in inflection and derivation.
1. Renouncing skepticism

1.3. Stability of contrast
Setting the stage

The skeptic’s argument:

- The syntactic and semantic contrast between cells in an inflectional paradigm is *stable* across lexemes: e.g. the opposition between present and past is the same for all verbs.

\[
\frac{\text{laughs}}{\text{laughed}} = \frac{\text{wash}}{\text{washed}} = \frac{\text{pay}}{\text{paid}}
\]

- On the other hand, the meaning and distribution of a derived lexeme is somewhat unpredictable, and hence the contrasts between lexemes standing in the same derivational relation is somewhat unstable across lexemes.

\[
\frac{\text{laugh}}{\text{laughable}} \neq \frac{\text{wash}}{\text{washable}} \neq \frac{\text{pay}}{\text{payable}}
\]

- As a result, derivational families can’t be structured in paradigms, because we can’t decide what counts as “filling the same cell”.

- Bonami and Paperno (submitted) explores the issue of *stability of contrasts* in inflection and derivation using a distributional approach.
Distributional semantics in a nutshell

The distributional hypothesis (see also Harris 1954; Firth 1957): The degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear. (Lenci, 2008, 3)

Contemporary computational linguistics operationalizes this idea to deduce semantic representations from large corpora.

Toy example: we start with a cooccurrence table:

<table>
<thead>
<tr>
<th></th>
<th>ride</th>
<th>eats</th>
</tr>
</thead>
<tbody>
<tr>
<td>dog</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>horse</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>car</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>
Distributional semantics in a nutshell II

Such cooccurrence counts are vectors:

<table>
<thead>
<tr>
<th></th>
<th>ride</th>
<th>eats</th>
</tr>
</thead>
<tbody>
<tr>
<td>dog</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>horse</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>car</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

In practice:

Realistic representations rely on cooccurrences with very large lexica in large corpora ⇒ many more dimensions.

For efficiency reasons, most current systems rely on prediction tasks rather than explicit cooccurrence counts to infer vector representations (see e.g. Mikolov et al., 2013).

These technical aspects can be ignored here.
One highly relevant application: proportional analogies through vector arithmetics (Mikolov et al., 2013)
Distributional semantics in a nutshell IV

- Proportional analogy works to the extent that differences between pairs of words are similar.

- These difference vectors represent the shift in distribution from one word to the next.

- Studying the similarity of these difference vectors, tells us about stability of contrasts.
In this paper, we made systematic comparisons between inflectional and derivational relations in French.

We looked at 174 pairings of an inflectional and a derivational relation.

<table>
<thead>
<tr>
<th>Inflectional relation</th>
<th>vs.</th>
<th>Derivational relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>INF verb ~ 3SG imperfect verb</td>
<td>vs.</td>
<td>INF verb ~ SG action noun</td>
</tr>
<tr>
<td>PL agent noun ~ SG agent noun</td>
<td>vs.</td>
<td>PL agent noun ~ present participle of verb</td>
</tr>
</tbody>
</table>

We showed that in 172 out of 174 cases, contrasts are significantly more stable ($p < 0.001$) within the inflectional relation than within the derivational relation.
Bonami and Paperno (submitted) confirms received wisdom: when making strictly parallel comparisons, inflectional contrasts are systematically less diverse than derivational contrasts.

Note though that the difference is a matter of quantity: inflectional constrats are not absolutely stable.

In addition, these results are compatible with a situation where inflection and derivation only tend to occupy two extremes of a gradient, with some overlap in the middle.

We now compare systematically the similarity among shift vectors for 471 morphological relations documented in our dataset.
Indeed, while derivational relations are on average less stable than inflectional relations, there is no categorical cutting point.
We have looked at three skeptical arguments against derivational paradigms based on three purported categorical differences between inflection and derivation:

1. Productivity (and the status of gaps)
2. Variation
3. Stability of contrasts

In all three cases, we have concluded that

- The parameter in question is gradient by nature (Dressler, 1989)
- Although there might be a general tendency for inflection and derivation to occupy opposite ends of the gradient, there is overlap in the middle.

It is striking that this conclusion is reached mostly through realizing that *inflection* is not as well-behaved as previously thought.
2. An agnostic notion of paradigm
Structural prejudices I

We are attuned to thinking of inflectional paradigms as structured by orthogonal oppositions:

<table>
<thead>
<tr>
<th>SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>M buono</td>
<td>buoni</td>
</tr>
<tr>
<td>F buona</td>
<td>buone</td>
</tr>
</tbody>
</table>

Paradigm of the Italian adjective BUONO ‘good’

We are also attuned to thinking of derivational families as trees of base-derivative relations:
However, proponents of derivational paradigms repeatedly warned us as to the limitations of derivation trees:

See, among many others, Jackendoff (1975); van Marle (1984); Corbin (1987); Bochner (1993); Becker (1993); Bauer (1997); Booij (1997, 2010); Tribout (2010); Roché et al. (2011); Lignon et al. (2014); Strnadová (2014); Hathout and Namer (in preparation)
At the same time, studies of implicative relations in inflection have highlighted the fact that predictability relations need not be morphosyntactically motivated.

Matthews (1972); Wurzel (1984); Aronoff (1994); Brown (1998); Pirrelli and Battista (2000); Bonami and Boyé (2002); Blevins (2003, 2006, 2016); Ackerman et al. (2009); Ackerman and Malouf (2013); Stump and Finkel (2013)

Hence, in this line of research, all pairwise relations between cells in a paradigm are equally worthy of attention.
Structural prejudices IV

This suggests that both inflectional paradigms and structured derivational families can be seen as dense networks of gradient predictability relations.

Bonami and Strnadová (2018): Such networks can be organized as paradigms if we can

- Identify syntactic/semantic contrasts (Štekauer, 2014) recurring from family to family.
- Align families on the basis of these contrasts.
Some definitions (Bonami and Strnadvová, 2018)

► Morphological family
Any set of morphologically related words.

► Paradigmatic system
Collection of morphological families exhibiting the same set of contrasts.

► Paradigm
One member of a paradigmatic system.

► Series
Set of words that enter the same set of contrasts in their respective families (Hathout and Namer, in preparation).
Some definitions (Bonami and Strnadová, 2018)

- **Morphological family**
  Any set of morphologically related words.

- **Paradigmatic system**
  Collection of morphological families exhibiting the same set of contrasts.

- **Paradigm**
  One member of a paradigmatic system.

- **Series**
  Set of words that enter the same set of contrasts in their respective families (Hathout and Namer, in preparation).

Derivational example:

```
Verb
  Action_N
  Agent_N
laver
lavage
laveur
former
formation
formateur
gonfler
gonflement
gonfleur
```
Some definitions (Bonami and Strnadová, 2018)

- **Morphological family**
  Any set of morphologically related words.

- **Paradigmatic system**
  Collection of morphological families exhibiting the same set of contrasts.

- **Paradigm**
  One member of a paradigmatic system.

- **Series**
  Set of words that enter the same set of contrasts in their respective families (Hathout and Namer, in preparation).
Remarks

- Two primitives for the definitions:
  - Morphological relatedness
  - Set of relevant syntactic/semantic contrasts
- We do not define paradigmatic systems as exhaustive, neither vertically nor horizontally.
  - No claim that families are bounded, or that exhaustive families have the exact same shape.
  - On the other hand, we can cut bounded slices in piles of partial families.
  - Classical inflectional paradigms are such slices.
- Gaps (defectivity) or synonymy within a paradigm (overabundance) can be dealt with using slightly more complex definitions.
  - Higher-order notion of paradigms as aligned families of sets of words.
- Aligning relations can be fine-grained or coarse-grained
  - Multiple ways of choosing relevant contrasts for different purposes.
In a paradigmatic system, the same contrasts may be encoded in different ways for different paradigms.

This is true both for inflectionally and derivationally-related words.

<table>
<thead>
<tr>
<th>NOM.SG</th>
<th>GEN.PL</th>
<th>TOPONYM</th>
<th>DEMONYM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) hrad</td>
<td>hradů</td>
<td>France</td>
<td>Français</td>
</tr>
<tr>
<td>(b) žena</td>
<td>žen</td>
<td>Russie</td>
<td>Russe</td>
</tr>
<tr>
<td>(c) tátá</td>
<td>tátů</td>
<td>Albanie</td>
<td>Albanais</td>
</tr>
<tr>
<td>(d) stavení</td>
<td>stavení</td>
<td>Corse</td>
<td>Corsican</td>
</tr>
</tbody>
</table>

Partial inflectional paradigms of a few Czech nouns

Partial paradigms of French toponyms and related demonyms
Fruitful analogies: Orthogonality of content and marking

- In a paradigmatic system, the formally unmarked cell (if any) need not be the same for all paradigms.
- This is true both for inflectionally and derivationally-related words.

<table>
<thead>
<tr>
<th>NOM.SG</th>
<th>GEN.PL</th>
<th>TOPONYM</th>
<th>DEMONYM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) hrad</td>
<td>hradů</td>
<td>France</td>
<td>Français</td>
</tr>
<tr>
<td>(b) žena</td>
<td>žen</td>
<td>Russie</td>
<td>Russe</td>
</tr>
<tr>
<td>(c) tátů</td>
<td>tátů</td>
<td>Albanie</td>
<td>Albanais</td>
</tr>
<tr>
<td>(d) stavení</td>
<td>stavení</td>
<td>Corse</td>
<td>Corse</td>
</tr>
</tbody>
</table>

Partial inflectional paradigms of a few Czech nouns

Partial paradigms of French toponyms and related demonyms
Fruitful analogies: Heteroclisis

- In a paradigmatic system, some paradigms may use an exponence strategy that is a hybrid of two others.
- This is true both for inflectionally and derivationally-related words.

<table>
<thead>
<tr>
<th>NOM.SG</th>
<th>GEN.PL</th>
<th>'castle'</th>
<th>'woman'</th>
<th>'dad'</th>
<th>'building'</th>
</tr>
</thead>
<tbody>
<tr>
<td>hrad</td>
<td>hradů</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>žena</td>
<td>žen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tátá</td>
<td>tátů</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stavení</td>
<td>stavení</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TOPONYM</th>
<th>DEMONYM</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>Français</td>
</tr>
<tr>
<td>Russie</td>
<td>Russe</td>
</tr>
<tr>
<td>Albanie</td>
<td>Albanais</td>
</tr>
<tr>
<td>Corse</td>
<td>Corse</td>
</tr>
</tbody>
</table>

Partial inflectional paradigms of a few Czech nouns

Partial paradigms of French toponyms and related demonyms
Fruitful analogies: Syncretism

- In a paradigmatic system, some paradigms may fail to contrast formally words that contrast in content.
- This is true both for inflectionally and derivationally-related words.

<table>
<thead>
<tr>
<th>NOM.SG</th>
<th>GEN.PL</th>
<th>TOPONYM</th>
<th>DEMONYM</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) hrad</td>
<td>hradů</td>
<td>France</td>
<td>Français</td>
</tr>
<tr>
<td>(b) žena</td>
<td>žen</td>
<td>Russie</td>
<td>Russe</td>
</tr>
<tr>
<td>(c) tátu</td>
<td>táta</td>
<td>Albanie</td>
<td>Albanais</td>
</tr>
<tr>
<td>(d) stavení</td>
<td>stavení</td>
<td>Corse</td>
<td>Corse</td>
</tr>
</tbody>
</table>

Partial inflectional paradigms of a few Czech nouns

Partial paradigms of French toponyms and related demonyms
I have argued that conventional representations for inflectional paradigms and derivational families distract us from important structural similarities between the two. This is not to say that these representations do not teach us something interesting, e.g. for the study of exponence or lexical innovation.

I have proposed a general definition of paradigmatic systems that is

- Agnostic to the differences between inflection and derivation
- Crucially partial: Different partial paradigms can be studied for different purposes

I have shown how this definition can be used to draw fruitful analogies between phenomena in inflection and derivation.

Next step: discuss evidence that derivational paradigms have nontrivial structure.
3. Predictability of form in inflectional and derivational paradigms
The PARADIGM CELL FILLING PROBLEM: What licenses reliable inferences about the inflected (and derived) surface forms of a lexical item?
(Ackerman et al., 2009, 54)

- IMPLICATIVE STRUCTURE (Wurzel, 1984) is crucial.
- Since Ackerman et al. (2009), emerging tradition of assessing implicative structure through CONDITIONAL ENTROPY: a measure of how difficult it is to predict the form filling cell $B$ knowing the form filling cell $A$.
  - See Ackerman et al. (2009); Ackerman and Malouf (2013); Blevins (2016); Bonami and Boyé (2014); Bonami and Luís (2014); Sims (2015); Bonami and Beniamine (2016); Sims and Parker (2016); Beniamine (forthcoming).
  - Here we follow the methodology of Bonami and Beniamine (2016).
Predictability in paradigms II

- Although from the outset the PCFP was presented as a problem for both inflection and derivation, later empirical studies have focused on inflection.
- Bonami and Strnadová (2018) applies the same methods to derivational paradigmatic systems.
- Two families of results that justify the importance of (implicative) paradigm structure:
  - Differential predictability
  - Joint predictiveness
- We first present these on a simple inflectional example, and then show parallel results on a derivational example.
Differential predictability in inflection

- Reliability of prediction depends on minute relations between the forms filling two paradigm cells.
- Hence, reliability of prediction varies pair of cells by pair of cells.
- Illustration with French adjectives:

<table>
<thead>
<tr>
<th></th>
<th>F.SG</th>
<th>F.PL</th>
<th>M.SG</th>
<th>M.PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>F.SG</td>
<td>—</td>
<td>0</td>
<td>0.213</td>
<td>0.231</td>
</tr>
<tr>
<td>F.PL</td>
<td>0</td>
<td>—</td>
<td>0.213</td>
<td>0.231</td>
</tr>
<tr>
<td>M.SG</td>
<td>0.641</td>
<td>0.641</td>
<td>—</td>
<td>0.018</td>
</tr>
<tr>
<td>M.PL</td>
<td>0.666</td>
<td>0.666</td>
<td>0.041</td>
<td>—</td>
</tr>
</tbody>
</table>

Unary implicative entropy between paradigm cells in French adjectives (data from Bonami et al. 2014)
Differential predictability in derivation

We apply the same method to a dataset of 913 triples \( \langle \text{Verb, Action noun, Masculine agent noun} \rangle \) from French.

Derivational relations from the *Démonette* database (Hathout and Namer, 2014), phonemic transcriptions from the *GLÀFF* lexicon (Hathout et al., 2014).

<table>
<thead>
<tr>
<th>Family</th>
<th>Verb</th>
<th>Action noun</th>
<th>Agent noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>abaisser ‘lower’</td>
<td>a.bε.se</td>
<td>a.bε.smã;a.bεs.mã</td>
<td>a.be.sœε</td>
</tr>
<tr>
<td>abandonner ‘abandon’</td>
<td>a.bã.dɔ.ne</td>
<td>a.bã.dɔ̃</td>
<td>a.bã.dɔ.nœε</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

Results:

<table>
<thead>
<tr>
<th></th>
<th>Verb</th>
<th>Action_N</th>
<th>Agent_N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>—</td>
<td>1.115</td>
<td>0.709</td>
</tr>
<tr>
<td>Action_N</td>
<td>0.101</td>
<td>—</td>
<td>0.269</td>
</tr>
<tr>
<td>Agent_N</td>
<td>0.264</td>
<td>1.114</td>
<td>—</td>
</tr>
</tbody>
</table>

Unary implicative entropy for \( \langle \text{Verb, Action_N, Agent_N} \rangle \) triples
Differential predictability in derivation

Unary implicative entropy for (Verb, Action_N, Agent_N) triples

<table>
<thead>
<tr>
<th>Verb</th>
<th>Action_N</th>
<th>Agent_N</th>
</tr>
</thead>
<tbody>
<tr>
<td>laver</td>
<td>lavage</td>
<td>laveur</td>
</tr>
<tr>
<td>‘wash’</td>
<td>‘washing’</td>
<td>‘washer’</td>
</tr>
<tr>
<td>contrôler</td>
<td>contrôle</td>
<td>contrôleur</td>
</tr>
<tr>
<td>‘control’</td>
<td>‘control’</td>
<td>‘controller’</td>
</tr>
<tr>
<td>corriger</td>
<td>correction</td>
<td>correcteur</td>
</tr>
<tr>
<td>‘correct’</td>
<td>‘correction’</td>
<td>‘corrector’</td>
</tr>
<tr>
<td>former</td>
<td>formation</td>
<td>formateur</td>
</tr>
<tr>
<td>‘train’</td>
<td>‘training’</td>
<td>‘trainer’</td>
</tr>
<tr>
<td>couvrir</td>
<td>couverture</td>
<td>couvreur</td>
</tr>
<tr>
<td>‘write’</td>
<td>‘writing’</td>
<td>‘writer’</td>
</tr>
<tr>
<td>gonfler</td>
<td>gonflement</td>
<td>gonfleur</td>
</tr>
<tr>
<td>‘inflated’</td>
<td>‘inflating’</td>
<td>‘inflater’</td>
</tr>
</tbody>
</table>

Sample triples

- Action nouns are hardest to predict, because of the diversity of marking strategies (-age, -ment, -ion, -ure, conversion, etc.)
## Differential predictability in derivation

<table>
<thead>
<tr>
<th>Verb</th>
<th>Action_N</th>
<th>Agent_N</th>
</tr>
</thead>
<tbody>
<tr>
<td>laver</td>
<td>lavage</td>
<td>laveur</td>
</tr>
<tr>
<td>contrôler</td>
<td>contrôle</td>
<td>contrôleur</td>
</tr>
<tr>
<td>corriger</td>
<td>correction</td>
<td>correcteur</td>
</tr>
<tr>
<td>former</td>
<td>formation</td>
<td>formateur</td>
</tr>
<tr>
<td>couvrir</td>
<td>couverture</td>
<td>couvreur</td>
</tr>
<tr>
<td>gonfler</td>
<td>gonflement</td>
<td>gonfleur</td>
</tr>
</tbody>
</table>

Unary implicative entropy for (Verb, Action_N, Agent_N) triples

- **Verbs** are easiest to predict: the only challenging cases are stem suppletion and non-first conjugation.

- **Sample triples**

<table>
<thead>
<tr>
<th>Verb</th>
<th>Action_N</th>
<th>Agent_N</th>
</tr>
</thead>
<tbody>
<tr>
<td>laver</td>
<td>lavage</td>
<td>laveur</td>
</tr>
<tr>
<td>contrôler</td>
<td>contrôle</td>
<td>contrôleur</td>
</tr>
<tr>
<td>corriger</td>
<td>correction</td>
<td>correcteur</td>
</tr>
<tr>
<td>former</td>
<td>formation</td>
<td>formateur</td>
</tr>
<tr>
<td>couvrir</td>
<td>couverture</td>
<td>couvreur</td>
</tr>
<tr>
<td>gonfler</td>
<td>gonflement</td>
<td>gonfleur</td>
</tr>
</tbody>
</table>

- **Verbs** are easiest to predict: the only challenging cases are stem suppletion and non-first conjugation.
Differential predictability in derivation

<table>
<thead>
<tr>
<th>Verb</th>
<th>Action_N</th>
<th>Agent_N</th>
</tr>
</thead>
<tbody>
<tr>
<td>laver</td>
<td>lavage</td>
<td>laveur</td>
</tr>
<tr>
<td>contrôler</td>
<td>contrôle</td>
<td>contrôleur</td>
</tr>
<tr>
<td>corriger</td>
<td>correction</td>
<td>correcteur</td>
</tr>
<tr>
<td>former</td>
<td>formation</td>
<td>formateur</td>
</tr>
<tr>
<td>couvrir</td>
<td>couverture</td>
<td>couvreur</td>
</tr>
<tr>
<td>gonfler</td>
<td>gonflement</td>
<td>gonfleur</td>
</tr>
</tbody>
</table>

▶ Action nouns are good predictors of agent nouns, since they almost always use the same stem.

Unary implicative entropy for (Verb, Action_N, Agent_N) triples

<table>
<thead>
<tr>
<th>Verb</th>
<th>Action_N</th>
<th>Agent_N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>— 1.115</td>
<td>0.709</td>
</tr>
<tr>
<td>Action_N</td>
<td>0.101 —</td>
<td>0.269</td>
</tr>
<tr>
<td>Agent_N</td>
<td>0.264 1.114 —</td>
<td></td>
</tr>
</tbody>
</table>

Sample triples
Differential predictability in derivation

<table>
<thead>
<tr>
<th></th>
<th>Verb</th>
<th>Action_N</th>
<th>Agent_N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>—</td>
<td>1.115</td>
<td>0.709</td>
</tr>
<tr>
<td>Action_N</td>
<td>0.101</td>
<td>—</td>
<td>0.269</td>
</tr>
<tr>
<td>Agent_N</td>
<td>0.264</td>
<td>1.114</td>
<td>—</td>
</tr>
</tbody>
</table>

Unary implicative entropy for (Verb, Action_N, Agent_N) triples

On the other hand, verbs are not so good predictors of agent nouns, because, even in the absence of suppletion, one has to guess whether the -at- augment should be used.
Joint predictiveness in inflection

Bonami and Beniamine (2016) on Romance conjugation: on average, knowing multiple forms of the same lexeme makes the PCFP a lot easier.

For French adjectives:

<table>
<thead>
<tr>
<th>Predictor Count</th>
<th>Implicative Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 predictor</td>
<td>0.2966</td>
</tr>
<tr>
<td>2 predictors</td>
<td>0.1443</td>
</tr>
<tr>
<td>3 predictors</td>
<td>0.0044</td>
</tr>
</tbody>
</table>

Average implicative entropy

This provides a strong argument for paradigms as first class citizens of the morphological universe: there is useful knowledge on the system that can only be attained by attending to (sub)paradigms.
Joint predictiveness in derivation

Predicting from two members of a morphological family is a lot easier than predicting from just one.

<table>
<thead>
<tr>
<th>Predictor Count</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 predictor</td>
<td>0.595</td>
</tr>
<tr>
<td>2 predictors</td>
<td>0.196</td>
</tr>
</tbody>
</table>

Average implicative entropy
Joint predictiveness in derivation II

In particular, predicting the form of verbs from knowledge of the two nouns is trivial.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Predicted</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb, Action_N</td>
<td>Agent_N</td>
<td>0.138</td>
</tr>
<tr>
<td>Verb, Agent_N</td>
<td>Action_N</td>
<td>0.444</td>
</tr>
<tr>
<td>Agent_N, Action_N</td>
<td>Verb</td>
<td>0.006</td>
</tr>
</tbody>
</table>

All the remaining uncertainty is caused by a handful of -ionner verbs (Lignon and Namer, 2010).

- (Action_N, Agent_N) ⇒ Verb
- (percussion, percuteur) ⇒ percuter
- (inspection, inspecteur) ⇒ inspecter
- (perquisition, perquisiteur) ⇒ perquisitionner
- (fonction, foncteur) ⇒ fonctionner

Sample triples
Just like inflectional paradigms, derivational paradigms exhibit differential predictability and joint predictiveness.

- Although most commonly the verb is the formal base of the action noun and the agent noun, the nouns are much better predictors of the verbs than the other way around:
- Hence there is relevant information flowing from derivatives to base that speakers are likely to rely on.
- Joint predictiveness shows that global knowledge of the derivational paradigm is more informative than knowledge of individual words.
  - In particular, joint knowledge of two nouns leads to quasi-categorical knowledge on the verb.

This shows that there is irreducibly paradigmatic structure in the derivational lexicon.
4. Towards predictability of content
Predictability of content

- The PCFP is a production problem: how can speakers produce forms they do not know?
- There is a converse recognition problem: given knowledge of the lexicon and the morphological system, how can speakers assign the right meaning to an unknown word belonging to some paradigm?
- A concrete example:
  - Suppose I know the meaning of *pay, payer, payment*.
  - I now hear for the first time in context the word *payable*.
  - How easily and reliably does my knowledge of the English morphological system help me infer the meaning of *payable*?
- Three tasks:
  1. Identify the morphological family.
  2. Identify the paradigm cell.
  3. Predict meaning within that cell of that paradigm.

- (3) is the question of PREDICTABILITY OF CONTENT in paradigms.
Predictability of content, 2

▶ Just as with predictability of form:
  ▶ It could be that there are asymmetries in predictability of content.
  ▶ It could be that some words are good/bad predictors or good/bad predictees.
  ▶ It could be that joint knowledge of multiple words improves prediction dramatically.

▶ Predictability of content relates to the idea of stability of contrasts: we expect that more stable contrasts lead to more accurate prediction.

▶ We may operationalize predictability of content using the same distributional methods discussed in the first part of the talk.
An example

- The same resources and methodology used in Bonami and Paperno (submitted) can be put to use to compare stability of contrasts among verbs, action nouns and agent nouns.

<table>
<thead>
<tr>
<th>(V,N) relation</th>
<th>vs.</th>
<th>(N,N) relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG action noun ~ INF verb</td>
<td>vs.</td>
<td>SG action noun ~ SG agent noun</td>
</tr>
<tr>
<td>PL agent noun ~ present participle of verb</td>
<td>vs.</td>
<td>PL agent noun ~ SG action noun</td>
</tr>
<tr>
<td>...</td>
<td>vs.</td>
<td>...</td>
</tr>
</tbody>
</table>

- Result: (V,N) contrasts are more stable than (N,N) contrasts.
  - This is unsurprising, given that in most cases the verb is the formal base for both nouns.
  - On the other hands it confirms the validity of the methodology.
A new research question

- Is it always the case that relations between formal bases and their derivatives are semantically more predictable than relations among derivatives?
  - If not, this is more evidence for paradigmatic organization.
  - Think of *social, socialism, socialist* vs. *commune, communism, communist*

- To explore this, we need large scale documentation of derivational families.

- Demonext project (PI F. Namer, 2018–2022): stay tuned!
Thanks

- Collaborators:
  - Sacha Beniamine (U. Paris Diderot)
  - Matías Guzman Naranjo (Düsseldorf U.)
  - Timothee Mickus (U. Paris Diderot)
  - Denis Paperno (CNRS - Nancy)
  - Jana Strnadová (Google France)

- Institutions:
  - Labex EFL, Strand 2: Experimental grammar
  - ANR Project Demonext (PI Fiammetta Namer)
  - Laboratoire de linguistique formelle (U. Paris Diderot & CNRS)
References


References II


References III


Sagot, B. (2010). ‘The Leff, a freely available and large-coverage morphological and syntactic lexicon for French’. In *Proceedings of LREC 2010*.


6. Appendices

6.1. A. Bonami and Paperno (submitted)
Semantic contrasts as shift vectors

- We rely on **distributional semantics**: the meaning of a word is approximated by a high-dimensional vector representing its distribution in a corpus.

- Within such a framework, we can examine how vectors representing derivationally-related words relate to each other (Marelli and Baroni, 2015).

- Simple way of doing this: the contrast in meaning between two words is the difference between their two vectors; i.e., the vector representing what it takes to go from the meaning of one word to the meaning of the other.

![Diagram showing lavait → laver as a shift vector]

- We will call this vector the **shift vector**.
Semantic contrasts as shift vectors II

- Word vectors corresponding to the same paradigm cell will be similar in some dimensions and different in others.

- The word vectors may be very different but the shift vectors still be very similar.

- Stability of semantic contrasts amounts to similarity of shift vectors.

**NB:** We are not examining distance between word meanings but distance between shifts in meaning (compare Wauquier 2016).
The hypothesis

- We look at triples of morphologically-related forms, one of which is used as the pivot for comparison.
The hypothesis

- We look at triples of morphologically-related forms, one of which is used as the pivot for comparison.
- We compute shift vectors between the pivot and the other forms.
The hypothesis

- We look at triples of morphologically-related forms, one of which is used as the pivot for comparison.
- We compute shift vectors between the pivot and the other forms.
- We then expect the shift vectors for derivationally-related pairs to be more diverse than those for inflectionally-related pairs.
The execution, I

Vector space constructed from the FrWac corpus (Baroni et al., 2009) using word2vec (Mikolov et al., 2013).

- CBOW algorithm, window size 5, negative sampling with 10 samples, 400 dimensions

Paradigmatic system of 6576 (partial) families and 59 cells constructed from:

1. Derivational relations between verbs, action nouns and agent nouns from Démonette (Hathout and Namer, 2014)
2. Hand-constructed set of derivational relations between verbs and -able adjectives
3. Inflectional relations from the GLÀFF (Hathout et al., 2014)

We then look for triples of cells where:

1. There is a derivational relation between the first (pivot) and second cell and an inflectional relation between the first and third.
2. We have enough data to select 100 triples of words such that
   2.1 there is a single word in each cell,
   2.2 no word has homonyms,
   2.3 all words have a frequency above 50,
   2.4 the frequency ratio between the nonpivot cells is between $\frac{1}{5}$ and 5,
   2.5 the median frequency ratio is 1 or very close to 1.
The execution, II

- We found 174 partial paradigmatic systems verifying these requirements.
- Note that two different systems may provide evidence on the same derivational relation:

<table>
<thead>
<tr>
<th>pivot</th>
<th>comparison 1</th>
<th>comparison 2</th>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>changer</td>
<td>changeur</td>
<td>changeait</td>
<td>0.356</td>
</tr>
<tr>
<td>prolonger</td>
<td>prolongateur</td>
<td>prolongeait</td>
<td>0.380</td>
</tr>
<tr>
<td>entendre</td>
<td>entendeur</td>
<td>entendait</td>
<td>0.389</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Sample system 1: (V.inf, Agent_N.sg, V.ipfV.3sg)

<table>
<thead>
<tr>
<th>pivot</th>
<th>comparison 1</th>
<th>comparison 2</th>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>possesseurs</td>
<td>possesseur</td>
<td>possédez</td>
<td>0.236</td>
</tr>
<tr>
<td>finisseurs</td>
<td>finisseur</td>
<td>finissez</td>
<td>0.244</td>
</tr>
<tr>
<td>dégustateurs</td>
<td>dégustateur</td>
<td>dégustez</td>
<td>0.229</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Sample system 2: (Agent_N.pl, Agent_N.sg, V.prs.2pl)
For each of the 174 systems:

- We compute the two shift vector averages.
- We compute the Euclidian distance between each individual vector and the average vector.
- We perform a t-test to assess whether there is a significant difference in distance to the average between the shift vectors for the two compared cells.
Data selection for experiment 2

- Vector space constructed from the FrWac corpus (Baroni et al., 2009) using word2vec (Mikolov et al., 2013).
  - CBOW algorithm, window size 5, negative sampling with 10 samples, 400 dimensions

- Paradigmatic system of 6576 (partial) families and 59 cells constructed from:
  1. Derivational relations between verbs, action nouns and agent nouns from Démonette (Hathout and Namer, 2014)
  2. Hand-constructed set of derivational relations between verbs and -able adjectives
  3. Inflectional relations from the GLÀFF (Hathout et al., 2014)

- We then look for pairs of cells where we have enough data to select at least 10 pairs of words such that
  1. no word has homonyms,
  2. all words have a frequency above 50,
  3. the frequency ratio between the nonpivot cells is between $\frac{1}{5}$ and 5.