

It takes a family

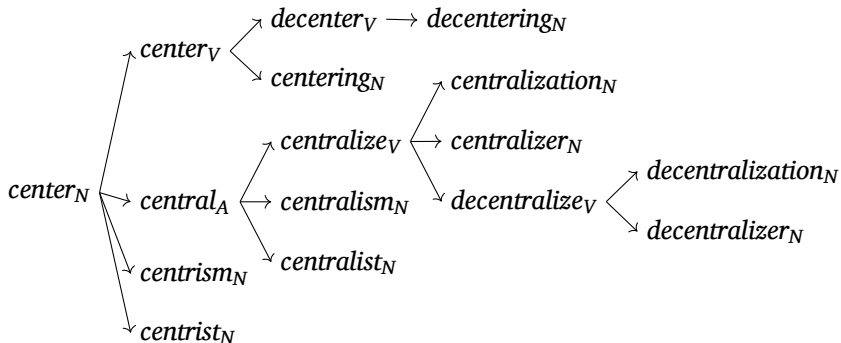
Morphological families shape
the interpretation of derived words

| | | |
|----------------|-----------------|---------------|
| Olivier Bonami | Louise McNally | Denis Paperno |
| U. Paris Cité | U. Pompeu Fabra | Utrecht U. |

Originally presented at LLcD — Paris, 10/09/2024

The rooted tree approach to word formation

- ▶ General presumption of mainstream approaches to derivation: a lexeme is either simplex or has a single, determinate parent.
 1. Morpheme-based approaches: morpheme addition.
 2. Word-based approaches: lexeme formation rule application.
- ▶ As a consequence, derivational families have the structure of a rooted tree. (Stump, 2019)



Classical arguments for the paradigmatic approach I

- Phenomena at odds with the rooted-tree approach:

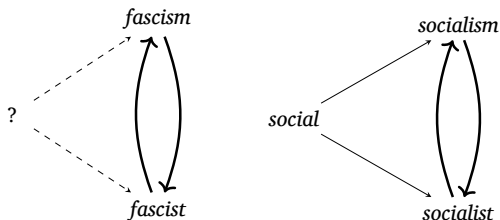
Back-formation What looks like an output arises first and motivates what looks like the corresponding input

bartend ← *bartender*

Conversion The orientation of conversion pairs can be undecidable (Marchand, 1963; Tribout, 2020):

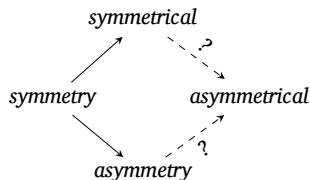
judge_N \longleftrightarrow *judge_V*

Cross-formation Morphological relation between two complex items, despite absent or poorly motivated base (Becker, 1993).

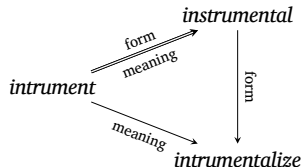


Classical arguments for the paradigmatic approach II

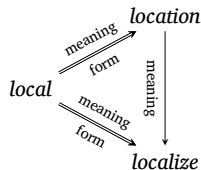
Multimotivation A derived item is equally well motivated by two derivation paths, leading one to be unwilling to choose one over the other (Corbin, 1976).



Form-content mismatches The formal base of a derived item seems different from its semantic base (Hathout and Namer, 2014b).



Family support Multiple members of its family contribute to shaping the semantics of an item (Strnadová, 2014).



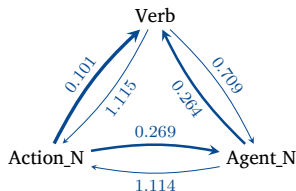
Making the arguments systematic

- ▶ This line of argument has failed to convince many morphologists.
- ▶ Arguably this is because the arguments are **anecdotal**.
 - ▶ Sure, there are some examples that the rooted tree approach cannot account for, but they are rare enough to be considered abnormal.
 - ▶ The lexicon is messy, morphological theory need not account for every lexical accident.
- ▶ Recent work develops a new line of argumentation, looking for quantitative evidence of system-level paradigmatic organization.
 - ▶ **Lateral prediction**: a lexeme's properties are better predicted by a member of its family other than its base.
 - ▶ **Joint prediction**: Joint knowledge of multiple members of its family leads to sizably better predictions of a lexeme's properties.

| | Prediction of form | Prediction of meaning |
|--------------------|-----------------------------|----------------------------------|
| Lateral prediction | Bonami and Strnadová (2019) | Bonami and Guzman Naranjo (2023) |
| Joint prediction | Bonami and Strnadová (2019) | |

Lateral prediction of form

- ▶ Bonami and Strnadová (2019) examine how interpredictable the forms of verbs, action nouns and agent nouns are in French, applying to data from Démonette (Hathout and Namer, 2014a) and the GLAFF (Hathout, Sajous, et al., 2014) tools from Qumin Beniamine (2018).



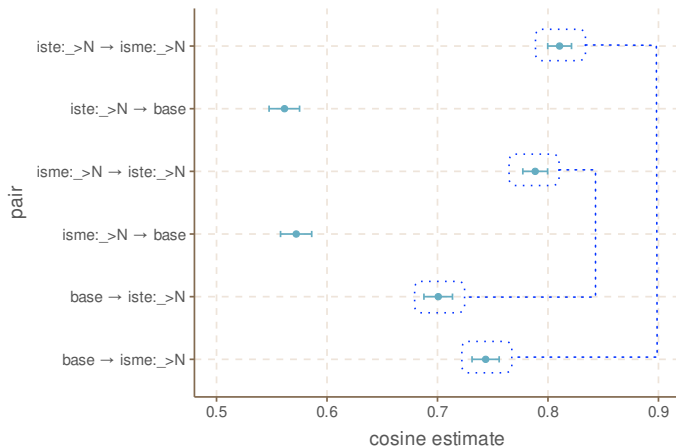
| Verb | Action_N | Agent_N |
|------------------|-------------------|-------------------|
| laver | lavage | laveur |
| 'wash' | 'washing' | 'washer' |
| contrôler | contrôle | contrôleur |
| 'control' | 'control' | 'controller' |
| corriger | correction | correcteur |
| 'correct' | 'correction' | 'corrector' |
| former | formation | formateur |
| 'train' | 'training' | 'trainer' |
| couvrir | couverture | couvreur |
| 'write' | 'writing' | 'writer' |
| gonfler | gonflement | gonfleur |
| 'inflate' | 'inflating' | 'inflater' |

Sample triples

- ▶ For easily explainable reasons, action nouns are better predictors of agent nouns than their base verbs.

Lateral prediction of meaning

- ▶ Bonami and Guzman Naranjo (2023) make exactly the same kind of point for semantics, using distributional vectors as proxies for meanings.
- ▶ They train statistical models to predict the vectors in one morphological category from the vectors in another.
- ▶ At least for *-isme* and *-iste*, it is indeed the case that lateral prediction is more reliable than prediction from the base.



Joint prediction of form

- ▶ Bonami and Beniamine (2016) showed that, in inflectional paradigms, knowing more than one form makes prediction of unknown forms much more reliable.
- ▶ Bonami and Strnadová (2019) show the same thing for derivational paradigms, again focusing on French verbs and deverbal nouns.

| Predictors | Predicted | Entropy |
|-------------------|-----------|---------|
| Verb | Agent_N | 0.709 |
| Action_N | Agent_N | 0.269 |
| Verb, Action_N | Agent_N | 0.138 |
| Verb | Action_N | 1.115 |
| Agent_N | Action_N | 1.114 |
| Verb, Agent_N | Action_N | 0.444 |
| Agent_N | Verb | 0.709 |
| Action_N | Verb | 0.101 |
| Agent_N, Action_N | Verb | 0.006 |

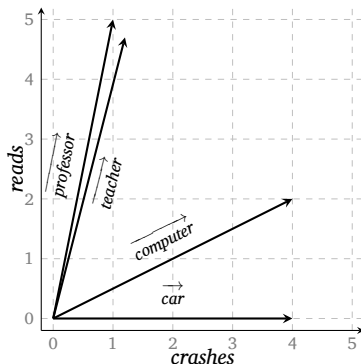
The missing piece of the puzzle

| | Prediction of form | Prediction of meaning |
|--------------------|-----------------------------|----------------------------------|
| Lateral prediction | Bonami and Strnadová (2019) | Bonami and Guzman Naranjo (2023) |
| Joint prediction | Bonami and Strnadová (2019) | This study |

Modelling strategy I

- We rely on **distributional vectors** as proxies for word meanings.

(For a recent state of the art see Lenci 2018; Boleda 2020)

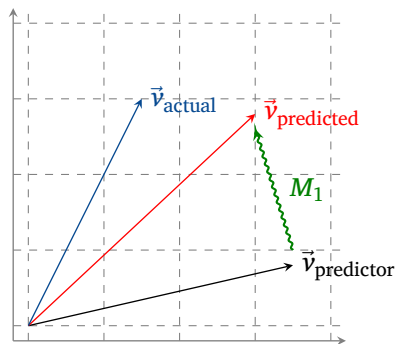


- Modern vector spaces are derived from very large corpora by training a neural network to predict text (e.g. Mikolov et al. 2013)
- Dense vectors of real numbers, usually with 100 to 400 dimensions.

Modelling strategy II

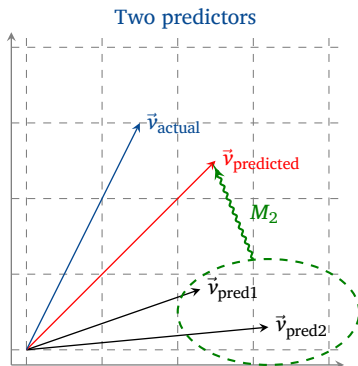
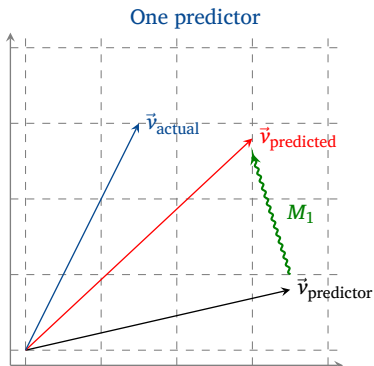
- ▶ One can train a statistical model to predict the vector of a word from that of a related word.
- ▶ This has been used by a variety of authors to assess the ‘regularity’ of the semantic relation between pairs of word that stand in the same morphological relation.

(a.o. Marelli and Baroni, 2015; Bonami and Guzman Naranjo, 2023)



Modelling strategy III

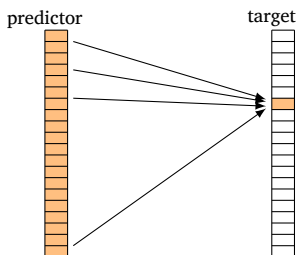
- In our case we can compare models relying on a single predictor with models relying on two predictors:



- If the model with two predictors performs significantly better, we have found evidence of nontrivial joint prediction.

Model architecture I

- Marelli and Baroni (2015): each dimension in the target vector is predicted by a linear model taking all dimensions as input.

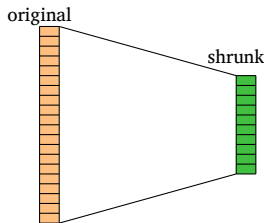


Model architecture II

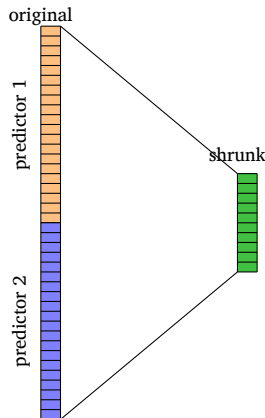
► In this study:

1. We first reduce vector size to 10 dimensions using Principal Component Analysis, both for single word vectors and for pairs of vectors.

One predictor

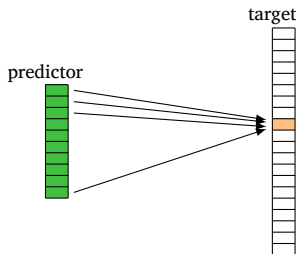


Two predictors



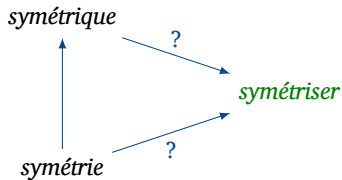
Model architecture III

2. We then use these lower dimension predictor vectors to predict the full vector of the target.



3. We use Boosting Trees rather than linear models.
- We train all models with 10-fold cross-validation, and report aggregated results.

Study 1: Complex verbs and their bases



The dataset

- ▶ 420 verbs in *-iser* or *-ifier* morphologically related to both a noun and an adjective.

| Verb | Noun | Adjective |
|--------------------------------|----------------------------|-------------------------------|
| <i>symétriser</i> ‘symmetrize’ | <i>symétrie</i> ‘symmetry’ | <i>symétrique</i> ‘symmetric’ |

- ▶ Subset of the dataset used by Bonami and Thuilier (2019).
- ▶ All lexemes attested at least 20 times in the FrCow web corpus (Schäfer and Bildhauer, 2012; Schäfer, 2015).
- ▶ The two suffixes have the same range of meanings (Plag, 1999; Namer, 2009; Guzmán Naranjo and Bonami, 2023).
- ▶ It is often unclear whether the noun or adjective should be taken as the base. (Bonami and Thuilier, 2019).

| Verb | Noun | Adjective |
|--|----------------------------|------------------------------|
| <i>totaliser</i> ‘total’ | <i>total</i> ‘total’ | <i>total</i> ‘total’ |
| <i>corsifier</i> ‘make Corsican’ | <i>Corse</i> ‘Corsica’ | <i>corse</i> ‘Corsican’ |
| <i>Staliniser</i> ‘make Stalinist’ | <i>Staline</i> ‘Stalin’ | <i>stalinien</i> ‘Stalinist’ |
| <i>cardinaliser</i> ‘name cardinal _N ’ ~ ‘make cardinal _A ’ | <i>cardinal</i> ‘cardinal’ | <i>cardinal</i> ‘cardinal’ |

The vectors

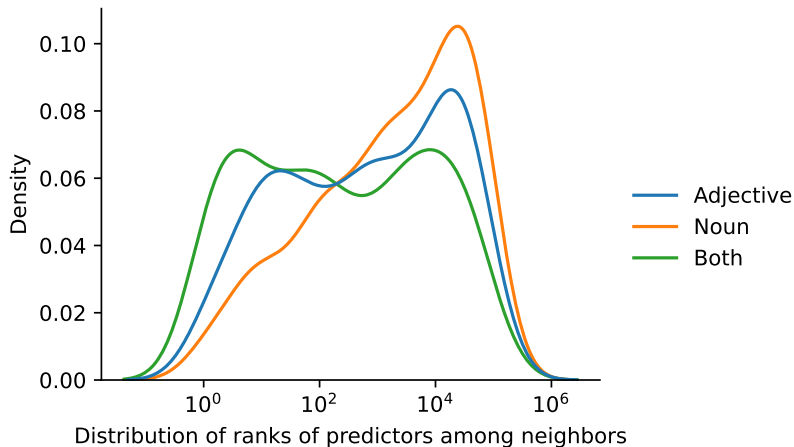
- ▶ Vector space from Guzmán Naranjo and Bonami (2023).
- ▶ Based on the FRCOW corpus (Schäfer, 2015; Schäfer and Bildhauer, 2012).
- ▶ Vectors for lexemes rather than wordforms.
- ▶ This entailed building a doctored version of the corpus:
 - ▶ Every word is replaced by a tagged lemma:
 - ▶ *un dîner* \leadsto un_art dîner_nom
 - ▶ *Paul dînera* \leadsto paul_pn dîner_ver
- ▶ Vector space computed using the Gensim (Řehůřek, 2010) implementation of word2vec (Mikolov et al., 2013).

Hyperparameters: 2 training epochs, 5 negative samples, window size 5, vector size 100.

Results

- ▶ Measure of performance: rank of the actual vector among vectors for verbs in the neighborhood of the predicted vector.
- ▶ The actual verb vector is more distant in rank from the predicted vector when predicting from either nouns or adjectives than when predicting from both:

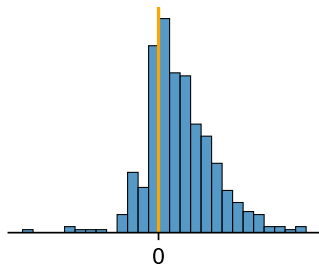
Low rank \Leftrightarrow High performance



Results

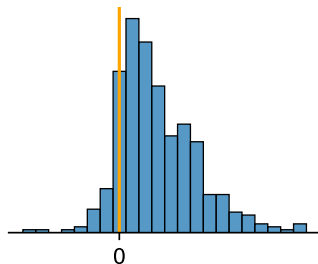
- Pairwise comparisons: difference between log ranks of prediction with single predictor and two predictors.

Adjectives vs. Both



Positive difference for 75% of data

Nouns vs. Both



Positive difference for 88% of data

- In both cases, a Wilcoxon signed-rank test finds a highly significant difference between the performance of the two models ($p < 10^{-20}$).

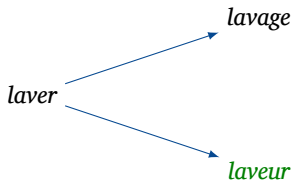
Interim discussion

- ▶ We find clear evidence that predicting a verb from the related noun or adjective leads to worse performance than predicting it from both.
- ▶ In principle, this is compatible with the following scenario:
 - ▶ Derived verbs indeed have a unique base, that we have failed to identify.

| Verb | Noun | Adjective | Likely base |
|-------------------------------------|------------------------------|------------------------------|-------------|
| <i>platoniser</i> 'Platonize' | <i>Platon</i> 'Plato' | <i>platonique</i> 'Platonic' | Noun |
| <i>cartésianiser</i> 'Cartesianize' | <i>Descartes</i> 'Descartes' | <i>cartésien</i> 'Cartesian' | Adjective |

- ▶ If this is true, our single predictor models could be confused because they are trained in part on spurious instances of derivation. For example, the pair (*Descartes*, *cartésianiser*) should not be fed to the noun-based model.
- ▶ This alternative is hard to dismiss, because we do not know how to identify bases.

Study 2: Verb-related agent/instrument nouns



The dataset

- ▶ 740 nouns in *-eur* morphologically related to both a verb and an action noun.

| AGT | ACT | V |
|------------------------|-------------------------|---------------------|
| <i>laveur</i> ‘washer’ | <i>lagage</i> ‘washing’ | <i>laver</i> ‘wash’ |

- ▶ Subset of the dataset used by Bonami and Strnadová (2019), itself derived from Démonette 1.2 (Hathout and Namer, 2014a).
- ▶ All lexemes attested at least 20 times in the FrCow web corpus.
- ▶ The suffix derives agent or instrument nouns, with many items ambiguous between the two.

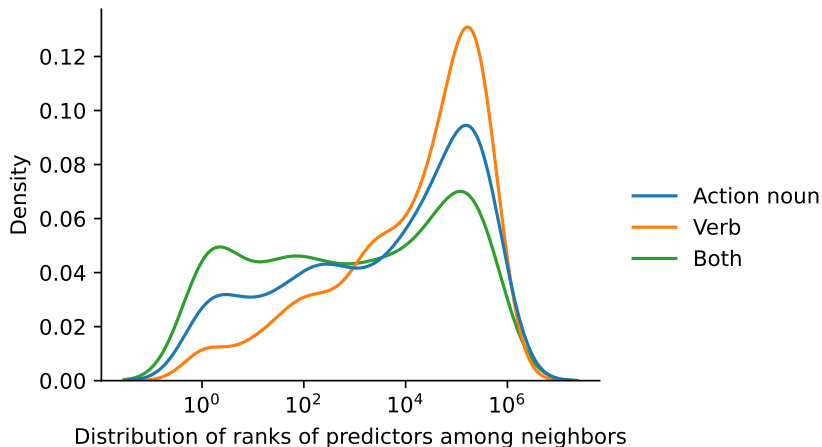
| AGT | ACT | V |
|------------------------------|-------------------------------|--------------------------|
| <i>directeur</i> ‘director’ | <i>direction</i> ‘direction’ | <i>diriger</i> ‘direct’ |
| <i>congélateur</i> ‘freezer’ | <i>congélation</i> ‘freezing’ | <i>congeler</i> ‘freeze’ |
| <i>sondeur</i> ‘pollster’ | <i>sondage</i> ‘poll’ | <i>sonder</i> ‘poll’ |
| ~ ‘sounder’ | ~ ‘sounding’ | ~ ‘probe for depth’ |

- ▶ The verb is the base in the traditional sense, most of the time.
 - ▶ 68% of agent nouns stand in a transparent formal relation with the verb.
 - ▶ 74% of action nouns contain a suffix absent from the agent noun.
 - ▶ Very few cases with clear evidence to the contrary.

| | | |
|---------------------------|----------------------------|-------------------------------|
| <i>foncteur</i> ‘functor’ | <i>fonction</i> ‘function’ | <i>fonctionner</i> ‘function’ |
|---------------------------|----------------------------|-------------------------------|

Results

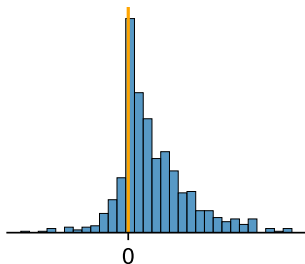
- ▶ We use exactly the same methods as in the first study.
- ▶ Model performance is markedly worse when predicting from either verbs or action nouns than when predicting from both:



Results

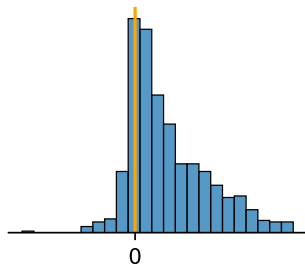
- Pairwise comparisons: difference between log ranks of prediction with single predictor and two predictors.

Action nouns vs. Both



Positive difference for 79% of data

Verbs vs. Both



Positive difference for 81% of data

- In both cases, a Wilcoxon signed-rank test finds a highly significant difference between the performance of the two models ($p < 10^{-30}$).

Conclusions

- ▶ We looked at two datasets with a different makeup in terms of traditional base-derivative relations.











- ▶ In both cases, joint information on two members of a lexeme's derivational family leads to a significantly better prediction of that lexeme's meaning than knowledge of only one of the two members.
- ▶ This is the strongest kind of evidence for paradigmatic organization (Bonami and Beniamine, 2016): no way to account for such effects assuming that derived lexemes relate to a single member of their family.
- ▶ This study complements previous results on family effects in form prediction (Bonami and Strnadová, 2019) and on lateral prediction of meaning (Bonami and Guzman Naranjo, 2023).
- ▶ Future work:
 - ▶ Check these results against human behavior.
 - ▶ Mine the results to figure out why joint prediction helps.
 - ▶ Draw theoretical consequences for the modelling of morphological relatedness.







Thank you!

Questions?








References I

- 
- Bauer, Laurie (1997). “Derivational Paradigms.” In: *Yearbook of Morphology 1996*. Ed. by Geert Booij and Jaap van Marle. Dordrecht: Kluwer, pp. 243–256 (cit. on p. 3).
- 
- Becker, Thomas (1993). “Back-formation, cross-formation, and ‘bracketing paradoxes’ in paradigmatic morphology.” In: *Yearbook of Morphology 1993*. Ed. by Geert Booij and Jaap van Marle. Dordrecht: Kluwer, pp. 1–25 (cit. on p. 4).
- 
- Beniamine, Sacha (2018). “Typologie quantitative des systèmes de classes flexionnelles.” PhD thesis. Université Paris Diderot (cit. on p. 7).
- 
- Blevins, James P. (2001). “Paradigmatic derivation.” In: *Transactions of the Philological Society* 99.2, pp. 211–222 (cit. on p. 3).
- 
- Bochner, Harry (1993). *Simplicity in Generative Morphology*. Berlin: Mouton de Gruyter (cit. on p. 3).
- 
- Boleda, Gemma (2020). “Distributional Semantics and Linguistic Theory.” In: *Annual Review of Linguistics* 6.1, pp. 213–234 (cit. on p. 11).
- 
- Bonami, Olivier and Sacha Beniamine (2016). “Joint predictiveness in inflectional paradigms.” In: *Word Structure* 9.2, pp. 156–182 (cit. on pp. 9, 27).
- 
- Bonami, Olivier and Matías Guzmán Naranjo (2023). “Distributional evidence for derivational paradigms.” In: *The semantics of derivational morphology: theory, methods, evidence*. Ed. by Sven Kotowski and Ingo Plag. Berlin: De Gruyter, pp. 219–258 (cit. on pp. 6, 8, 10, 12, 27).









References II

- 
- Bonami, Olivier and Jana Strnadová (2019). “Paradigm structure and predictability in derivational morphology.” In: *Morphology* 29.2, pp. 167–197. eprint: <http://rdcu.be/Io3H> (cit. on pp. 3, 6, 7, 9, 10, 24, 27).
- 
- Bonami, Olivier and Juliette Thuilier (2019). “A statistical approach to affix rivalry: French *-iser* and *-ifier*.” In: *Word Structure* 12.1, pp. 4–41 (cit. on p. 18).
- 
- Bonami, Olivier and Gert Webelhuth (2021). “Periphrasis and morphosyntactic mismatch in Czech.” In: *One-to-many relations in morphology, syntax and semantics*. Ed. by Berthold Crysmann and Manfred Sailer. Berlin: Language Science Press, pp. 85–115. URL: <https://langsci-press.org/catalog/book/262>.
- 
- Corbin, Danielle (1976). “Peut-on faire l’hypothèse d’une dérivation en morphologie?” In: *Grammaire transformationnelle : syntaxe et lexique*. Ed. by Jean-Claude Chevalier. Lille: Presses Universitaires de Lille, pp. 47–91 (cit. on p. 5).
- 
- Guzmán Naranjo, Matías and Olivier Bonami (2023). “A distributional assessment of rivalry in word formation.” In: *Word Structure* 16.1, pp. 86–113 (cit. on pp. 18, 19).
- 
- Hathout, Nabil and Fiammetta Namer (2014a). “Démonette, a French derivational morpho-semantic network.” In: *Linguistic Issues in Language Technology* 11.5, pp. 125–168 (cit. on pp. 7, 24).

References III

-  Hathout, Nabil and Fiammetta Namer (2014b). “Discrepancy between form and meaning in Word Formation: the case of over- and under-marking in French.” In: *Morphology and meaning (Selected papers from the 15th International Morphology Meeting, Vienna, February 2012)*. Ed. by Franz Rainer et al. Current Issues in Linguistic Theory. Amsterdam: John Benjamins, pp. 177–190 (cit. on p. 5).
-  — (2022). “ParaDis: a family and paradigm model.” In: *Morphology* 32.2, pp. 153–195 (cit. on p. 3).
-  Hathout, Nabil, Franck Sajous, and Basilio Calderone (2014). “GLÀFF, a Large Versatile French Lexicon.” In: *Proceedings of LREC 2014* (cit. on p. 7).
-  Lenci, Alessandro (2018). “Distributional models of word meaning.” In: *Annual Review of Linguistics* 4, pp. 151–171 (cit. on p. 11).
-  Marchand, Hans (1963). “On a question of contrary analysis with derivationally connected but morphologically uncharacterized words.” In: *English Studies* 44, pp. 176–187 (cit. on p. 4).
-  Marelli, Marco and Marco Baroni (2015). “Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics.” In: *Psychological Review* 122, pp. 485–515 (cit. on pp. 12, 14).
-  Marle, Jaap van (1984). *On the Paradigmatic Dimension of Morphological Creativity*. Dordrecht: Foris (cit. on p. 3).
-  Mikolov, Tomas et al. (2013). “Efficient Estimation of Word Representations in Vector Space.” In: *CoRR* abs/1301.3781. URL: <http://arxiv.org/abs/1301.3781> (cit. on pp. 11, 19).

References IV

- 
- Namer, Fiammetta (2009). *Morphologie, Lexique et Traitement Automatique des Langues*. London: Hermès Science Publishing (cit. on p. 18).
- 
- Plag, Ingo (1999). *Morphological productivity*. Berlin: Mouton de Gruyter (cit. on p. 18).
- 
- Pounder, Amanda (2000). *Process and paradigms in word-formation morphology*. Vol. 131. Walter de Gruyter (cit. on p. 3).
- 
- Řehůřek, Radim (2010). “Software Framework for Topic Modelling with Large Corpora.” In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pp. 45–50 (cit. on p. 19).
- 
- Robins, R. H. (1959). “In defense of WP.” In: *Transactions of the Philological Society* 58, pp. 116–144 (cit. on p. 3).
- 
- Schäfer, Roland (2015). “Processing and Querying Large Web Corpora with the COW14 Architecture.” In: *Proceedings of Challenges in the Management of Large Corpora*, pp. 28–34 (cit. on pp. 18, 19).
- 
- Schäfer, Roland and Felix Bildhauer (2012). “Building Large Corpora from the Web Using a New Efficient Tool Chain.” In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, pp. 486–493 (cit. on pp. 18, 19).
- 
- Štekauer, Pavol (2014). “Derivational Paradigms.” In: *The Oxford Handbook of Derivational Morphology*. Ed. by Rochelle Lieber and Pavol Štekauer. Oxford: Oxford University Press, pp. 354–369 (cit. on p. 3).

References V



Strnadová, Jana (2014). “Les réseaux adjectivaux: Sur la grammaire des adjectifs dénominaux en français.” PhD thesis. Université Paris Diderot et Univerzita Karlova V Praze (cit. on p. 5).



Stump, Gregory (2019). “Some sources of apparent gaps in derivational paradigms.” In: *Morphology* 29, pp. 271–292 (cit. on p. 2).



Tribout, Delphine (2020). “Nominalization, verbalization or both? Insights from the directionality of noun-verb conversion in French.” In: *Zeitschrift für Wortbildung/Journal of Word Formation* 4.2, pp. 187–207 (cit. on p. 4).

Discussion I

- ▶ We find clear evidence that predicting a verb from the related noun or adjective leads to worse performance than predicting it from both.
- ▶ In principle, this is compatible with the following scenario:
 - ▶ Derived verbs indeed have a unique base, that we have failed to identify.
 - ▶ Our single predictor models are confused because they are trained both on true instances of derivation (where the base indeed has this PoS) and spurious instances (where the base has the other PoS).
- ▶ This alternative is hard to evaluate, because we do not know how to identify bases.

Discussion II

► One tentative idea:

- We expect the base to be formally more similar to the derivative than the nonbase.

| Verb | Noun | Adjective | Likely base |
|-------------------------------------|------------------------------|------------------------------|-------------|
| <i>platoniser</i> 'Platonize' | <i>Platon</i> 'Plato' | <i>platonique</i> 'Platonic' | Noun |
| <i>cartésianiser</i> 'Cartesianize' | <i>Descartes</i> 'Descartes' | <i>cartésien</i> 'Cartesian' | Adjective |

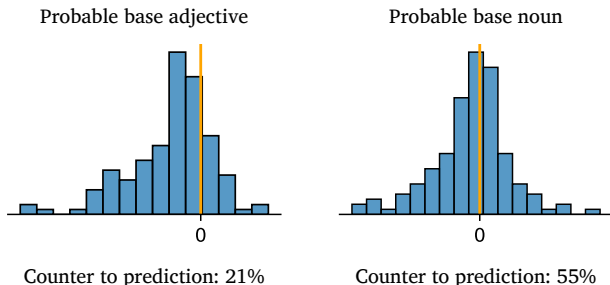
- Among models with a single predictor, we expect the model using the 'true base' to perform better.
- This makes predictions that can be assessed using our data:

| $ED(V,A) - ED(V,N)$ | # of datapoints | Prediction |
|---------------------|-----------------|----------------------------------|
| < 0 | 147 | N-based model performs better. |
| $= 0$ | 135 | None. |
| > 0 | 138 | Adj-based model performs better. |

ED = edit distance

Discussion III

Difference between log ranks for adjective-based and noun-based models



- ▶ This is really hard to interpret:
 - ▶ We get the expected result for likely adjective bases, but the unexpected result for likely nominal bases.
 - ▶ Does this falsify the hypothesis that there is a true base? Or is our method for finding the true base not the right one?
- ▶ If anything this confirms that the rooted tree hypothesis lacks empirical teeth without a reproducible method for finding bases.