

---

# A quantitative comparison between word-formation & inflection: A look at paradigms in French

Gilles Boyé

Gauvain Schalchli

Université Bordeaux-Montaigne & UMR5263    Université Bordeaux-Montaigne & UMR5263

---

## 1 Introduction

The frontier between inflection and word-formation has been described in many ways from the unified approach of morpheme-based morphology (Hockett 1954; Lieber 1982; Halle 1973) to the total split of lexeme-based morphology (e.g. Beard 1995; Anderson 1982) and some intermediary positions against the total split (Booij 1994). In this paper, we propose to reevaluate the situation in a realistic word-based approach with a discussion about paradigmaticity in word-formation and inflection based on a quantitative comparison on French data.

## 2 Word-formation vs inflection in theory

The separation of inflection and word-formation made by lexeme-based theories is usually linked with their typical orientations:

- inflection expresses syntactic properties and its exponents sit at the periphery of words
- word-formation aims at producing lexical items and does not depend on syntactic contexts

In this type of frameworks, inflection describes the relation between words attached to the same lexeme and word-formation the relation between lexemes morphologically related. Booij (1994) focuses on a type of inflection that does not realise properties directly linked to contextual agreement (e.g. number in nouns, tense in verbs) and argues that this is a sort of intermediary morphology between agreement inflection (contextual inflection) and word-formation (arbitrary coinage) where the speaker can pick the feature he wants but inside a set of choices limited by inflection (arbitrary inflection<sup>1</sup>).

Other arguments about split morphology revolve around three axes that are more characteristic of inflection:

- (1) a. productivity: general availability of inflectional forms
- b. semantic regularity: relations between the content of inflectional forms are generally robust
- c. paradigmatic organisation: many to many relations are typical of inflectional systems

The preceding points in (1) have not received the same level of scrutiny when comparing inflection and word-formation. 1a has received a lot of attention from many authors (cf. Dal, 2003, for an overview). Recently Bonami & Paperno (2018) evaluated the relative semantic regularity of word-formation and inflectional semantic relations (1b). In this paper, we propose to compare the paradigmatic organisation of word-formation and inflection with a quantitative study on French.

---

<sup>1</sup>In Booij's terms: *inherent inflection*.

### 3 Word-formation vs inflection in French

In a realistic approach, sampling data is essential to evaluate the difference and the similarities between word-formation and inflection.

To study the quantitative difference between word-formation and inflection, we used Lexique3, a French lexicon compiled from corpora by New (2006). Our dataset of derivation chains was compiled by using Dérif (Namer, 2003) on the lemmas of Lexique3 and recouping bases for all derivation trees. In Lexique3, Dérif found 17564 derived words associated with 8435 bases and 986 derivations. For inflection, we collected all forms of verbs with at least two co-forms<sup>2</sup> in Lexique3. Our inflectional data set contains 16194 forms for 5835 verbs.

Following Bonami & Boyé (2014), we consider that paradigmatic generalizations originate with series of analogies between forms associated with the same pair of cells. As defended by Bonami & Beniamine (2016), we believe that joint-predictiveness is a central force in paradigm organization and that the number of co-forms for a given pair of cells is an indication of how reliable the generalizations made from this series will be.

Table 1 presents the number of co-forms for the 51 most common word-formation on the left and the number of co-forms for the 51 cells of the verbal paradigm on the right. The shades give an approximation of the number of co-forms (dark orange  $\geq 2000$ , orange  $\geq 300$ , light orange  $\geq 30$  and gray orange  $< 30$ ).<sup>3</sup>

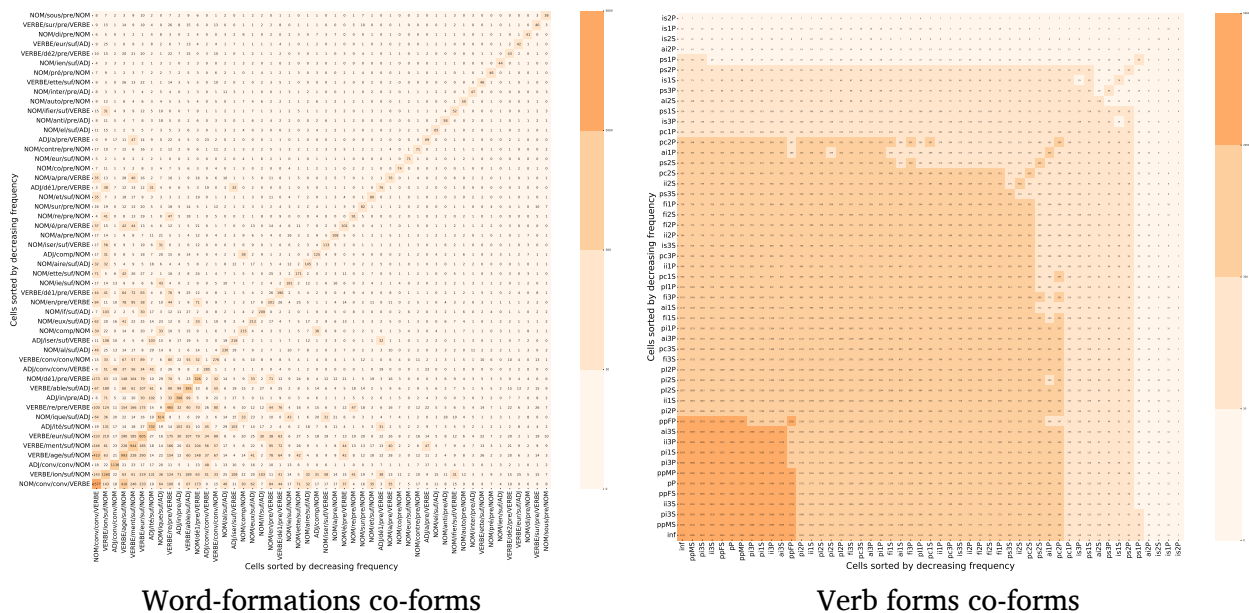


Figure 1: Co-occurrence based on Lexique3

As expected, there are huge differences between the two maps in Figure 1:

- the inflection data has large zones of densely populated areas (dark orange, orange) while the word-formation is mostly empty (gray orange).
- inflection zones are concentric but the word-formation zones are patchy
- the dark orange shade does not appear in word-formation

<sup>2</sup>We call co-forms a set of forms occupying different paradigm cells for the same lexeme.

<sup>3</sup>In the word-formation figure, the diagonal represents the number derived words in the base series: base⇒derived word.

- the series effect is marked in word-formation by the darker diagonal counting the relation between a derived word and its base while there is no such difference in shade for inflection

However the observations above must not distract us from the similarities between parts of inflection and part of word-formation:

- the peripheral part of inflection is as empty as the majority of derivation (gray orange)
- the orange of word-formation corresponds to the outer of the orange part of inflection

## 4 Conclusion

From these observations, we propose that the comparison between word-formation and inflection should not be done at a global level but rather on specific sets of data. Some word-formation processes seem to form paradigms in a way similar to core inflection and some part of inflection behave on a par with typical word-formation.

In this paper, we compared word-formation and verbal inflection in French and found some intersection between the two. We think that word-formation paradigmatic phenomena could be closer to that of inflectional systems with much smaller paradigms, for example, in French, nouns and adjectives inflectional paradigms have respectively 2 and 4 cells, bringing them much closer to the size of “derivational paradigms”. In this perspective, French deadjectival adverbs in *-ment* or pairs of masculine/feminine animate nouns would be particularly suited for a paradigmatic analysis in word-formation, preserving the intuitions of proponents of inflection and derivation in one analysis.

## References

- Anderson, Stephen R. 1982. Where’s Morphology? *Linguistic Inquiry* 13(4). 571–612.
- Beard, Robert. 1995. *Lexeme morpheme-based morphology*. State University of New York Press.
- Bonami, Olivier & Sarah Beniamine. 2016. Joint predictiveness in inflectional paradigms. *Word Structure* 9(2). 156–182.
- Bonami, Olivier & Gilles Boyé. 2014. De formes en thèmes. *Foisonnements morphologiques. Etudes en hommage à Françoise Kerleroux* 17–45.
- Bonami, Olivier & Denis Paperno. 2018. Inflection vs. derivation in a distributional vector space. *Lingue e linguaggio* 17(2). 173–196.
- Booij, Geert. 1994. Against split morphology. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1993*, 27–49. Kluwer Academic Publishers.
- Dal, Georgette. 2003. Productivité morphologique: définitions et notions connexes. *Langue française* 3–23.
- Halle, Morris. 1973. Prolegomena to a theory of word formation. *Linguistic Inquiry* 4(1). 3–16. <http://www.jstor.org/stable/4177749>.
- Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10. 210–234.
- Lieber, Rochelle. 1982. *On the organization of the lexicon*. Indiana University Linguistics Club.
- Namer, Fiammetta. 2003. Automatiser l’analyse morpho-sémantique non affixale : le système dérif. *Cahiers de Grammaire* 28. 31–48.
- New, Boris. 2006. Lexique 3: Une nouvelle base de données lexicales. In *Actes de la Conférence Traitement Automatique des Langues Naturelles (TALN)*, 892–900.