Baseless derivation: the behavioural reality of derivational paradigms

Maria Copot Université Paris Cité
Olivier Bonami Université Paris Cité

1 Background

The historically dominant conceptualisation of the structure of derivational families is the ROOTED TREE, where each lexeme is either at the root of a derivational tree or has a unique parent. In this view, illustrated in Figure 1, lexemes are linked by monodirectional relationships, and there is a always single path relating two lexemes.

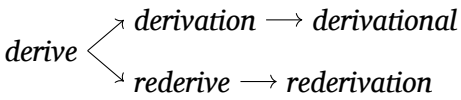

Figure 1: A rooted tree representation of part of the derivational family of *derive*.

This view incurs a number of theoretical and descriptive issues. The rooted tree view forces one to choose a single base for each derived lexeme, which is problematic, for instance, when a derived lexeme has multiple potential bases (does *asymmetrical* come from *asymmetric* or from *symmetrical*?), or when a lexeme's semantics do not come from its formal base, but from another member of the derivational family (Hathout & Namer, 2014): the formal process *X*∼*Xics* generally signifies a relationship between an object and the discipline that studies it (*graph*∼*graphics*, *gene*∼*genetics*), but the story is more complicated for the pair *language*∼*linguistics* - *linguistics* has taken its meaning from *language*, but its base form from *linguist*, since *linguistics* is not the study of linguists. Moreover, the rooted tree's requirement that relationships be monodirectional is ill-suited to capture cases of backformation, where a formally simpler lexeme can be shown to have been derived from a formally more complex one) or cross-fromation (Becker, 1993), where two derived lexemes are more predictive of each other's form and meaning that their common base is, or when the base is absent).

An alternative to the rooted tree view of derivational families is a paradigmatic approach (Robins, 1959; Becker, 1993; Bochner, 1993; Bauer, 1997; Štekauer, 2014; Bonami & Strnadová, 2019; Hathout & Namer, 2022). Seeing derivational families as paradigmatic involves foregrounding the multiple, bidirectional relationships that exist between related lexemes, positing no single lexeme as more basic than, or logically previous to, any other.

2 Motivation

The two views of morphology make different predictions about which relationships between word forms are accessible to speakers. A paradigmatic approach supposes that speakers keep track of all bidirectional relationships between word forms. The rooted tree view of word formation is traditionally associated with post-bloomfieldian morphemic approach to morphology, but such approaches haven't traditionally engaged in making predictions about the cognitive reality of relationships of predictability within words (though see Jun & Albright 2016; Cotterell et al. 2019 for related examples in inflection). Word-based approaches to word formation

such as Aronoff (1976) and Stump (2019) also suggest that the canonical situation is for words of a language to be organised in a rooted tree structure, a view that carries the implication that predictability relationships between words should also follow said structure.

Nevertheless, empirical evidence brought to bear on this matter consists largely of discussion on the merits of specific linguistic examples. Evidence that targets larger parts of the morphological system does exist: for example, Bonami & Strnadová (2019) map out the relationships of form predictability in a subset of related verbs and agent and action deverbal nouns in French, and find that despite the verb supposedly being the base form in this triplet, the action noun is on average just as predictable from the agent noun as it is from the verb, and the action noun is a better predictor of the agent noun than the verb is. Bonami & Guzman Naranjo (2023) find that similar paradigmatic relationships exist for meaning: they train statistical models to predict the distributional vector of a lexeme from the vector of a derivationally related lexeme. They find that the meaning of a lexeme in a given derivational cell is at least somewhat predictable from the meaning of a derivationally related lexeme in a different cell, so implicative relationships of meaning exist in derivational families. Moreover, the meaning of the formal base is not always the best predictor of the meaning of a derivationally related lexeme: they find that lexemes linked by the pattern *Xisme*∼*Xiste* are better predictors of each other's meaning compared to predicting the meaning of each from the base.

In this talk we compare the rooted tree and the paradigmatic view with the yardstick of cognitive reality. We report on a behavioural experiment to investigate whether speakers are aware of the individual implicative relationships that recent work claims to be structuring derivational paradigms. The experiment aimed to test whether speakers' mental representation of derivational families resembles more closely the rooted tree view or the paradigmatic view. The paradigmatic conceptualisation of derivational families predicts that speakers would be aware of and exploit all available patterns of predictability, regardless of whether the base is the predictor, or at all involved in the prediction. The rooted tree view would posit that speakers only keep track of relationships of predictability where the base is the predictor.

3 Methodology

We performed and acceptability judgement task on French data. We presented speakers with a sentence containing two derivationally related pseudowords and we asked them to rate the acceptability of the second. The more expected the second word is based on the form of the first, the better it was to be be rated. Figure 2 shows a sample item.

J'adore le monde de la catonisation. Je veux être { *catonisateur* / *catonisiteur* / *catoniseur* } quand je serai grand.

I love the world of ACTION_NOUN. I want to be { AGENT_NOUN-1 / AGENT_NOUN-2 / AGENT_NOUN-3 } when I grow up.

Figure 2: Sample experimental item, followed by an English translation. Only one of the forms filling the second slot is presented to each participant. The three forms have different levels of predictability conditional on the knowledge that their ACTION NOUN is *catonisation*

Six directed pairs of cells (all permutations of VERB, AGENT NOUN, ACTION NOUN) were chosen for the experiment on the basis of previous work on identifying derivational paradigms

in French (Bonami & Strnadová, 2019). The verb, present in the items only in the infinitive form, was assumed to be the base of both action and agent deverbal nouns by traditional accounts. The cell pairs involve making predictions from the base, towards the base, or between two nonbase cells.

Under a rooted tree view, the predictability of the second word form given the first should only matter when the first form is the verb, since these would be the only cases in which speakers are expected to keep track of predictability relationships. Under a paradigmatic view, word form predictability should have a positive effect in all directed cell pairs.

Participants were shown a video of someone speaking out the items, and signalling which word they should provide an acceptability judgement for. We chose to present stimuli in this way so as not to allow cues from orthography to influence participants' responses. 60 French native speakers were recruited on Prolific.co, and shown 54 crucial items each (9 from each directed cell pair) and 24 distractors (words in inflectional relationships). The same sentence frame could appear with three different pseudoword pairs, of different levels of predictability - the level of predictability of the pseudowords that each sentence appeared with was randomised. Within items for crucial cell pairs, the three levels of predictability were uniformly distributed.

4 Results

We fitted a Bayesian mixed effects beta regression to the judgements, predicting them based on the cell pair they instantiate, a phonological well-formedness score of the second form obtained from a separate norming experiment, and the predictability of the second word form based on the first, calculated with the Minimal Generalisation Learner (Albright, 2002; Albright & Hayes, 2003) on data from Démonette (Namer et al., 2019). The conditional effects are pictured below.

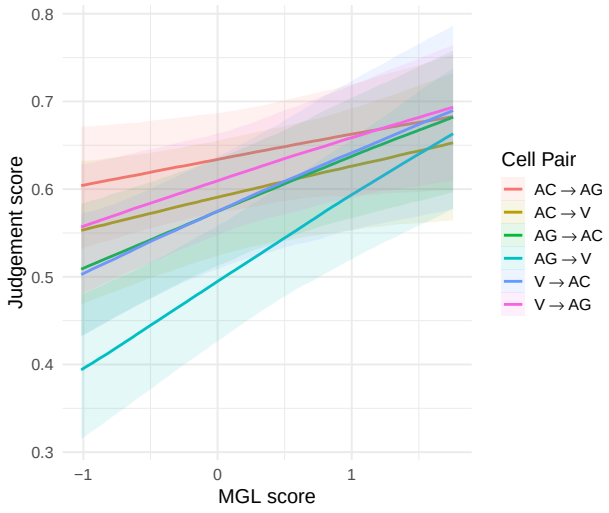

Figure 3: Conditional effect plots of the model

5 Discussion

The results fit well with a paradigmatic conception of derivational paradigms within the mental lexicon. Predictability of the second form given the first always has a positive impact on judgement, regardless of the cells involved, in all directions of prediction. A rooted tree view would expect predictability to only have a positive effect when predicting from the base, which

in this case is the verb. Particularly striking is that the case in which predictability matters the most is when predicting from the agent noun towards the verb, which is unexpected under a rooted tree view. It is important to note that the two cell pairs in which the verb is not at all involved also show a positive effect of predictability: a rooted tree view would predict that in this case, if speakers are missing knowledge of the base form and prediction needs to go through the base, speakers would either be lost or would attempt to reconstruct the base, which is problematic since for many of our items where the base was neither the predicted form nor the predictor, the base form was ambiguous.

References

Albright, Adam C. 2002. *The identification of bases in morphological paradigms*: University of California, Los Angeles dissertation.

Albright, Adam C. & Bruce P. Hayes. 2003. Rules vs. analogy in english past tenses: A computational/experimental study. *Cognition* 90. 119–161.

Aronoff, M. 1976. *Word formation in generative grammar* Linguistic inquiry monographs. Penguin Random House LLC. https://books.google.fr/books?id=syIXAQAAMAAJ.

Bauer, Laurie. 1997. Derivational paradigms. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1996*, 243–256. Dordrecht: Kluwer.

Becker, Thomas. 1993. Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology. In *Yearbook of morphology 1993*, 1–25. Springer.

Bochner, Harry. 1993. *Simplicity in generative morphology*. Berlin, Boston: De Gruyter Mouton. doi:doi:10.1515/9783110889307.

Bonami, Olivier & Matías Guzman Naranjo. 2023. Distributional evidence for derivational paradigms. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: theory, methods, evidence*, 219–258. Berlin: De Gruyter.

Bonami, Olivier & Jana Strnadová. 2019. Paradigm structure and predictability in derivational morphology. *Morphology* 29. doi:10.1007/s11525-018-9322-6.

Cotterell, Ryan, Christo Kirov, Mans Hulden & Jason Eisner. 2019. On the complexity and typology of inflectional morphological systems. *Transactions of the Association for Computational Linguistics* 7. 327–342. doi:10.1162/tacl_a_00271. https://aclanthology.org/Q19-1021.

Hathout, Nabil & Fiammetta Namer. 2014. Discrepancy between form and meaning in word formation: the case of over- and under-marking in french doi:10.1075/cilt.327.12hat.

Hathout, Nabil & Fiammetta Namer. 2022. Paradis: a family and paradigm model. *Morphology* 32(2). 153–195. doi:10.1007/s11525-021-09390-w.

Jun, Jongho & Adam C. Albright. 2016. *Speakers' knowledge of alternations is asymmetrical: Evidence from seoul korean verb paradigms*. Berlin: Cambridge University Press.

Namer, Fiammetta, Lucie Barque, Olivier Bonami, Pauline Haas, Nabil Hathout & Delphine Tribout. 2019. Demonette2 — Une base de données dérivationnelles du français à grande échelle : premiers résultats. In *Actes de TALN*, Toulouse, France. https://halshs.archives-ouvertes.fr/halshs-02275652/document.

Robins, Robert. 1959. In defence of WP. *Transactions of the Philological Society* 58. 116 – 144. doi:10.1111/j.1467-968X.1959.tb00301.x.

Stump, Gregory. 2019. Some sources of apparent gaps in derivational paradigms. *Morphology* 29. 271–292. doi:10.1007/s11525-019-09339-4.

Štekauer, Pavol. 2014. Derivational paradigms. In Rochelle Lieber & Pavol Štekauer (eds.), *The oxford handbook of derivational morphology*, 354–369. Oxford: Oxford University Press.