A story-based approach to derivational paradigms

Daniele Sanacore — Université Toulouse - Jean Jaurès
Nabil Hathout — Université Toulouse - Jean Jaurès
Fiammetta Namer — Université de Lorraine

Derivational family slicing. The paradigmatic nature of derivation has been much discussed in the last decades, in the perspective of unifying inflection and derivation (Van Marle, 1985; Stump, 1991; Bauer, 1997; Boyé & Schalchli, 2016; Hathout & Namer, 2019). To this end, many authors have proposed to extend paradigms to derivation (Bochner, 1993; Booij, 2010; Jackendoff & Audring, 2018; Hathout & Namer, 2022). Despite the growing number of papers adopting a paradigmatic approach to derivation, the nature of derivational paradigms is still debated. Bonami & Strnadová (2019) propose that derivational paradigms are alignments of "slices" of derivational families (that we may call "paradigmatic families") having the same content relations. Like Bauer (2019) and Antoniova & Štekauer (2016), they consider that the structure of paradigms is determined by meaning. On the other hand, the question of the delimitation of the derivational paradigms has hardly been discussed. In this talk, we focus on this question. We propose a methodology for the slicing of derivational families into paradigmatic families that can be aligned in order to form derivational paradigms. In this abstract, we illustrate this methodology with French examples.

Stories that tell morphosemantic relations. Our procedure starts from a derivational family. As an example, consider the French family of the artifact noun *pot* 'pot' in (1).

(1) F1 = {*pot*, *poterie*, *potier*, *rempoter*, *rempotage*}
'pot', 'pottery', 'potter', 'to repot', 'repotting'

In order to identify all the relevant semantic relations in F1, we first consider all F1 subsets of size ≥ 2. We refer to this cover as cov(F1) as in (2).

(2) cov(F1) = {{*pot*, *poterie*}, {*pot*, *potier*}, ⋯, {*rempoter*, *rempotage*}, {*pot*, *potier*, *poterie*}, ⋯, {*pot*, *poterie*, *potier*, *rempoter*, *rempotage*}}

A first difficulty when it comes to identifying the semantic relations that connect the lexemes in a subset or are involved in the characterization of these relations is the lack of resources which a systematic description of the lexical relations present in the lexicon could be drawn from. Note that resources such as WordNet (Fellbaum, 1999), FrameNet (Ruppenhofer et al., 2003) or even JeuxDeMots (Lafourcade & Joubert, 2013) would not be suitable because the range of the relations they provide is too limited. Another option would be to interview speakers to obtain such descriptions, by asking them to tell us a story that contains the words in the subset (as in some radio games). Unfortunately, we do not have the means to carry out such large-scale surveys. For this reason, our proposal is purely methodological[1]. In order to illustrate our method and unfold its different stages, we propose to implement it on some stories that we will produce ourselves.

For each subset in cov(F1), we produce a set of stories that contain instances of the lexemes included in the subset, like the ones in (3) for the subset {*pot*, *poterie*}. Stories may be made up of one sentence (3a) or many ones (3b).

(3) a. h11 = Hier, Marc a fabriqué un pot magnifique dans le cours de poterie.[2]

[1] The availability of generative models like *ChatGPT* makes it possible to envisage a large-scale production of the stories we need.

[2] 'Yesterday, Marc made a beautiful pot in the pottery class.'

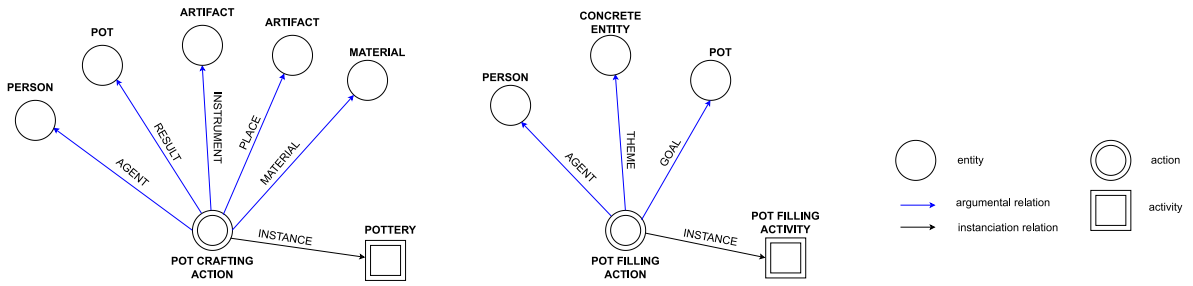

Figure 1: Family meaning bundles (FMB) built from stories about the lexemes in word family F1. The FMB on the left hand side describes the action of crafting pots. The one on the right hand side describes the action of transferring to a new pot. The entities are represented by their ontological categories. In the FMB on the left hand side, the action of making pots is an instance of the activity pottery. In the bundle on the right, the action of transferring into a new pot is an instance of an activity of the same nature, which may be usually performed in gardening.

b. h12 = Ambre a toujours voulu apprendre à fabriquer des pots. Le mois dernier elle s'est inscrite à un cours de poterie proposé par la mairie.[3]

Our hypothesis is that lexemes that are strongly linked by semantic relations will regularly co-occur in stories. For instance, the lexemes *pot*, *potier* and *poterie* are strongly semantically related and will regularly co-occur in stories. On the other hand, we expect that lexemes that are semantically distant may co-occur in fewer stories and that their relations in these stories will be quite episodic. For example, a story where someone takes a plant out of a pot made up for the subset {*pot*, *rempoter*}, like the one in (4) will hardly involve a potter or pottery (the activity of pot crafting).

(4) h13 = Nous avons dû rempoter notre Aloe Vera parce que son ancien pot est devenu trop petit.[4]

Abstracting semantic bundles from the stories. To turn stories like h11, h12 or h13 told in a textual form into more formal objects that can be more easily manipulated and compared, we propose to transcribe them as semantic networks similar to those proposed by Sowa (2014). The operation is performed separately for each subset of cov(F1). Then, we replace the vertices of the network by labels that indicate the ontological class of the corresponding referents. The graphs are then linearized and the relations they contain are clustered. The resulting clusters are graphs that describe the semantic content that can be associated to the lexemes of the subset. We will call this graph "subset meaning bundle." (SMB). The operation is repeated for all subsets of cov(F1) to obtain a set of subset meaning bundles that we then align on the basis of the ontological categories of the entities they contain. The more general meaning bundles obtained in this way may be called "family meaning bundles" (FMB). These can in turn be aligned in the same way as before to build semantic paradigms similar to those proposed by Hathout & Namer (2022). We will call them "lexical meaning bundles." (LMB) Figure 1 shows two family meaning bundles that describe actions and activities (Roché, 2017; Fradin, 2020) that involve pots: one where a pot is the result of the action and the other where it is a goal.

Note that stories about subsets of lexemes from a family like F1 do not necessarily mention all the concepts contained in one of the FMB. For example, the story in (3a) does not speak of

[3] 'Ambre has always wanted to learn how to make pots. Last month she signed up for a pottery course offered by the city hall.'

[4] 'We had to repot our Aloe Vera because its old pot became too small.'

instruments or materials. Conversely, some concepts included in the FMB may have no realization in the word families they originate from. For example, the material in the FMB on the left hand side in Figure 1 is not realized by a lexeme included in F1.

Family meaning bundle alignment. The method illustrated on the family of *pot* can be applied to the other derivational families. Each family yields one or several family meaning bundles that may be aligned with FMB from other families. The alignment is based on the ontological nature of their entities and events. For example the family of *brique* (5) includes lexemes denoting artifacts (*brique*), people that make these artifacts (*briquetier*) and places where the artifacts are made (*briqueterie*). It yields a FMB that may be aligned with the FMB on the left hand side in Figure 1 which has these same vertices. Similarly, the family of *bouteille* in (6) yields a FMB describing a bottle filling (*embouteiller*; *embouteillage*) that may be aligned with the one on the right hand side in Figure 1. In this way, lexical meaning bundles are semantic paradigms that delimit and structure the derivational paradigms.

(5) F2 = {*brique*, *briquetier*, *briqueterie*}
'brick', ' brickmaker', 'brick factory'

(6) F3 = {*bouteille*, *embouteiller*, *embouteillage*}
'bottle', 'to bottle', 'bottling'

Meaning bundle projection. We now can slice word families into paradigmatic families by projecting on them the lexical meaning bundles. For example, the projection of LMB having the same structure as the FMB in Figure 1 on the derivational family F1 results in the paradigmatic families in (7). We see that two paradigmatic families overlap and share the lexeme *pot*. Paradigmatic families in (7a) and (7b) highlight two facets of the meaning of *pot*: its production and its use.

(7) a. f_1 = (*pot*, *potier*, *poterie*)
b. f_2 = (*pot*, *rempoter*, *rempotage*)

Paradigmatic families align in semantically delimited derivational paradigms. Table 1 presents the derivational paradigm related to artifact making, and the Table 2 the one related to moving entities into recipients. The ontological and relational labels of the semantic bundle serve as indexes of the paradigm columns. We can see in Table 1 that some concepts (vertices) in the meaning bundles may not be realized morphologically in some families. It is the case for the materials which are not morphologically realized in the families of *pot* and *brique* but is in the family of *fer-blanc* 'tinplate', *ferblanterie* 'tinware', *ferblantier* 'tinsmith'. We also see that one concept in a lexical meaning bundle may correspond to more than one lexeme in a family. Both the verb *rempoter* and the action noun *rempotage* correspond to the event node pot filling action in the right FMB in Figure 1. Similarly, several lexemes in a paradigmatic family may have the same form, as in the case of *poterie* and *ferblanterie* (activity and artifact).

artifact	**person**	**activity**	**place**	**material**
pot/poterie	*potier*	*poterie*	-	-
brique	*briquetier*	-	*briqueterie*	-
ferblanterie	*ferblantier*	*ferblanterie*	-	*fer-blanc*
...	...	...	...	...

Table 1: Derivational paradigm of artifact making families

recipient	**filling$_V$**	**filling$_N$**
pot	*rempoter*	*rempotage*
bouteille	*embouteiller*	*embouteillage*
...	...	...

Table 2: Derivational paradigm of recipient filling families

References

Antoniova, Vesna & Pavol Štekauer. 2016. Derivational paradigms within selected conceptual fields–contrastive research. *Facta Universitatis, Series: Linguistics and Literature* 61–75.

Bauer, Laurie. 1997. Derivational paradigms. In *Yearbook of morphology 1996*, 243–256. Springer.

Bauer, Laurie. 2019. Notions of paradigm and their value in word-formation. *Word Structure* 12(2). 153–175.

Bochner, Harry. 1993. *Simplicity in generative morphology*. De Gruyter Mouton.

Bonami, Olivier & Jana Strnadová. 2019. Paradigm structure and predictability in derivational morphology. *Morphology* 29(2). 167–197.

Booij, Geert. 2010. Construction morphology. *Language and linguistics compass* 4(7). 543–555.

Boyé, Gilles & Gauvain Schalchli. 2016. The status of paradigms. *The Cambridge handbook of morphology* 206–234.

Fellbaum, Christiane (ed.). 1999. *Wordnet: an electronic lexical database*. Cambridge, MA: MIT Press.

Fradin, Bernard. 2020. Characterizing derivational paradigms. In *Paradigmatic relations in word formation*, 49–84. Brill.

Hathout, Nabil & Fiammetta Namer. 2019. Paradigms in word formation: what are we up to? *Morphology* 29(2). 153–165.

Hathout, Nabil & Fiammetta Namer. 2022. Paradis: a family and paradigm model. *Morphology* 1–43.

Jackendoff, Ray & Jenny Audring. 2018. Relational Morphology in the Parallel Architecture. In *The Oxford Handbook of Morphological Theory*, 390–408. Oxford University Press.

Lafourcade, Mathieu & Alain Joubert. 2013. Bénéfices et limites de l'acquisition lexicale dans l'expérience jeuxdemots. In Nuria Gala & Michael Zock (eds.), *Ressources lexicales: Contenu, construction, utilisation, évaluation*, vol. 30 Linguisticae Investigationes, Supplementa, 187–216. Amsterdam: John Benjamins.

Roché, M. 2017. Les familles dérivationnelles: comment ça marche. *Toulouse: Université Toulouse* 2.

Ruppenhofer, Josef, Collin F. Baker & Charles J. Fillmore. 2003. The framenet database and software tools. In *Proceedings of the tenth euralex international congress*, 371–375. Copenhagen, Denmark.

Sowa, John F. 2014. *Principles of semantic networks: Explorations in the representation of knowledge*. Morgan Kaufmann.

Stump, Gregory T. 1991. A paradigm-based theory of morphosemantic mismatches. *Language* 675–725.

Van Marle, Jaap. 1985. *On the paradigmatic dimension of morphological creativity*. De Gruyter.