...

Adèle Hénot-Mortier

"Small" vectors in big spaces, and the semantics of French *-et(te)*-suffixation

...

https://doi.org/..., Received ...; accepted ...

Abstract: This paper investigates the behavior of the French "diminutive" suffix *-et(te)* through the lens of Distributional Semantics. Using various word embedding models assigning (sub)words to high-dimensional vectors, we show that the representation of *-et(te)*-suffixation depends on the gender of the base, that of the derivative, and crucially, on their interaction. Specifically, some models capture the empirical observation that gender-matching base-derivative pairs in *-et(te)* are semantically *closer* than mismatching pairs, and that, within each group (matching/mismatching), pairs formed with the feminine suffix *-ette* tend to be semantically closer than those formed with its masculine counterpart *-et*. In terms of semantic *stability*, *-et(te)*-suffixation is shown to be more stable across gender-matching base-derivative pairs than gender-mismatching ones. Although their robustness should be further assessed, these findings are consistent with the empirical picture. Furthermore, they support the hypothesis that *-et(te)* is ambiguous between a transparent, gender-matching diminutive modifier, and a head characterized by a looser, "similative" semantics.

Keywords: diminutives, distributed morphology, distributional semantics, word embedding models

PACS: ...

Communicated by: ...
Dedicated to ...

Adèle Hénot-Mortier, Massachusetts Institute of Technology

1 The French diminutive suffix *-et*/*-ette*

Diminutive and augmentative affixes generally express a difference of size and/or an affective relation towards the base word (Stump 1993, Jurafsky 1996 i.a.). The semantics of the French "diminutive" suffix *-et*/*-ette*, is however notoriously hard to describe in a uniform fashion (Hasselrot, 1957; Weber, 1963; Hasselrot, 1972; Milner, 1989; Delhay, 1995; Fradin, 2003). Based on Stump (1998)'s criteria, and on Hénot-Mortier (2024)'s observations, this Section characterizes *-et(te)* in terms of feature preservation and semantic transparency, productivity, range and domain, and its interaction with other morphology. These criteria are shown to depend on the gender of the derivative (masculine *-et* vs. feminine *-ette*) and its interaction with the gender of the base.

1.1 Feature preservation

French assigns grammatical gender (masculine or feminine) to all nominals. As shown in (1), feminine ("F") bases generally yield feminine diminutives, while masculine ("M") bases generally yield masculine diminutives (Bally, 1932). F-diminutives are marked with *-ette*, while M-diminutives are marked with *-et*.

(1) a. maison$_{\mathrm{F}}$ → maisonn-*ette*$_{\mathrm{F}}$
'house' → 'small (cute) house'

b. balcon$_{\mathrm{M}}$ → balconn-*et*$_{\mathrm{M}}$
'balcony' → 'small (cute) balcony'

Milner (1989) however observed that *-ette* may attach to masculine bases (producing F-diminutives) and *-et* to feminine bases (producing M-diminutives). We dub this phenomenon *gender-mismatch.* Milner observed that gender-mismatch often leads to a looser semantic relationship between the base and the derived form. The pairs in (2) and (3a) exemplify this observation. Gender-mismatching derivatives tend to share some similarity with the base, usually in terms of shape or function. Certain bases, like *boule* ('ball'), can be suffixed with both *-et* and *-ette*, leading to both a gender-mismatching, non-diminutive pair (3a) and a gender-matching, diminutive pair (3b).

(2) char$_{\mathrm{M}}$ → char-*ette*$_{\mathrm{F}}$
char$_{\mathrm{M}}$ *→ char-*et*$_{\mathrm{M}}$
'chariot' → 'cart'

(3) a. boule$_\mathrm{F}$ → boul-*et*$_\mathrm{M}$
‘ball’ → ‘cannonball’/‘ball’(and chain)

b. boule$_\mathrm{F}$ → boul-*ette*$_\mathrm{F}$
‘ball’ → ‘small ball’

This suggests that gender-matching and mismatching forms are derived from distinct morphological processes: the former, featurally and semantically transparent (the derivative is a diminutive of the base and “inherits” its gender), the latter, featurally and semantically more opaque (the derivative is some “approximation” of the base,[1] and gets its own gender).

1.2 Productivity

The feminine variant of the suffix is relatively productive, while the masculine variant is sensibly less so, as shown by the English loans in (4). The successful derivations are all gender-matching and associated with a meaning of the form *small/insignificant M* with *M* the meaning of the base noun.

(4) a. *start-up*$_\mathrm{F}$ → *start-up-ette*$_\mathrm{F}$
b. *deadline*$_\mathrm{F}$ → *deadline-ette*$_\mathrm{F}$
c. *punchline*$_\mathrm{F}$ → *punchlin-ette*$_\mathrm{F}$
d. *brunch*$_\mathrm{M}$?→ *brunch-et*$_\mathrm{M}$
e. *workshop*$_\mathrm{M}$?→ *workshop-et*$_\mathrm{M}$
f. *feedback*$_\mathrm{M}$ *→ *feedback-et*$_\mathrm{M}$
g. *meeting*$_\mathrm{M}$ *→ *meetingu-et*$_\mathrm{M}$
h. *challenge*$_\mathrm{M}$ *→ *challeng-et*$_\mathrm{M}$

This is substantiated by quantitative data: an analysis of the (native) French lexicon shows that -*ette*-suffixation is two times more prevalent that -*et*-suffixation in gender-matching cases, despite the fact that feminine and masculine nominals were in equal proportions in the dataset at stake (Hénot-Mortier, 2024). This gender difference suggests that -*et* and -*ette* underlyingly differ, at the very least for a subset of the matching cases. For gender-mismatching forms, Hénot-Mortier observes that the rates of -*et* and -*ette* forms are roughly equal, which contrasts with the gender-matching group. This discrepancy between gender-matching and mismatching forms is in line with the idea that these forms do not result from the same morphological process.

1 See Brucale and Mocciaro (2023); Stosic and Amiot (2023) for discussions on approximation and diminution in morphology.

1.3 Range and domain

-et/*-ette* can apply to various syntactic categories, but sometimes preserves and sometimes changes, this input category. As noted in e.g. Dal (1997), *-et(te)* surfaces on nominals, but also on adjectival bases (creating new adjectives, see (5)), and verbal bases (creating deverbal nouns, see (6)). In the adjective case, gender is determined *via* agreement with the modified noun. In the verb-to-noun case, derivatives are preferentially feminine; this again points to a difference between *-ette* (feminine and domain-general), vs. *-et* (masculine and more domain-specific).

(5) a. mignon → mignon-*et(te)*
‘cute’ → ‘cutesy’

b. gentil → gentill-et(te)
‘nice’ → ‘niceish’

(6) a. balayer → balay-$ette_F$
‘sweep’ → ‘brush’

b. siffler → siffl-et_M
‘whistle’ → ‘whistle’

1.4 Interaction with other morphology

The feminine variant *-ette* may surface after some derivational morphology (7), but always before inflection (e.g. pluralization). The masculine variant *-et* also appears before inflection; no evidence was found in support of *-et* appearing after derivation.[2]

(7) *révolu-tion*$_F$ → *révolu-tion-ette*$_F$
‘revolution’ → ‘small/insignificant revolution’

-et/*-ette* cannot be staked to produce an intensified diminutive meaning; however, a similar effect can be achieved using *-inet*/*-inette*, which, in French at least, does not appear decomposable into two independent suffixes (*-in*+*et(te)*); see (9).[3] If *-inet(te)* is indeed one single morpheme in French, then the unavailability of iterated *-et(te)*-suffixation may stem from competition. Indeed, assuming iterated *-et(te)*-suffixation and *-inet(te)*-suffixation have the same semantic effect, iterated *-et(te)*-suffixation should be dispreferred due to being structurally more complex.

(8) a. *maison-ette*$_F$ *→ *maison-et(te)-ette*$_F$
?→ *maison-inette*$_F$
‘small house’ → ‘very small house’

2 However, this might be because many derived nominals are feminine, and diminutives are generally gender-matching.

3 This differs from Italian; see De Belder et al. (2014) i.a.

b. *bomb-ette*$_F$ *→ *bomb-et(te)-ette*$_F$
→ *bomb-inette*$_F$
'small bomb' → 'very small bomb'

(9) *maison*$_F$ *→ *maison-ine*$_F$
'house' *→ 'small house'

In summary, although *-et(te)* is a suffix that can apply to different categories of base, it is most productive in the nominal domain. Productivity and semantic transparency in this domain seem to be modulated by two factors: whether *-et(te)* changes the gender of the base; and whether the derivative is a feminine form in *-ette* or a masculine form in *-et*. In the rest of this paper, we will therefore focus on the semantic effect of *-et(te)* on nominals, and how it gets modulated by gender. Specifically, we will investigate if distributional models of language capture the observation that gender-mismatching *-et(te)* is less likely to be transparent than gender-matching *-et(te)*, and if, within each group (matching/mismatching), forms in *-ette* are more transparent than forms in *-et*. Before doing so, we review a recent approach to this suffix in the Distributed Morphology framework.

2 An approach to *-et(te)* in the Distributed Morphology Framework

2.1 Background on Distributed Morphology

Distributed Morphology (henceforth DM, Halle and Marantz 1993, 1994) offers a framework to connect the morphosyntactic and semantic behavior of *-et(te)*. In DM, syntactic operations like Merge, Move, and Agree, apply to formatives. Formatives are bundles of syntactic features, and get mapped to exponents (phonological forms) post-syntactically. Importantly in DM, roots (which are formatives) are category-neutral and must undergo categorization by the action of *heads*: n (nominalizer), a (adjectivizer), v (verbalizer) etc; see Figure 1.

Subsequent work within the DM framework points out the crucial distinction between creating words from roots and creating words from existing words, that is, from roots that are already merged with some word-creating head (Marantz 1997; Harley and Noyer 1998; Marantz 2001; Arad 2005, 2005 i.a.). This instantiates the Lexical Decomposition Hypothesis.[4] Behind this division is the idea that merging

4 This hypothesis is often associated with DM, but not restricted to this framework (Borer, 2005, 2013; Fathi and Lowenstamm, 2016).

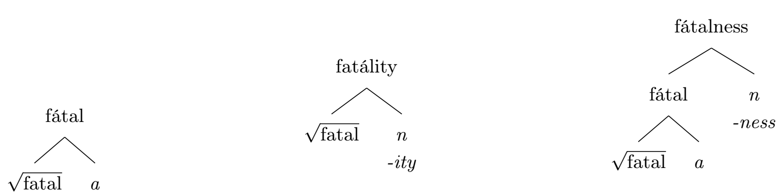
a head creates a new “opaque” morphological object whose internal structure and properties cannot be accessed by subsequent morphological operations. In particular, the first head to be merged takes a morpho-phonologically and semantically underspecified root and assigns it a form and a meaning that will be the only elements available to further transformations; it is thus expected to “set” the semantic and morpho-phonological features of the newly created word.

(a) The noun ‘hammer’ obtained by merging a nominalizing head (*n*) with the category-neutral root √hammer.

(b) The verb ‘to hammer’, obtained by merging a verbalizing head (*v*) with the category-neutral root √hammer.

Fig. 1: Categorizing a root in different ways in DM

This can successfully explain the existence of opacity effects witnessed in *both* the semantic and the phonological domain. For instance, affixes such as *-ity*, which appear to “shift” the stress of the base they attach to, also tend produce semantically less predictable meanings than otherwise similar affixes, such as *-ness*, which do not shift stress (Aronoff, 1976). This pattern can be captured by assuming that the former kind of affix merges at the root level (where several possible forms and meanings are still accessible and modifiable), while the latter kind merges with already created words, whose phonology and semantics are more fixed; see Figure 2.[5]



(a) The adjective ‘fátal’, bearing initial stress.

(b) The noun ‘fatálity’: shifted stress; means an occurrence of death, or helplessness in the face of fate.

(c) The noun ‘fátalness’: initial (non-shifted) stress; means the quality of being fatal.

Fig. 2: Deriving *fatálity* (root-level) and *fátalness* (adjective-level) from the root √fatal

5 In DM, a root-level affix like *-ity* therefore does not genuinely “shift” stress; it just produces a different stress assignment than the one that would have arisen from the merger of a root-level adjectivizing head *a* at the same site.

2.2 Diminutives in DM

Why is DM of interest in the context of -*et(te)*-suffixation then? By assuming diminutives are either modifiers or nominalizing heads, and may merge at the root or at the word level, Wiltschko and Steriopolo (2007) derived four different clusters of properties characterizing diminutives.[6] First, the "level" parameter (root vs. word) controls the affix' productivity (root derivations are less productive), the range of syntactic categories the affix appears to apply to,[7] and whether or not it can appear above derivational markers (only word level derivations can). Second, the "type-of-merge" parameter (head vs. modifier) controls whether or not an affix can modify the morpho-syntactic features of its input (e.g. gender[8]). Interestingly, languages like Russian feature diminutive suffixes varying across *both* parameters (Steriopolo, 2017). Italian, Modern Hebrew, Polish, Spanish, and Tunisian Arabic were also argued to exhibit language-internal variation of the level parameter (De Belder et al., 2014). Some languages (e.g. Italian) can recruit the *same* surface suffix for either root- or word-level derivations.

To the above syntactic characterizations, Hénot-Mortier (2024), building on earlier work (Arad, 2003; De Belder et al., 2014), adds the following semantic ones.

First, root-level categorization should be less semantically transparent than word-level derivations, whether they amount to categorization or modification. To see this, consider a root $\sqrt{r}$, with an underspecified predicative meaning $\tilde{M}$. Set theoretically, $\tilde{M}$ can be understood as a union of more specified predicative meanings. For instance, $\sqrt{\text{ball}}$ may denote the set of individuals which are vaguely ball-like (spherical, solid, mesoscopic objects).[9] The nominalization of a root $\sqrt{r}$, $n(\sqrt{r})$, is assumed to have meaning M, arbitrarily precisifying $\tilde{M}$. M can be understood as a subset of $\tilde{M}$. For instance, $n(\sqrt{\text{ball}})$ may select, out of all the ball-like individuals, the ones that could be narrowly described as balls. The derivation of this meaning is shown in Figure 3a.

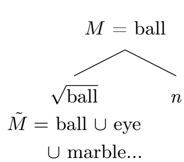
6 See also De Belder et al. (2014) for similar though less parametrized approaches.

7 Root-level derivations, unlike word-level ones, can seemingly apply to many such categories; while in effect, they apply to the same uncategorized root.

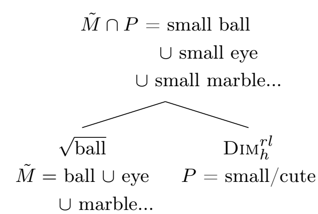
8 Our background assumption will be that gender features are neither represented in category-neutral roots, nor in a dedicated projection. Instead, they will be assumed to be hosted in nominalizing heads (Lecarme, 2002; Kihm, 2005; Ferrari, 2005; Acquaviva, 2009; Kramer, 2015, 2016; Deal, 2016). Crucially, we take this to include evaluative heads (including diminutive heads).

9 It may also include the denotation of more abstract verbal predicates (e.g. the action of *rolling*), or adjectives (e.g. the state of being *round*), but those meanings will most likely be filtered by subsequent nominalization/diminutization.

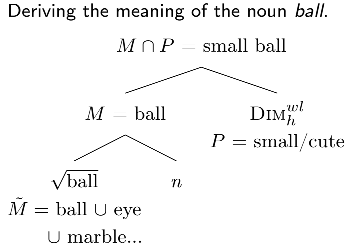
Let A_h^{rl} be a root-level head with meaning P, assumed to be predicative. For instance, if A is diminutive, P should refer to a set of small/cute individuals. Assuming that Merge semantically amounts to predicate intersection, $\mathrm{A}_h^{rl}(\sqrt{r})$ is expected to intersect P with the underspecificed root's meaning $\tilde{M}$. For instance, $\textsc{Dim}_h^{rl}(\sqrt{\text{ball}})$ would end up denoting small/cute, ball-like elements, including small marbles or small eyes, which would not exactly qualify as small balls. This is shown in Figure 3b. More generally, the meaning of $\mathrm{A}_h^{rl}(\sqrt{r})$ is not transparently related to that of the nominalized root, M.
Let A_h^{wl} be a word-level head with meaning P. Applying A_h^{wl} to the *nominalized* root $n(\sqrt{r})$ (with meaning M) produces the meaning $P \cap M$, i.e. the set of elements M that are P – a subset of M. For instance, $\textsc{Dim}_h^{wl}(n(\sqrt{\text{ball}}))$ would end up denoting small/cute individuals that are balls *stricto sensu*; see Figure 3c. More generally, the meaning of $\mathrm{A}_h^{wl}(n(\sqrt{r}))$ is transparently related to that of the nominalized root, M. The same would hold if a word-level *modifier* with meaning P got merged instead. This implies that root-level categorization is less semantically transparent than word-level derivations in general.



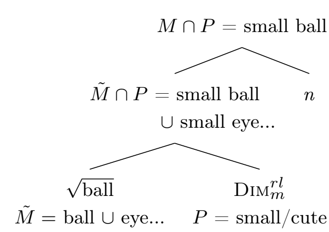
(a) Deriving the meaning of the noun *ball*.



(b) Root-level head-diminutization of *ball*.



(c) Word-level head-diminutization of *ball*.



(d) Root-level modification of *ball*.

Fig. 3: Semantics of nominalization and diminutization, taking the root $\sqrt{\text{ball}}$ as example.

The second semantic prediction formulated by Hénot-Mortier is that root-level modification should be more semantically transparent than root-level categorization. To see this, let A_m^{rl} be a root-level modifier with meaning P. Applying A_m^{rl} to $\sqrt{r}$

yields meaning $P \cap \tilde{M} \subseteq \tilde{M}$, which retains some degree of underspecification. For instance, the denotation of $\text{Dim}_m^{rl}(\sqrt{r})$ would be expected to include small eyes and small marbles. When $\text{A}_m^{rl}(\sqrt{r})$ gets subsequently nominalized, the semantic output is then expected to be a precisification of $P \cap \tilde{M}$ which can be reasonably assumed to correspond to $P \cap M$.[10] In other words, nominalization and modification should commute at the semantic level, which is exemplified in Figure 3d. Overall, this implies that root-level modification should be as semantically transparent as word-level derivations; and more transparent than root-level categorization. The next section reviews how these general predictions apply to the case of *-et(te)*.

2.3 A DM account of the semantic and morphosyntactic behavior of *-et(te)*

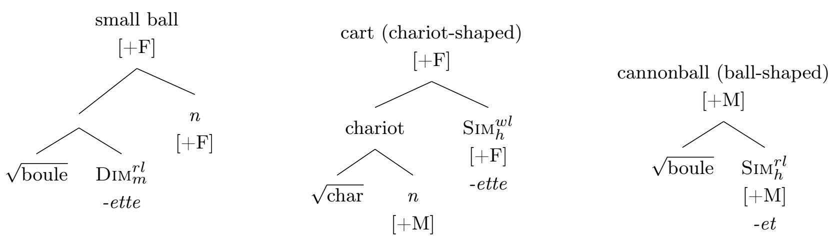
Based on key empirical datapoints, Section 1 showed that in the nominal domain, *-et(te)* varies across two "dimensions". It can be gender-matching (majority of the cases) or gender-mismatching (non-negligible minority of the cases). And the resulting meaning can be purely diminutive, or rather "similative". Through a more systematic analysis of the French lexicon, Hénot-Mortier (2024) confirms these empirical observations and provides statistical evidence indicating that gender-matching base-derivative pairs are more common, and also more likely to stand in a diminutive relation, than gender-mismatching ones (in line with Milner's original observation). Also, independently of gender-(mis)match, pairs resulting in masculine derivatives in *-et* appear overall less likely to be diminutive than those resulting in feminine derivatives in *-ette*.[11]

Hénot-Mortier (2024) accounted for this pattern by proposing that *-et(te)* may be the realization of two main kinds of suffixes. The first kind is a purely diminutive root-level modifier (labeled Dim_m^{rl}), which is productive, as well as morphosyntactically and semantically transparent. This suffix is expected to produce the vast majority of gender-matching, purely diminutive forms such as *boulette*; see (3b) and Figure 4a for a derivation.

10 This presupposes that the kind of semantic precisification performed by a categorization operation is restriction-stable. For instance, if $P \cap \tilde{M}$ denotes the set of small ball-like elements (small balls, small eyes, small marbles etc.), it is natural to assume that the precisification effect of nominalization should restrict this set to that of small balls *stricto sensu*.

11 This difference was proved significant in the matching case; and constituted a trend in the mismatching case – characterized by very small sample sizes.

The second kind of suffix is a non purely diminutive gendered head labeled $\textsc{Sim}_h$. This kind of head is assumed to have a nominalizing effect, and carries gender features that get "imposed"[12] on the base, likely leading to base-derivative gender mismatches. Semantically, $\textsc{Sim}_h$ conveys a similarity in shape or function between base and derivative. This similarity may accidentally display diminutive traits, for a restricted number of base-derivative pairs derived *via* $\textsc{Sim}_h$. Lastly, Hénot-Mortier proposed that the merger site of $\textsc{Sim}_h$ crucially depends on the gender feature this head carries: word-level if feminine, and root-level, if masculine. Together, these assumptions predict that the feminine, word-level Sim head, labeled $\textsc{Sim}_h^{wl}$, should apply quite productively, and give rise to semantically transparent feminine derivatives (even if the base was originally nominalized as masculine). By contrast, the masculine root-level Sim head, labeled $\textsc{Sim}_h^{rl}$, should apply less productively, bleed the insertion of a higher *n*-head, and give rise to semantically more opaque masculine derivatives. Figures 4b and 4c show how the two variants of $\textsc{Sim}_h$ apply to their base; note that $\textsc{Sim}_h^{rl}$ in particular, acts as a semantically contentful nominalizer, imposing masculine on the newly formed word. Also note that $\textsc{Sim}_h^{wl}$ and $\textsc{Sim}_h^{rl}$ may apply to bases that would have a matching gender if independently nominalized, which means that a minority of gender-matching forms are expected to display a Sim semantics (as opposed to a Dim semantics).



(a) Deriving the gender-matching, diminutive form *boulette* ('small ball', (3b)). Derivation similar to Figure 3d.

(b) Deriving the M-to-F non-diminutive form *charette* ('cart', (2)).

(c) Deriving the F-to-M non-diminutive form *boulet* ('cannonball', (3a)).

Fig. 4: Three derivations leading to surface similar *-et(te)* form, with different genders and semantics.

12 This means that the gender features carried by a head like $\textsc{Sim}_h$ may *overwrite* gender features introduced by lower heads, e.g. *n*, if present. We thank an anonymous reviewer for identifying this *caveat*. This assumption is implicitly made in e.g. Wiltschko and Steriopolo (2007), to account for the properties of the German diminutive *-chen*, a noun-specific suffix appearing above nominalizing morphology (indicating word-level status), which nevertheless imposes [+neuter; +count] on the derivative (suggesting head status).

This analysis was shown to be in line with the empirical data. Let us briefly summarize the argument. First, gender-matching forms in *-ette* (FF-group) can either result from the application of $\textsc{Dim}_m^{rl}$ or $\textsc{Sim}_h^{wl}$, while gender-matching forms in *-et* (MM-group) can result from the application of either $\textsc{Dim}_m^{rl}$ or $\textsc{Sim}_h^{rl}$. This predicts both groups to often be diminutive, due to their compatibility with the productive, transparent, diminutive suffix $\textsc{Dim}_m^{rl}$. This also predicts the FF-group to be slightly more transparent than the MM group, given that the feminine $\textsc{Sim}_h^{wl}$ applies to words, while the masculine $\textsc{Sim}_h^{wl}$ applies to underspecificed roots. Second, gender-mismatching forms in *-ette* (MF-group) can only arise from $\textsc{Sim}_h^{wl}$, while gender-mismatching forms in *-et* (FM-group) can only arise from $\textsc{Sim}_h^{rl}$. The unavailability of $\textsc{Dim}_m^{rl}$ in these two groups predicts them to only be diminutive "by accident". This cashes out the observation that gender-mismatching pairs are overall less diminutive than gender-matching ones. Additionally, the subtle contrast between feminine and masculine forms derived in the matching groups is expected to be preserved in the mismatching groups: the MF-group should be slightly more transparent than the FM-group, due to $\textsc{Sim}_h^{wl}$ (feminine) being merged at the word level, unlike $\textsc{Sim}_h^{rl}$ (masculine).

In brief, we expect *-et(te)* to have a different semantics in the gender-matching vs. mismatching groups, because the former, unlike the latter, is compatible with $\textsc{Dim}_m^{rl}$. In both gender-matching and gender-mismatching groups, we also expect a small difference between *-ette* (more transparent) and *-et* (less transparent). These predictions are summarized in Table 1.

Base-derivative group	**Possible derivation**	**Productivity**	**Semantic transparency**
FF	$\textsc{Dim}_m^{rl}$	✓	Diminutive on precise base
	$\textsc{Sim}_h^{wl}$	✓	Similative on precise base
MM	$\textsc{Dim}_m^{rl}$	✓	Diminutive on precise base
	$\textsc{Sim}_h^{rl}$	✗	Similative on vague base
MF	$\textsc{Sim}_h^{wl}$	✓	Similative on precise base
FM	$\textsc{Sim}_h^{rl}$	✗	Similative on vague base

Tab. 1: Predicted semantic effect of *-et(te)*-suffixation, as a function of the genders of the base and of the derivative (Hénot-Mortier, 2024).

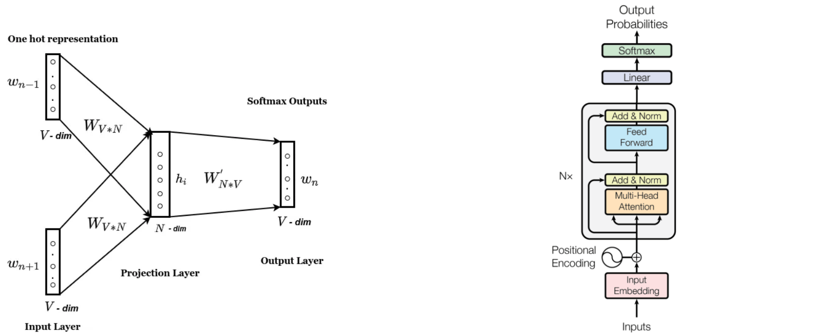
The next section introduces computational tools useful for quantitatively assessing semantic productivity in morphology, and shows how these tools can be specifically applied to *-et(te)*-suffixation.

3 Computational tools for morphosemantics

3.1 Distributional Semantics and Language Models

According to the Distributional Hypothesis (henceforth DH, Harris, 1954), words occurring in similar syntactic environments should be semantically close. For instance, *cat* and *dog* both distribute like nouns, and often cooccur with words like *cute*, *paws*, *sniffing*, *petting* etc. Their environments will thus be more similar to each other, than, say, those of *cat* and *banana*, or *cat* and *jump*. There are two important things to note about the DH. First, it makes predictions about semantic closeness (a relative measure), but not about absolute meanings. Second, the DH is not a clear biconditional hypothesis; elements occurring in the same syntactic environments will also be more likely to belong to the same syntactic category and share features like gender. Thus, fully equating syntactic distribution with semantic meaning in fact incorporates a variety of non-purely semantic factors inside the concept of "meaning".

To define semantic closeness under the DH, one must model what environments are, and how to compare them. Language Models (henceforth LM), understood as architectures trained to efficiently predict words from other words (*environment*), can help. As a byproduct of their prediction objective, LMs produce efficient vector representations of words or subwords. Figure 5 details two LM architectures: the Continuous Bag of Words (CBOW; (Mikolov et al., 2013a,b) predicting a center word from an unordered list of surrounding words using a simple feedforward Neural Network; and the more recent and complex Encoder Transformer, based on the concept of self-attention (Vaswani et al., 2017; Devlin et al., 2019).



(a) The CBOW architecture used to produce Word2Vec embeddings (Mikolov et al., 2013a,b).

(b) Building block of the BERT architecture (Devlin et al., 2019), taken from Vaswani et al. (2017).

Fig. 5: Two LM architectures predicting (sub)words from context (sub)words.

In both cases, the LM is passed a list of words or subwords (*environment*),[13] converts them into vectors, and performs algebraic operations on them to produce an output vector corresponding to a complex mixture of the inputs. This vector is rescaled to define a probability distribution over the lexicon, from which the word to predict is sampled. Because LMs are trained to map environments to suitable "filler" words, the output vector obtained before the sampling step is expected to both summarize relevant aspects of the environment and to capture the filler's meaning in a condensed, efficient fashion. Such vectors are commonly called word-vectors and the space they form, (word) embedding.[14]

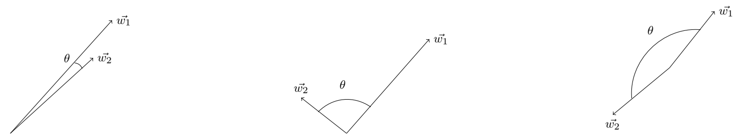
The fact LMs assign (sub)words to vectors incorporating information about the environment they occur in enables comparisons between (sub)word meanings *via* their corresponding vectors, as per the DH. A commonly-used measure of similarity between word-vectors is cosine similarity (henceforth CosSim), which is proportional to the angle between two vectors: if $\text{CosSim}(\vec{w_1}, \vec{w_2}) = +1$, $\vec{w_1}$ and $\vec{w_2}$ share direction and orientation (their angle is $0°$), and the corresponding words should be semantically close (see Figure 6a); if $\text{CosSim}(\vec{w_1}, \vec{w_2}) = 0$, $\vec{w_1}$ and $\vec{w_2}$ are orthogonal (their angle is $90°$), and the corresponding words should be semantically unrelated (see Figure 6b). CoSims between word-vectors are rarely close to -1 (extreme case shown in Figure 6c). Unlike Euclidean distance, CosSim is insensitive to the vectors' magnitudes; it only measures if they display the same *proportions* of their various features.

Arithmetic operations on the embedding space may also be assigned an interpretation. For instance, subtracting the word-vector of a base from that of its derivative (see (10)), may represent the contrast in meaning between the two

13 In the CBOW case, the input environment is modeled as a sequence of one-hot encoded word-vectors. The model is insensitive to the order of the words in the sequence. In the case of BERT-like Encoders, the environment is encoded as a sequence of tokens (integer labels assigned to words or subwords based on frequent character co-occurrences). These tokens are themselves encoded as vectors incorporating information about their position in the sequence, making Encoders order-sensitive. The environment contains a special `MASK` token understood as a placeholder for the (sub)word to predict.

14 In the CBOW model, the embedding space (which corresponds to the rows of matrix W in Figure 5a) is "static": it is computed once and for all for each word of a given training dataset. In the case of BERT, word-vectors are extracted from the last or penultimate layer of the model, after feeding it a specific input sequence. The penultimate layer in particular has been argued to capture high-level linguistic information, while remaining fairly task-independent (Xiao, 2018). Because the output depends on which sequence is passed to the model, any given word can be assigned a different vector depending on which sequence containing it is fed to BERT. For that reason, BERT embeddings are called "contextualized".

forms, i.e. the semantics (in the broad sense) of the morphological operation itself.[15] We will call such vectors “affixal vectors”. On top of pure semantics, affixal vectors may incorporate information about the morphosyntactic features that the affix modifies (e.g. gender), or about the change of syntactic category it induces.



(a) $\theta \simeq 0°$
CosSim$(\vec{w_1}, \vec{w_2}) \simeq +1$

(b) $\theta \simeq 90°$
CosSim$(\vec{w_1}, \vec{w_2}) \simeq 0$

(c) $\theta \simeq 180°$
CosSim$(\vec{w_1}, \vec{w_2}) \simeq -1$

Fig. 6: CosSim as a function of the angle between two vectors.

(10) *Affixal vectors.* If d is derived from b *via* a-affixation, then the vector representation of a given the pair (b, d), is $\overrightarrow{a} = \overrightarrow{d} - \overrightarrow{b}$.

3.2 Previous work of LMs and morphology

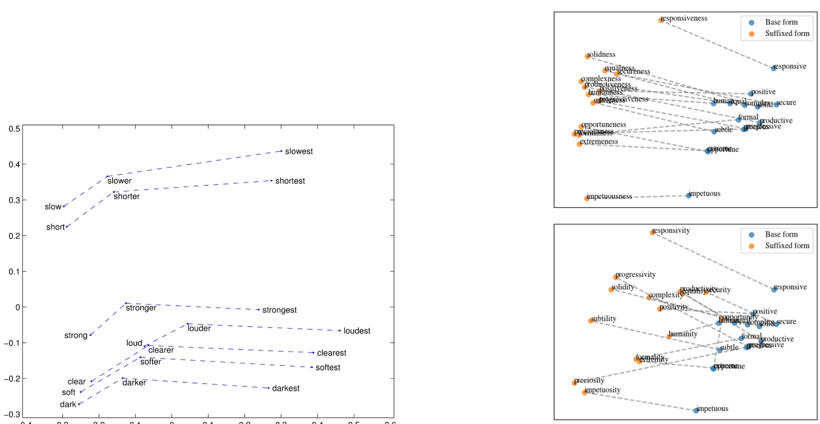
The fact that embeddings encode morphological information has been confirmed for various languages and morphological phenomena. Pennington et al. (2014) showed that English embeddings encode operations such as comparative and superlative formation as stable geometric translations (i.e. stable affixal vectors). Figure 7a shows how affixal vectors associated with *-er* (comparative) and *-st* (superlative, applied on top of a comparative[16]), represented as dashed lines, appear parallel across base-derivative pairs – this near-parallelism implies high CosSims among these vectors. Similarly, Musil et al. (2019) showed that Czech embeddings encode various derivational operations as different clusters of vectors.

Embeddings also appear to encode specific *properties* of morphological operations. First, competition between operations: Naranjo and Bonami (2023) show that rival operations in French yields affixal vectors that are indistinguishable; Hénot-Mortier (2022) shows that, when two similar morphological processes may

15 There exists more complex ways to represent affixation; Marelli and Baroni (2015) for instance propose that affixes correspond to linear operations (i.e. matrices) mapping a base to its derivative.

16 In line with the Containment Hypothesis (Bobaljik, 2012).

in principle be blocked by the existence of a lexical item with a similar meaning,[17] the blocked operation happens to be closer in the embedding space to the blocking transformation, than the licensed operation is. More broadly, embeddings have been used to compare the different contributions of distinct morphological processes; Bonami and Guzmán Naranjo (2023) for instance, use word-vector to measure the predictive power of paradigmatic relations in French, as opposed to that of derivational relations.
Second, stability: Bonami and Paperno (2018) showed, based on French data, that embeddings capture the observation that inflection is more "regular" than derivation. Based on similar methods, Benbaji-Elhadad et al. (2022, 2024) show that Hebrew embeddings encode denominal verbs (word-level derivatives) as closer to their base than verbs directly derived from the same root; and also show that pairs of English suffixes (like *-ness*/*-ity*) differ in terms of stability (see Figure 7b).



(a) Comparative and superlatives operations form parallel "lines" in GloVe (Pennington et al., 2014).

(b) *-ness* (word-level, parallel lines) and *-ity* (often root-level, less parallel lines) suffixation in fastText (Benbaji-Elhadad et al., 2024).

Fig. 7: Affixal vectors in 2-dimensional word embeddings.

In a similar vein, Schäfer (2023) shows that *-ly* in English has different signatures in some embedding spaces – in line with the idea that *-ly* is sometimes inflectional/word-level and sometimes derivational/root-level. All these works (Bonami and Paperno 2018; Benbaji-Elhadad et al. 2022, 2024; Schäfer 2023) exploit two main metrics to quantify semantic transparency: locality, as measured

17 For instance *glory* blocks **gloriosity*, but not *gloriousness* (Aronoff, 1976).

by base-derivative CosSim (11), and affix stability, as measured by the CosSim between individual affixal vectors and their centroid (12).[18]

(11) *Transparency as base-derivative locality*. Let a and a' be two operations with roughly similar meanings, but a is productive and transparent (inflection/word-level operation), while a' less so (derivation/root-level operation). If d is derived from b *via* a, and d' is derived from b' *via* a', then d is expected to be on average semantically closer to b than d' to b', i.e.: $\text{CosSim}(\overrightarrow{b}, \overrightarrow{d}) > \text{CosSim}(\overrightarrow{b'}, \overrightarrow{d'})$.

(12) *Transparency as affix stability*. Let a and a' be as in (11). Let $(\overrightarrow{a}_i)_{i\in I}$ and $(\overrightarrow{a'}_i)_{i\in I}$ be two same-size families of affixal vectors computed by difference (following (10)) over many base-derivative pairs resulting from a- or a'-affixation. Because a is assumed more transparent than a', $(\overrightarrow{a}_i)_{i\in I}$ should be less "spread" than $(\overrightarrow{a'}_i)_{i\in I}$. Let $\overrightarrow{c}$ and $\overrightarrow{c'}$ be the respective centroids (vector averages) of $(\overrightarrow{a}_i)_{i\in I}$ and $(\overrightarrow{a'}_i)_{i\in I}$. The $(\overrightarrow{a}_i)_{i\in I}$ vectors should be overall closer to their centroid $\overrightarrow{c}$ than the $(\overrightarrow{a'}_i)_{i\in I}$ to $\overrightarrow{c'}$, i.e., for most $i \in I$: $\text{CosSim}(\overrightarrow{c}, \overrightarrow{a_i}) > \text{CosSim}(\overrightarrow{c'}, \overrightarrow{a'_i})$.

We will follow a similar approach to study the effect of *-et(te)*-suffixation across base-derivative pairs, and interactions with gender change. The next section proceeds to adapt (11) and (12) to this case study.

3.3 Predictions for *-et(te)*

Section 3.2 just modeled transparency as increased base-derivative similarity within the embedding space (11), and increased affixal vector stability (12). Moreover, Section 2.2 concluded that root-level categorization should be less semantically transparent than both root-level modification and word-level derivations. In the case of *-et(te)*, Hénot-Mortier's (2024) account then leads to the prediction that $\text{D}\textsc{im}_m^{rl}$ and $\text{S}\textsc{im}_h^{wl}$ should be more local (as per (11)) and stable (as per (12)) than $\text{S}\textsc{im}_h^{rl}$. But the core semantics of Sɪᴍ and Dɪᴍ should also be taken into account: because Dɪᴍ is assumed to be more semantically specific than Sɪᴍ, it is expected to be more local and stable than Sɪᴍ in the vector space.

18 Bonami and Paperno (2018) use the Euclidean distance instead of CosSim in (12) – reversing inequalities. Following Benbaji-Elhadad et al. (2024), we will use CosSim instead, for consistency with (11), and also because we want to translate semantic stability as vector parallelism, as illustrated by Figure 7. Euclidean distance would not guarantee that this equivalence holds.

These observations (compiled in Table 1) translate into the following topological characterizations: the gender-matching groups of base-derivative pairs (FF and MM) should involve more local and stable transformations than the gender-mismatching ones (FM and MF). Among the two gender-matching groups, the FF group should be characterized by slightly more locality and stability than the MM group. Among the two gender-mismatching groups, the MF group should be characterized by slightly more locality and stability than the FM group.
To determine if these predictions are captured by word embeddings, we consider a dataset of base-derivative pairs divided into four equally-sized groups (FF, MM, MF, FM). For each base-derivative pair (b, d) within that dataset, vectors for the base ($\overrightarrow{b}$) and the derivative ($\overrightarrow{d}$) are extracted from a LM, and the corresponding affixal vector ($\overrightarrow{a} = \overrightarrow{d} - \overrightarrow{b}$) can be computed. For any group G (FF, MM, MF, or FM), $\overrightarrow{c^{\mathrm{G}}}$ is defined as the centroid of the affixal vectors belonging to group G. Given that framework, the general predictions in (11-12) are rephrased in (13-14).

(13) *Base-derivative locality*. Groups should be ordered according to how semantically close the derivative in *-et(te)* is from its base. The following two predictions should hold on average:

a. Effect of gender-(mis)match: FF and MM derivations are more local than MF and FM ones. For (b, d) a gender-matching pair, and (b', d') a gender-mismatching one: $\mathrm{CosSim}(\overrightarrow{b}, \overrightarrow{d}) > \mathrm{CosSim}(\overrightarrow{b'}, \overrightarrow{d'})$

b. Effect of suffix gender: FF (resp. MF) derivations are more local than MM (resp. FM) ones. For (b, d) and (b', d') two base-derivative pairs, both matching or both mismatching, with d a F-form in *-ette* and d' a M-form in *-et*: $\mathrm{CosSim}(\overrightarrow{b}, \overrightarrow{d}) > \mathrm{CosSim}(\overrightarrow{b'}, \overrightarrow{d'})$

(14) *Affix stability*. Groups should be ordered according to how close their affixal vectors are to their respective centroids. The following two predictions should hold on average:

a. Effect of gender-(mis)match: FF and MM derivations are more stable than MF and FM ones. For $(b, d) \in \mathrm{G} = \mathrm{FF}$ or MM, $(b', d') \in \mathrm{G}' =$ MF or FM, and given $\overrightarrow{a} = \overrightarrow{d} - \overrightarrow{b}$ and $\overrightarrow{a'} = \overrightarrow{d'} - \overrightarrow{b'}$ their respective affixal vectors: $\mathrm{CosSim}(\overrightarrow{c^{\mathrm{G}}}, \overrightarrow{a}) > \mathrm{CosSim}(\overrightarrow{c^{\mathrm{G}'}}, \overrightarrow{a'})$

b. Effect of suffix gender: FF (resp. MF) derivations are more stable than MM (resp. FM) ones. For $(b, d) \in \mathrm{G}$ and $(b', d') \in \mathrm{G}'$ two base-derivative pairs, both matching or both mismatching, with d a F-form in *-ette* and d' a M-form in *-et*, and given $\overrightarrow{a} = \overrightarrow{d} - \overrightarrow{b}$ and $\overrightarrow{a'} = \overrightarrow{d'} - \overrightarrow{b'}$ their respective affixal vectors: $\mathrm{CosSim}(\overrightarrow{c^{\mathrm{G}}}, \overrightarrow{a}) > \mathrm{CosSim}(\overrightarrow{c^{\mathrm{G}'}}, \overrightarrow{a'})$

These predictions come with a *caveat*: word embeddings encode not only semantic information, but also purely (morpho)syntactic distributional regularities. In particular, the vector representations of gender-mismatching operations (MF and FM groups) are likely to capture gender change along one or more relevant dimensions. Such information may play a role in (13a) and (14a). Controlling for the contribution of gender change while preserving other semantic dimensions (especially those related to "diminutiveness") is a difficult problem. With this limitation in mind, the next section describes how the predictions in (13) and (14) were tested across several LMs, and presents the resulting findings.

4 Assessing semantic transparency across gender groups

4.1 Dataset and models used

Data, models and analyses are publicly available on OSF. For each gender group (FF, MM, MF, FM), 44 base-derivative pairs were sampled from Hénot-Mortier's dataset (Hénot-Mortier, 2025).[19] 44 was the size of the smallest of the four groups (MF) in this dataset. Sampling was meant to ensure that subsequent dimensionality reduction operations applied to the embeddings would not "favor" more productive groups over the others, which could have constituted a confounding factor.[20] It also ensured that spread differences between groups of vectors (central to prediction (14)) would not be driven by differences in group size.[21] The results presented here appear robust to sampling.

Four pretrained LMs were tested: fastText (Grave et al., 2018); FlauBERT-Large-Cased (Le et al., 2020), CamemBERT-Large (Martin et al., 2020) and Canine-c (Clark et al., 2021). Importantly, these LMs exhibit varying degrees of sensitivity to subword information. fastText, an elaboration of the CBOW

19 Bases corresponding to proper names (along with their derivative) were excluded, because they may be likely to occur in environments varying significantly from those of regular nouns.

20 If a group of vectors (e.g. the FF group) is significantly larger than the others, then, its vectors may end up almost fully determining which dimensions "matter" when evaluating similarities: namely, the dimensions along which the vectors *of that particular group* vary the most. Consequently, this larger group may be analyzed as more variable than the others, for reasons that only stem from initial baserate differences.

21 Indeed, a comparatively smaller sample tends to yield a noisier estimate of the mean, which can lead to less stable measures of similarity between its datapoints and that mean.

model, takes into account not only individual words, but also a range of fixed-length substrings contained within them ("*n*-grams", which may sometimes coincide with actual morphemes, e.g. the 4-gram *ette*). As a result, FASTTEXT is sensitive to subword information, but in a way that may dilute the influence of proper morphological decomposition.[22] FLAUBERT and CAMEMBERT are Encoder LMs based on the BERT architecture (Devlin et al., 2019).[23] These models are sensitive to subword information through their tokenization procedure, i.e. the algorithm mapping input word sequences into integer sequences (later converted into vectors). These LMs' tokenization procedures depend on frequent character cooccurrences and as such display some overlap with morphological parsing; however they do not guarantee that every word-final *et(te)* sequence will be mapped to a token. This will be further investigated and discussed in Section 5. Lastly, CANINE-C displays key architectural similarities with FLAUBERT and CAMEMBERT, but is multilingual, and tokenizerless, i.e. operates purely at the character level. Despite its architectural elaboration, it is therefore not directly sensitive to subword information.[24]

4.2 Assessing base-derivative similarity

Word-vectors for the bases and derivatives of our dataset were extracted from the four LMs under consideration. FASTTEXT is readily available as a fixed embedding mapping a large sample of French words to individual vectors. FLAUBERT, CAMEMBERT and CANINE-C by contrast, can produce "contextualized" token-vectors after being fed a specific input sequence (typically one or more sentences). To every token in the input sequence, these LMs assign a vector sensitive to the other tokens in the input. Vectors for the target bases and derivatives were extracted from these LMs' penultimate (for FLAUBERT and CAMEMBERT) or last (for CANINE-C) layer, using the sequence *C'est un(e) ...* ('It's $\mathrm{a}_{\mathrm{F/M}}$ ...') as left context.[25] If the target base or derivative was split into multiple tokens by the LM, their corresponding vectors were averaged to form that of the entire word.

22 I thank a reviewer for pointing out this specificity of the model.

23 CamemBERT is actually building on ROBERTA (Liu et al., 2019), a more "robust" version of BERT.

24 This sensitivity may still be *derived* internally by the model, as it computes complex contextualized representations of its character-level inputs *via* self-attention. Such LMs thus constitute an interesting limiting case when assessing if linguistic generalizations can be genuinely extrapolated from poorly encoded inputs.

25 This context was meant to reduce potential syntactic ambiguities (by forcing the target word to be a nominal), and lexical ambiguities (by imposing a fixed gender *via* the article).

For each LM, corresponding word-vectors underwent Cosine-Kernel Principal Component Analysis (henceforth CosKPCA). This was intended to remove uninformative dimensions while optimally retaining pairwise measures of CosSim between word-vectors.[26] Table 2 summarizes the characteristics of the original and reduced vector spaces across LMs. The table also details how well the reduced spaces preserve CosSims between vectors of our dataset, by means of a correlation coefficient between (i) pairwise CosSims computed between vectors in the reduced space, and (ii) pairwise CosSims computed similarly in the original space.

Model	Initial Dimension	Reduced Dimension	Pearson correlation between the pairwise CosSims of raw vs. reduced word vectors
fastText	300	74	95.4%
FlauBERT	1024	32	97.8%
CamemBERT	1024	56	98.1%
Canine-c	768	31	99.2%

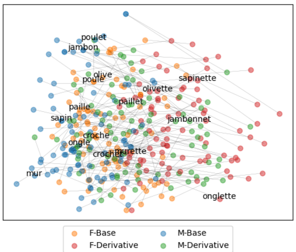
Tab. 2: Characteristics of the reduced word embeddings.

Figure 8 shows 2-dimensional CosKPCAs of our datapoints across LMs. These plots are expected to encode variation along gender and suffixation since those were the core factors the dataset was built on. Although not all models display a crisp 2D clustering based on these features, their topology is never fully random.[27] fastText and Canine-c exhibit a poorer clustering and preserve the geometry of the original space less well than the other LMs (correlation with original space $\sim$ 50%, vs. $\sim$ 75% for the other LMs). This likely reflects the fact that such models receive little to no morphological cues in their input and so rely primarily on contextual cues and linguistically unbiased subword information (n-grams or characters) to derive word-vectors. By contrast, FlauBERT and CamemBERT appear to encode

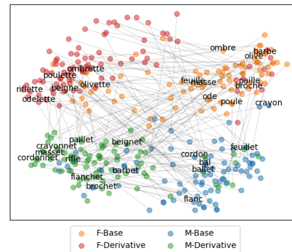
26 The optimal dimensionality was determined by (i) for each possible reduced dimensionality d, computing the Pearson correlation between the pairwise CosSims between word-vectors in the reduced d-dimensional space, and those computed in the original space; (ii) defining as optimal the smallest value of d such that the difference between the correlations computed for $d-1$ and for d is below .05% (inflection point).

27 Permutation tests performed using cosine-Silhouette (a clustering quality score; Rousseeuw, 1987) as dependent measure showed that all embeddings were associated with a significantly higher Silhouette than if their vectors were randomly reassigned labels, (1000 such random reassignments were evaluated against the gold-standard one). This was true in dimension 2, and also in the higher reduced dimensions used for testing.

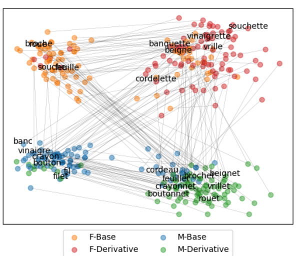
suffixation along the x-axis and gender along the y-axis. This indicates that such LMs may benefit from tokenization, potentially splitting derivatives into their base and the -*et(te)* suffix (point further discussed in Section 5).



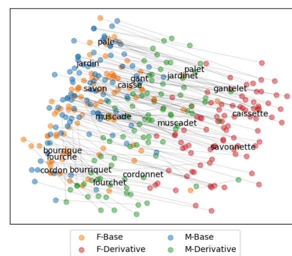
(a) fastText; CosSim correlation 48.5%.



(b) FlauBERT; CosSim correlation 81.3%.



(c) CamemBERT; CosSim correlation 74.2%.



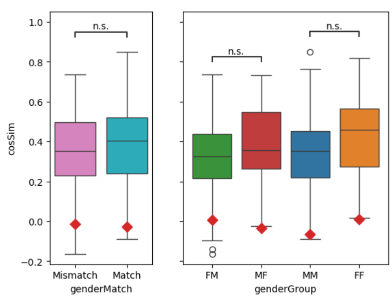
(d) Canine-c; CosSim correlation 55.9%.

Fig. 8: 2D reductions of base and derivative word-vectors. Lines between pairs of word-vectors associate bases with their derivative, and also represent 2D affixal vectors.

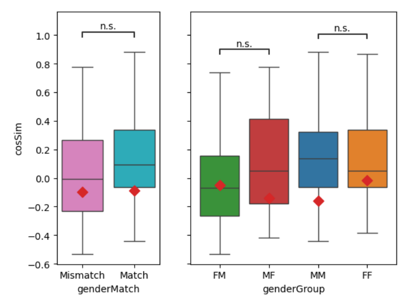
In the CamemBERT plot especially, vectors mapping bases to derivatives (and representing -*et(te)* affixation) form parallel "clusters" themselves, whose directions and positions depend on the genders of the base and derivative. $\overrightarrow{ette}^{\mathrm{FF}}$ and $\overrightarrow{et}^{\mathrm{MM}}$ are both horizontal, right-pointing vectors, while $\overrightarrow{ette}^{\mathrm{MF}}$ and $\overrightarrow{et}^{\mathrm{FM}}$ are respectively upward and downward-pointing diagonal vectors. This confirms the intuition that affixal vectors encode gender-(mis)match as part of their distributional semantics (for CamemBERT, within their y-component). This is in fact a direct consequence of the distribution of masculine and feminine forms in the vector space. This pattern will surface again in the next section.

Following prediction (13), the four gender groups were tested for "locality" differences as measured by base-derivative CosSim. Testing was conducted in the reduced, but still high-dimensional spaces (third column of Table 2). CosSim distributions are summarized as box plots in Figure 9, for each LM. Each plot is divided into a left section showing the distributions of CosSims for gender-matching

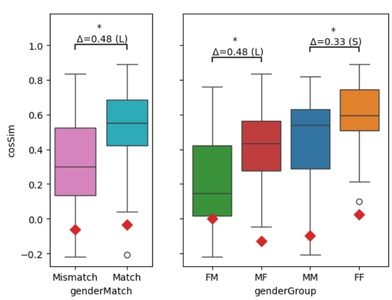
(FF+MM) vs. mismatching (MF+FM) groups, and a right section detailing these distributions for each individual group. Two general observations surface from these plots. First, although base-derivative CosSims are not always very high, they remain visibly higher than group baselines for all LMs except FlauBERT. Second, there are in principle 24 possible orderings of the four groups in terms of their median base-derivative CosSim. Disregarding significance levels, the right sections of the plots indicate that CamemBERT exhibits the expected ordering, while fastText and FlauBERT feature one misplaced groups. Canine-c appears to be the only model not supporting the expected trend. This is likely due to the fact that this LMs is primarily character-sensitive; given that forms in *-et* display more character overlap with their base than forms in *-ette*, the FM and MM groups get assigned comparatively higher measures of CosSim than the MF and FF groups.



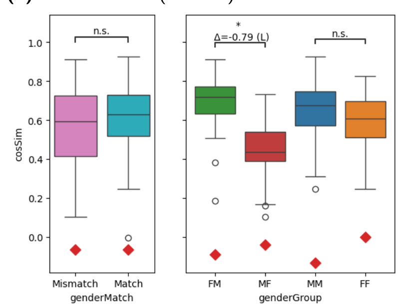
(a) fastText ($d = 74$)



(b) FlauBERT ($d = 32$)



(c) CamemBERT ($d = 56$)



(d) Canine-c ($d = 31$)

Fig. 9: Distributions of base-derivative CoSims across groups and LMs. Red diamond markers indicate CosSim baselines, computed for each group by averaging CosSims between all bases and all derivatives from that group (regardless of pairing).

Turning to the significance of the observed trend, results are more mixed. For each model, differences between groups were tested using two-sided Mann-Whitney

U-tests, corrected for multiple comparisons *via* the Benjamini-Yekutieli procedure (Benjamini and Yekutieli, 2001). Effect sizes correspond to Cliff's Δ. The effect of gender-mismatch (prediction (13a)) was only significant in CamemBERT (large effect); see left sections of the plots. This LM also captured the effect of suffix gender (prediction (13b)) with large or small effect sizes; see right sections of the plots. Canine-c displays a large effect in the *unexpected* direction in the mismatching case, treating F-to-M forms as more similar to their base than M-to-F forms. This can be explained by the character-sensitivity of the model, and constitutes the only piece of evidence going directly against prediction (13b) across LMs.

In summary, one of the four LMs, CamemBERT, exhibits a base-derivative CosSim pattern aligning with the prediction that base-derivative pairs resulting from more productive and purely diminutive -*et(te)*-suffixation should exhibit more semantic closeness. Two other LMs follow a similar trend, while the last LMs, Canine-c, partially contradicts it. We now turn to the second prediction, pertaining to the stability of -*et(te)*-suffixation across gender groups.

4.3 Assessing the stability of affixal vectors

For each LM and each base-derivative pair in our dataset, an affixal vector was computed following equation (10). Initial computations were performed in the raw spaces, subsequently reduced by fitting a CosKPCA to the set of affixal vectors. Table 3 summarizes the characteristics of the reduced spaces, similarly to Table 2.

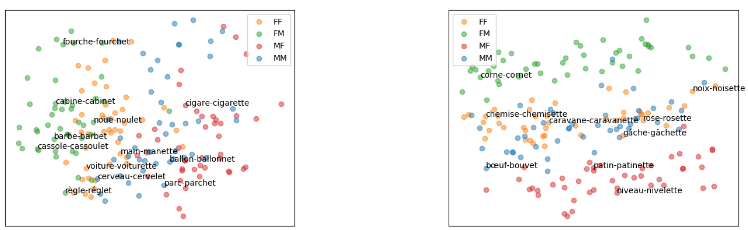
Model	Initial Dimension	Reduced Dimension	Pearson correlation between the pairwise CosSims of raw vs. reduced affixal vectors
fastText	300	80	97.8%
FlauBERT	1024	40	98.4%
CamemBERT	1024	41	98.6%
Canine-c	768	38	98.8%

Tab. 3: Characteristics of the reduced suffix embeddings.

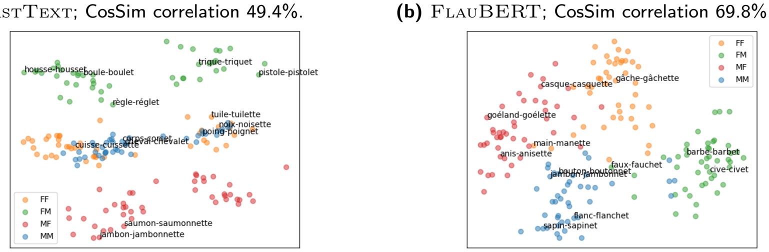
Figure 10 shows 2D CosKPCAs of the affixal vectors from our dataset. Vectors associated with the same gender group cluster together, in 2D as well as in higher dimensions.[28] For all models, gender-mismatching affixal vectors appear

28 Footnote 27 details how this was tested quantitatively.

to “sandwich” gender-matching ones roughly along one dimension: the x-axis for fastText and Canine-c, the y-axis for FlauBERT and CamemBERT. This dimension thus captures gender interactions between base and derivative. In the case of CamemBERT, this clustering is visibly consistent with the direction of the “lines” in Figure 8c. In that figure, gender-matching operations were associated with horizontal lines, while gender-mismatching operations were associated with upward- (MF) or downward- (FM) pointing diagonal lines. This three-way distinction is mirrored in Figure 10c, transforming lines into points in the space. More broadly, affixal vectors capturing shifts in syntactic environment (here, in relation to gender) is in line with the DH. Also in line with the DH is the observation that affixal vectors in fastText, FlauBERT and CamemBERT do not capture so well *absolute* properties of their environment, e.g. the fact that $\overrightarrow{ette}^{\mathrm{FF}}$ is a gender-matching *feminine* operation, while $\overrightarrow{et}^{\mathrm{MM}}$ is its *masculine* counterpart.



(a) fastText; CosSim correlation 49.4%. **(b)** FlauBERT; CosSim correlation 69.8%.

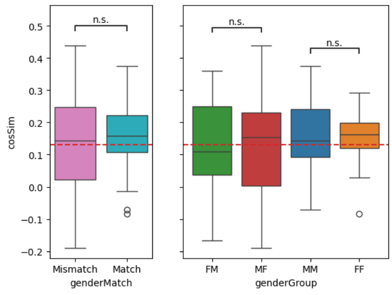


(c) CamemBERT; CosSim correlation 73.7%. **(d)** Canine-c; CosSim correlation 70.8%.

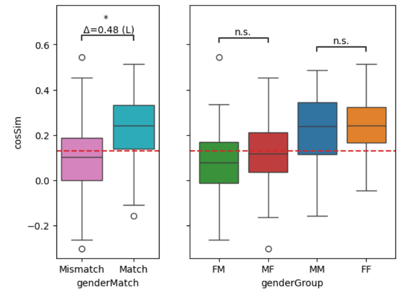
Fig. 10: 2D reductions of affixal vectors corresponding to *-et(te)*-suffixation between various base-derivative pairs. Affixal vectors are colored according to the gender of the base and that of the derivative they are computed from.

For each LM, the dimension that does *not* differentiate between gender-matching and mismatching clusters in 2D (respectively y, x, x and y in Figure 10) corresponds to a mix of other factors, and is also where most of the within-group variation occurs. Canine-c constitutes a bit of an outlier in that respect, as it leverages this dimension (y-axis) to differentiate between FF and MM operations.[29] With FLAUBERT and CAMEMBERT, this dimension happens to correlate with the number of tokens contained in bases and derivatives used to compute each affixal vector.[30] Across LMs, the four 2D clusters of affixal vectors do not seem to vastly differ in terms of spread, casting doubt on prediction (14). We proceed to test for differences in that respect in higher dimensions.

Testing prediction (14) in the higher-dimensional spaces described in the third column of Table 3 yields very mixed results, summarized in Figure 11. The box plots in that figure should be interpreted similarly to those in Figure 9, except the dependent measure is now the CosSim between each affixal vector and the centroid of its gender group, in accordance with prediction (14). Baselines correspond to the mean affix-centroid CosSim that would be measured for some vector sample disregarding the existence of gender groups. For all models but CAMEMBERT (and arguably FLAUBERT), measures of affix-centroid CosSims distribute around the baselines, suggesting an absence of spread differences between gender groups. Still, the effect of gender-mismatch (prediction (14a)) turned out significant for FLAUBERT and CAMEMBERT (large effects); see left sections of the plots. The effect of suffix gender was not captured by any of the LMs at stake; see right sections of the plots. Prediction (14b) is therefore not supported.



(a) FASTTEXT ($d = 80$)

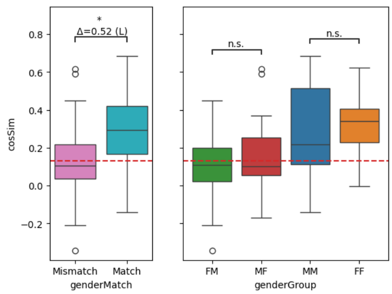


(b) FLAUBERT ($d = 40$)

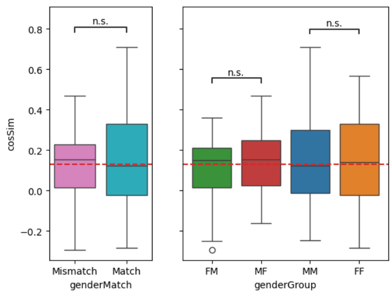
---

29 This can be explained by the fact this LM is primarily character-sensitive and so more likely to encode differences between *-et* and *-ette*, even when both are gender-matching, just because these two suffixes involve different characters.

30 This comes from the fact that in these BERT-like embeddings, the word-vector of a multi-token word was computed as the mean of the vectors of its constitutive tokens. The influence of the number of tokens on vector representations, which can be understood as an imperfect measure of morphological complexity, is difficult to control for without potentially affecting other, more contentful dimensions of meaning.



(c) CAMEMBERT ($d = 41$)



(d) CANINE-C ($d = 38$)

Fig. 11: Distributions of affix-centroid CoSims across groups and LMs. Red dashed lines are baselines computed by averaging CosSims between 44 randomly sampled affixal vectors and their centroid (a grand average was then computed over 1000 sampling iterations).

In summary, the results obtained for the BERT-like LMs suggest that gender-matching and mismatching operations may stem from morphological processes characterized by different degrees of semantic transparency. The stability differences observed in these LMs go in the direction predicted by morphological theory and individual empirical datapoints. No LM exhibits clear trends in the opposite direction. But the existence of a difference in semantic transparency *within* the matching and mismatching groups remains quite unclear. Considering Hénot-Mortier (2024)'s account, this support the DIM vs. SIM distinction, but not the word- vs. root-level distinction within SIM, that was expected to tease apart MF from FM and FF from MM groups.

5 Discussion and concluding remarks

Word embeddings extracted from LMs capture core morphological regularities, consistent with the past literature on similar phenomena. For instance, they distinguish between bases and their derivatives, and differentiate forms according to their grammatical gender. Additionally, morphological operations, seen as geometric translations (vectors) changing a base into its derivative, were shown to encode gender-change, in that gender-matching and mismatching operations were assigned different representations. Differences between affix representations could be visually related to the divisions established by the LMs at the word level, between bases and derivatives of various genders. This was especially clear with the BERT-like LMs, which come with a tokenization procedure breaking down words into subwords, for instance (and ideally) *boulette* into *boul+ette*. The fact such LMs clearly distinguished between bases and derivatives and between different variants of -*et(te)* is thus relatively unsurprising, given that, by design, these LMs *can* assign a diminutized form a representation corresponding to the average of the contextualized representations of its base and suffix (see (15a) for a simplified illustration). This in turn predicts corresponding affixal vectors computed by subtraction (as per (10)) to potentially include pre-isolated affixal information (see the $\overrightarrow{ette}_{c=\mathrm{boul}}$ term in (15b)). Still, the fact that BERT-like LMs assign coherent, fine-grained representations to the relevant tokens remains remarkable.

(15) a. The $_{c=}$ notation signals a representation is sensitive to context c.
$\overrightarrow{boulette} = \overrightarrow{boul + ette} = \frac{1}{2}(\overrightarrow{boul}_{c=\mathrm{ette}} + \overrightarrow{ette}_{c=\mathrm{boul}})$

b. $\overrightarrow{ette} \overset{(10)}{=} \overrightarrow{boul + ette} - \overrightarrow{boule} = \frac{1}{2}(\overrightarrow{boul}_{c=\mathrm{ette}} + \overrightarrow{ette}_{c=\mathrm{boul}}) - \overrightarrow{boule}_{c=\emptyset}$

By contrast, LMs based solely on character decomposition (Canine-c) or arbitrary n-grams (fastText) do not come with the same preprocessing advantages. This may explain why these LMs can appear too sensitive to surface patterns (e.g. the orthographic difference between -*et* and -*ette*), and seem comparatively less influenced by deeper distributional differences – ultimately leading to a failure to capture the target predictions. The conceptual question that arises from these findings and the observed contrasts across LMs, is whether the distinction between gender-matching and mismatching operations is entirely fueled by featural/syntactic differences characterizing the environment in which these operations occur, or, instead, is driven by differences that remain more semantic in nature. This question is a very hard one to address, given that embeddings, (and beyond them, the DH), do not essentially distinguish changes in context driven by pure meaning (i.e. linked to the external world, and extra-linguistic facts), from purely syntactic shifts.

Beyond core morphological regularities, we also attempted to test higher-order properties of -*et(te)*-suffixation, in particular, differences in base-derivative closeness, and in the stability of suffixation operations, that we took to be indicative of morphosemantic transparency. In that domain, our predictions were less well-supported, in particular when we turned to measures of stability. There are a few factors that might explain these results and may be better addressed in future work (with the help of bigger datasets).

First, we suggested that the representation of affixal vectors is influenced by whether, and how, LMs tokenize the corresponding bases and derivatives. Focusing on datapoints recruiting a stable number of tokens (say one for the base and two for the derivative), or only retaining pairs characterized by morphologically accurate tokenizations, may successfully filter noisy representations and neutralize interfering factors. However, overly curating the dataset may also threaten the validity of our initial predictions, in that the number and nature of the tokens a word (especially, a derivative) involves, may in itself be indicative of the word's degree of transparency. For instance, if many derivatives of the FF-group were tokenized into their base and the suffix *-ette*, the resulting affixal vectors may be expected to be fairly stable, as they would all incorporate a shared vector for *-ette* – *modulo* arguably small contextual variations (see (15b)). Further analyzing the tokenizations assigned to bases and derivatives across gender groups may thus constitute an indirect measure of the transparency of each group. This line of inquiry is sketched in Tables 4 and 5, which compile the proportions of derivatives whose FlauBERT and CamemBERT tokenizations include or nearly include a *-et(te)* token.

Gender group	% of derivatives with exact -et(te) token		% of derivatives with near-exact -et(te) token	
	Full dataset	Working dataset	Full dataset	Working dataset
FF	51.0	63.6	68.6	75.0
MM	29.8	29.5	33.0	34.1
MF	45.5	45.5	65.9	65.9
FM	38.3	38.6	48.9	50.0

Tab. 4: Proportions of (near)-correctly tokenized *-et(te)* suffixes across gender groups (FlauBERT tokenizer). The full dataset refers to Hénot-Mortier's (2024) imbalanced dataset, while the working dataset is the one used to conduct our analyses (44 datapoints per group). Near-exact *-et(te)* tokens are defined as the *et(te)* substring, preceded by at most one additional character.

Gender group	% of derivatives with exact -et(te) token		% of derivatives with near-exact -et(te) token	
	Full dataset	Working dataset	Full dataset	Working dataset
FF	18.6	4.5	35.2	20.5
MM	13.8	20.5	58.5	59.1
MF	22.7	22.7	59.1	59.1
FM	10.7	11.4	25.5	27.3

Tab. 5: Proportions of (near)-correctly tokenized *-et(te)* suffixes across gender groups (CamemBERT tokenizer).

These tables together show that FLAUBERT isolates *-ette* more reliably than *-et*, and clearly outperforms CAMEMBERT at these tasks. Given that CAMEMBERT displayed crisper clusterings and eventually better matched our predictions than FLAUBERT, we can conclude that tokenization need not be fully morphologically accurate to help capture the expected generalizations. Still, tokenization *per se* appears very helpful, given that CAMEMBERT even more evidently outperformed the tokenizerless yet otherwise equally sophisticated CANINE-C LM.
Secondly, differing frequencies within and across base-derivative pairs may constitute a confounding factor. Gender-mismatching base-derivative pairs are in fact likely to be overall less frequent in corpora; which may cause them to be associated with noisier representations (and tokenizations), for reasons independent from morphology. This, in turn, may affect base-derivative closeness and affixal vector stability. Testing for a significant correlation between base/derivative frequencies and CosSims may help clarify the confounding potential of word frequencies.
Lastly, our study focused on relative predictions, comparing CosSims between groups, and abstracting away from the supposedly diminutive semantics assigned to *-et(te)*. We did so, because word embeddings, being based on the DH, are relational models which cannot really identify words or operations as "diminutive" *per se*. However, if forms known to be unequivocally diminutive could be shown to distribute in certain restricted subspaces, it may become possible to analyze the semantics of outliers and determine if word embeddings truly encode "diminutiveness", in addition to the general notion of transparency that we assessed here.

To conclude, we leveraged computational tools to show that some models of Distributional Semantics capture the empirical observation that gender-matching base-derivative pairs in *-et(te)* are more locally distributed than mismatching pairs, and that, within each group (matching/mismatching), pairs in *-ette* are more locally distributed than those in *-et*. In terms of semantic stability, *-et(te)*-suffixation was shown to be more stable across gender-matching pairs than gender-mismatching ones. Although their robustness should be further assessed, these findings are consistent with the empirical picture and support the hypothesis that *-et(te)* is ambiguous between a transparent, gender-matching diminutive modifier, and a head characterized by a looser, "similative" semantics. In terms of model design and selection, our results suggest that architectural complexity (found in BERT-like Encoders) *must* be combined with a biased[31] word decomposition process (even if morphologically inaccurate) to derive linguistically-informed representations.

31 I.e. selecting specific, variable-length substrings, as opposed to exhaustive n-grams or character-level decompositions

Acknowledgment: I thank the audiences and reviewers of the 4th International Symposium of Morphology, the 16th Conference on Syntax, Phonology and Language Analysis, and the QMUL Syntax-Semantics Research Group meeting of 18/02/26 for relevant questions, comments, datapoints and suggestions. Special thanks go to David Adger, Artemis Alexiadou, Olivier Bonami, and Linnaea Stockall.

References

Acquaviva, P. (2009). Roots and lexicality in Distributed Morphology. In *York Essex Morphology Meeting 2*, pages 1–21.

Arad, M. (2003). Locality constraints on the interpretations of roots. *Natural Language and Linguistic Theory*, 21(4):737–778.

Arad, M. (2005). *Roots and Patterns: Hebrew Morpho-syntax*. Springer, Dordrecht.

Aronoff, M. (1976). *Word Formation in Generative Grammar*. Linguistic Inquiry monographs. MIT press.

Bally, C. (1932). *Linguistique générale et linguistique française*.

Benbaji-Elhadad, I., Doron, O., and Hénot-Mortier, A. (2022). Word-embeddings distinguish denominal and root-derived verbs in semitic. In Moortgat, M. and Wijnholds, G., editors, Proceedings *End-to-End Compositional Models of Vector-Based Semantics,* NUI Galway, 15-16 August 2022, volume 366 of *Electronic Proceedings in Theoretical Computer Science*, pages 35–49. Open Publishing Association.

Benbaji-Elhadad, I., Doron, O., and Hénot-Mortier, A. (2024). Distinguishing levels of morphological derivation in word-embedding models. In *Proceedings of the 53rd Annual Meeting of the North East Linguistics Society*, Georg August University of Göttingen, Germany. GLSA.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4).

Bobaljik, J. D. (2012). *Universals in Comparative Morphology: Suppletion, Superlatives, and the Structure of Words*. The MIT Press.

Bonami, O. and Guzmán Naranjo, M. (2023). *Distributional evidence for derivational paradigms*, page 219–258. De Gruyter.

Bonami, O. and Paperno, D. (2018). Inflection vs. derivation in a distributional vector space. *Lingue e linguaggio*, (2):173–196.

Borer, H. (2005). *Structuring sense. Volume 1: In name only*. OUP, Oxford.

Borer, H. (2013). *Structuring sense. Volume 1: Taking form*. OUP, Oxford.

Brucale, L. and Mocciaro, E. (2023). Approximation through suffixation: -ḍḍu/-a in sicilian. *Zeitschrift für Wortbildung / Journal of Word Formation*, 7(1).

Clark, J. H., Garrette, D., Turc, I., and Wieting, J. (2021). CANINE: pre-training an efficient tokenization-free encoder for language representation. *CoRR*, abs/2103.06874.

Dal, G. (1997). *Grammaire du suffixe -et(te)*. PhD thesis, Paris.

De Belder, M., Faust, N., and Lampitelli, N. (2014). *On an inflectional and a derivational diminutive*, pages 149–163. Oxford University Press.

Deal, A. R. (2016). Plural exponence in the Nez Perce DP: a DM analysis. *Morphology*, 26(3-4):313–339.

Delhay, C. (1995). Le diminutif: la diminution sans comparaison? *Faits de langues*, 3(5):63–72.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Burstein, J., Doran, C., and Solorio, T., editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Fathi, R. and Lowenstamm, J. (2016). The gender assignment pattern of French nouns: Selection and allomorphy. *Morphology*, 26(3–4):477–509.

Ferrari, F. (2005). *A syntactic analysis of the nominal systems of Italian and Luganda: how nouns can be formed in the syntax*. PhD thesis, New York University.

Fradin, B. (2003). Le traitement de la suffixation en -et. *Langages*, 37(152):51–77.

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., and Mikolov, T. (2018). Learning word vectors for 157 languages. In Calzolari, N., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Hasida, K., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S., and Tokunaga, T., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Halle, M. and Marantz, A. (1993). Distributed Morphology and the Pieces of Inflection. In Hale, K. and Keyser, S. J., editors, *The View from Building 20*, pages 111–176, Cambridge, MA. MIT Press.

Halle, M. and Marantz, A. (1994). Some Key Features of Distributed Morphology. In *Papers on Phonology and Morphology*, MITWPL 21, pages 275–288. MIT Working Papers in Linguistics, Cambridge, MA.

Harley, H. and Noyer, R. (1998). Licensing in the non-lexicalist lexicon: Nominalizations, vocabulary items and the encyclopedia. In Harley, H., editor, *MIT Working Papers in Linguistics 32: Papers from the UPenn/MIT roundtable on argument structure and aspect*, pages 119–137. MITWPL, Cambridge, MA.

Harris, Z. S. (1954). Distributional structure. *WORD*, 10(2–3):146–162.

Hasselrot, B. (1957). Etudes sur la formation diminutive dans les langues romanes. In *Acta Universitatis Upsaliensis 1957 : 11*, Uppsala. Lundequistska Bokhandeln.

Hasselrot, B. (1972). Etude sur la vitalité de la formation diminutive française au xxe siècle. In *Acta Universitatis Upsaliensis, Studia Romanica Upsaliensia 8*, Uppsala. Almqvist & Wiksells.

Hénot-Mortier, A. (2022). Evidence for an encoding of morphological blocking effects within two English word embedding models. In *Architectures and Mechanisms of Language Processing 2022 (AMLaP 28)*.

Hénot-Mortier, A. (2024). Highs and lows of the french diminutive suffix -et(te). *Lingue e linguaggio*, (2):245–266.

Hénot-Mortier, A. (2025). French nominal diminutives in -et(te).

Jurafsky, D. (1996). Universal Tendencies in the Semantics of the Diminutive. *Language*, 72(3):533.

Kihm, A. (2005). *Noun class, gender and the lexicon-syntax-morphology interfaces: A comparative study of Niger-Congo and Romance languages*, pages 459–512. OUP, Oxford.

Kramer, R. (2015). *The Morphosyntax of Gender*. Oxford Studies in Theoretical Linguistics 58. Oxford University Press, Oxford.

Kramer, R. (2016). The location of gender features in the syntax. *Language and Linguistics Compass*, 10(11):661–677.

Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., Crabbé, B., Besacier, L., and Schwab, D. (2020). Flaubert: Unsupervised language model pre-training for french. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 2479–2490, Marseille, France. European Language Resources Association.

Lecarme, J. (2002). *Gender "polarity": Theoretical aspects of Somali nominal morphology*, pages 109–141. Cascadilla Press, Somerville, Massachusetts.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach.

Marantz, A. (1997). No escape from syntax. *University of Pennsylvania Working Papers in Linguistics*, 4:201–225.

Marantz, A. (2001). *Words*.

Marelli, M. and Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, 122(3):485–515.

Martin, L., Muller, B., Ortiz Suárez, P. J., Dupont, Y., Romary, L., de la Clergerie, É., Seddah, D., and Sagot, B. (2020). CamemBERT: a tasty French language model. In Jurafsky, D., Chai, J., Schluter, N., and Tetreault, J., editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219, Online. Association for Computational Linguistics.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In Burges, C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

Milner, J.-C. (1989). Genre et dimension dans les diminutifs français. *Linx*, 21(1):191–201.

Musil, T., Vidra, J., and Mareček, D. (2019). Derivational morphological relations in word embeddings.

Naranjo, M. G. and Bonami, O. (2023). A distributional assessment of rivalry in word formation. *Word Structure*, 16(1):87–114.

Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Schäfer, M. (2023). *Splitting -ly's: Using word embeddings to distinguish derivation and inflection*, page 259–300. De Gruyter.

Steriopolo, O. (2017). Syntactic variation in diminutive suffixes: Russian, kolyma yukaghir, and itelmen. *Languages*, 2(4).

Stosic, D. and Amiot, D. (2023). Can suffixal augmentation express approximation? the case of evaluative nouns. *Zeitschrift für Wortbildung / Journal of Word Formation*, 7(1).

Stump, G. T. (1993). How peculiar is evaluative morphology? *Journal of Linguistics*, 29(1):1–36.

Stump, G. T. (1998). *The Handbook of Morphology*, chapter Inflection, pages 13–43. Blackwell, London.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Weber, M. (1963). *Contributions à l'étude du diminutif en français moderne : essai de systématisation*. PhD thesis.

Wiltschko, M. and Steriopolo, O. (2007). Parameters of variation in the syntax of diminutives. In Radisic, M., editor, *Proceedings of the 2007 Canadian Linguistics Association Annual Conference*.

Xiao, H. (2018). bert-as-service. https://github.com/hanxiao/bert-as-service.