"Small" vectors in big spaces, and the semantics of French-*et(te)*-suffixation[1]

Adèle Hénot-Mortier (Queen Mary University of London)

June 3, 2026

ZAS Berlin

[1] I thank the audiences and reviewers of ISMo 4, the SinFonIJA 16, and the 18/02/26 QMUL Syntax-Semantics Research Group meeting for relevant questions, comments, datapoints and suggestions.

Introduction

French diminutives in *-et(te)*

- Diminutive and augmentative affixes generally express a difference of **size** and/or an **affective relation** towards the base word.[2]
- The French "diminutive" suffix *-et(te)* for instance, appears to express smallness, cuteness or endearment in many instances.

(1) a. maison → maisonn-ette
'house' → 'small (cute) house'

b. balcon → balconn-et
'balcony' → 'small (cute) balcony'

[2] Stump 1993, Jurafsky 1996 i.a.

French diminutives in *-et(te)*

- Diminutive and augmentative affixes generally express a difference of **size** and/or an **affective relation** towards the base word.[2]
- The French “diminutive” suffix *-et(te)* for instance, appears to express smallness, cuteness or endearment in many instances.

(1) a. maison → maisonn-ette
‘house’ → ‘small (cute) house’

b. balcon → balconn-et
‘balcony’ → ‘small (cute) balcony’

[2] Stump 1993, Jurafsky 1996 i.a.

Empirical challenges posed by *-et(te)*

- Still, the **semantics** of *-et(te)* is notoriously hard to capture in a uniform fashion.[3]
- From a **morphosyntactic** viewpoint, *-et(te)* also exhibits varying properties regarding:
  - the preservation of gender features;
  - productivity;
  - range and domain;
  - and interaction with other morphology.[4]
- Let's briefly review the data.

[3] Hasselrot, 1957; Weber, 1963; Hasselrot, 1972; Milner, 1989; Delhay, 1995; Fradin, 2003.
[4] Criteria inspired by Stump (1998)

Empirical challenges posed by *-et(te)*

- Still, the **semantics** of *-et(te)* is notoriously hard to capture in a uniform fashion.[3]
- From a **morphosyntactic** viewpoint, *-et(te)* also exhibits varying properties regarding:
  - the preservation of gender features;
  - productivity;
  - range and domain;
  - and interaction with other morphology.[4]
- Let's briefly review the data.

[3] Hasselrot, 1957; Weber, 1963; Hasselrot, 1972; Milner, 1989; Delhay, 1995; Fradin, 2003.
[4] Criteria inspired by Stump (1998)

Empirical challenges posed by *-et(te)*

- Still, the **semantics** of *-et(te)* is notoriously hard to capture in a uniform fashion.[3]
- From a **morphosyntactic** viewpoint, *-et(te)* also exhibits varying properties regarding:
  - the preservation of gender features;
  - productivity;
  - range and domain;
  - and interaction with other morphology.[4]
- Let's briefly review the data.

[3] Hasselrot, 1957; Weber, 1963; Hasselrot, 1972; Milner, 1989; Delhay, 1995; Fradin, 2003.
[4] Criteria inspired by Stump (1998)

Empirical challenges posed by *-et(te)*

- Still, the **semantics** of *-et(te)* is notoriously hard to capture in a uniform fashion.[3]
- From a **morphosyntactic** viewpoint, *-et(te)* also exhibits varying properties regarding:
  - the preservation of gender features;
  - productivity;
  - range and domain;
  - and interaction with other morphology.[4]
- Let's briefly review the data.

[3] Hasselrot, 1957; Weber, 1963; Hasselrot, 1972; Milner, 1989; Delhay, 1995; Fradin, 2003.
[4] Criteria inspired by Stump (1998)

Empirical challenges posed by *-et(te)*

- Still, the **semantics** of *-et(te)* is notoriously hard to capture in a uniform fashion.[3]
- From a **morphosyntactic** viewpoint, *-et(te)* also exhibits varying properties regarding:
    - the preservation of gender features;
    - productivity;
    - range and domain;
    - and interaction with other morphology.[4]
- Let's briefly review the data.

[3] Hasselrot, 1957; Weber, 1963; Hasselrot, 1972; Milner, 1989; Delhay, 1995; Fradin, 2003.
[4] Criteria inspired by Stump (1998)

Empirical challenges posed by *-et(te)*

- Still, the **semantics** of *-et(te)* is notoriously hard to capture in a uniform fashion.[3]
- From a **morphosyntactic** viewpoint, *-et(te)* also exhibits varying properties regarding:
    - the preservation of gender features;
    - productivity;
    - range and domain;
    - and interaction with other morphology.[4]
- Let's briefly review the data.

---

[3] Hasselrot, 1957; Weber, 1963; Hasselrot, 1972; Milner, 1989; Delhay, 1995; Fradin, 2003.
[4] Criteria inspired by Stump (1998)

Empirical challenges posed by *-et(te)*

- Still, the semantics of *-et(te)* is notoriously hard to capture in a uniform fashion.[3]
- From a morphosyntactic viewpoint, *-et(te)* also exhibits varying properties regarding:
  - the preservation of gender features;
  - productivity;
  - range and domain;
  - and interaction with other morphology.[4]
- Let's briefly review the data.

[3] Hasselrot, 1957; Weber, 1963; Hasselrot, 1972; Milner, 1989; Delhay, 1995; Fradin, 2003.
[4] Criteria inspired by Stump (1998)

Preservation of gender features: matching case

- French assigns **grammatical gender** (**F**eminine or **M**asculine) to all nominals.
- As shown in (1), feminine ("F") bases generally yield feminine diminutives, while masculine ("M") bases generally yield masculine diminutives.[5]

(1) a. $\text{maison}_{\mathbf{F}} \overset{\mathbf{FF}}{\rightarrow} \text{maisonn-ette}_{\mathbf{F}}$
'house' $\rightarrow$ 'small (cute) house'

b. $\text{balcon}_{\mathbf{M}} \overset{\mathbf{MM}}{\rightarrow} \text{balconn-et}_{\mathbf{M}}$
'balcony' $\rightarrow$ 'small (cute) balcony'

[5] Bally, 1932.

Preservation of gender features: matching case

- French assigns **grammatical gender** (Feminine or Masculine) to all nominals.
- As shown in (1), feminine ("F") bases generally yield feminine diminutives, while masculine ("M") bases generally yield masculine diminutives.[5]

(1) a. $\text{maison}_{\mathbf{F}} \overset{\mathbf{FF}}{\rightarrow} \text{maisonn-ette}_{\mathbf{F}}$
‘house’ → ‘small (cute) house’

b. $\text{balcon}_{\mathbf{M}} \overset{\mathbf{MM}}{\rightarrow} \text{balconn-et}_{\mathbf{M}}$
‘balcony’ → ‘small (cute) balcony’

[5] Bally, 1932.

Preservation of gender features: mismatching case

- But **gender mismatches** occur: *-ette* may attach to masculine bases (2) and *-et* to feminine bases (3a)[6].

(2) $\text{char}_{\mathbf{M}} \overset{\mathbf{MF}}{\rightarrow} \text{char-ette}_{\mathbf{F}}$
$\text{char}_{\mathbf{M}} \;*\!\overset{\mathbf{MM}}{\rightarrow} \text{char-et}_{\mathbf{M}}$
'chariot' → 'cart'

(3) a. $\text{boule}_{\mathbf{F}} \overset{\mathbf{FM}}{\rightarrow} \text{boul-et}_{\mathbf{M}}$
'ball' → 'cannonball'/'ball'(and chain)

b. $\text{boule}_{\mathbf{F}} \overset{\mathbf{FF}}{\rightarrow} \text{boul-ette}_{\mathbf{F}}$
'ball' → 'small ball'

- Milner (1989) observed that mismatches often lead to a **looser semantic relationship between base and derivative**–usually a relation of mere similarity (shape, function).[7]
- Semantic differences between gender-matching and mismatching forms will be the main focus of the talk.

[6] Milner, 1989.
[7] Supported by a more systematic and quantitative analysis of the French lexicon–see Appendix.

Preservation of gender features: mismatching case

- But **gender mismatches** occur: *-ette* may attach to masculine bases (2) and *-et* to feminine bases (3a)[6].

(2)
$$\begin{array}{lcl} \text{char}_{\mathbf{M}} & \overset{\mathbf{MF}}{\rightarrow} & \text{char-ette}_{\mathbf{F}} \\ \text{char}_{\mathbf{M}} & *\overset{\mathbf{MM}}{\rightarrow} & \text{char-et}_{\mathbf{M}} \\ \text{`chariot'} & \rightarrow & \text{`cart'} \end{array}$$

(3) a.
$$\begin{array}{lcl} \text{boule}_{\mathbf{F}} & \overset{\mathbf{FM}}{\rightarrow} & \text{boul-et}_{\mathbf{M}} \\ \text{`ball'} & \rightarrow & \text{`cannonball'/`ball'(and chain)} \end{array}$$

b.
$$\begin{array}{lcl} \text{boule}_{\mathbf{F}} & \overset{\mathbf{FF}}{\rightarrow} & \text{boul-ette}_{\mathbf{F}} \\ \text{`ball'} & \rightarrow & \text{`small ball'} \end{array}$$

- Milner (1989) observed that mismatches often lead to a **looser semantic relationship between base and derivative**–usually a relation of mere similarity (shape, function).[7]
- Semantic differences between gender-matching and mismatching forms will be the main focus of the talk.

[6] Milner, 1989.
[7] Supported by a more systematic and quantitative analysis of the French lexicon–see Appendix.

Preservation of gender features: mismatching case

- But **gender mismatches** occur: *-ette* may attach to masculine bases (2) and *-et* to feminine bases (3a)[6].

(2) $\text{char}_{\mathbf{M}} \overset{\mathbf{MF}}{\rightarrow} \text{char-ette}_{\mathbf{F}}$
$\text{char}_{\mathbf{M}} \; *\overset{\mathbf{MM}}{\rightarrow} \text{char-et}_{\mathbf{M}}$
'chariot' → 'cart'

(3) a. $\text{boule}_{\mathbf{F}} \overset{\mathbf{FM}}{\rightarrow} \text{boul-et}_{\mathbf{M}}$
'ball' → 'cannonball'/'ball'(and chain)

b. $\text{boule}_{\mathbf{F}} \overset{\mathbf{FF}}{\rightarrow} \text{boul-ette}_{\mathbf{F}}$
'ball' → 'small ball'

- Milner (1989) observed that mismatches often lead to a **looser semantic relationship between base and derivative**–usually a relation of mere similarity (shape, function).[7]
- Semantic differences between gender-matching and mismatching forms will be the main focus of the talk.

[6] Milner, 1989.
[7] Supported by a more systematic and quantitative analysis of the French lexicon–see Appendix.

Productivity

- The native lexicon appears to contain twice as many gender-matching -*ette* forms than gender-matching -*et* forms,[8] meaning **-*ette* is more productive than -*et*** in the matching case.
- In the mismatching case, **-*et* and -*ette* are equally prevalent**.
- Diminutized English loans in (4) are gender-matching, and more likely to be well-formed with -*ette*.

(4) a. start-up$_{\mathsf{F}}$ $\xrightarrow{\mathsf{FF}}$ start-up-ette$_{\mathsf{F}}$
b. deadline$_{\mathsf{F}}$ $\xrightarrow{\mathsf{FF}}$ deadline-ette$_{\mathsf{F}}$
c. punchline$_{\mathsf{F}}$ $\xrightarrow{\mathsf{FF}}$ punchlin-ette$_{\mathsf{F}}$
d. brunch$_{\mathsf{M}}$?$\xrightarrow{\mathsf{MM}}$ brunch-et$_{\mathsf{M}}$
e. workshop$_{\mathsf{M}}$?$\xrightarrow{\mathsf{MM}}$ workshop-et$_{\mathsf{M}}$

- This all suggests **productivity depends on the interaction between the gender of the suffix/derivative and that of the base**.

[8] Hénot-Mortier, 2024.

Productivity

- The native lexicon appears to contain twice as many gender-matching *-ette* forms than gender-matching *-et* forms,[8] meaning ***-ette* is more productive than *-et*** in the matching case.
- In the mismatching case, ***-et* and *-ette* are equally prevalent**.
- Diminutized English loans in (4) are gender-matching, and more likely to be well-formed with *-ette*.

(4)
a. start-up$_{\text{F}}$ $\xrightarrow{\text{FF}}$ start-up-ette$_{\text{F}}$
b. deadline$_{\text{F}}$ $\xrightarrow{\text{FF}}$ deadline-ette$_{\text{F}}$
c. punchline$_{\text{F}}$ $\xrightarrow{\text{FF}}$ punchlin-ette$_{\text{F}}$
d. brunch$_{\text{M}}$?$\xrightarrow{\text{MM}}$ brunch-et$_{\text{M}}$
e. workshop$_{\text{M}}$?$\xrightarrow{\text{MM}}$ workshop-et$_{\text{M}}$

- This all suggests **productivity depends on the interaction between the gender of the suffix/derivative and that of the base.**

[8] Hénot-Mortier, 2024.

Productivity

- The native lexicon appears to contain twice as many gender-matching -*ette* forms than gender-matching -*et* forms,[8] meaning **-*ette* is more productive than -*et*** in the matching case.
- In the mismatching case, **-*et* and -*ette* are equally prevalent**.
- Diminutized English loans in (4) are gender-matching, and more likely to be well-formed with -*ette*.

(4)
a. start-up$_{\mathsf{F}}$ $\xrightarrow{\mathsf{FF}}$ start-up-ette$_{\mathsf{F}}$
b. deadline$_{\mathsf{F}}$ $\xrightarrow{\mathsf{FF}}$ deadline-ette$_{\mathsf{F}}$
c. punchline$_{\mathsf{F}}$ $\xrightarrow{\mathsf{FF}}$ punchlin-ette$_{\mathsf{F}}$
d. brunch$_{\mathsf{M}}$?$\xrightarrow{\mathsf{MM}}$ brunch-et$_{\mathsf{M}}$
e. workshop$_{\mathsf{M}}$?$\xrightarrow{\mathsf{MM}}$ workshop-et$_{\mathsf{M}}$

- This all suggests **productivity depends on the interaction between the gender of the suffix/derivative and that of the base.**

[8] Hénot-Mortier, 2024.

Productivity

- The native lexicon appears to contain twice as many gender-matching -*ette* forms than gender-matching -*et* forms,[8] meaning **-*ette* is more productive than -*et*** in the matching case.
- In the mismatching case, **-*et* and -*ette* are equally prevalent**.
- Diminutized English loans in (4) are gender-matching, and more likely to be well-formed with -*ette*.

(4)
a. start-up$_{\mathsf{F}}$ $\xrightarrow{\mathsf{FF}}$ start-up-ette$_{\mathsf{F}}$
b. deadline$_{\mathsf{F}}$ $\xrightarrow{\mathsf{FF}}$ deadline-ette$_{\mathsf{F}}$
c. punchline$_{\mathsf{F}}$ $\xrightarrow{\mathsf{FF}}$ punchlin-ette$_{\mathsf{F}}$
d. brunch$_{\mathsf{M}}$?$\xrightarrow{\mathsf{MM}}$ brunch-et$_{\mathsf{M}}$
e. workshop$_{\mathsf{M}}$?$\xrightarrow{\mathsf{MM}}$ workshop-et$_{\mathsf{M}}$

- This all suggests **productivity depends on the interaction between the gender of the suffix/derivative and that of the base**.

[8] Hénot-Mortier, 2024.

- *-et(te)* applies to nominals, but also to adjectives and verbs[9].

(5) a. mignon → mignon-et(te)
'cute' → 'cutesy'

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

(6) a. balayer → balay-ette$_{\mathbf{F}}$
'sweep' → 'brush'

b. siffler → siffl-et$_{\mathbf{M}}$
'whistle' → 'whistle'

- With adjectives, gender is determined *via* nominal agreement.
- With verbs, derivatives are preferentially feminine, pointing again to a cross-categorial productivity difference between *-et* and *-ette*.

[9] Dal, 1997.

- *-et(te)* applies to nominals, but also to adjectives and verbs[9].

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

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

(6) a. balayer → balay-ette$_{\mathbf{F}}$
‘sweep’ → ‘brush’

b. siffler → siffl-et$_{\mathbf{M}}$
‘whistle’ → ‘whistle’

- With adjectives, gender is determined *via* nominal agreement.
- With verbs, derivatives are preferentially feminine, pointing again to a cross-categorial productivity difference between *-et* and *-ette*.

[9] Dal, 1997.

Range and domain

- *-et(te)* applies to nominals, but also to adjectives and verbs[9].

(5) a. mignon → mignon-et(te)
'cute' → 'cutesy'

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

(6) a. balayer → balay-ette$_{\mathbf{F}}$
'sweep' → 'brush'

b. siffler → siffl-et$_{\mathbf{M}}$
'whistle' → 'whistle'

- With adjectives, gender is determined *via* nominal agreement.
- With verbs, derivatives are preferentially feminine, pointing again to a cross-categorial productivity difference between *-et* and *-ette*.

[9] Dal, 1997.

Interaction with other morphology

- *-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.[10]

(7) révolu-tion$_{\mathbf{F}}$ $\overset{\mathbf{FF}}{\rightarrow}$ révolu-tion-ette$_{\mathbf{F}}$
‘revolution’ $\rightarrow$ ‘small/insignificant revolution’

- *-et*/*-ette* **cannot be stacked** 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)*).[11]

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

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

Interaction with other morphology

- *-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.[10]

(7) révolu-tion$_{\mathbf{F}}$ $\xrightarrow{\mathbf{FF}}$ révolu-tion-ette$_{\mathbf{F}}$
‘revolution’ $\rightarrow$ ‘small/insignificant revolution’

- *-et*/*-ette* **cannot be stacked** 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)*).[11]

[10] However, this might be because many derived nominals are feminine, and diminutives are generally gender-matching.
[11] This differs from Italian; see De Belder et al. (2014) i.a.

Interim summary

- Although *-et(te)* is a suffix that can apply to different categories, it is **most productive in the nominal domain**.
- In this domain, productivity and semantic transparency are modulated by two factors:
    1. Whether *-et(te)* changes the gender of the base: **gender-matching forms are likely to result from a featurally and semantically transparent process**, while mismatching forms result from a featurally and semantically more opaque process.[12] Schematically: $\boxed{\text{FM, MF} < \text{MM, FF}}$.
    2. Whether the derivative is a F-form in *-ette* or a M-form in *-et*: ***-ette* forms are more likely to result from the transparent process than *-et* forms**. Schematically: $\boxed{\text{FM} < \text{MF}}$ and $\boxed{\text{MM} < \text{FF}}$.

[12] See Brucale and Mocciaro (2023) and Stosic and Amiot (2023) for discussions on approximation and diminution in morphology.

Interim summary

- Although *-et(te)* is a suffix that can apply to different categories, it is **most productive in the nominal domain**.
- In this domain, productivity and semantic transparency are modulated by two factors:
    1. Whether *-et(te)* changes the gender of the base: **gender-matching forms are likely to result from a featurally and semantically transparent process**, while mismatching forms result from a featurally and semantically more opaque process.[12] Schematically: $\boxed{\text{FM, MF < MM, FF}}$.
    2. Whether the derivative is a F-form in *-ette* or a M-form in *-et*: ***-ette* forms are more likely to result from the transparent process than *-et* forms**. Schematically: $\boxed{\text{FM < MF}}$ and $\boxed{\text{MM < FF}}$.

[12] See Brucale and Mocciaro (2023) and Stosic and Amiot (2023) for discussions on approximation and diminution in morphology.

Interim summary

- Although *-et(te)* is a suffix that can apply to different categories, it is **most productive in the nominal domain**.
- In this domain, productivity and semantic transparency are modulated by two factors:
    1. Whether *-et(te)* changes the gender of the base: **gender-matching forms are likely to result from a featurally and semantically transparent process**, while mismatching forms result from a featurally and semantically more opaque process.[12] Schematically: $\boxed{\text{FM, MF < MM, FF}}$.
    2. Whether the derivative is a F-form in *-ette* or a M-form in *-et*: ***-ette* forms are more likely to result from the transparent process than *-et* forms**. Schematically: $\boxed{\text{FM < MF}}$ and $\boxed{\text{MM < FF}}$.

[12] See Brucale and Mocciaro (2023) and Stosic and Amiot (2023) for discussions on approximation and diminution in morphology.

Interim summary

- Although *-et(te)* is a suffix that can apply to different categories, it is **most productive in the nominal domain**.
- In this domain, productivity and semantic transparency are modulated by two factors:
    1. Whether *-et(te)* changes the gender of the base: **gender-matching forms are likely to result from a featurally and semantically transparent process**, while mismatching forms result from a featurally and semantically more opaque process.[12] Schematically: $\boxed{\text{FM, MF < MM, FF}}$.
    2. Whether the derivative is a F-form in *-ette* or a M-form in *-et*: ***-ette* forms are more likely to result from the transparent process than *-et* forms**. Schematically: $\boxed{\text{FM < MF}}$ and $\boxed{\text{MM < FF}}$.

[12] See Brucale and Mocciaro (2023) and Stosic and Amiot (2023) for discussions on approximation and diminution in morphology.

What's next

- We will now focus on the **semantic effect of *-et(te)* on nominals, and how it gets modulated by gender**.
- We will investigate if distributional models of language capture the observations that gender-mismatching *-et(te)* is less likely to be transparent than gender-matching *-et(te)*, and that, within each group (matching/mismatching), forms in *-ette* appear more transparent than forms in *-et*.
- Before this, let's review a recent approach to *-et(te)* in the Distributed Morphology framework.

What's next

- We will now focus on the **semantic effect of *-et(te)* on nominals, and how it gets modulated by gender**.
- We will investigate if distributional models of language capture the observations that gender-mismatching *-et(te)* is less likely to be transparent than gender-matching *-et(te)*, and that, within each group (matching/mismatching), forms in *-ette* appear more transparent than forms in *-et*.
- Before this, let's review a recent approach to *-et(te)* in the Distributed Morphology framework.

What's next

- We will now focus on the **semantic effect of *-et(te)* on nominals, and how it gets modulated by gender**.
- We will investigate if distributional models of language capture the observations that gender-mismatching *-et(te)* is less likely to be transparent than gender-matching *-et(te)*, and that, within each group (matching/mismatching), forms in *-ette* appear more transparent than forms in *-et*.
- Before this, let's review a recent approach to *-et(te)* in the Distributed Morphology framework.

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

Distributed Morphology in a nutshell

- Distributed Morphology (now **DM**[13]) offers a framework to connect the morphosyntactic and semantic behavior of *-et(te)*.
- In DM, syntactic operations (MERGE, MOVE, AGREE), apply to bundles of syntactic features (*formatives*) which get mapped to phonological forms (*exponents*) post-syntactically.
- Roots (which are formatives) are **category-neutral** before merging with categorizing *heads*: *n* (nominalizer), *a* (adjectivizer), *v* (verbalizer) etc; see Fig. 1.

(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

[13] Halle and Marantz 1993, 1994

Distributed Morphology in a nutshell

- Distributed Morphology (now **DM**[13]) offers a framework to connect the morphosyntactic and semantic behavior of *-et(te)*.
- In DM, syntactic operations (MERGE, MOVE, AGREE), apply to bundles of syntactic features (*formatives*) which get mapped to phonological forms (*exponents*) post-syntactically.
- Roots (which are formatives) are **category-neutral** before merging with categorizing *heads*: *n* (nominalizer), *a* (adjectivizer), *v* (verbalizer) etc; see Fig. 1.

(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

[13] Halle and Marantz 1993, 1994

Distributed Morphology in a nutshell

- Distributed Morphology (now **DM**[13]) offers a framework to connect the morphosyntactic and semantic behavior of *-et(te)*.
- In DM, syntactic operations (MERGE, MOVE, AGREE), apply to bundles of syntactic features (*formatives*) which get mapped to phonological forms (*exponents*) post-syntactically.
- Roots (which are formatives) are **category-neutral** before merging with categorizing *heads*: *n* (nominalizer), *a* (adjectivizer), *v* (verbalizer) etc; see Fig. 1.

(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

[13] Halle and Marantz 1993, 1994

Distributed Morphology in a nutshell

- Distributed Morphology (now **DM**[13]) offers a framework to connect the morphosyntactic and semantic behavior of *-et(te)*.
- In DM, syntactic operations (MERGE, MOVE, AGREE), apply to bundles of syntactic features (*formatives*) which get mapped to phonological forms (*exponents*) post-syntactically.
- Roots (which are formatives) are **category-neutral** before merging with categorizing *heads*: *n* (nominalizer), *a* (adjectivizer), *v* (verbalizer) etc; see Fig. 1.

(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

[13] Halle and Marantz 1993, 1994

The key role of categorization in DM

- Subsequent work within the DM framework[14] points out the crucial distinction between **creating words from roots vs. from existing words** (i.e. categorized roots).
- Categorization is assumed to produce an 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 "sets" the semantic and morpho-phonological features of the newly created word.**

[14] Marantz 1997; Harley and Noyer 1998; Marantz 2001; Arad 2005, 2005 i.a.

The key role of categorization in DM

- Subsequent work within the DM framework[14] points out the crucial distinction between **creating words from roots vs. from existing words** (i.e. categorized roots).
- Categorization is assumed to produce an 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 "sets" the semantic and morpho-phonological features of the newly created word.**

[14] Marantz 1997; Harley and Noyer 1998; Marantz 2001; Arad 2005, 2005 i.a.

The key role of categorization in DM

- Subsequent work within the DM framework[14] points out the crucial distinction between **creating words from roots vs. from existing words** (i.e. categorized roots).
- Categorization is assumed to produce an 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 "sets" the semantic and morpho-phonological features of the newly created word.**

[14] Marantz 1997; Harley and Noyer 1998; Marantz 2001; Arad 2005, 2005 i.a.

Diminutives differ across 2 dimensions

- Based on cross-linguistic data, Wiltschko and Steriopolo (2007) argue that diminutive affixes vary cross two main dimensions:[15]
  - The "**level**" dimension (root vs. word) controls the affix' productivity, the range of syntactic categories the affix appears to apply to, and whether or not it can appear above derivational markers.
  - The "**type-of-merge**" dimension (head vs. modifier) controls whether or not an affix can modify the morphosyntactic features of its input (e.g. gender).
- Russian features diminutive suffixes varying across *both* parameters (Steriopolo, 2017).
- Italian can recruit the *same* surface suffix for either root- or word-level derivations (De Belder et al., 2014).

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

Semantic predictions

- Hénot-Mortier (2024), building on earlier work,[16] adds the following semantic characterizations:
  - **Root-level categorization should be less semantically transparent than word-level derivations (categorization/modification)**. Roughly, this is because root-level categorization arbitrarily sets a semantics based on an underspecified root meaning, while other operations modify meanings that have already been set in this way.
  - **Root-level modification should be more semantically transparent than root-level categorization**, roughly because modification proceeds by intersection while categorization amounts to a more arbitrary kind of precisification.

$$\text{Root-level categorization} < \text{Root-level modification}, \text{Word-level modification}, \text{Word-level categorization}$$

[16] Arad, 2003; De Belder et al., 2014.

Semantic predictions

- Hénot-Mortier (2024), building on earlier work,[16] adds the following semantic characterizations:
  - **Root-level categorization should be less semantically transparent than word-level derivations (categorization/modification)**. Roughly, this is because root-level categorization arbitrarily sets a semantics based on an underspecified root meaning, while other operations modify meanings that have already been set in this way.
  - **Root-level modification should be more semantically transparent than root-level categorization**, roughly because modification proceeds by intersection while categorization amounts to a more arbitrary kind of precisification.

Root-level categorization	<	Root-level modification	,	Word-level modification	,	Word-level categorization

[16] Arad, 2003; De Belder et al., 2014.

Semantic predictions

- Hénot-Mortier (2024), building on earlier work,[16] adds the following semantic characterizations:
  - **Root-level categorization should be less semantically transparent than word-level derivations (categorization/modification)**. Roughly, this is because root-level categorization arbitrarily sets a semantics based on an underspecified root meaning, while other operations modify meanings that have already been set in this way.
  - **Root-level modification should be more semantically transparent than root-level categorization**, roughly because modification proceeds by intersection while categorization amounts to a more arbitrary kind of precisification.

Root-level categorization < Root-level modification , Word-level modification , Word-level categorization

[16] Arad, 2003; De Belder et al., 2014.

Semantic predictions

- Hénot-Mortier (2024), building on earlier work,[16] adds the following semantic characterizations:
  - **Root-level categorization should be less semantically transparent than word-level derivations (categorization/modification)**. Roughly, this is because root-level categorization arbitrarily sets a semantics based on an underspecified root meaning, while other operations modify meanings that have already been set in this way.
  - **Root-level modification should be more semantically transparent than root-level categorization**, roughly because modification proceeds by intersection while categorization amounts to a more arbitrary kind of precisification.

Root-level categorization	<	Root-level modification	,	Word-level modification	,	Word-level categorization

[16] Arad, 2003; De Belder et al., 2014.

Morphosemantics of *-et(te)*

- Recall that in the nominal domain, *-et(te)* can be **gender-matching** (most cases) or **mismatching** (non-negligible minority of the cases). And the resulting meaning can be purely **diminutive** (most gender-matching cases), or rather "**similative**" (most gender-mismatching cases).
- Hénot-Mortier (2024) proposed that *-et(te)* may be the realization of two main kinds of suffixes:
  - a purely **diminutive, transparent root-level modifier** DIM;
  - and a **similative head SIM whose merger site depends on the gender features it carries**.
- Following implicit assumptions in Wiltschko and Steriopolo (2007), **gender features are assumed to be hosted in nominalizing heads**.[17] Crucially, we take this to include SIM.
- Let's quickly review the mechanics behind DIM and SIM.

[17] Lecarme, 2002; Ferrari, 2005; Kihm, 2005; Acquaviva, 2009; Kramer, 2015; Deal, 2016; Kramer, 2016.

Morphosemantics of *-et(te)*

- Recall that in the nominal domain, *-et(te)* can be **gender-matching** (most cases) or **mismatching** (non-negligible minority of the cases). And the resulting meaning can be purely **diminutive** (most gender-matching cases), or rather "**similative**" (most gender-mismatching cases).
- Hénot-Mortier (2024) proposed that *-et(te)* may be the realization of two main kinds of suffixes:
  - a purely **diminutive, transparent root-level modifier** DIM;
  - and a **similative head SIM whose merger site depends on the gender features it carries**.
- Following implicit assumptions in Wiltschko and Steriopolo (2007), **gender features are assumed to be hosted in nominalizing heads**.[17] Crucially, we take this to include SIM.
- Let's quickly review the mechanics behind DIM and SIM.

[17] Lecarme, 2002; Ferrari, 2005; Kihm, 2005; Acquaviva, 2009; Kramer, 2015; Deal, 2016; Kramer, 2016.

Morphosemantics of *-et(te)*

- Recall that in the nominal domain, *-et(te)* can be **gender-matching** (most cases) or **mismatching** (non-negligible minority of the cases). And the resulting meaning can be purely **diminutive** (most gender-matching cases), or rather "**similative**" (most gender-mismatching cases).
- Hénot-Mortier (2024) proposed that *-et(te)* may be the realization of two main kinds of suffixes:
  - a purely **diminutive, transparent root-level modifier** DIM;
  - and a **similative head SIM whose merger site depends on the gender features it carries**.
- Following implicit assumptions in Wiltschko and Steriopolo (2007), **gender features are assumed to be hosted in nominalizing heads**.[17] Crucially, we take this to include SIM.
- Let's quickly review the mechanics behind DIM and SIM.

[17] Lecarme, 2002; Ferrari, 2005; Kihm, 2005; Acquaviva, 2009; Kramer, 2015; Deal, 2016; Kramer, 2016.

Morphosemantics of *-et(te)*

- Recall that in the nominal domain, *-et(te)* can be **gender-matching** (most cases) or **mismatching** (non-negligible minority of the cases). And the resulting meaning can be purely **diminutive** (most gender-matching cases), or rather “**similative**” (most gender-mismatching cases).
- Hénot-Mortier (2024) proposed that *-et(te)* may be the realization of two main kinds of suffixes:
  - a purely **diminutive, transparent root-level modifier** DIM;
  - and a **similative head SIM whose merger site depends on the gender features it carries**.
- Following implicit assumptions in Wiltschko and Steriopolo (2007), **gender features are assumed to be hosted in nominalizing heads**.[17] Crucially, we take this to include SIM.
- Let’s quickly review the mechanics behind DIM and SIM.

[17] Lecarme, 2002; Ferrari, 2005; Kihm, 2005; Acquaviva, 2009; Kramer, 2015; Deal, 2016; Kramer, 2016.

Morphosemantics of *-et(te)*

- Recall that in the nominal domain, *-et(te)* can be **gender-matching** (most cases) or **mismatching** (non-negligible minority of the cases). And the resulting meaning can be purely **diminutive** (most gender-matching cases), or rather "**similative**" (most gender-mismatching cases).
- Hénot-Mortier (2024) proposed that *-et(te)* may be the realization of two main kinds of suffixes:
  - a purely **diminutive, transparent root-level modifier** DIM;
  - and a **similative head SIM whose merger site depends on the gender features it carries**.
- Following implicit assumptions in Wiltschko and Steriopolo (2007), **gender features are assumed to be hosted in nominalizing heads**.[17] Crucially, we take this to include SIM.
- Let's quickly review the mechanics behind DIM and SIM.

[17] Lecarme, 2002; Ferrari, 2005; Kihm, 2005; Acquaviva, 2009; Kramer, 2015; Deal, 2016; Kramer, 2016.

Morphosemantics of *-et(te)*

- Recall that in the nominal domain, *-et(te)* can be **gender-matching** (most cases) or **mismatching** (non-negligible minority of the cases). And the resulting meaning can be purely **diminutive** (most gender-matching cases), or rather “**similative**” (most gender-mismatching cases).
- Hénot-Mortier (2024) proposed that *-et(te)* may be the realization of two main kinds of suffixes:
  - a purely **diminutive, transparent root-level modifier** Dɪᴍ;
  - and a **similative head Sɪᴍ whose merger site depends on the gender features it carries**.
- Following implicit assumptions in Wiltschko and Steriopolo (2007), **gender features are assumed to be hosted in nominalizing heads**.[17] Crucially, we take this to include Sɪᴍ.
- Let’s quickly review the mechanics behind Dɪᴍ and Sɪᴍ.

[17] Lecarme, 2002; Ferrari, 2005; Kihm, 2005; Acquaviva, 2009; Kramer, 2015; Deal, 2016; Kramer, 2016.

- $\textsc{Dim}_m^{rl}$ is a purely **diminutive root-level modifier**, which is productive, as well as **morphosyntactically and semantically transparent**.
- $\textsc{Dim}_m^{rl}$ is expected to produce the vast majority of gender-matching, purely diminutive forms–derivation below.

Fig. 2: Deriving the gender-matching, diminutive form *boulette* ('small ball', (3b)).

Diminutive *-et(te)*

- $\textsc{Dim}^{rl}_{m}$ is a purely **diminutive root-level modifier**, which is productive, as well as **morphosyntactically and semantically transparent**.
- $\textsc{Dim}^{rl}_{m}$ is expected to produce the vast majority of gender-matching, purely diminutive forms–derivation below.

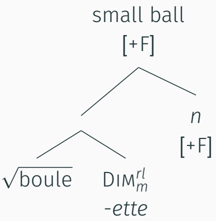


Fig. 2: Deriving the gender-matching, diminutive form *boulette* ('small ball', (3b)).

Similative *-et(te)*

- $\textsc{Sim}_h$ is a **gendered head**, assumed to have a nominalizing effect, and to convey a **similarity in shape or function** between base and derivative (with accidental diminutive traits).
- $\textsc{Sim}_h$ carries gender features that get **“imposed”**[18] **on the base**, likely leading to gender mismatches.
- The merger site of $\textsc{Sim}_h$ is gender-dependent: word-level if feminine ($\textsc{Sim}_h^{wl}$), and root-level, if masculine ($\textsc{Sim}_h^{rl}$).

(a) The MF non-diminutive form *charette* (2). (b) The FM non-diminutive form *boulet* (3a).

Fig. 3: Gender-mismatching derivations with $\textsc{Sim}_h$.

[18] This assumption is implicitly made in e.g. Wiltschko and Steriopolo (2007), to account for the properties of the German diminutive *-chen*.

- SIM$_h$ is a **gendered head**, assumed to have a nominalizing effect, and to convey a **similarity in shape or function** between base and derivative (with accidental diminutive traits).
- SIM$_h$ carries gender features that get **"imposed"**[18] **on the base**, likely leading to gender mismatches.
- The merger site of SIM$_h$ is gender-dependent: word-level if feminine (SIM$_h^{wl}$), and root-level, if masculine (SIM$_h^{rl}$).

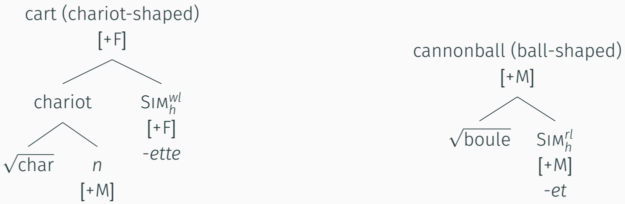
(a) The MF non-diminutive form *charette* (2). (b) The FM non-diminutive form *boulet* (3a).

Fig. 3: Gender-mismatching derivations with SIM$_h$.

[18] This assumption is implicitly made in e.g. Wiltschko and Steriopolo (2007), to account for the properties of the German diminutive *-chen*.

Similative *-et(te)*

- SIM$_h$ is a **gendered head**, assumed to have a nominalizing effect, and to convey a **similarity in shape or function** between base and derivative (with accidental diminutive traits).
- SIM$_h$ carries gender features that get **"imposed"**[18] **on the base**, likely leading to gender mismatches.
- The merger site of SIM$_h$ is gender-dependent: word-level if feminine (SIM$_h^{wl}$), and root-level, if masculine (SIM$_h^{rl}$).



(a) The MF non-diminutive form *charette* (2). **(b)** The FM non-diminutive form *boulet* (3a).

Fig. 3: Gender-mismatching derivations with SIM$_h$.

[18] This assumption is implicitly made in e.g. Wiltschko and Steriopolo (2007), to account for the properties of the German diminutive *-chen*.

Semantic predictions refined

- 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 $\textsc{Dim}^{rl}_{m}$.
- Within both gender-matching and mismatching groups, we also expect a **small difference between *-ette* (more transparent) and *-et* (less transparent)**. See below.

Base-derivative group	Possible derivation	Productivity	Semantic transparency
FF	$\textsc{Dim}^{rl}_{m}$	✓	Diminutive☺ on precise☺ base
	$\textsc{Sim}^{wl}_{h}$	✓	Similative☹ on precise☺ base
MM	$\textsc{Dim}^{rl}_{m}$	✓	Diminutive☺ on precise☺ base
	$\textsc{Sim}^{rl}_{h}$	✗	Similative☹ on vague☹ base
MF	$\textsc{Sim}^{wl}_{h}$	✓	Similative☹ on precise☺ base
FM	$\textsc{Sim}^{rl}_{h}$	✗	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.

Computational tools for morpho-semantics

The Distributional Hypothesis

- According to the Distributional Hypothesis (now DH[19]), **words occurring in similar syntactic environments should be semantically close.** Two caveats.
- The DH makes predictions about semantic closeness (a **relative measure**), but not about absolute meanings.
- 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".

[19] Harris, 1954.

- According to the Distributional Hypothesis (now DH[19]), **words occurring in similar syntactic environments should be semantically close.** Two caveats.
- The DH makes predictions about semantic closeness (a **relative measure**), but not about absolute meanings.
- 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".

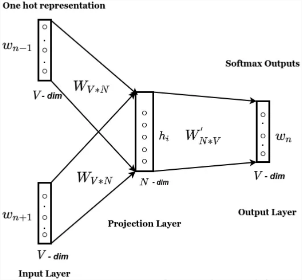
[19] Harris, 1954.

The Distributional Hypothesis

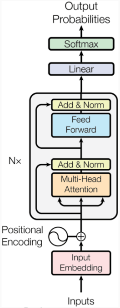
- According to the Distributional Hypothesis (now DH[19]), **words occurring in similar syntactic environments should be semantically close.** Two caveats.
- The DH makes predictions about semantic closeness (a **relative measure**), but not about absolute meanings.
- The DH is not a clear biconditional 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".

[19] Harris, 1954.

- Language Models (henceforth **LMs**), understood as architectures trained to efficiently predict words from other words can **act as models of the DH**.
- As a byproduct of their prediction objective, LMs produce efficient vector representations of words or subwords.



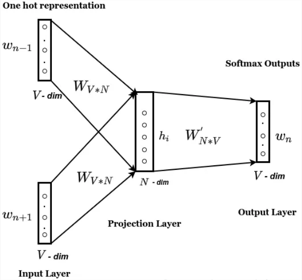
(a) The Continuous Bag Of Words architecture used to produce Word2Vec embeddings (Mikolov et al., 2013a, 2013b).



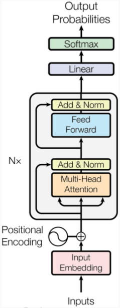
(b) Building block of the BERT Encoder architecture (Devlin et al., 2019), taken from Vaswani et al. (2017).

Fig. 4: Two LM architectures predicting (sub)words from context (sub)words.

- Language Models (henceforth **LMs**), understood as architectures trained to efficiently predict words from other words can **act as models of the DH**.
- As a byproduct of their prediction objective, LMs produce efficient vector representations of words or subwords.



(a) The Continuous Bag Of Words architecture used to produce Word2Vec embeddings (Mikolov et al., 2013a, 2013b).

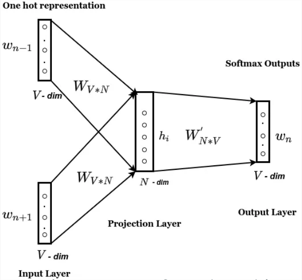


(b) Building block of the BERT Encoder architecture (Devlin et al., 2019), taken from Vaswani et al. (2017).

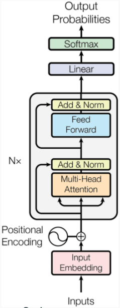
Fig. 4: Two LM architectures predicting (sub)words from context (sub)words.

Language Models and the DH

- Language Models (henceforth **LMs**), understood as architectures trained to efficiently predict words from other words can **act as models of the DH**.
- As a byproduct of their prediction objective, LMs produce efficient vector representations of words or subwords.



(a) The Continuous Bag Of Words architecture used to produce Word2Vec embeddings (Mikolov et al., 2013a, 2013b).



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

Fig. 4: Two LM architectures predicting (sub)words from context (sub)words.

Learning word vectors from LMs

- To predict a target word from its context, a LM is typically passed a list of words or subwords (the context), 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 contexts to suitable “filler” words, **the output vector obtained before the sampling step is expected to both summarize relevant aspects of the context 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**.

Learning word vectors from LMs

- To predict a target word from its context, a LM is typically passed a list of words or subwords (the context), 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 contexts to suitable "filler" words, **the output vector obtained before the sampling step is expected to both summarize relevant aspects of the context 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**.

Learning word vectors from LMs

- To predict a target word from its context, a LM is typically passed a list of words or subwords (the context), 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 contexts to suitable “filler” words, **the output vector obtained before the sampling step is expected to both summarize relevant aspects of the context 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**.

Static vs. contextualized embeddings

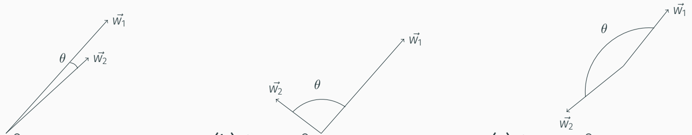
- In the Continuous Bag Of Words model, the embedding space is "**static**": it is computed once and for all for each word of a given training dataset.
- In the case of BERT-like Encoders, word-vectors are extracted from the last or penultimate layer of the model, after feeding it a specific input sequence. 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**".

Static vs. contextualized embeddings

- In the Continuous Bag Of Words model, the embedding space is "**static**": it is computed once and for all for each word of a given training dataset.
- In the case of BERT-like Encoders, word-vectors are extracted from the last or penultimate layer of the model, after feeding it a specific input sequence. 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**".

Measuring similarity in word embeddings

- Cosine similarity (henceforth **CosSim**) is commonly used to similarity between vectors. CosSim is proportional to the angle between two vectors.



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

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

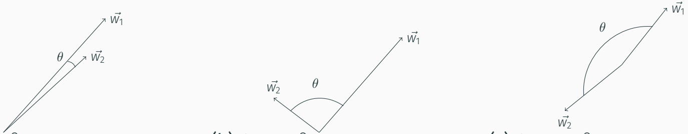
(c) $\theta \simeq 180°$
CosSim$(\vec{w_1}, \vec{w_2}) \simeq -1$

Fig. 5: CosSim as a function of the angle between two vectors.

- Unlike Euclidean distance, CosSim is insensitive to the vectors' magnitudes; it only measures if they display the same *proportions* of their various features.

Measuring similarity in word embeddings

- Cosine similarity (henceforth **CosSim**) is commonly used to similarity between vectors. CosSim is proportional to the angle between two vectors.



(a) $\theta \simeq 0°$
$\text{CosSim}(\vec{w}_1, \vec{w}_2) \simeq +1$

(b) $\theta \simeq 90°$
$\text{CosSim}(\vec{w}_1, \vec{w}_2) \simeq 0$

(c) $\theta \simeq 180°$
$\text{CosSim}(\vec{w}_1, \vec{w}_2) \simeq -1$

Fig. 5: CosSim as a function of the angle between two vectors.

- 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 in word embeddings

- 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 (8)), **may represent the contrast in meaning between the two forms**, i.e. the semantics (in the broad sense) of the morphological operation itself.

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

- 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.

Arithmetic operations in word embeddings

- 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 (8)), **may represent the contrast in meaning between the two forms**, i.e. the semantics (in the broad sense) of the morphological operation itself.

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

- 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.

Arithmetic operations in word embeddings

- 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 (8)), **may represent the contrast in meaning between the two forms**, i.e. the semantics (in the broad sense) of the morphological operation itself.

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

- 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.

Empirical support for subtractive affixal vectors

- The fact 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)**.

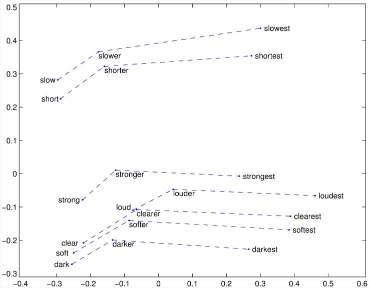


Fig. 6: Comparative and superlatives operations form parallel "lines" in GloVe (Pennington et al., 2014).

Empirical support for subtractive affixal vectors

- The fact 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)**.

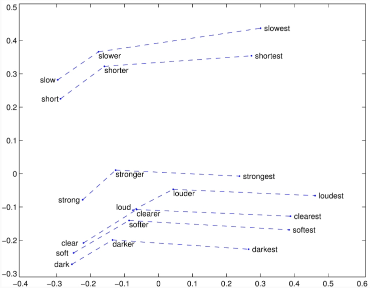


Fig. 6: Comparative and superlatives operations form parallel “lines” in GloVe (Pennington et al., 2014).

Empirical support for subtractive affixal vectors

- Benbaji-Elhadad et al. (2022, 2024) show that Hebrew embeddings encode denominal verbs (word-level derivatives) as **closer** to their base than root-derived counterparts.
- They also show, building on Aronoff (1976) that pairs of English suffixes like -*ness*/-*ity* differ in terms of **stability**–see below.[20]

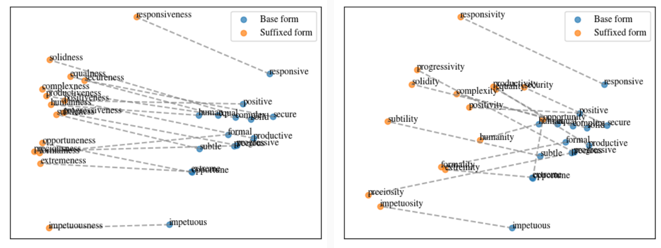


Fig. 7: -*ness* (word-level, parallel lines) and -*ity* (often root-level, less parallel lines) suffixation in fastText (Benbaji-Elhadad et al., 2024).

[20] See also Bonami and Paperno 2018; Musil et al. 2019; Bonami and Guzmán Naranjo 2023; Naranjo and Bonami 2023; Schäfer 2023 for similar methodologies applied to different languages and data.

Quantifying transparency in word embeddings

- Existing work exploits two main metrics to quantify semantic transparency: locality, as measured by base-derivative CosSim (9), and affix stability, as measured by the CosSim between individual affixal vectors and their centroid (10).
- Let's go through these two metrics one by one.

Base-derivative locality

- Transparent derivations like pure diminutization are systematic and **minimal**; they do not radically change the lexical meaning of the base.
- In the vector space, **a base and its transparent derivative should then be close** in terms of CosSim.

(9) *Transparency as base-derivative locality.* Let a and a' be two operations with roughly similar meanings, but a is productive and transparent, while a' less so.
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.:
$\boxed{\text{CosSim}(\overrightarrow{b}, \overrightarrow{d}) > \text{CosSim}(\overrightarrow{b'}, \overrightarrow{d'})}$

Base-derivative locality

- Transparent derivations like pure diminutization are systematic and **minimal**; they do not radically change the lexical meaning of the base.
- In the vector space, **a base and its transparent derivative should then be close** in terms of CosSim.

(9) *Transparency as base-derivative locality.* Let a and a' be two operations with roughly similar meanings, but a is productive and transparent, while a' less so.
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.:
$$\boxed{\mathrm{CosSim}(\overrightarrow{b}, \overrightarrow{d}) > \mathrm{CosSim}(\overrightarrow{b'}, \overrightarrow{d'})}$$

Base-derivative locality

- Transparent derivations like pure diminutization are systematic and **minimal**; they do not radically change the lexical meaning of the base.
- In the vector space, **a base and its transparent derivative should then be close** in terms of CosSim.

(9) *Transparency as base-derivative locality.* Let a and a' be two operations with roughly similar meanings, but a is productive and transparent, while a' less so.
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.:

$$\boxed{\text{CosSim}(\overrightarrow{b}, \overrightarrow{d}) > \text{CosSim}(\overrightarrow{b'}, \overrightarrow{d'})}$$

Affix stability

- Transparent derivations lead to the **same change** across base-derivative pairs.
- In the vector space, the affixal vectors instantiating the same transparent transformation should have same direction and orientation, i.e. exhibit **high CosSims with their centroid (mean)**.

(10) *Transparency as affix stability*. Let a and a' be two operations with roughly similar meanings, but a is productive and transparent, while a' less so.
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 (8)) over many base-derivative pairs resulting from a- or a'-affixation. Let $\overrightarrow{c}$ and $\overrightarrow{c'}$ be the centroids 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 $i \in I$:
$\boxed{\mathrm{CosSim}(\overrightarrow{c}, \overrightarrow{a_i}) > \mathrm{CosSim}(\overrightarrow{c'}, \overrightarrow{a'_i})}$.

Affix stability

- Transparent derivations lead to the **same change** across base-derivative pairs.
- In the vector space, the affixal vectors instantiating the same transparent transformation should have same direction and orientation, i.e. exhibit **high CosSims with their centroid (mean)**.

(10) *Transparency as affix stability*. Let a and a' be two operations with roughly similar meanings, but a is productive and transparent, while a' less so.
Let $(\vec{a}_i)_{i \in I}$ and $(\vec{a'}_i)_{i \in I}$ be two same-size families of affixal vectors computed by difference (following (8)) over many base-derivative pairs resulting from a- or a'-affixation. Let $\vec{c}$ and $\vec{c'}$ be the centroids of $(\vec{a}_i)_{i \in I}$ and $(\vec{a'}_i)_{i \in I}$.
The $(\vec{a}_i)_{i \in I}$ vectors should be overall closer to their centroid $\vec{c}$ than the $(\vec{a'}_i)_{i \in I}$ to $\vec{c'}$, i.e., for most $i \in I$:
$\boxed{\mathrm{CosSim}(\vec{c}, \vec{a_i}) > \mathrm{CosSim}(\vec{c'}, \vec{a'_i})}$.

Affix stability

- Transparent derivations lead to the **same change** across base-derivative pairs.
- In the vector space, the affixal vectors instantiating the same transparent transformation should have same direction and orientation, i.e. exhibit **high CosSims with their centroid (mean)**.

(10) *Transparency as affix stability*. Let a and a' be two operations with roughly similar meanings, but a is productive and transparent, while a' less so.
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 (8)) over many base-derivative pairs resulting from a- or a'-affixation. Let $\overrightarrow{c}$ and $\overrightarrow{c'}$ be the centroids 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 $i \in I$:
$\boxed{\text{CosSim}(\overrightarrow{c}, \overrightarrow{a_i}) > \text{CosSim}(\overrightarrow{c'}, \overrightarrow{a'_i})}$.

Effect of gender-mismatch on locality and stability

Root-level categorization	<	Root-level modification	,	Word-level modification	,	Word-level categorization

- In the case of *-et(te)*, since DIM has a more specific meaning than SIM, it is expected to be more local and stable than SIM.
- I.e., gender-matching groups of base-derivative pairs (FF and MM; mostly DIM^{rl}_{m}-derived) should involve more local and stable transformations than the gender-mismatching ones (FM and MF, mostly SIM_{h}-derived). Schematically:

$$\boxed{\text{FM, MF} \overset{(9\text{-}10)}{<} \text{MM, FF}}$$

- We thus expect an effect of **gender-mismatch** on locality and stability.

Effect of gender-mismatch on locality and stability

Root-level categorization	<	Root-level modification	,	Word-level modification	,	Word-level categorization

- In the case of *-et(te)*, since DIM has a more specific meaning than SIM, it is expected to be more local and stable than SIM.
- I.e., gender-matching groups of base-derivative pairs (FF and MM; mostly DIM_m^{rl}-derived) should involve more local and stable transformations than the gender-mismatching ones (FM and MF, mostly SIM_h-derived). Schematically:
  $\boxed{\text{FM, MF} \overset{(9\text{-}10)}{<} \text{MM, FF}}$
- We thus expect an effect of **gender-mismatch** on locality and stability.

Effect of suffix gender on locality and stability

Root-level categorization	<	Root-level modification	,	Word-level modification	,	Word-level categorization

- Additionally, $\textsc{Sim}_h^{wl}$ (feminine variant, word level) should be more local and stable than $\textsc{Sim}_h^{rl}$ (masculine variant, root level).
- I.e., Among the two gender-matching groups, the FF group should be characterized by slightly more locality and stability than the MM group. And among the two gender-mismatching groups, the MF group should be characterized by slightly more locality and stability than the FM group. Schematically:
$\boxed{\text{FM} \overset{(9\text{-}10)}{<} \text{MF and MM} \overset{(9\text{-}10)}{<} \text{FF}}$
- We thus expect an effect of **suffix gender** (keeping constant the gender-mismatch parameter) on locality and stability.

Effect of suffix gender on locality and stability

$$\begin{array}{c} \text{Root-level} \\ \text{categorization} \end{array} < \begin{array}{c} \text{Root-level} \\ \text{modification} \end{array}, \begin{array}{c} \text{Word-level} \\ \text{modification} \end{array}, \begin{array}{c} \text{Word-level} \\ \text{categorization} \end{array}$$

- Additionally, $\textsc{Sim}_h^{wl}$ (feminine variant, word level) should be more local and stable than $\textsc{Sim}_h^{rl}$ (masculine variant, root level).
- I.e., Among the two gender-matching groups, the FF group should be characterized by slightly more locality and stability than the MM group. And among the two gender-mismatching groups, the MF group should be characterized by slightly more locality and stability than the FM group. Schematically: $\boxed{\text{FM} \overset{(9\text{-}10)}{<} \text{MF and MM} \overset{(9\text{-}10)}{<} \text{FF}}$
- We thus expect an effect of **suffix gender** (keeping constant the gender-mismatch parameter) on locality and stability.

Testing the locality predictions

Dataset

- For each group of base-derivative pairs (FF, MM, MF, FM), 44 pairs were sampled from Hénot-Mortier's dataset (Hénot-Mortier, 2025).[21]
- Same-size sampling ensured that spread differences between groups of vectors (central to stability predictions) would not be driven by differences in group size.
- The results presented here appear robust to sampling.

[21] Data, models and analyses are publicly available on OSF.

Dataset

- For each group of base-derivative pairs (FF, MM, MF, FM), 44 pairs were sampled from Hénot-Mortier’s dataset (Hénot-Mortier, 2025).[21]
- Same-size sampling ensured that spread differences between groups of vectors (central to stability predictions) would not be driven by differences in group size.
- The results presented here appear robust to sampling.

[21] Data, models and analyses are publicly available on OSF.

Dataset

- For each group of base-derivative pairs (FF, MM, MF, FM), 44 pairs were sampled from Hénot-Mortier's dataset (Hénot-Mortier, 2025).[21]
- Same-size sampling ensured that spread differences between groups of vectors (central to stability predictions) would not be driven by differences in group size.
- The results presented here appear robust to sampling.

[21] Data, models and analyses are publicly available on OSF.

Models

- Four pretrained LMs were tested. Importantly, these LMs exhibit **varying degrees of sophistication and sensitivity to subword information**.
  - FASTTEXT (Grave et al., 2018) is fairly light, static, and **substring** (“*n*-gram”) **sensitive**, i.e. incorporates morphemic information is a “diluted” way.
  - FLAUBERT and CAMEMBERT (Le et al., 2020; Martin et al., 2020) are more sophisticated **token-sensitive** Encoders trained on French data. Tokens are integers corresponding to words or subwords, determined based on frequent character cooccurrences. They often overlap with actual morphemes.
  - CANINE-C (Clark et al., 2021) is an Encoder too, but multilingual and tokenizerless, i.e. operates purely at the character level. Despite its elaboration, it is **architecturally blind to subword information**.[22]

[22] This sensitivity may still be *derived* via self-attention.

Models

- Four pretrained LMs were tested. Importantly, these LMs exhibit **varying degrees of sophistication and sensitivity to subword information**.
    - fastText (Grave et al., 2018) is fairly light, static, and **substring** (“*n*-gram”) **sensitive**, i.e. incorporates morphemic information is a “diluted” way.
    - FlauBERT and CamemBERT (Le et al., 2020; Martin et al., 2020) are more sophisticated **token-sensitive** Encoders trained on French data. Tokens are integers corresponding to words or subwords, determined based on frequent character cooccurrences. They often overlap with actual morphemes.
    - Canine-c (Clark et al., 2021) is an Encoder too, but multilingual and tokenizerless, i.e. operates purely at the character level. Despite its elaboration, it is **architecturally blind to subword information**.[22]

[22] This sensitivity may still be *derived* via self-attention.

Models

- Four pretrained LMs were tested. Importantly, these LMs exhibit **varying degrees of sophistication and sensitivity to subword information**.
  - fastText (Grave et al., 2018) is fairly light, static, and **substring** ("*n*-gram") **sensitive**, i.e. incorporates morphemic information is a "diluted" way.
  - FlauBERT and CamemBERT (Le et al., 2020; Martin et al., 2020) are more sophisticated **token-sensitive** Encoders trained on French data. Tokens are integers corresponding to words or subwords, determined based on frequent character cooccurrences. They often overlap with actual morphemes.
  - Canine-c (Clark et al., 2021) is an Encoder too, but multilingual and tokenizerless, i.e. operates purely at the character level. Despite its elaboration, it is **architecturally blind to subword information**.[22]

[22] This sensitivity may still be *derived* via self-attention.

Vector extraction

- For each LM, vectors corresponding to bases and derivatives from each group (FF, MM, MF, FM) were extracted.[23]
- The vectors underwent Cosine-Kernel Principal Component Analysis (henceforth CosKPCA), to **remove uninformative dimensions** while optimally retaining pairwise measures of CosSim between word-vectors. Reduced dimensionalities range between 31 and 74, and retained more than 95% or the original spaces' geometry.
- The next two slides show visualization in 2D, obtained via CosKPCA as well.

[23] See Appendix for how this was done for each LM.

Vector extraction

- For each LM, vectors corresponding to bases and derivatives from each group (FF, MM, MF, FM) were extracted.[23]
- The vectors underwent Cosine-Kernel Principal Component Analysis (henceforth CosKPCA), to **remove uninformative dimensions** while optimally retaining pairwise measures of CosSim between word-vectors. Reduced dimensionalities range between 31 and 74, and retained more than 95% or the original spaces' geometry.
- The next two slides show visualization in 2D, obtained via CosKPCA as well.

[23] See Appendix for how this was done for each LM.

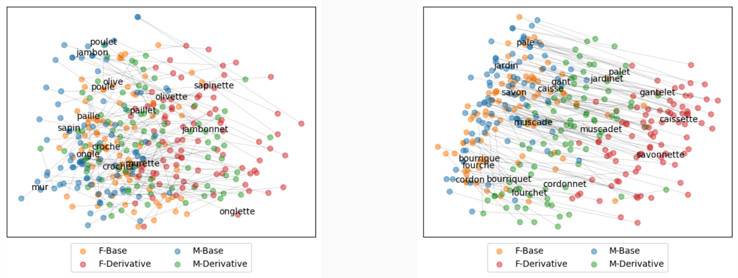
Vector extraction

- For each LM, vectors corresponding to bases and derivatives from each group (FF, MM, MF, FM) were extracted.[23]
- The vectors underwent Cosine-Kernel Principal Component Analysis (henceforth CosKPCA), to **remove uninformative dimensions** while optimally retaining pairwise measures of CosSim between word-vectors. Reduced dimensionalities range between 31 and 74, and retained more than 95% or the original spaces' geometry.
- The next two slides show visualization in 2D, obtained via CosKPCA as well.

[23] See Appendix for how this was done for each LM.

2D visualizations (tokenizerless LMs)

- 2D reduction retains $\simeq$ 50% of the original spaces' geometry.
- Clusters are quite fuzzy, but not fully random.[24]
- "Lines" (affixal vectors) from bases to their corresponding derivative are fairly stable (left-to-right) especially with CANINE-C.



(a) FASTTEXT; CosSim correlation 48.5%.

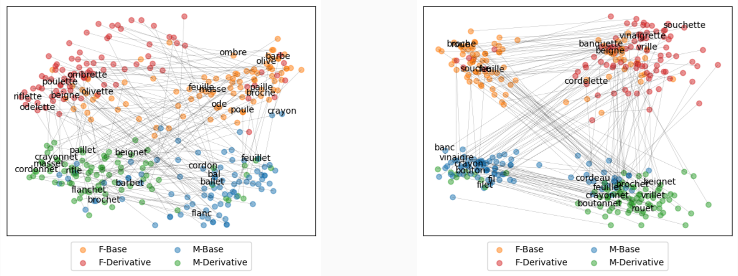
(b) CANINE-C; CosSim correlation 55.9%.

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

[24] Appendix shows how this was tested in 2D and in the high-dimensional reduced spaces.

2D visualizations (token-based LMs)

- More geometry retained, clearer clusters, distributed along gender (*y*) and derivatineness (*x*), which may indicate LMs leverage a tokenization of derivatives into base+*et(te)*.
- "Lines" (affixal vectors) also form clusters (horizontal vs. downward-diagonal vs. upward-diagonal).

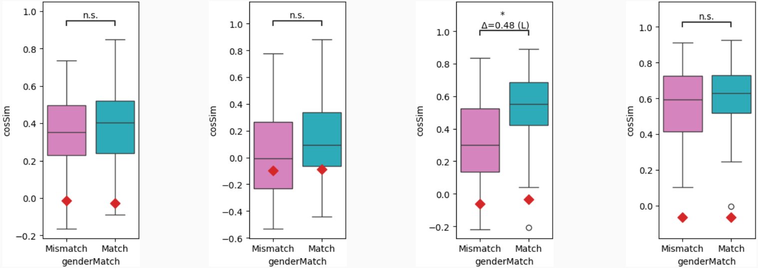


(a) FLAUBERT; CosSim correlation 81.3%.

(b) CAMEMBERT; CosSim correlation 74.2%.

Fig. 9: 2D reductions of base and derivative word-vectors. Lines maps bases with their derivative, and also represent 2D affixal vectors.

Base-derivative locality depending on gender-(mis)match

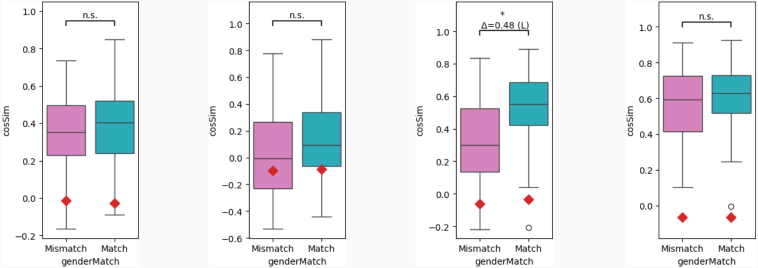


(a) FASTTEXT 😐 (b) FLAUBERT 😐 (c) CAMEMBERT 🙂 (d) CANINE-C 😐

Fig. 10: Distributions of base-derivative CoSims across gender-matching vs. mismatching groups. Red markers are CoSim baselines computed by averaging CoSims between all bases and all derivatives (no pairing) in each group.

- Only CAMEMBERT exhibits the expected asymmetry between gender-matching (more transparent, more local) and mismatching groups (2-sided Mann-Whitney U-test, large effect as per Cliff's Δ).
- No evidence against this result across LMs.

Base-derivative locality depending on gender-(mis)match

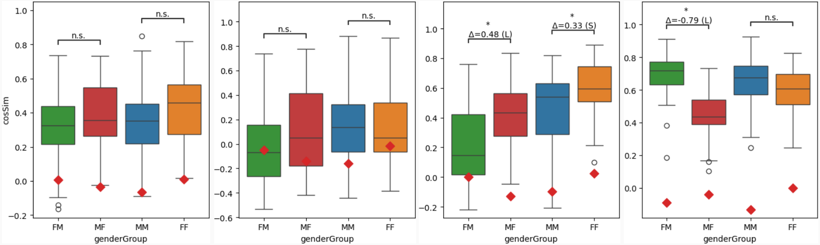


(a) FASTTEXT 😐 (b) FLAUBERT 😐 (c) CAMEMBERT 🙂 (d) CANINE-C 😐

Fig. 10: Distributions of base-derivative CoSims across gender-matching vs. mismatching groups. Red markers are CoSim baselines computed by averaging CoSims between all bases and all derivatives (no pairing) in each group.

- Only CAMEMBERT exhibits the expected asymmetry between gender-matching (more transparent, more local) and mismatching groups (2-sided Mann-Whitney U-test, large effect as per Cliff's Δ).
- No evidence against this result across LMs.

Base-derivative locality depending on suffix gender

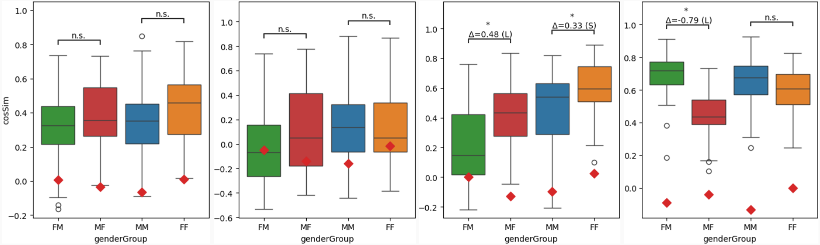


(a) FASTTEXT 😐 (b) FLAUBERT 😐 (c) CAMEMBERT 🙂 (d) CANINE-C ☹

Fig. 11: Distributions of base-derivative CoSims across groups and LMs. Red markers are CosSim baselines. *p*-values are Benjamini-Yekutieli corrected.

- Only CAMEMBERT exhibits the expected asymmetries between feminine forms in *-ette* (more transparent, more local) and masculine forms in *-et* (large and small effects).
- CANINE-C exhibits opposite evidence in gender-mismatching groups–probably due to oversensitivity to character overlap (larger overlap between base and base+*et* than between base and base+*ette*).

Base-derivative locality depending on suffix gender



(a) FASTTEXT 😐 (b) FLAUBERT 😐 (c) CAMEMBERT 🙂 (d) CANINE-C ☹️

Fig. 11: Distributions of base-derivative CoSims across groups and LMs. Red markers are CosSim baselines. *p*-values are Benjamini-Yekutieli corrected.

- Only CAMEMBERT exhibits the expected asymmetries between feminine forms in *-ette* (more transparent, more local) and masculine forms in *-et* (large and small effects).
- CANINE-C exhibits opposite evidence in gender-mismatching groups–probably due to oversensitivity to character overlap (larger overlap between base and base+*et* than between base and base+*ette*).

Take-home from locality testing

- When it comes to the closeness between base and derivative word vectors, CamemBERT verifies our to predictions by displaying:
  - An **effect of gender-(mis)match** in the expected direction: gender-matching operations are more local than gender-mismatching ones.
  - A nested effect of **suffix gender**: forms in *-ette* are closer to their base than forms in *-et*, keeping the gender-mismatch parameter constant.
- This LM thus supports our empirical and theoretical distinctions.
- Two other LMs follow a similar trend, while the last LMs, Canine-c, partially contradicts it, likely due to this model's oversensitivity to character overlaps.
- We now turn to the second prediction, pertaining to the stability of *-et(te)*-suffixation across gender groups.

Take-home from locality testing

- When it comes to the closeness between base and derivative word vectors, CamemBERT verifies our to predictions by displaying:
    - An **effect of gender-(mis)match** in the expected direction: gender-matching operations are more local than gender-mismatching ones.
    - A nested effect of **suffix gender**: forms in *-ette* are closer to their base than forms in *-et*, keeping the gender-mismatch parameter constant.
- This LM thus supports our empirical and theoretical distinctions.
- Two other LMs follow a similar trend, while the last LMs, Canine-c, partially contradicts it, likely due to this model's oversensitivity to character overlaps.
- We now turn to the second prediction, pertaining to the stability of *-et(te)*-suffixation across gender groups.

Take-home from locality testing

- When it comes to the closeness between base and derivative word vectors, CamemBERT verifies our to predictions by displaying:
  - An **effect of gender-(mis)match** in the expected direction: gender-matching operations are more local than gender-mismatching ones.
  - A nested effect of **suffix gender**: forms in *-ette* are closer to their base than forms in *-et*, keeping the gender-mismatch parameter constant.
- This LM thus supports our empirical and theoretical distinctions.
- Two other LMs follow a similar trend, while the last LMs, Canine-c, partially contradicts it, likely due to this model's oversensitivity to character overlaps.
- We now turn to the second prediction, pertaining to the stability of *-et(te)*-suffixation across gender groups.

Take-home from locality testing

- When it comes to the closeness between base and derivative word vectors, CamemBERT verifies our to predictions by displaying:
  - An **effect of gender-(mis)match** in the expected direction: gender-matching operations are more local than gender-mismatching ones.
  - A nested effect of **suffix gender**: forms in *-ette* are closer to their base than forms in *-et*, keeping the gender-mismatch parameter constant.
- This LM thus supports our empirical and theoretical distinctions.
- Two other LMs follow a similar trend, while the last LMs, Canine-c, partially contradicts it, likely due to this model's oversensitivity to character overlaps.
- We now turn to the second prediction, pertaining to the stability of *-et(te)*-suffixation across gender groups.

Take-home from locality testing

- When it comes to the closeness between base and derivative word vectors, CamemBERT verifies our to predictions by displaying:
  - An **effect of gender-(mis)match** in the expected direction: gender-matching operations are more local than gender-mismatching ones.
  - A nested effect of **suffix gender**: forms in *-ette* are closer to their base than forms in *-et*, keeping the gender-mismatch parameter constant.
- This LM thus supports our empirical and theoretical distinctions.
- Two other LMs follow a similar trend, while the last LMs, Canine-c, partially contradicts it, likely due to this model's oversensitivity to character overlaps.
- We now turn to the second prediction, pertaining to the stability of *-et(te)*-suffixation across gender groups.

Take-home from locality testing

- When it comes to the closeness between base and derivative word vectors, CamemBERT verifies our to predictions by displaying:
    - An **effect of gender-(mis)match** in the expected direction: gender-matching operations are more local than gender-mismatching ones.
    - A nested effect of **suffix gender**: forms in *-ette* are closer to their base than forms in *-et*, keeping the gender-mismatch parameter constant.
- This LM thus supports our empirical and theoretical distinctions.
- Two other LMs follow a similar trend, while the last LMs, Canine-c, partially contradicts it, likely due to this model's oversensitivity to character overlaps.
- We now turn to the second prediction, pertaining to the stability of *-et(te)*-suffixation across gender groups.

Testing the stability predictions

Affixal vector computation

- For each LM and each base-derivative pair in our dataset, an affixal vector was computed following equation (8), repeated below. These computations were performed in the raw spaces.

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

- Affixal vectors were then efficiently reduced via CosKPCA.[25] Reduced dimensionalities ranged between 38 and 80.

[25] Same methodology as with the bases and derivatives–see Appendix.

Affixal vector computation

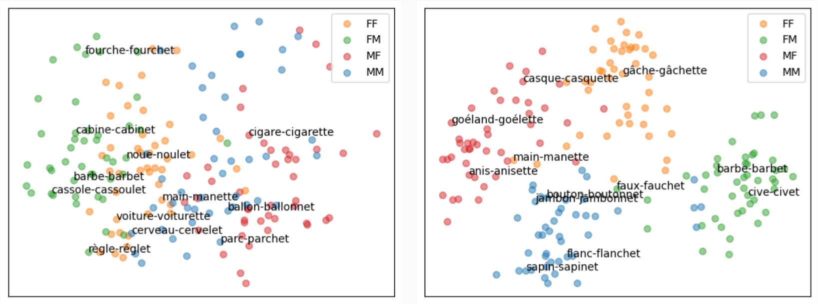
- For each LM and each base-derivative pair in our dataset, an affixal vector was computed following equation (8), repeated below. These computations were performed in the raw spaces.

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

- Affixal vectors were then efficiently reduced via CosKPCA.[25] Reduced dimensionalities ranged between 38 and 80.

[25] Same methodology as with the bases and derivatives–see Appendix.

2D visualizations (tokenizerless LMs)

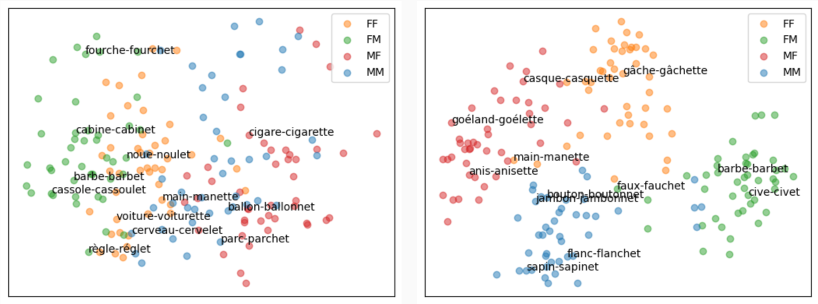


(a) fastText; CosSim correlation 49.4%. (b) Canine-c; CosSim correlation 70.8%.

Fig. 12: 2D reductions of affixal vectors for *-et(te)* across groups.

- Gender-matching operations (blue, orange) are sandwiched between the mismatching ones.
- No clear spread differences...
- Canine-c distinguishes comparatively well between MM and FF operations, likely due to superficial character differences between *-et* and *-ette*.

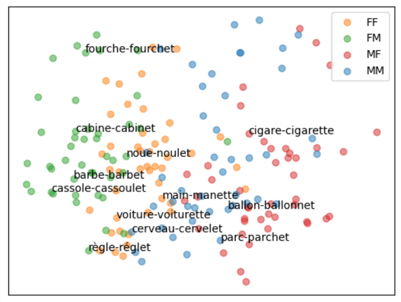
2D visualizations (tokenizerless LMs)



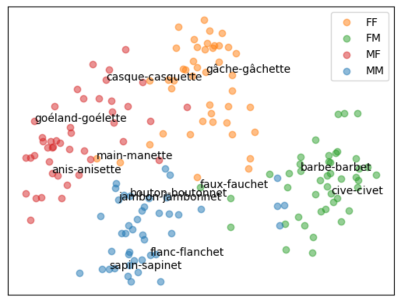
(a) FASTTEXT; CosSim correlation 49.4%. (b) CANINE-C; CosSim correlation 70.8%.

Fig. 12: 2D reductions of affixal vectors for *-et(te)* across groups.

- Gender-matching operations (blue, orange) are sandwiched between the mismatching ones.
- No clear spread differences…
- CANINE-C distinguishes comparatively well between MM and FF operations, likely due to superficial character differences between *-et* and *-ette*.



(a) FASTTEXT; CosSim correlation 49.4%.

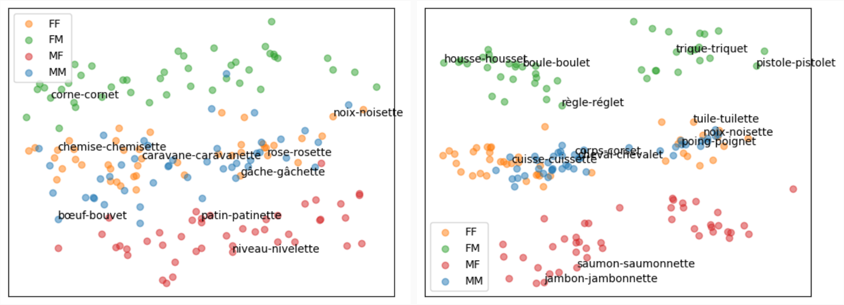


(b) CANINE-C; CosSim correlation 70.8%.

Fig. 12: 2D reductions of affixal vectors for *-et(te)* across groups.

- Gender-matching operations (blue, orange) are sandwiched between the mismatching ones.
- No clear spread differences…
- CANINE-C distinguishes comparatively well between MM and FF operations, likely due to superficial character differences between *-et* and *-ette*.

2D visualizations (token-based LMs)

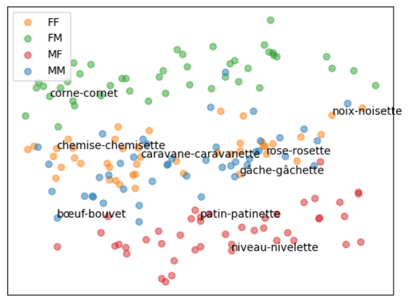


(a) FLAUBERT; CosSim correlation 69.8%. (b) CAMEMBERT; CosSim correlation 73.7%.

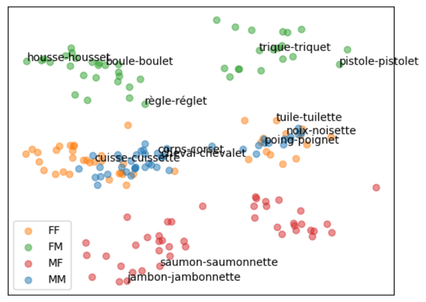
Fig. 13: 2D reductions of affixal vectors for *-et(te)* across groups.

- Gender-matching operations (blue, orange) are sandwiched between the mismatching ones.
- For these token-based LMs, *x* roughly correlates with the number of tokens in the base and derivative.
- Again, no clear spread differences...

2D visualizations (token-based LMs)



(a) FLAUBERT; CosSim correlation 69.8%.

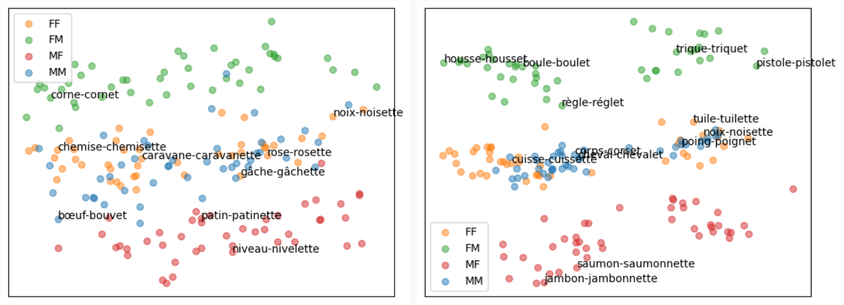


(b) CAMEMBERT; CosSim correlation 73.7%.

Fig. 13: 2D reductions of affixal vectors for *-et(te)* across groups.

- Gender-matching operations (blue, orange) are sandwiched between the mismatching ones.
- For these token-based LMs, x roughly correlates with the number of tokens in the base and derivative.
- Again, no clear spread differences...

2D visualizations (token-based LMs)

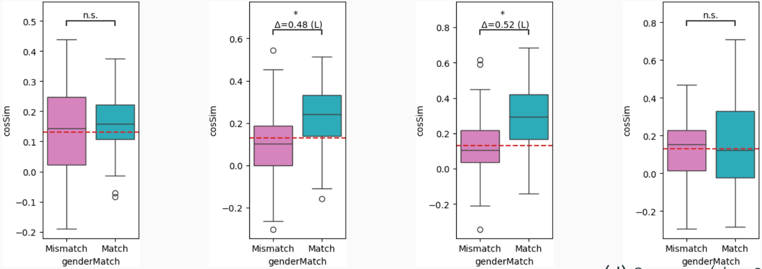


(a) FLAUBERT; CosSim correlation 69.8%. (b) CAMEMBERT; CosSim correlation 73.7%.

Fig. 13: 2D reductions of affixal vectors for *-et(te)* across groups.

- Gender-matching operations (blue, orange) are sandwiched between the mismatching ones.
- For these token-based LMs, x roughly correlates with the number of tokens in the base and derivative.
- Again, no clear spread differences...

Affix stability depending on gender-(mis)match



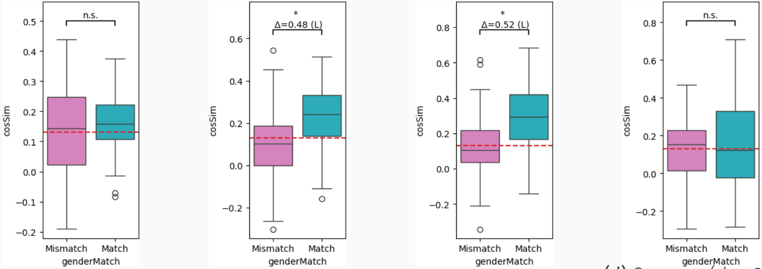
(a) FASTTEXT ($d = 80$) (b) FLAUBERT ($d = 40$) (c) CAMEMBERT ($d = 41$) (d) CANINE-C ($d = 38$)

Fig. 14: Distributions of affix-centroid CosSims across groups and LMs. Red dashed lines are baselines assuming no difference between groups.[26]

- The token-based models (FLAUBERT, CAMEMBERT) exhibit a difference in the expected direction between gender-matching affixal vectors (more stable) and gender-mismatching ones (large effects).
- No evidence going in the opposite direction.

[26] Computed by averaging CosSims between 44 randomly sampled affixal vectors and their centroid. A grand average was then computed over 1000 sampling iterations.

Affix stability depending on gender-(mis)match



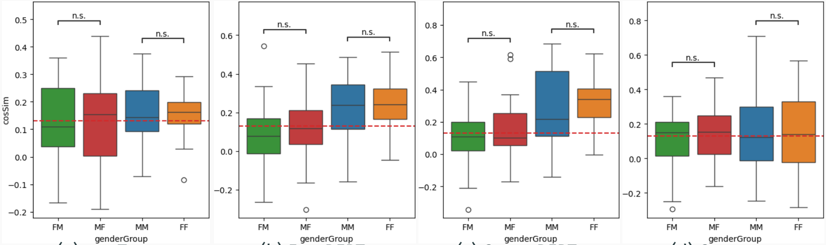
(a) FASTTEXT ($d = 80$) (b) FLAUBERT ($d = 40$) (c) CAMEMBERT ($d = 41$) (d) CANINE-C ($d = 38$)

Fig. 14: Distributions of affix-centroid CosSims across groups and LMs. Red dashed lines are baselines assuming no difference between groups.[26]

- The token-based models (FLAUBERT, CAMEMBERT) exhibit a difference in the expected direction between gender-matching affixal vectors (more stable) and gender-mismatching ones (large effects).
- No evidence going in the opposite direction.

[26] Computed by averaging CosSims between 44 randomly sampled affixal vectors and their centroid. A grand average was then computed over 1000 sampling iterations.

Affix stability depending on suffix gender

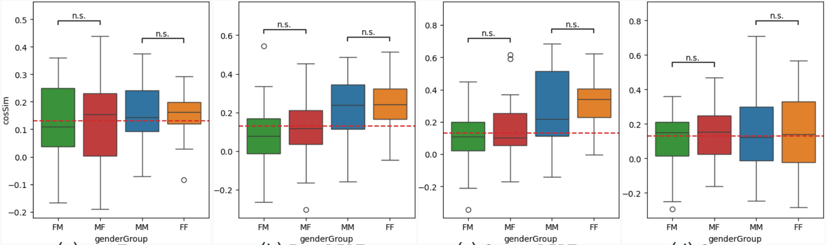


(a) FASTTEXT (b) FLAUBERT (c) CAMEMBERT (d) CANINE-C

Fig. 15: Distributions of affix-centroid CoSims across groups and LMs.

- All models but CAMEMBERT (and arguably FLAUBERT) distribute along the baseline, meaning group distinctions are not relevant.
- No model exhibits the expected pattern (FM<MF and MM<FF) at a significant level.

Affix stability depending on suffix gender



(a) FASTTEXT (b) FLAUBERT (c) CAMEMBERT (d) CANINE-C

Fig. 15: Distributions of affix-centroid CoSims across groups and LMs.

- All models but CAMEMBERT (and arguably FLAUBERT) distribute along the baseline, meaning group distinctions are not relevant.
- No model exhibits the expected pattern (FM<MF and MM<FF) at a significant level.

Interim summary

- When it comes to the stability of affixal vectors, the two tokenizer-bsaed LMs exhibit an **effect of gender-(mis)match** in the expected direction: gender-matching operations are more stable (i.e. their affixal vectors are overall closer to their centroid) than gender-mismatching operations.
- This supports the **distinction between a transparent "gender-matching" modifier Dim and a less transparent, "gender-imposing" head Sim**.
- However, **no effect of suffix gender was detected**: forms in *-ette* are no more stable than forms in *-et*, keeping the gender-mismatch parameter constant.
- Let's now discuss the implications of these overall mixed results.

Interim summary

- When it comes to the stability of affixal vectors, the two tokenizer-bsaed LMs exhibit an **effect of gender-(mis)match** in the expected direction: gender-matching operations are more stable (i.e. their affixal vectors are overall closer to their centroid) than gender-mismatching operations.
- This supports the **distinction between a transparent “gender-matching” modifier DIM and a less transparent, “gender-imposing” head SIM**.
- However, **no effect of suffix gender was detected**: forms in *-ette* are no more stable than forms in *-et*, keeping the gender-mismatch parameter constant.
- Let’s now discuss the implications of these overall mixed results.

Interim summary

- When it comes to the stability of affixal vectors, the two tokenizer-bsaed LMs exhibit an **effect of gender-(mis)match** in the expected direction: gender-matching operations are more stable (i.e. their affixal vectors are overall closer to their centroid) than gender-mismatching operations.
- This supports the **distinction between a transparent "gender-matching" modifier Dim and a less transparent, "gender-imposing" head Sim**.
- However, **no effect of suffix gender was detected**: forms in *-ette* are no more stable than forms in *-et*, keeping the gender-mismatch parameter constant.
- Let's now discuss the implications of these overall mixed results.

Interim summary

- When it comes to the stability of affixal vectors, the two tokenizer-bsaed LMs exhibit an **effect of gender-(mis)match** in the expected direction: gender-matching operations are more stable (i.e. their affixal vectors are overall closer to their centroid) than gender-mismatching operations.
- This supports the **distinction between a transparent "gender-matching" modifier Dim and a less transparent, "gender-imposing" head Sim**.
- However, **no effect of suffix gender was detected**: forms in *-ette* are no more stable than forms in *-et*, keeping the gender-mismatch parameter constant.
- Let's now discuss the implications of these overall mixed results.

Discussion and outlook

LMs capture core features like gender and affixation

- 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.

LMs capture core features like gender and affixation

- 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.

LMs capture core features like gender and affixation

- 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.

Focus on the token-based LMs

- FlauBERT and CamemBERT, which come with a tokenization procedure breaking down words into subwords, for instance (and ideally) *boulette* into *boul+ette*, displayed the clearest representations.
- 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.
- This in turn predicts corresponding **affixal vectors computed by subtraction to potentially include pre-isolated affixal information**.
- Still, the fact that **BERT-like LMs assign coherent, fine-grained representations to the relevant tokens remains remarkable**.

Focus on the token-based LMs

- FlauBERT and CamemBERT, which come with a tokenization procedure breaking down words into subwords, for instance (and ideally) *boulette* into *boul+ette*, displayed the clearest representations.
- 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.
- This in turn predicts corresponding **affixal vectors computed by subtraction to potentially include pre-isolated affixal information**.
- Still, the fact that **BERT-like LMs assign coherent, fine-grained representations to the relevant tokens remains remarkable**.

Focus on the token-based LMs

- FlauBERT and CamemBERT, which come with a tokenization procedure breaking down words into subwords, for instance (and ideally) *boulette* into *boul+ette*, displayed the clearest representations.
- 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.
- This in turn predicts corresponding **affixal vectors computed by subtraction to potentially include pre-isolated affixal information**.
- Still, the fact that **BERT-like LMs assign coherent, fine-grained representations to the relevant tokens remains remarkable**.

Focus on the token-based LMs

- FlauBERT and CamemBERT, which come with a tokenization procedure breaking down words into subwords, for instance (and ideally) *boulette* into *boul+ette*, displayed the clearest representations.
- 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.
- This in turn predicts corresponding **affixal vectors computed by subtraction to potentially include pre-isolated affixal information**.
- Still, the fact that **BERT-like LMs assign coherent, fine-grained representations to the relevant tokens remains remarkable**.

Focus on the tokenizerless LMs

- 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 **oversensitive to surface patterns**–ultimately leading to a failure to capture the target predictions.
- This is particularly striking in the case of Canine-c, which is as sophisticated as the other BERT-like models, but suffers significantly from the absence of tokenization.
- This suggest that, when dealing with fine-grained morphological predictions, "attention is ***not*** all you need"; **you basically need a (LM-external) parser**! Let's now discuss further the nature and impact of tokenization.

Focus on the tokenizerless LMs

- 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 **oversensitive to surface patterns**–ultimately leading to a failure to capture the target predictions.
- This is particularly striking in the case of CANINE-C, which is as sophisticated as the other BERT-like models, but suffers significantly from the absence of tokenization.
- This suggest that, when dealing with fine-grained morphological predictions, "attention is ***not*** all you need"; **you basically need a (LM-external) parser**! Let's now discuss further the nature and impact of tokenization.

Focus on the tokenizerless LMs

- 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 **oversensitive to surface patterns**–ultimately leading to a failure to capture the target predictions.
- This is particularly striking in the case of Canine-c, which is as sophisticated as the other BERT-like models, but suffers significantly from the absence of tokenization.
- This suggest that, when dealing with fine-grained morphological predictions, “attention is ***not*** all you need”; **you basically need a (LM-external) parser**! Let’s now discuss further the nature and impact of tokenization.

Nature and impact of tokenization

- The relative success of FlauBERT and CamemBERT suggests that the representation of affixal vectors is influenced by whether, and how, LMs tokenize the corresponding bases and derivatives.
- Focusing on datapoints characterized by morphologically accurate tokenizations may successfully filter noisy representations.
- 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.
- Further analyzing the tokenizations assigned to bases and derivatives across gender groups may thus constitute an indirect measure of the transparency of each group.

Nature and impact of tokenization

- The relative success of FlauBERT and CamemBERT suggests that the representation of affixal vectors is influenced by whether, and how, LMs tokenize the corresponding bases and derivatives.
- Focusing on datapoints characterized by morphologically accurate tokenizations may successfully filter noisy representations.
- 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.
- Further analyzing the tokenizations assigned to bases and derivatives across gender groups may thus constitute an indirect measure of the transparency of each group.

Nature and impact of tokenization

- The relative success of FlauBERT and CamemBERT suggests that the representation of affixal vectors is influenced by whether, and how, LMs tokenize the corresponding bases and derivatives.
- Focusing on datapoints characterized by morphologically accurate tokenizations may successfully filter noisy representations.
- 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.
- Further analyzing the tokenizations assigned to bases and derivatives across gender groups may thus constitute an indirect measure of the transparency of each group.

Nature and impact of tokenization

- The relative success of FlauBERT and CamemBERT suggests that the representation of affixal vectors is influenced by whether, and how, LMs tokenize the corresponding bases and derivatives.
- Focusing on datapoints characterized by morphologically accurate tokenizations may successfully filter noisy representations.
- 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.
- Further analyzing the tokenizations assigned to bases and derivatives across gender groups may thus constitute an indirect measure of the transparency of each group.

LMs need tokenization to "do morphology"

- For FlauBERT and CamemBERT, we identified the number of correctly tokenized derivatives across gender groups.[27]
- FlauBERT appeared to isolate *-ette* more reliably than *-et*, and **clearly outperformed 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.

[27] See Appendix for data

LMs need tokenization to "do morphology"

- For FLAUBERT and CAMEMBERT, we identified the number of correctly tokenized derivatives across gender groups.[27]
- FLAUBERT appeared to isolate *-ette* more reliably than *-et*, and **clearly outperformed 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.

[27] See Appendix for data

LMs need tokenization to “do morphology”

- For FlauBERT and CamemBERT, we identified the number of correctly tokenized derivatives across gender groups.[27]
- FlauBERT appeared to isolate *-ette* more reliably than *-et*, and **clearly outperformed 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.

[27] See Appendix for data

LMs need tokenization to "do morphology"

- For FlauBERT and CamemBERT, we identified the number of correctly tokenized derivatives across gender groups.[27]
- FlauBERT appeared to isolate *-ette* more reliably than *-et*, and **clearly outperformed 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.

[27] See Appendix for data

Caveat: frequencies

- **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.

Caveat: frequencies

- **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.

Caveat: frequencies

- **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.

Caveat: relativity of the predictions

- 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***.
- The conceptual (and hard) question that remains is whether the observed 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.

Caveat: relativity of the predictions

- 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***.
- The conceptual (and hard) question that remains is whether the observed 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.

Caveat: relativity of the predictions

- 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***.
- The conceptual (and hard) question that remains is whether the observed 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.

Conclusion

- We showed that LMs can 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*.
- *-et(te)*-suffixation was also **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 *must* be combined with a biased word decomposition process to derive linguistically-informed representations.**

Conclusion

- We showed that LMs can 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*.
- *-et(te)*-suffixation was also **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 *must* be combined with a biased word decomposition process to derive linguistically-informed representations.

Conclusion

- We showed that LMs can 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*.
- *-et(te)*-suffixation was also **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 *must* be combined with a biased word decomposition process to derive linguistically-informed representations.**

Thank you!

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

Harris, Z. S. (1954).Distributional structure. *WORD*, *10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

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

Weber, M. (1963). *Contributions à l'étude du diminutif en français moderne : Essai de systématisation* [Doctoral dissertation].

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

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

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. https://doi.org/10.1016/0377-0427(87)90125-7

Milner, J.-C. (1989).Genre et dimension dans les diminutifs français. *Linx*, *21*(1), 191–201. https://doi.org/10.3406/linx.1989.1141

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

Stump, G. T. (1993).How peculiar is evaluative morphology? *Journal of Linguistics*, *29*(1), 1–36. https://doi.org/10.1017/s0022226700000037

Halle, M., & Marantz, A. (1994). Some Key Features of Distributed Morphology. In *Papers on phonology and morphology* (pp. 275–288). MIT Working Papers in Linguistics.

Delhay, C. (1995).Le diminutif: La diminution sans comparaison? *Faits de langues*, *3*(5), 63–72. https://doi.org/10.3406/flang.1995.976

Jurafsky, D. (1996).Universal Tendencies in the Semantics of the Diminutive. *Language*, *72*(3), 533. https://doi.org/10.2307/416278

Dal, G. (1997). *Grammaire du suffixe -et(te)* [Doctoral dissertation]. Didier Erudition.

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

Harley, H., & Noyer, R. (1998). **Licensing in the non-lexicalist lexicon: Nominalizations, vocabulary items and the encyclopedia.** In H. Harley (Ed.), *Mit working papers in linguistics 32: Papers from the upenn/mit roundtable on argument structure and aspect* (pp. 119–137). MITWPL.

Stump, G. T. (1998). **The handbook of morphology.** In A. Spencer & A. Zwicky (Eds.). Blackwell.

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

Lecarme, J. (2002). **Gender "polarity": Theoretical aspects of Somali nominal morphology.** In P. Boucher & M. Plénat (Eds.), *Many morphologies* (pp. 109–141). Cascadilla Press.

Arad, M. (2003).**Locality constraints on the interpretations of roots.** *Natural Language and Linguistic Theory*, *21*(4), 737–778. https://doi.org/10.1023/a:1025533719905

- Fradin, B. (2003).Le traitement de la suffixation en -et. *Langages*, *37*(152), 51–77. https://doi.org/10.3406/lgge.2003.2439
- Arad, M. (2005). *Roots and Patterns: Hebrew Morpho-syntax.* Springer.
- Ferrari, F. (2005). *A syntactic analysis of the nominal systems of Italian and Luganda: How nouns can be formed in the syntax* [Doctoral dissertation, New York University].
- Kihm, A. (2005). Noun class, gender and the lexicon-syntax-morphology interfaces: A comparative study of Niger-Congo and Romance languages. In G. Cinque & R. S. Kayne (Eds.), *The Oxford handbook of comparative syntax* (pp. 459–512). OUP.
- Wiltschko, M., & Steriopolo, O. (2007). Parameters of variation in the syntax of diminutives. In M. Radisic (Ed.), *Proceedings of the 2007 canadian linguistics association annual conference*.
- Acquaviva, P. (2009).Roots and lexicality in Distributed Morphology. *York Essex Morphology Meeting 2*, 1–21.

De Belder, M., Faust, N., & Lampitelli, N. (2014). **On an inflectional and a derivational diminutive.** In A. Alexiadou, H. Borer, & F. Schäfer (Eds.), *The syntax of roots and the roots of syntax* (pp. 149–163). Oxford University Press.

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

Kramer, R. (2015). ***The Morphosyntax of Gender.*** Oxford University Press.

Deal, A. R. (2016).**Plural exponence in the Nez Perce DP: A DM analysis.** *Morphology*, *26*(3-4), 313–339. https://doi.org/10.1007/s11525-015-9277-9

Kramer, R. (2016).**The location of gender features in the syntax.** *Language and Linguistics Compass*, *10*(11), 661–677. https://doi.org/10.1111/lnc3.12226

Steriopolo, O. (2017). **Syntactic variation in diminutive suffixes: Russian, kolyma yukaghir, and itelmen.** *Languages*, *2*(4). https://doi.org/10.3390/languages2040023

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). **Attention is all you need.** In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Bonami, O., & Paperno, D. (2018). **Inflection vs. derivation in a distributional vector space.** *Lingue e linguaggio*, (2), 173–196. https://doi.org/10.1418/91864

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018, May). **Learning word vectors for 157 languages.** In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, & T. Tokunaga (Eds.), *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*. European Language Resources Association (ELRA). https://aclanthology.org/L18-1550/

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, June). **BERT: Pre-training of deep bidirectional transformers for language understanding.** In J. Burstein, C. Doran, & T. Solorio (Eds.), *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)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Musil, T., Vidra, J., & Mareček, D. (2019). **Derivational morphological relations in word embeddings.** https://arxiv.org/abs/1906.02510

Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., Crabbé, B., Besacier, L., & Schwab, D. (2020).**Flaubert: Unsupervised language model pre-training for french.** *Proceedings of The 12th Language Resources and Evaluation Conference*, 2479–2490. https://www.aclweb.org/anthology/2020.lrec-1.302

Martin, L., Muller, B., Ortiz Suárez, P. J., Dupont, Y., Romary, L., de la Clergerie, É., Seddah, D., & Sagot, B. (2020, July). **CamemBERT: A tasty French language model.** In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 7203–7219). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.645

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

Benbaji-Elhadad, I., Doron, O., & Hénot-Mortier, A. (2022). **Word-embeddings distinguish denominal and root-derived verbs in semitic.** In M. Moortgat & G. Wijnholds (Eds.), Proceedings *end-to-end compositional models of vector-based semantics*, nui galway, 15-16 august 2022 (pp. 35–49, Vol. 366). Open Publishing Association. https://doi.org/10.4204/EPTCS.366.6

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

Bonami, O., & Guzmán Naranjo, M. (2023, February). **Distributional evidence for derivational paradigms.** In *The semantics of derivational morphology* (pp. 219–258). De Gruyter. https://doi.org/10.1515/9783111074917-008

Brucale, L., & Mocciaro, E. (2023).**Approximation through suffixation: -ḍḍu/-a in sicilian.** *Zeitschrift für Wortbildung / Journal of Word Formation*, *7*(1). https://doi.org/10.21248/zwjw.2023.1.89

Naranjo, M. G., & Bonami, O. (2023).**A distributional assessment of rivalry in word formation.** *Word Structure*, *16*(1), 87–114. https://doi.org/10.3366/word.2023.0222

Schäfer, M. (2023, February). **Splitting -ly's: Using word embeddings to distinguish derivation and inflection.** In *The semantics of derivational morphology* (pp. 259–300). De Gruyter. https://doi.org/10.1515/9783111074917-009

Stosic, D., & Amiot, D. (2023).**Can suffixal augmentation express approximation? the case of evaluative nouns.** *Zeitschrift für Wortbildung / Journal of Word Formation*, *7*(1). https://doi.org/10.21248/zwjw.2023.1.33

Benbaji-Elhadad, I., Doron, O., & Hénot-Mortier, A. (2024).**Distinguishing levels of morphological derivation in word-embedding models.** *Proceedings of the 53rd Annual Meeting of the North East Linguistics Society*.

Hénot-Mortier, A. (2024).**Highs and lows of the french diminutive suffix -et(te).** *Lingue e linguaggio*, (2), 245–266. https://doi.org/10.1418/115294

Hénot-Mortier, A. (2025, May). ***French nominal diminutives in -et(te).*** Zenodo. https://doi.org/10.5281/zenodo.15353651

Systematic analysis f forms in *-et(te)*

- 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*.[28]

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

Vector extraction procedures

- fastText vectors were readily available as a fixed embedding mapping a large sample of French words to individual vectors.
- FlauBERT, CamemBERT and Canine-c were fed the sequence *C'est un(e) base|derivative* ('It's $a_{F/M}$ base|derivative'), producing "contextualized" token-vector(s) for each base and derivative.
- Contextualized token-vectors were extracted from the penultimate (for FlauBERT and CamemBERT) or last (for Canine-c) layer.
- 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.

Vector reduction

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

Model	Initial dimension	Reduced dimensions	Pearson correlation
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 Correlation coefficients are between (i) pairwise CosSims computed between vectors in the reduced space, and (ii) pairwise CosSims computed similarly in the original space.

Testing for clustering

- To check that F-bases, F-derivatives, M-bases, and M-derivatives formed different clusters, permutation tests were performed using cosine-Silhouette (a clustering quality score; Rousseeuw, 1987) as dependent measure.
- 1000 times, word vectors were assigned a random label (F-base, F-derivative, M-base, or M-derivative). Each time, the cosine-Silhouette score of the random groups was computed.
- This led to a sample of Silhouette scores of size 1000, representing the baseline under a random distribution of the ground-truth clusters.
- All four embeddings were associated with a significantly higher Silhouette than the distribution of random baselines.
- This was true in dimension 2, and also in the higher reduced dimensions used for testing.

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. 3: 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	% of derivatives with	% of derivatives with