Offset vectors and affix meaning in English nominalizations

Martin Schäfer
martin.schaefer@uni-leipzig.de

Universität Leipzig

Fribourg, September 5, 2025

Offset vectors

vector of derivative			vector of base	= **offset vector**
inclusiveness	(4 3 1)	−	**inclusive** (2 1 4)	= (2 2 -3)
inclusivity	(3 4 1)	−	**inclusive** (2 1 4)	= (1 3 -3)

Introduction

- Shafaei-Bajestan et al. (2024, p. 381), on English plural inflection: "the semantics of shift vectors is changing in close association with the semantics of the singular and plural words."
- Schäfer (2025), on the English *-ity*/*-ness* affix rivalry: the distributional vectors of adjectival bases successfully predict the affix choice.

Introduction

- Shafaei-Bajestan et al. (2024, p. 381), on English plural inflection: "the semantics of shift vectors is changing in close association with the semantics of the singular and plural words."
- Schäfer (2025), on the English *-ity*/*-ness* affix rivalry: the distributional vectors of adjectival bases successfully predict the affix choice.

My research questions:

1. Are the offset vectors of *-ity* base-derivative pairs distinct from the *-ness* pairs?
2. Are there further patterns associated with specific subsets of bases within the offset vectors?

The *-ity* and *-ness* affix rivalry:

(1) *-ity*

a. insular: insularity
b. eatable: eatability
c. sentimental: sentimentality

(2) *-ness*

a. red: redness
b. messy: messiness
c. pleasant: pleasantness

`Note: the study is restricted to adjectival bases used in Schäfer 2024b, data etc. at Schäfer 2024a`

Study 1: *-ity/-ness* offset vectors [methods]

- adj/*-ity/-ness* derivatives:
  tagged ukWaC corpus ∩ fastText vectorsets
- 1 million item fastText vectors **WITHOUT** subword information
- no doublets

→ set of 3014 base-derivative pairs

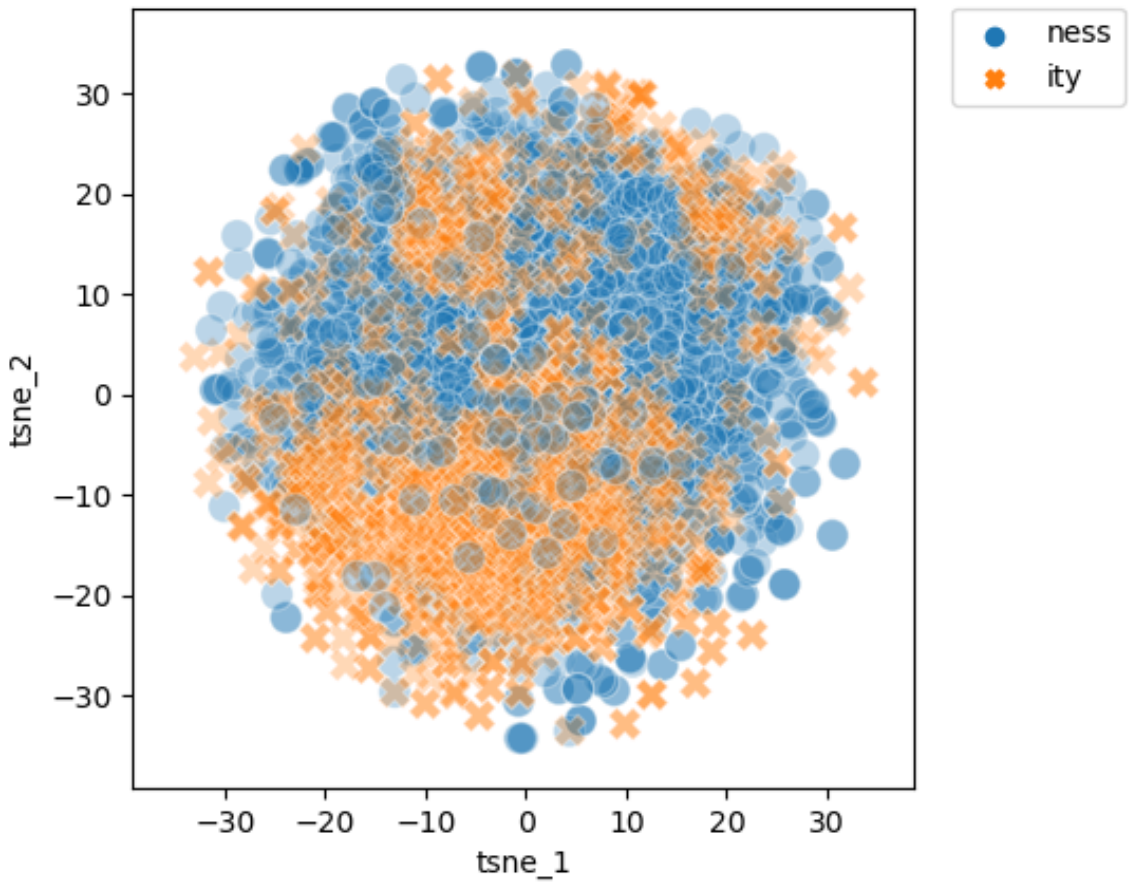
`ukWaC corpus: Baroni et al. (2009); fastText vectorsets: Mikolov, Grave, et al. (2017)`

Study 1: *-ity/-ness* offset vectors [methods]

- offset vectors
- downstream-analysis:
  - t-SNE for visualization
  - Linear Discriminant Analysis (LDA) for statistical corroboration

```
= same downstream analysis as Shafaei-Bajestan et al.
(2024) and Schäfer (2025).
```

Study 1: *-ity*/*-ness* offset vectors [results]



`average weighted F1 score: 0.838 (0.019 std)`
`[baseline classifier 0.395]`

Study 1: *-ity/-ness* offset vectors [discussion]

- clear difference, contrasting with results for French deadjectival derivations in Guzmán Naranjo and Bonami (2023)
- F1 score comparable to score for the bases
- no categorical difference, considerable variation, similar to results by Shafaei-Bajestan et al. (2024)

Study 1: *-ity/-ness* offset vectors [discussion]

- clear difference, contrasting with results for French deadjectival derivations in Guzmán Naranjo and Bonami (2023)
- F1 score comparable to score for the bases
- no categorical difference, considerable variation, similar to results by Shafaei-Bajestan et al. (2024)

Open issues:

- link to genre or text type
- possible frequency effects

Study 2: inside the *-ity*/*-ness* offset vectors

- both sets of offset vectors display considerable variation
- an obvious question is whether this variation is patterned in non-random ways even within the form pairs
- semantics of bases?

Study 2: inside the *-ity*/*-ness* offset vectors [Methods]

- Analogy task of Mikolov, Chen, et al. (2013)
- When adding the average offset vector to the base vector, is the target vector, that is, the actual *-ness* or *-ity* form associated with the base vector, contained in the nearest neighbors of the synthetic vector?

Study 2: inside the *-ity*/*-ness* offset vectors [Methods]

- Analogy task of Mikolov, Chen, et al. (2013)
- When adding the average offset vector to the base vector, is the target vector, that is, the actual *-ness* or *-ity* form associated with the base vector, contained in the nearest neighbors of the synthetic vector?

`[average -`*`ness`*` offset-vector ] + [vector of `*`smooth`*`] =`
`[synthetic vector for `*`smoothness`*`]`

How close is this synthetic vector to the actual vector for *smoothness*?

Study 2: inside the *-ity*/*-ness* offset vectors [Methods]

(3) five average vectors

a. **all**: average offset vector across all pairs
b. **ity**: average offset vector across all *-ity* pairs
c. **ness**: average offset vector across all *-ness* pairs
d. **ble**: average offset vector across all 547 *-ble* bases that take only *-ity*
e. **ed**: average offset vector across all 173 *-ed* bases

Study 2: inside the *-ity*/*-ness* offset vectors [Methods]

(4) four test sets:

a. **other *-ity***: 25 bases with no discernable morphological pattern that have only *-ity* derivatives (*sublime*, *secure*).

b. **other *-ness***: 25 bases of the same type that have only *-ness* derivatives (*harsh*, *smart*).

c. ***-ble* [*-ity*-only]**: 25 *-ble* bases that only have *-ity* derivatives (*lovable*, *notable*).

d. ***-ed* [*-ness*-only]**: 25 *-ed* bases that only have *-ness* derivatives (*directed*, *guarded*)

Study 2: inside the *-ity*/*-ness* offset vectors [results]

Table 1: test set (a), **other *-ity***

Rank	all	ity	ness	ble	ed
Rank 2	6	4	9	5	8
Rank 3	5	3	4	3	5
Rank 4	4	4	4	6	2
Rank 5	2	3	0	1	1
Rank 6-10	3	3	3	3	3
Rank 11-50	3	6	2	5	3
Rank >50	2	2	3	2	3

Study 2: inside the *-ity*/*-ness* offset vectors [results]

Table 2: test set (b), **other *-ness***

Rank	all	ity	ness	ble	ed
Rank 2	7	6	10	6	8
Rank 3	4	2	1	2	3
Rank 4	1	2	2	3	2
Rank 5	3	3	2	2	2
Rank 6-10	3	2	4	3	4
Rank 11-50	2	4	1	3	1
Rank >50	5	6	5	6	5

Study 2: inside the *-ity*/*-ness* offset vectors [results]

Table 3: test set (c), ***-ble* [*-ity* only]**

Rank	all	ity	ness	ble	ed
Rank 2	16	16	17	16	16
Rank 3	1	1	1	3	2
Rank 4	2	0	3	0	1
Rank 5	1	1	1	1	3
Rank 6-10	4	4	2	4	2
Rank 11-50	0	2	1	1	0
Rank >50	1	1	0	0	1

Study 2: inside the *-ity*/*-ness* offset vectors [results]

Table 4: test set (d), ***-ed* [-ness only]**

Rank	all	ity	ness	ble	ed
Rank 2	5	4	6	4	7
Rank 3	3	1	4	2	3
Rank 4	2	3	1	2	2
Rank 5	2	2	3	2	2
Rank 6-10	2	3	2	4	2
Rank 11-50	5	5	3	5	4
Rank >50	6	7	6	6	5

Study 2: inside the *-ity*/*-ness* offset vectors [discussion]

different test sets:

- ***ble* [only *ity*]** vs ***-ed* [only *-ness*]**
  - ***-ble* [*-ity* only]** test set: perhaps prototypical bases, in line with them forming the largest distinct subgroup of bases
  - ***-ed* [only *-ness*]**: perhaps different types of properties (less abstract?)

$\rightarrow$ systematic differences between different types of bases

Study 2: inside the *-ity*/*-ness* offset vectors [discussion]

different composed vectors:

- clear differences between *-ity* and *-ness*
- overall better performance of the *-ness* related average vectors can perhaps be linked to its greater productivity and its less distinct lexicalization effects (Bauer, Lieber, and Plag, 2013)
- ***-ble* [*-ity* only]** and ***-ed*** average vectors: optimized for their respective bases

Study 2: inside the *-ity*/*-ness* offset vectors [discussion]

Other:

- for some items clear evidence of lexicalization effects: lowest ranked examples (always across all 5 probes): *minority*, *otherness*, and *signedness*
- bad performance in comparison to Shafaei-Bajestan et al. (2024); plausible explanation: the less stable nature of derivational vs inflectional relationship, see Bonami and Paperno (2018).

Conclusion

(1) Are the offset vectors across the *ity/ness* non-doublet bases distinct from each other or not?

- clear but non-categorical differences

Conclusion

(2) Are there sub-regularities within the offset vectors of both affixes?

- (difference between the **ity** and **ness** vectors)
- sub-regularities based on morphological properties of the bases
  - ***-ble*** **[only *-ity*]** test set maximally different from the ***-ed*** **[*-ness* only]** test set
  - corresponding average vectors always performing best in the corresponding test sets
  - plausibly linked to prototypicality and semantic differences

Other possibilities

- Other operations between vectors
- Other conceptualizations of affixation (Marelli and Baroni, 2015): affix as matrix, affixation as matrix multiplication

Thank you!

References I

- Baroni, Marco et al. (2009). "The WaCky wide web: a collection of very large linguistically processed web-crawled corpora". In: *Language Resources and Evaluation* 43.3, pp. 209–226. ISSN: 1574-0218. DOI: 10.1007/s10579-009-9081-4.
- Bauer, Laurie, Rochelle Lieber, and Ingo Plag (2013). *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.
- Bonami, Olivier and Denis Paperno (2018). "Inflection vs. derivation in a distributional vector space". In: *Lingue e Linguaggio* 17.2, pp. 173–195.
- Guzmán Naranjo, Matías and Olivier Bonami (2023). "A distributional assessment of rivalry in word formation". In: *Word Structure* 16.1, pp. 87–114.

References II

- Marelli, Marco and Marco Baroni (2015). "Affixation in Semantic Space: Modeling Morpheme Meanings With Compositional Distributional Semantics". In: *Psychological Review* 122.3, pp. 485–515. DOI: 10.1037/a0039267.
- Mikolov, Tomas, Kai Chen, et al. (Jan. 2013). "Efficient Estimation of Word Representations in Vector Space". In: *ArXiv e-prints*. DOI: 10.48550/arXiv.1301.3781.
- Mikolov, Tomas, Edouard Grave, et al. (2017). "Advances in Pre-Training Distributed Word Representations". In: *CoRR*. DOI: 10.48550/arXiv.1712.09405.
- Schäfer, Martin (July 2024a). "A distributional semantics analysis of the two English suffixes -ity and -ness". In: DOI: 10.6084/m9.figshare.23538207.v1. URL: https://figshare.com/articles/online_resource/A_distributional_semantics_analysis_of_the_two_English_suffixes_-ity_and_-ness/23538207.

References III

- Schäfer, Martin (2024b). "The role of meaning in the rivalry of -ity and -ness: evidence from distributional semantics". In: *English Language and Linguistics*. Accepted for publication. URL: https://www.martinschaefer.info/publications/download/2024_ityNess_R2_martin_schaefer_accepted.pdf.
- — (2025). "The role of meaning in the rivalry of -ity and -ness: evidence from distributional semantics". In: *English Language and Linguistics*, pp. 1–46. DOI: 10.1017/S1360674324000443.
- Shafaei-Bajestan, Elnaz et al. (2024). "The pluralization palette: unveiling semantic clusters in English nominal pluralization through distributional semantics". In: *Morphology* 34.4, pp. 369–413. DOI: 10.1007/s11525-024-09428-9.