

A DISTRIBUTIONAL SEMANTICS APPROACH TO COMPETITION IN ITALIAN DEADJECTIVAL VERB FORMATION

Nicola Cirillo • Maria Pina De Rosa • Alessandro Maisto • Flavio Pisciotta - University of Salerno
{nicirillo,mpderosa,amaisto,fpisciotta}@unisa.it



References

1. Introduction

The study provides some insights into the **competition between derivational processes in deadjectival verb formation in Italian** by means of a corpus based analysis within the frame of Distributional Semantics (DS). Through DS we calculate the general vectors of each process and the similarity values between pairs of co-radical verbs. Our aim is to assess which derivational processes more often yield semantically similar verbs, and if **semantic similarity between verbs assumed on the basis of dictionary definitions is corroborated by the distribution of these verbs in corpora**.

3. Deadjectival verb formation

- The derivational processes of **suffixation, conversion (or zero-derivation), and parasynthesis**
- potentially compete for the expression of the core meanings to which the semantics of deadjectival verbs can be traced back to. Specifically,
 - the encoding of **causative/resultative** ('make/become (more) Adj') meanings, i.e. of change of state, is achieved by means of all three processes:
 - e.g. suff. *inglese* → *inglesizzare* *puro* → *purificare* *bianco* → *biancheggiare*
 - paras. *bello* → *abbellire* *giallo* → *ingiallire*
 - doppio → *sdoppiare*
 - conv. *sano* → *sanare* *scuro* → *scurire*
 - the encoding of **stative/similative** ('be / act Adj') meanings, i.e. implying no change of state, is achieved by **-eggiare** suffixation and conversion:
 - e.g. suff. *folle* → *folleggiare* conv. *paziente* → *pazientare*
 - have, over time, often been employed to form multiple verbs from a same adjectival base: while in some cases they express different meanings, in others they encode (fully or near-to) equivalent semantics, giving rise to **lexeme competition**:
 - e.g. *curvo* → a. suff. *curveggiare* 'to wind' (of rivers, paths, etc)
 - b. conv. *curvare* 'to bend'
 - c. paras. *incurvare* 'to bend'

2. Distributional Semantics

The core assumption of Distributional Semantics is that **semantically similar words show a similar distribution in linguistic contexts** (Harris, 1954). This hypothesis is operationalized by calculating the semantic similarity between two words in terms of **similarity between their vectors**, which are a **statistical representation of the contexts** in which the words appear (Lenci, 2018). To perform our analyses we employ Word2Vec (Mikolov et al., 2013), a widely used neural network model trained on itWaC, a 2-billion-word Italian corpus constructed from the Web (Baroni et al., 2009), and the Alacarte embedding algorithm (Khodak et al., 2018).

4. Data

- The dataset includes **784 verbs** (38.6% parasynthetics, 32.7% suffixed, 28.7% converted)
- collected** from the *Grande Dizionario della Lingua Italiana*, the reference historical dictionary of Italian, and the *Dizionario Italiano Sabatini Coletti*, a dictionary of contemporary usage;
 - occurring** with a frequency > 10 in itWaC (source for frequencies and vectors);
 - classified** based on derivational process and specific pattern, i.e. **i)** parasynthetic combination of a semantically void prefix and an inflectional ending: *ad-*, *in-*, *s-*, *de-*, *di-*, *ri-*, *ra-*, *rin-* Adj *-are*, *-ire* **ii)** suffix for suffixed verbs: *-izzare*, *-eggiare*, *-ificare* **iii)** inflectional ending for converted verbs: *-are*, *-ire*.

process	n_bases	n_types	n_tokens
conversion	215	225	4,420,657
parasynthesis	217	303	2,043,692
-izzare suffixation	195	196	1,405,112
-eggiare suffixation	35	35	27,099
-ificare suffixation	25	25	476,753

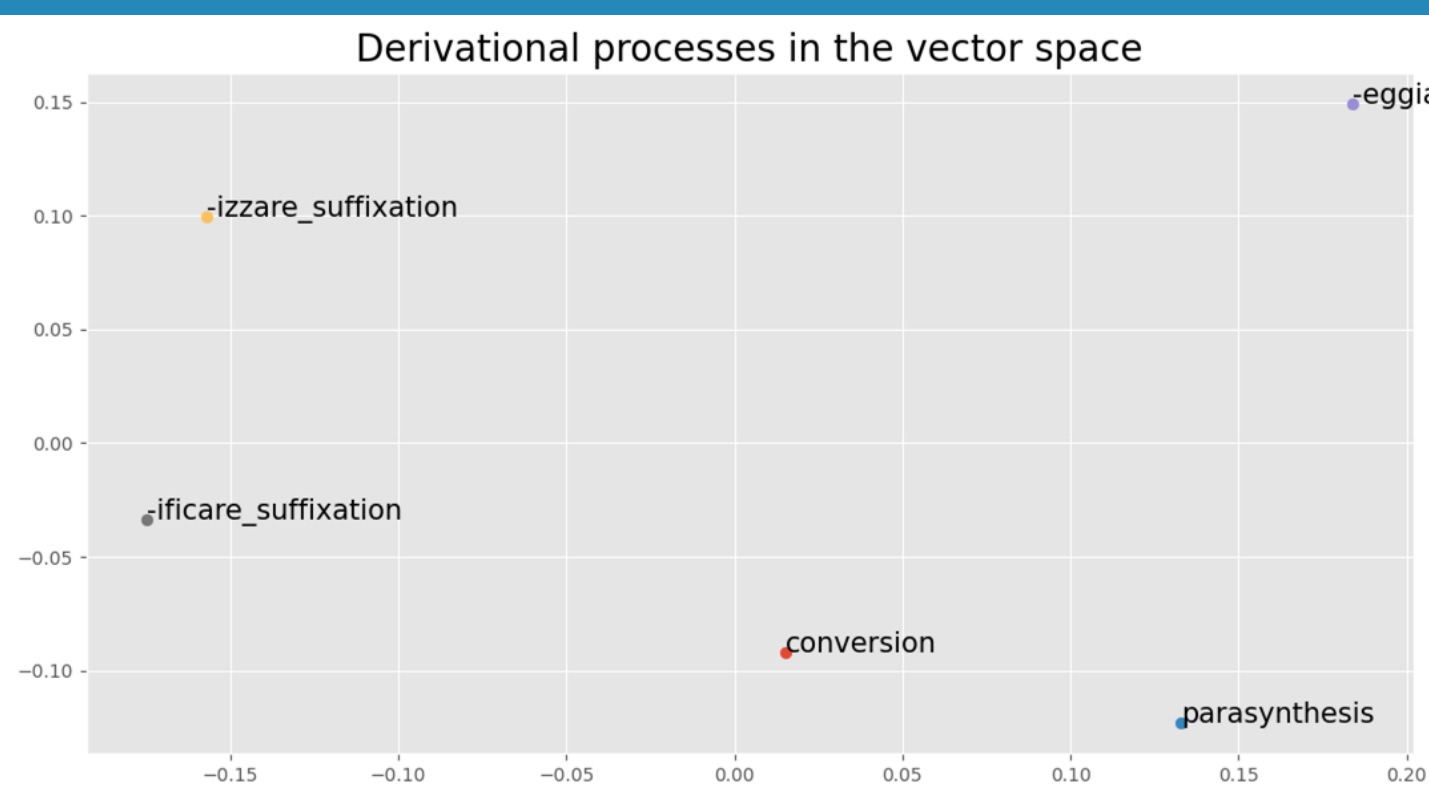
Research questions...

...and how we tackled them

- Which derivational processes express more similar semantics?
- How is semantic similarity between processes reflected into semantic similarity between verbs formed from a same adjective?
- Under which conditions do synonymous co-radical verbs formed by different processes coexist?

- We generate an average vector for each process and compare them with Cosine Similarity to find the most distributionally similar ones (Guzmán Naranjo & Bonami, 2023).
- We generate similarity values between pairs of co-radical verbs. Then we calculate the average of similarities for each pair of processes.
- We analyze the most similar neighbours of co-radical verbs pairs to detect difference in the meanings captured by the DS approach in a finer-grained fashion.

a) Process similarity



The model generally captures similarities between processes that are **consistent with our expectations...**

process_1	process_2	cosine_similarity
parasynthesis	conversion	0.9520631
-ificare suffixation	-izzare suffixation	0.9361985
-ificare suffixation	conversion	0.9207203
parasynthesis	-eggiare suffixation	0.9118946
conversion	-izzare suffixation	0.9070829
conversion	-eggiare suffixation	0.8941306
parasynthesis	-ificare suffixation	0.8764547
parasynthesis	-izzare suffixation	0.8601580
-izzare suffixation	-eggiare suffixation	0.8567521
-ificare suffixation	-eggiare suffixation	0.8323722

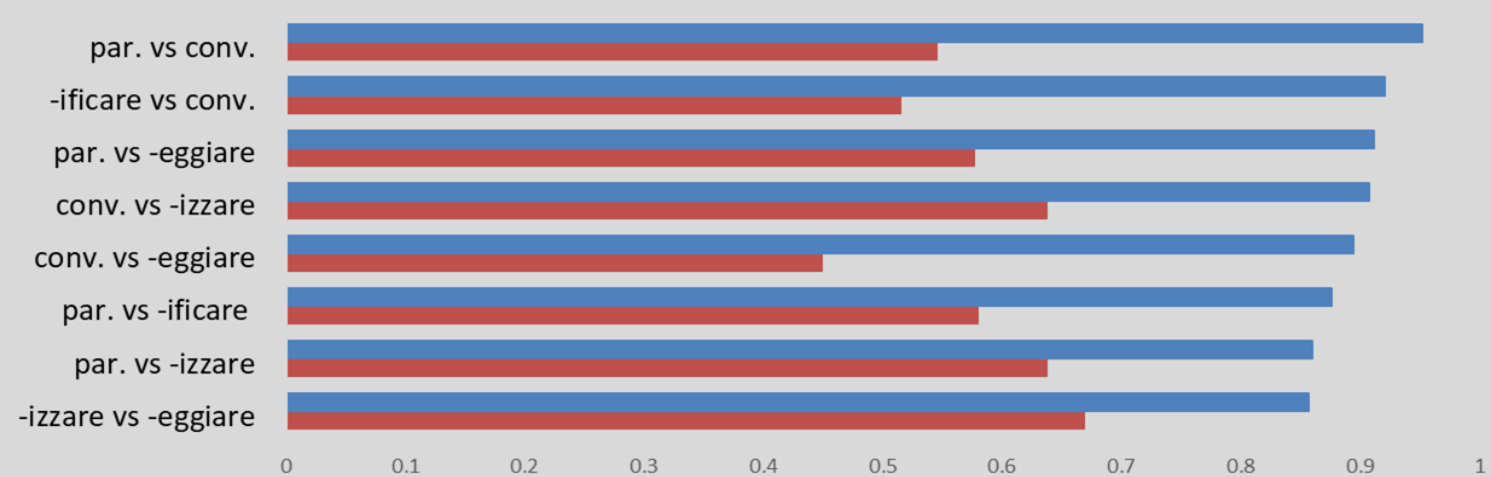
...with **some exceptions** (e.g., parasynthesis and *-eggiare* suffixation have a relatively high cosine similarity)

b) Co-radicals mean similarity

process_1	process_2	mean_cos_similarity	n_pairs
-izzare suffixation	-eggiare suffixation	0.6688296	9
conversion	-izzare suffixation	0.6376260	10
parasynthesis	-izzare suffixation	0.6369260	8
parasynthesis	-ificare suffixation	0.5790315	23
parasynthesis	-eggiare suffixation	0.5766763	30
conversion	parasynthesis	0.5451411	131
conversion	-ificare suffixation	0.5146794	12
conversion	-eggiare suffixation	0.4493575	11

Co-radical **pairs formed by means of the most similar processes** seem to be on average among the **least similar ones** (e.g., see conversion and parasynthesis).

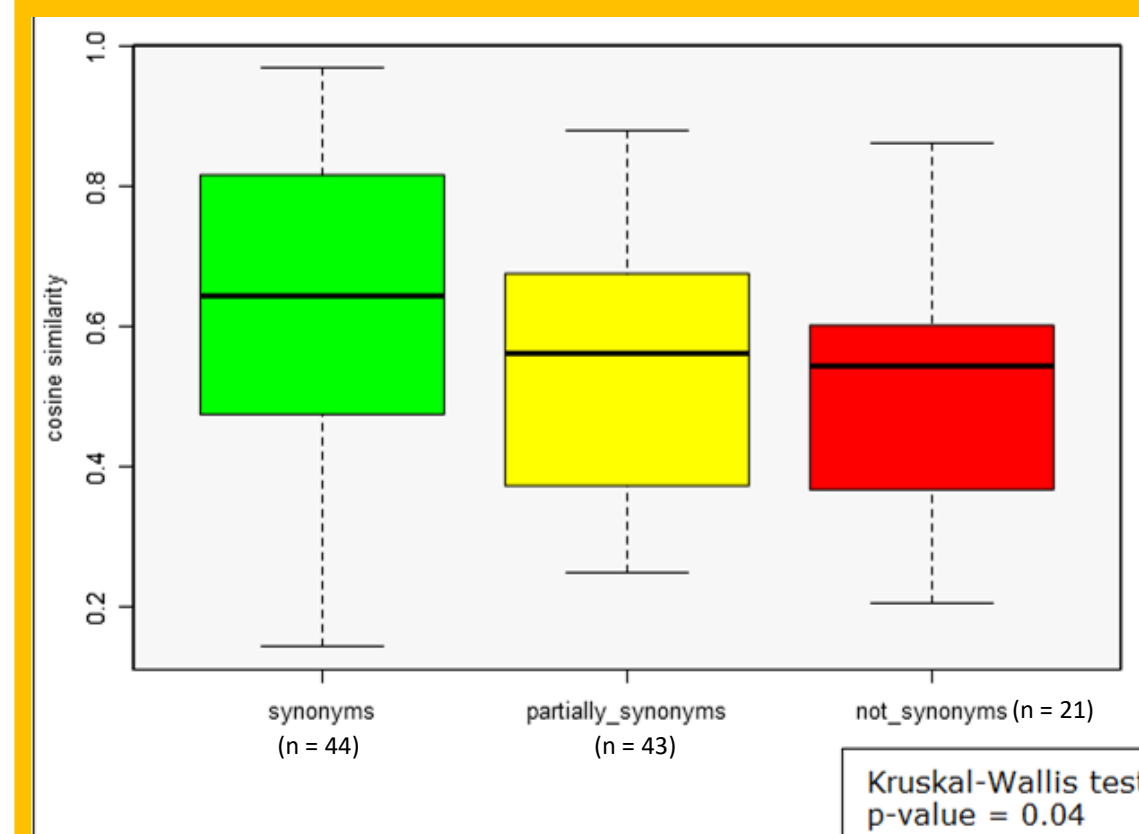
Processes cosine similarity (blue bars) vs Co-radical pairs mean cosine similarity (red bars)



Co-radical similarity: The case of converted and parasynthetic verbs

Through a comparison with dictionary definitions (GRADIT), we **assess the similarity scores** assigned to the converted and parasynthetic co-radical verbs, in order to test the reliability of data shown in b). We assigned three values:

- synonyms: the meanings fully match
- partially syn.: some of the meanings are shared by both verbs, others are specific to each of them
- not syn.: none of the meanings are shared.

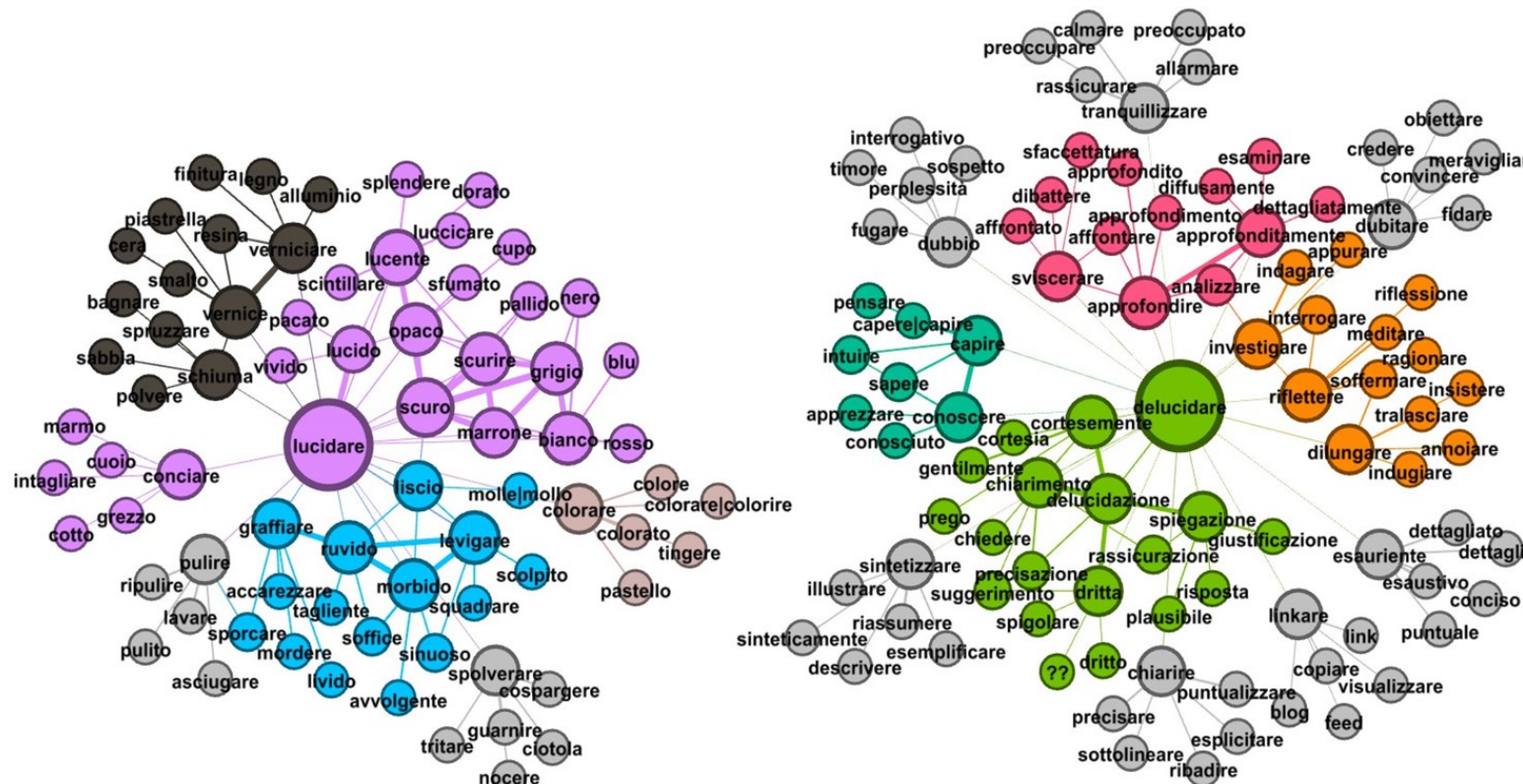


- There is a **significant difference** between the scores of the three groups, suggesting, on average, the **reliability of Word2Vec scores** (despite a large standard deviation).
- Although on average conversion and parasynthesis do not produce very similar pairs, there is **a group of semantically highly similar pairs** - among which, however, there is generally a high **difference in frequency** (median = 185.9%, calculated as absolute percentage difference).

c) Conversion vs parasynthesis: distributional behaviour of co-radicals

To find finer-grained differences between semantically similar co-radical verbs, we can plot their semantic graphs.

Semantic graphs were generated by extracting the **20 most similar neighbours of the two verbs**, as well as the 5 most similar neighbours of the 20 direct neighbours.



lucido 'shiny' 'lucid'
conv. *lucidare* 'to polish'
paras. *delucidare* 'to clarify'
similarity = 0.507414699
freq. difference = 167.72%

Not competing verbs:

- low cosine similarity
- distinct groups of neighbours



quieto 'calm'
→ conv. *quietare*
paras. *acquietare*
'to calm (down)'

similarity = 0.954316914
freq. difference = 7.59%

Competing verbs:

- high cosine similarity
- share many neighbours
- very low frequency difference but
- At least one niche for *quietare*: not only psych states, but also the semantic field of *rebellion, insurrection*, etc.
- Hints of register variation: literary words closer to *quietare*, such as *pascolare* 'to graze, to feed' and *figliuolo* 'son'.