Processing compounds: what frequency (alone) cannot explain

Marcello Ferro[1], Claudia Marzi[1], Vito Pirrelli[1], Christina Gagné[2], Thomas Spalding[2], Marco Marelli[3]

[1]Institute for Computational Linguistics CNR Pisa,
[2]University of Alberta,
[3]University of Milano-Bicocca

Abstract

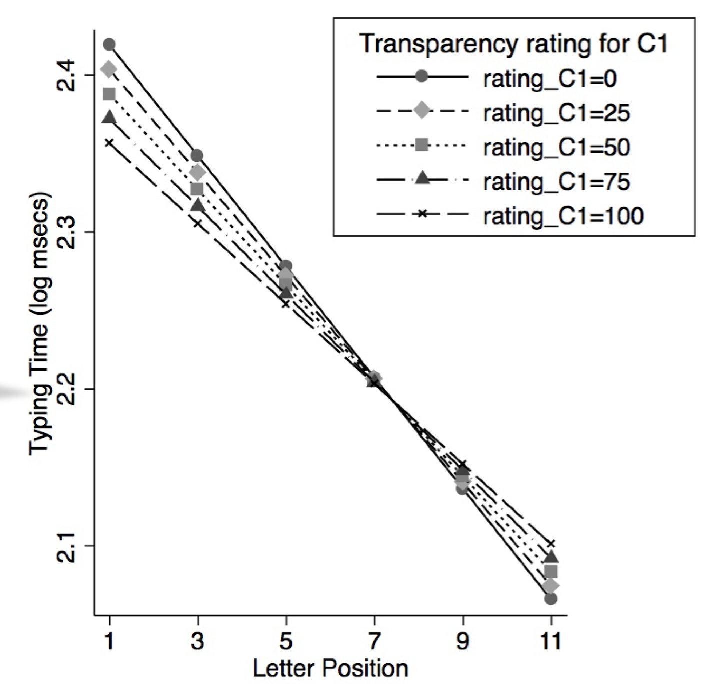
Evidence on **word typing** [3,4] clearly indicates that **morphemic structure is involved in written word production**.

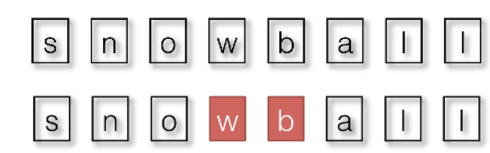
The production of compounds differs from that of monomorphemic words, and the **semantic transparency** of the compound constituents leads to different effects. In particular, elevation in typing latency at the morpheme boundary is larger **when the first constituent is transparent** than when it is **opaque**, but is unaffected by the transparency of the second constituent.

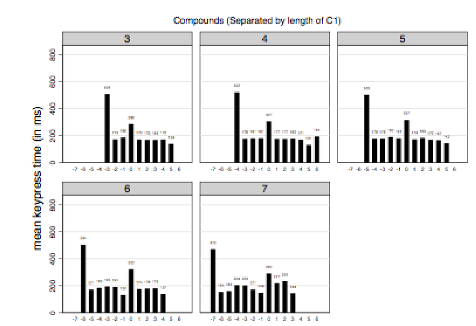
Furthermore, embedded pseudo-morphemes appear to influence the production of pseudo-compounds, but not in the same way that the embedded morphemes affect the production of compounds.

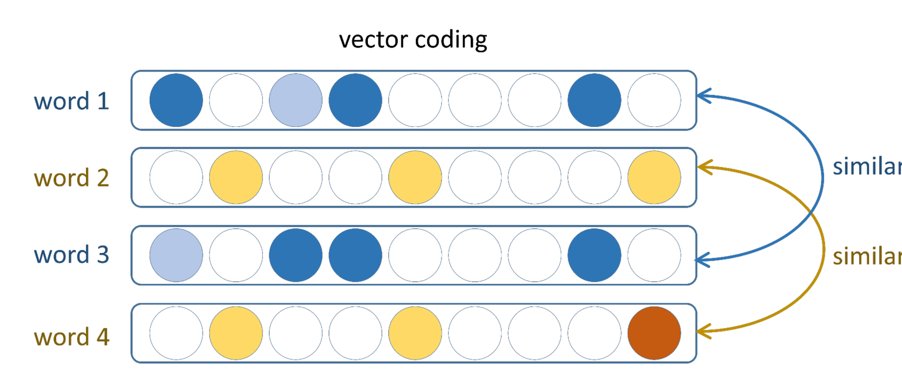
The evidence calls for a **highly interactive processing architecture**, where both compounds and constituents are stored and accessed through **overlaying patterns** of processing units. The level of activation of these patterns is a dynamic function of their degree of time-bound specialization for serial chunking, and their semantic contribution to the interpretation of the whole compound.

Here we show that a **recurrent neural network architecture** dynamically integrating these processing levels can simulate human evidence on compound typing as the outcome of **competing activation patterns**.





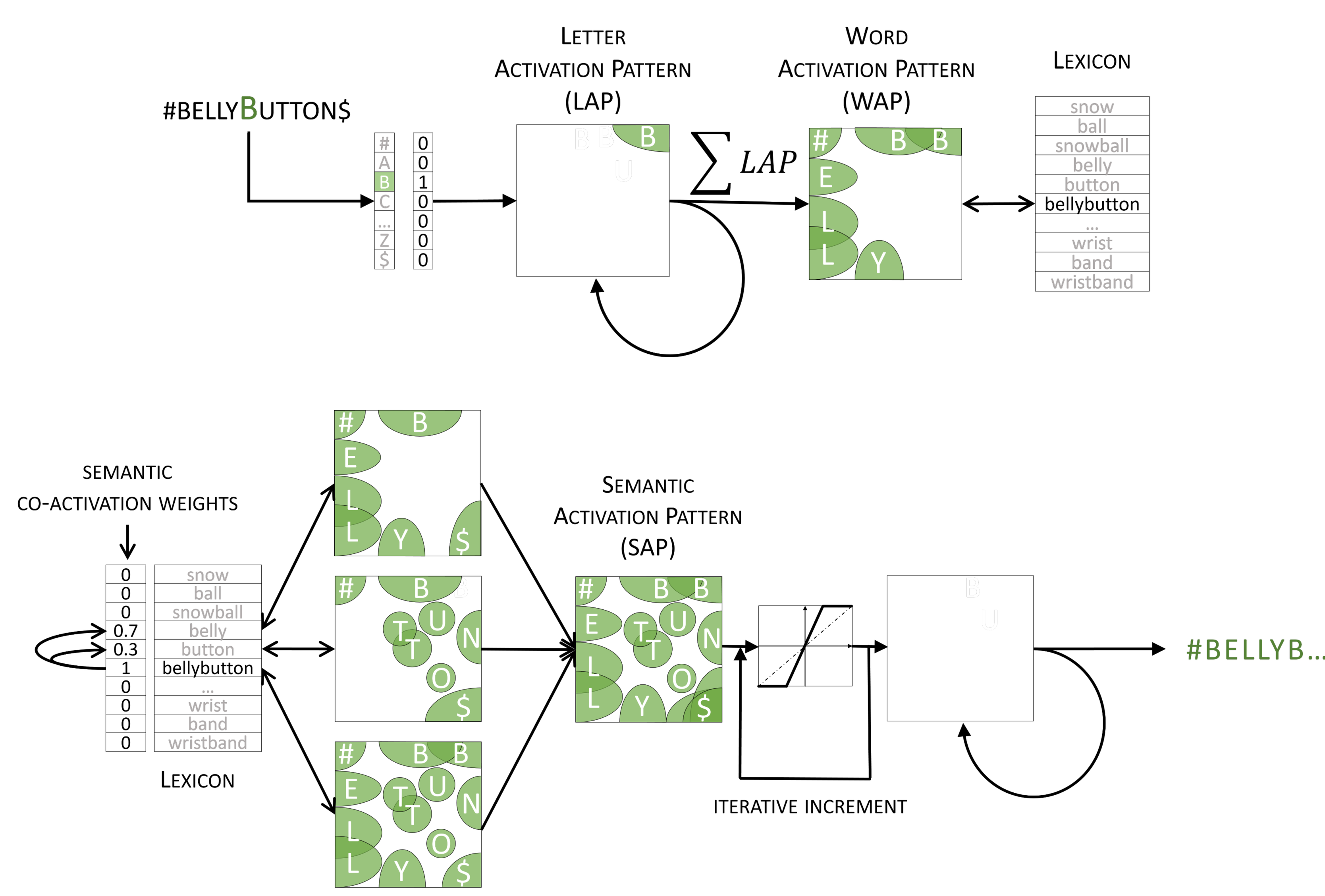




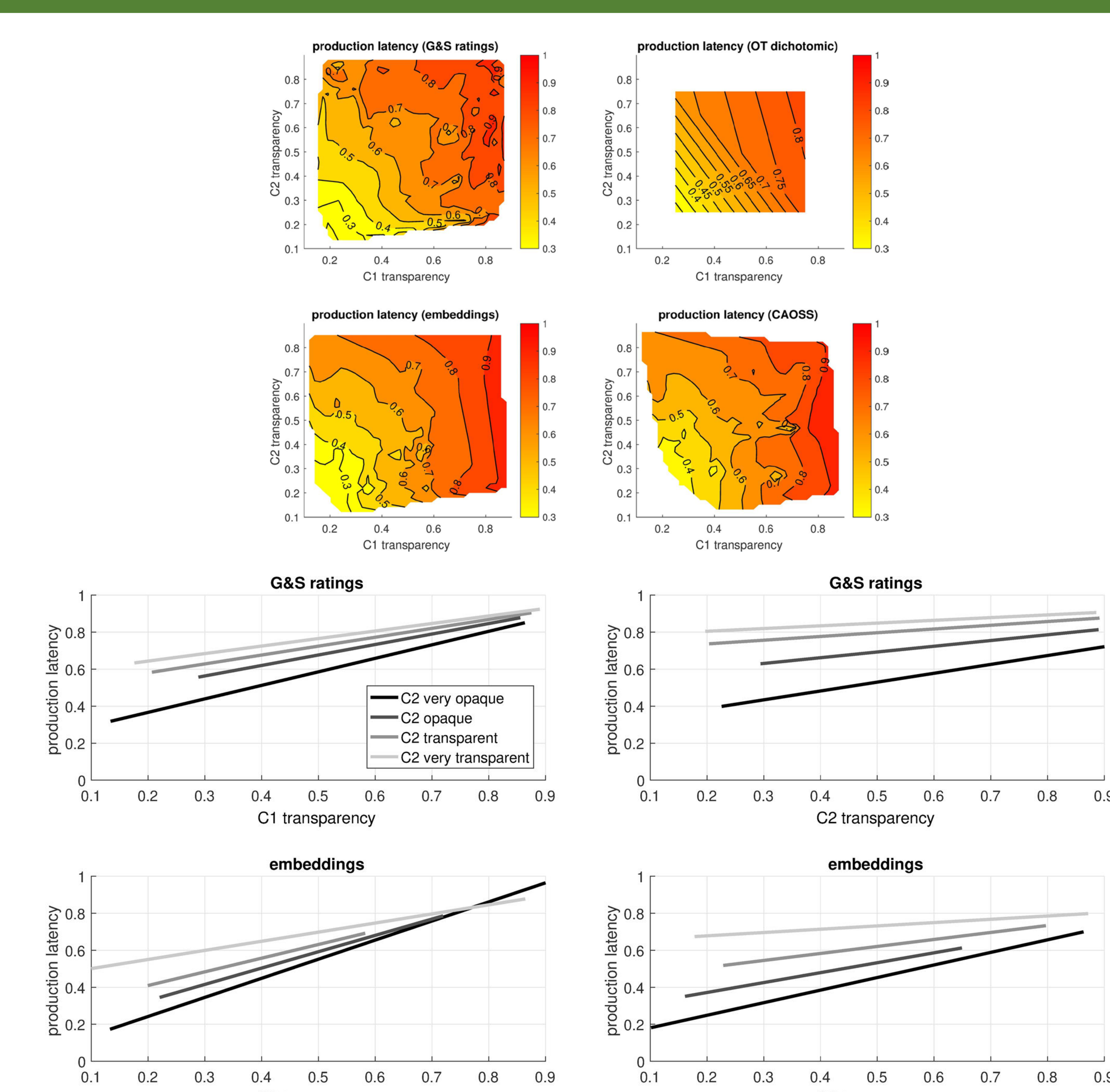
Method

- a **Temporal Self-organizing Map** (TSOM) [2,7,9] is trained on 136 compounds, from G&S original data set [3]
- for each compound *AB* (*e.g. snow ball*), $\overrightarrow{AB}$, $\vec{A}$ and $\vec{B}$ are the **vector-coded representations** of the meanings of *AB*, *A* and *B* respectively, based on corpus-driven **word embeddings** [8]
- for any $\overrightarrow{AB}$, its cosine distance from $\vec{A}$ and $\vec{B}$ is used to estimate the amount of **co-activation** of $\vec{A}$ and $\vec{B}$ in the TSOM **Semantic Activation Pattern**, when $\overrightarrow{AB}$ is input to the map
- the TSOM is then tested on the task of producing each compound from its **Semantic Activation Pattern** over twenty iterations, uniformly increasing per-node activation levels by small increments at each iteration
- per-letter accuracy in the task is monitored, timing the letter production latency by the number of iterations needed for each letter to be produced correctly
- for checking robustness of the effect, we repeated the same experiment with different proxies for semantic co-activation: dichotomous classification of compounds (opaque vs. transparent) [5], human transparency ratings [3], vector codings automatically generated by the CAOSS model [6].

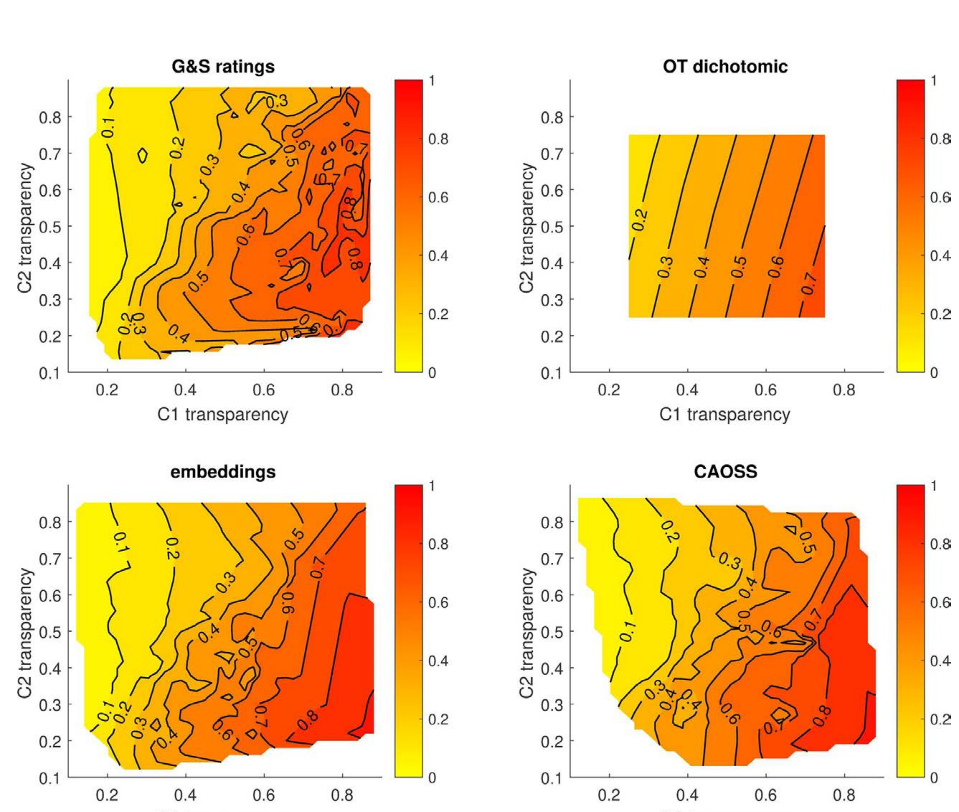
TSOM (word learning & production)

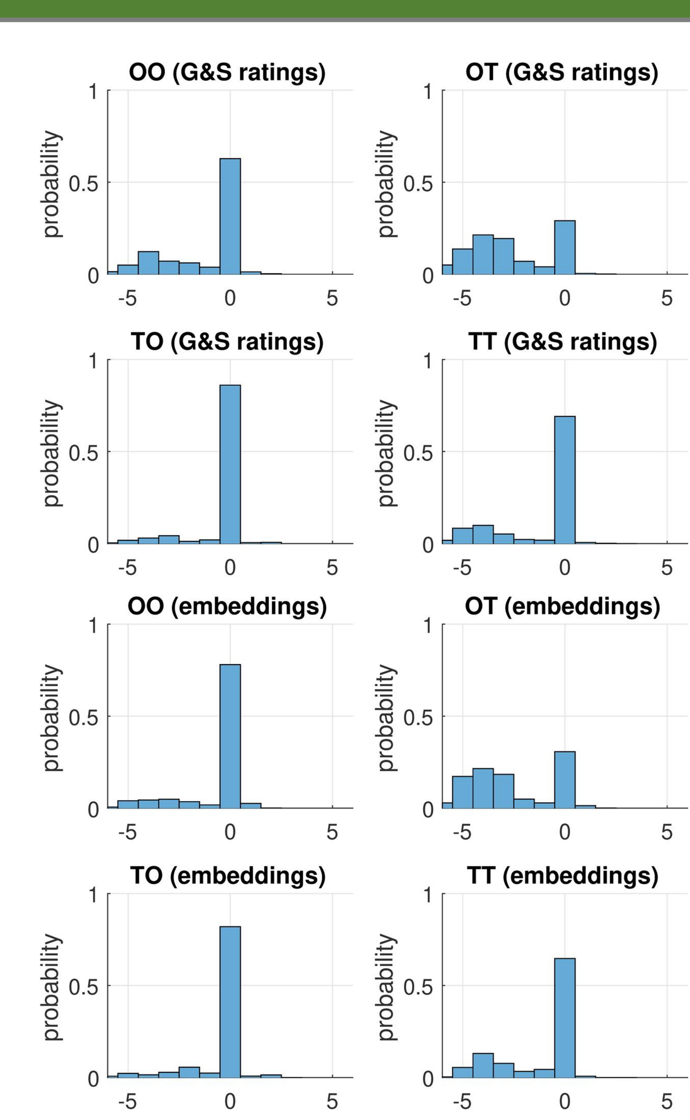


word production latency



letter production latency by transparency + error distribution by letter position





Findings

- in previous work [2], **TSOMs** were shown to reproduce effects of structural discontinuity in the production of **compounds** (as opposed to **mono-morphemic words**), based on the frequency-driven specialization of re-entrant temporal connections. No effects of semantic transparency were simulated
- here, semantic information (word embeddings) were integrated into the map architecture to simulate **gradient co-activation** of the meaning representations of constituents when more or less semantically transparent compounds are generated from their lexical nodes through Semantic Activation Patterns
- semantic transparency is shown to make **compound production slower**, particularly at the **constituent boundary**
- **increasing** semantic transparency of **both constituents** causes **diminishing** accuracy in **compound production**, with transparency of C1 playing a **more prominent** role than transparency of C2
- **letter production latency** increases at the **constituent boundary**, where a slowdown appears to be caused by the transparency of **C1**, **not** by the transparency of **C2**
- the effect was **consistently simulated** using **different semantic transparency scores** between constituents and compounds, based on word embeddings, transparency ratings, compositional word embeddings and dichotomous classifications, thereby providing indirect support to compositional perspectives on distributional semantics [1,6]
- linear models of letter production latency show that frequency effects are cancelled out by semantic effects, reproducing evidence of human typing

Essential references

1. Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, 122(3): 485-515.
2. Ferro, M., Cardillo, F.A., Pirrelli, V., Gagné, C.L., & Spalding, T.L. (2016). Written word production and lexical self-organisation: evidence from English (pseudo)compounds. In CLiC-it & EVALITA 2016, Ceur Workshop Proceedings, 1749 (25), Aachen, Germany.
3. Gagné, C.L., & Spalding, T.L. (2016). Effects of morphology and semantic transparency on typing latencies in English compound and pseudo-compound words. Journal of Experimental Psychology: Learning, Memory, and Cognition, 42: 1489-1495.
4. Gagné, C. L., & Spalding, T. L. (2014). Typing time as an index of morphological and semantic effects during English compound processing. Lingue e Linguaggio, XIII(2), 241-262.
5. Libben, G. (1998). Semantic Transparency in the Processing of compounds. Consequences for representation, processing and impairments. *Brain and Language*, 61: 30-44.
6. Marelli, M., Gagné, C.L., & Spalding, T.L. (2017) Compounding as Abstract Operation in Semantic Space: Investigating relational effects through a large-scale, data-driven computational model. Cognition, 166: 207-224.
7. Marzi, C., Ferro, M., Cardillo, F.A., & Pirrelli, V. (2016). Effects of frequency and regularity in an integrative model of word storage and processing, Italian Journal of Linguistics, 28 (1): 79-114.
8. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint ArXiv:1301.3781.
9. Pirrelli, V., Ferro, M., & Marzi, C. (2015). Computa-tional complexity of abstractive morphology, In Baerman M., Brown D., Corbett G. (Eds.) Under-standing and Measuring Morphological Complex-ity. 141-166, Oxford, United Kingdom: Oxford University Press.

www.comphyslab.eu