Towards a multiplex network model of word associations and similarity in the human mind

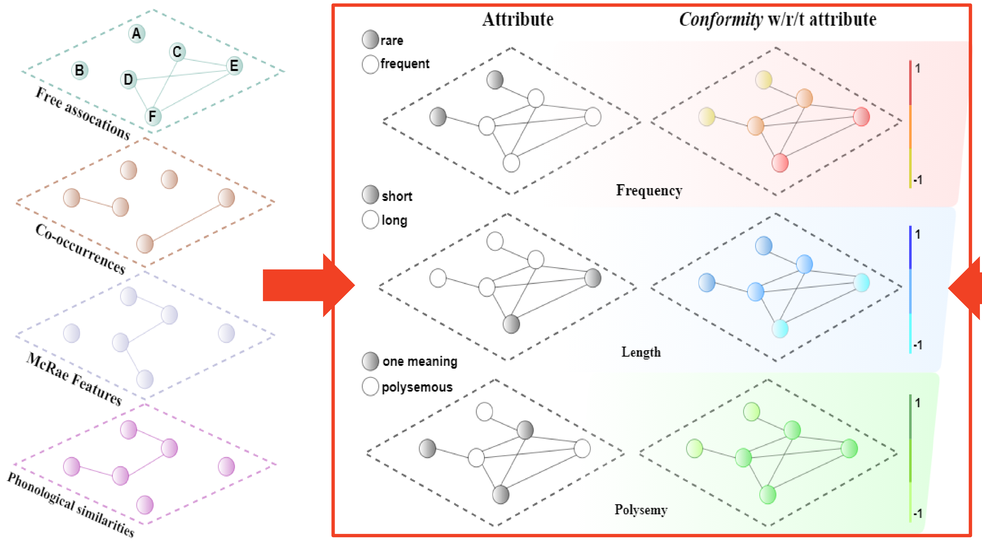
Salvatore Citraro [1, 3], Massimo Stella [2] and Giulio Rossetti [3]

1 Department of Computer Science, University of Pisa, Bruno Largo Pontecorvo, 3, Pisa, Italy
2 Department of Computer Science, University of Exeter, EX4 4QF Exeter, UK
3 KDD Lab. ISTI-CNR, via G. Moruzzi, 1, Pisa, Italy

Aims

Early language acquisition is a cognitive process mediating the learning of words according to heterogenous linguistic knowledge, e.g. semantics and phonology.

Even if **cognitive networks** are insightful for understanding how conceptual associations influence word acquisition, current approaches do not account for the interplay between structure and **word features**, i.e. exploitable node metadata.



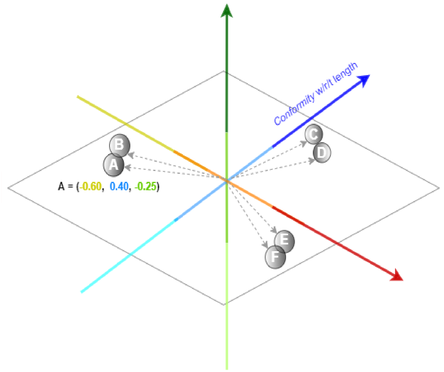
In this work, we **aim to investigate word learning** by merging relational structure and word features, like frequency, length and polysemy, in a **multilayer representation enhanced with vector spaces**.

Methodology

We reconstruct a multidimensional representation of **English toddlers' mental lexicon [1]**, building three *structural layers* – capturing free associations, feature sharing, word co-occurrence in child-directed speech and phonological similarities.

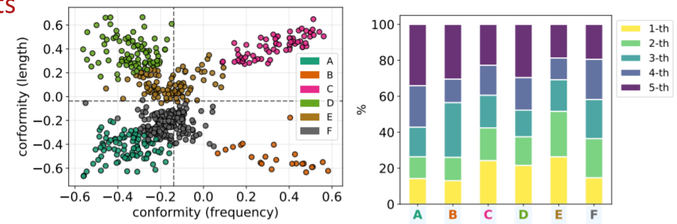
As reported in the Figure below, **layers are enhanced with similarity patterns** from a vectorial space, including **features** like **length, frequency** and **polysemy**.

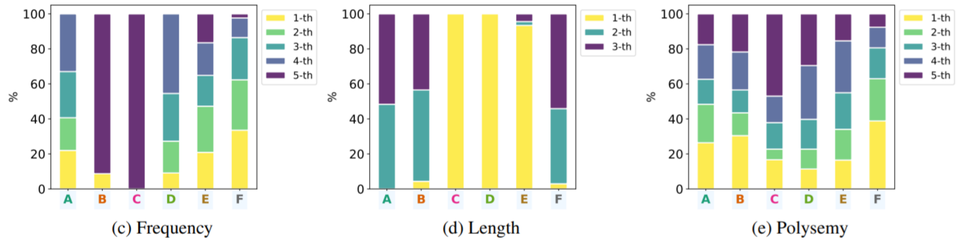
The enhanced multilayer network combines conceptual patterns between words either structural (i.e. connections) or similarity-based (i.e. possessing similar features).



By representing a word as a vector of its ***conformity* score [2]** w/r/t each attribute, we can identify **clusters** of words, differing in their tendencies to connect structurally with other nodes sharing similar features.

Results





Conformity vectors are reported above (top left).

They can identify six clusters (*k* chosen according to the elbow method). We evaluate them in terms of attribute values distribution (labels discretized by quantiles).

Structural+similarity patterns as assessed via conformity **unveil a cluster** of words of **short length** (d), **high frequency** (c) **and several meanings** (e) emerging early during cognitive development..

Conclusion

By building network layers spanning vector spaces across word metadata, we can identify a potential language kernel otherwise undetectable with viability.
Can we better characterize it in terms of cognitive interpretations or w/r/t semantic classes?

References

[1] Stella, Massimo, Nicole M. Beckage, and Markus Brede. "Multiplex lexical networks reveal patterns in early word acquisition in children." *Scientific reports* 7.1 (2017): 1-10.

[2] Giulio, Rossetti, Salvatore, Citraro, and Letizia Milli. "Conformity: a Path-Aware Homophily measure for Node-Attributed Networks." *IEEE Intelligent Systems* (2021).

Contacts

- salvatore.citraro@phd.unipi.it
- m.stella@exeter.ac.uk
- giulio.rossetti@isti.cnr.it

ISTITUTO DI SCIENZA E TECNOLOGIE DELL'INFORMAZIONE "A. FAEDO"