

Illés Anna Noémi

**ELTE Tehetséggondozási Tanács Utazási Támogatás – Beszámoló a Konferencia
Részvételről**

2026. május 21-24.

A pályázott tevékenység keretében a 2026. május 21-24. között Horvátországban megrendezésre került 17th Dubrovnik Conference on Cognitive Science (DUCOG) nemzetközi konferencián vettem részt. Az eseményt minden évben a Central European Cognitive Science Association (CECOG) szervezi a Centre for Advanced Studies intézményében, célja pedig a kognitív tudományok területén dolgozó kutatók és hallgatók közötti szakmai együttműködés elősegítése és megerősítése Közép-Európa térségében. A DUCOG idei témája az adaptáció („*Adaptations across different timescales*”), vagyis a környezeti ingerekhez való alkalmazkodás képessége volt, mely ugyan távolabbról kapcsolódik fő érdeklődési területemhez, az előadások változatossága számomra is rengeteg új információt hordozott és elősegítették ismereteim gazdagítását. Különösen érdekesnek találtam Kevin Mitchell előadását a tanulás generatív modelljéről, valamint Oláh Katalin előadását a mentalizáció és a felnőtt kötődési mintázatok kapcsolatáról – az utóbbi aspektus neurális sajátosságait különösen izgalmasnak és inspirálónak találtam és elgondolkodtatott arról, hogy szívesen vizsgálnám ezt a területet a jövőben.

Az idei rendezvényen poszter előadóként vettem részt a „*Comparison of Semantic Organization in a Concept Feature Norm of 6–9-Year-Old Children with Thematic and Taxonomic Language Models*” című angol nyelvű munkámmal, melyet alább csatoltam. A kutatásunk átfogó célja egy olyan fogalmi jellemző-alapú norma létrehozása, amely 6-9 éves gyermekek verbális leírásai alapján modellezi a fogalmi tudás szerveződését a felsorolt jellemzők közötti átfedések mintázata mentén, poszteremen keresztül pedig ezen projekt fő módszerspecifikus kérdéseit szerettem volna bemutatni és szakmai közönség elé tárni. A poszter elkészítésének folyamata rendkívül tanulságos volt számomra, ugyanis a vizualizációk és elemzések elkészítésén keresztül programozási készségeimet is lehetőségem volt gyakorolni. Az elkészült munka rövid, gördülékeny ismertetése angol nyelven idegen nyelvtudásomat is fejlesztette, ezen kívül segített abban, hogy magamban letisztázzam a kutatás fő pontjait, majd azokat az érdeklődő közönség ismereteihez igazítva hatékonyan át tudjam adni.

Ugyan poszteremre kevés olyan visszajelzést kaptam, melyeket integrálni tudunk kutatásunk módszertani eljárásának finomítása során, a konferencián résztvevőkkel rendkívül hasznos szakmai párbeszédet tudtam folytatni mind saját poszterem bemutatása, mind mások posztereinek áttekintése során. Lehetőségem volt új módszertani paradigmákat megismerni az előadókkal folytatott diskurzusokon keresztül, melyekre úgy vélem, én is értékes visszajelzéseket tudtam adni. Ezen párbeszédok segítettek abban, hogy egy olyan kutatói szemléletmód alapjait sajátítsam el, melyeken keresztül képessé válhatok a kutatások kritikus elemzésére saját szakmai tapasztalataimon keresztül. Ebben kiemelt jelentőségű volt a konferencia kis létszáma, melynek köszönhetően egy olyan légkör tudott kialakulni, ahol bátran

meg tudtam osztani saját ötleteimet, valamint tudtam kérdezni és visszajelzést adni a jelen lévő hallgatóknak és kutatóknak.

Úgy vélem, hogy az esemény ideális terepet jelentett a tudományos életbe való betekintésre, valamint személyes motivációm is nagyban erősítették a létrejött szakmai ismeretségek és párbeszéddek. Nagyon hálás vagyok, amiért lehetőségem volt egy ilyen jellegű szakmai eseményen részt venni, bízom benne, hogy a továbbiakban is lehetőségem lesz kutatásunkat nemzetközi konferencián bemutatni.

Budapest, 2026.06.02.



Illés Anna Noémi



Comparison of Semantic Organization in a Concept Feature Norm of 6–9-Year-Old Children with Thematic and Taxonomic Language Models



Anna Noémi Illés^{1,2}, Alex Ilyés^{1,3,4}, Borbála Brosig^{1,6}, Anna Bródy^{1,2}, Rebecca McFegan^{1,2}, Mónika Egyed^{1,2}, Szonja Szabó^{1,2}, Donát Keszthelyi^{1,2}, Éva Kovaliczky², Boróka Ferencz^{1,3,4}, Attila Keresztes^{1,4}

1. Brain Imaging Centre, HUN-REN Research Centre for Natural Sciences, Budapest, Hungary; 2. Faculty of Education and Psychology, ELTE Eötvös Loránd University, Budapest, Hungary; 3. Doctoral School of Psychology, ELTE Eötvös Loránd University, Hungary; 4. Institute of Psychology, ELTE Eötvös Loránd University, Hungary; 5. Department of Economics, Budapest University of Technology and Economics, Budapest, Hungary; 6. Semmelweis University, Doctoral School of Mental Health Sciences and Psychiatry, Budapest, Hungary

Contact info: illes.anna@ttk.hu

Introduction

- Concepts are mental representations defined by observable and latent semantic features that enable categorization [1, 2].
- In adults, concepts are suggested to be organized taxonomically based on pre-existing categories for efficient retrieval (Figure 1) [3].
- In contrast, in children before the age of seven, concepts are suggested to be organized primarily based on thematic relations defining situational co-occurrence (Figure 2) [4, 5].
- Eliciting verbal descriptions of concepts from participants can be quantified into semantic feature norms to model inter-concept relations [6, 7].
- Our goal is to create a concept feature norm based on Hungarian children producing concept definitions.

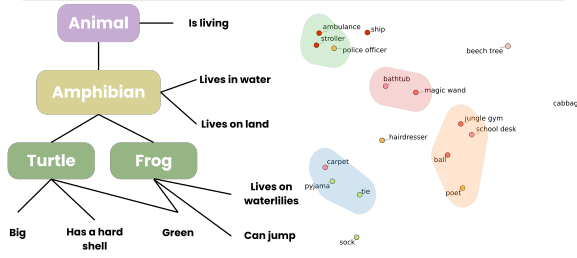


Figure 1. Taxonomic organization of 20 random concepts and their semantic features in our data. General attributes shared across subordinate concepts are stored at the superordinate level (e.g., “amphibian—lives in water”), while unique features are represented at the subordinate level (e.g., “turtle—has a hard shell”).

Preprocessing

- We first extracted tokens from utterances via the Hungarian model of *HuSpaCy* [11] and retained tokens with content words (verbs, nouns, adverbs, adjectives, numerals, adpositions, and postpositions) filtering minimal responses (e.g., “yes” or “no”), comments (e.g., “I know”), and naming the concepts themselves.
- Tokens were lemmatized and compiled into feature segments (e.g., “have_four_leg”, “live_water”), retaining specific prefixes (e.g., negation as “NEG”, “it has” as “HAS”).
- Punctuations, conjunctions, stopwords and low-information tokens (e.g., “the”) were either used as segment boundaries or discarded (Figure 6).
- FastText vectors [12] were averaged across lemmas within each segment and submitted to k-means clustering, producing 300-dimensional clusters (Figure 12) representing the proportional feature loading of each concept.

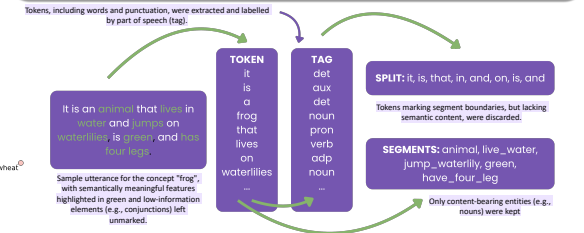


Figure 6. Feature extraction procedure through an example in our data.

Questions

Does comparing verbal definitions of concepts by 6-9-year-olds to large language corpora reveal an underlying thematic representational structure?

H1: When modelling concept representations based on co-occurring verbal feature descriptions by 6-9-year-olds, we expect that the pairwise similarities will correlate positively with pairwise similarities based on a word-embedding derived from word-to-word co-occurrences, and not with pairwise similarities based on a priori encyclopaedic categories derived from WordNet.

Furthermore, we aim to develop a methodological framework for extracting semantic features and mapping children’s conceptual representations from verbal descriptions.

Results

- We constructed representational dissimilarity matrices (RDM) via pairwise Pearson correlations between concept feature vectors, pairwise category agreements in the a priori taxonomic model, and pairwise similarities between word2vec vectors reflecting a thematic model (Figure 7-9). Then, we used Spearman correlation to compare similarities between the model RDMs and our data RDM.
- Comparing the RDM of our data with the taxonomic RDM yielded a significant positive result ($r = 0.18, p < .001$) (Figure 10).
- Comparing the RDM of our data with the thematic RDM yielded a significant positive result as well ($r = 0.23, p < .001$) (Figure 11).
- While both thematic and taxonomic structures positively correlated with our dataset, thematic structure showed stronger alignment. Weak correlations with both RDMs, however, indicate limited feature overlap across conceptual categories.

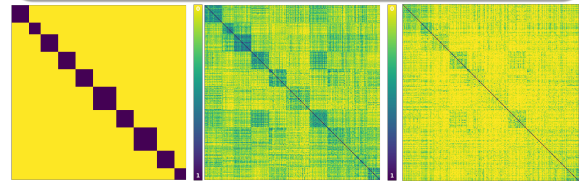


Figure 7-9. Representational Dissimilarity Matrices (RDMs) comparing pairwise similarities between: (1) the a priori taxonomic structure of 300 concepts derived from WordNet (2) a word2vec-derived thematic structure representing distance in vector-space of the 300 concepts (3) semantic similarity between concept vectors based on features extracted from children’s verbal responses.

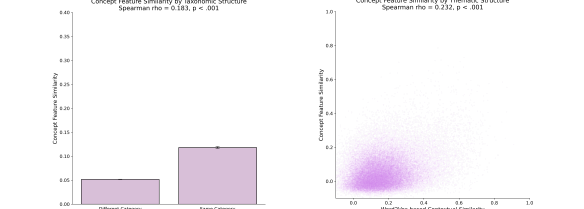


Figure 10-11. Pairwise correlation plots contrasting the dataset against a priori taxonomic and thematic structures. (1) Bar plot: within—(e.g., dog-duck) versus cross-category (e.g., dog-apple) similarity. (2) Scatter plot: contextual co-occurrence (x-axis) against shared semantic features per concept pair (y-axis).

Methods

- We selected 300 Hungarian concepts denoting common, familiar concepts from 10 a priori taxonomic categories (Figure 5.) based on the WordNet lexical database [8].
- Subsequently, we used a Hungarian word2vec word-embedding [9] to extract 200-dimensional vectors, with vector similarities representing thematic relations.
- We recruited 245 Hungarian children (130 females, $M_{age} = 7.76, SD_{age} = 1.46$) from fifteen elementary schools in Budapest and its agglomeration (Figure 4.), asking them to teach concept definitions to ‘Mimó’, an imaginary character, via a web-based application (Figures 3-4).
- We collected an average of 7.4 responses per concept ($SD = 2.4$, range = 1-25), with participants generating 9.54 features per concept on average ($SD = 2.58$, range = 4.5-17.0).
- Voice recordings were manually transcribed via a reliable protocol by 15 transcribers (mean pairwise Cohen’s kappa = .94, $SD = .02$, all above .88).
- Inter-rater reliability was assessed at the word level via a modified Word Error Rate approach [10], measuring pairwise comparisons between transcribers.



Figure 3. “Mimó”, a magical creature from a distant universe, is the central character of the study’s cover story, inviting children to assist in their quest to learn about the world.

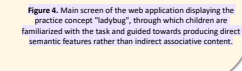


Figure 4. Main screen of the web application displaying the practice concept “ladybug”, through which children are familiarized with the task and guided towards producing direct semantic features rather than indirect associative content.



Figure 5. Sample concepts across the 10 taxonomic categories.

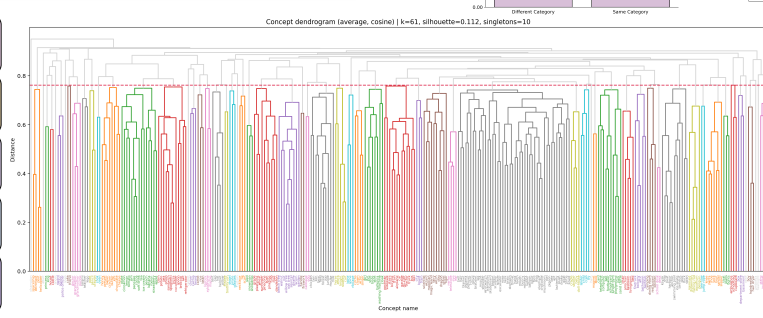


Figure 12. Hierarchical agglomerative clustering of concepts based on shared feature segments, with lower values indicating greater similarity. Colors represent 61 independent clusters (below red line). A silhouette score of 0.112 indicates fuzzy cluster boundaries, with singletons reflecting single-concept clusters.

Outstanding Questions

- How to optimize extraction of semantic features from free verbal productions?
- What are the inferential limits of children’s verbal descriptions of concepts?
 - E.g., do negations signal the absence of a feature, or can they be interpreted as implying a feature of opposite meaning?
- How to optimize clustering via quantified feature relations?
 - E.g., k-means, DBSCAN, agglomerative clustering etc.

Acknowledgments

The conference participation was supported by the Talent Support Council of ELTE and the Faculty Scholarship Committee of ELTE PPK.

References

[1] Young, D. P., & Yip, S. (2019). Semantic feature production norms for a large set of English and Spanish. *Behavior Research Methods*, 51(1), 187–206. <https://doi.org/10.3758/s13428-018-0811-3>

[2] Young, D. P., & Yip, S. (2019). Semantic feature production norms for a large set of English and Spanish. *Behavior Research Methods*, 51(1), 187–206. <https://doi.org/10.3758/s13428-018-0811-3>

[3] Goldstone, R. L., & Shiffrin, R. (1991). The automaticity of expertise: A case study of the development of expertise in chess. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(1), 1–16. <https://doi.org/10.1037/0278-7393.17.1.1>

[4] Goldstone, R. L., & Shiffrin, R. (1991). The automaticity of expertise: A case study of the development of expertise in chess. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(1), 1–16. <https://doi.org/10.1037/0278-7393.17.1.1>

[5] Goldstone, R. L., & Shiffrin, R. (1991). The automaticity of expertise: A case study of the development of expertise in chess. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(1), 1–16. <https://doi.org/10.1037/0278-7393.17.1.1>

[6] Goldstone, R. L., & Shiffrin, R. (1991). The automaticity of expertise: A case study of the development of expertise in chess. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(1), 1–16. <https://doi.org/10.1037/0278-7393.17.1.1>

[7] Goldstone, R. L., & Shiffrin, R. (1991). The automaticity of expertise: A case study of the development of expertise in chess. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(1), 1–16. <https://doi.org/10.1037/0278-7393.17.1.1>

[8] WordNet. (2010). Princeton University. <http://wordnet.princeton.edu/>

[9] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality (arXiv:1310.3745v2 [cs.LG]).

[10] Goldstone, R. L., & Shiffrin, R. (1991). The automaticity of expertise: A case study of the development of expertise in chess. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(1), 1–16. <https://doi.org/10.1037/0278-7393.17.1.1>

[11] Lipton, P., & Bell, S. (2011). The Neural Language Toolkit: Proceedings of the ACL-2011 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computing. https://doi.org/10.1142/9789811111111_0011

[12] Mikolov, T., Sutskever, I., & Dejean, A. (2013). Distributed Representations: Best Practices with Word2Vec. *EMNLP*, 1541–1546. https://doi.org/10.1142/9789811111111_0011

[13] Mikolov, T., Sutskever, I., & Dejean, A. (2013). Distributed Representations: Best Practices with Word2Vec. *EMNLP*, 1541–1546. https://doi.org/10.1142/9789811111111_0011

[14] Mikolov, T., Sutskever, I., & Dejean, A. (2013). Distributed Representations: Best Practices with Word2Vec. *EMNLP*, 1541–1546. https://doi.org/10.1142/9789811111111_0011

[15] Mikolov, T., Sutskever, I., & Dejean, A. (2013). Distributed Representations: Best Practices with Word2Vec. *EMNLP*, 1541–1546. https://doi.org/10.1142/9789811111111_0011