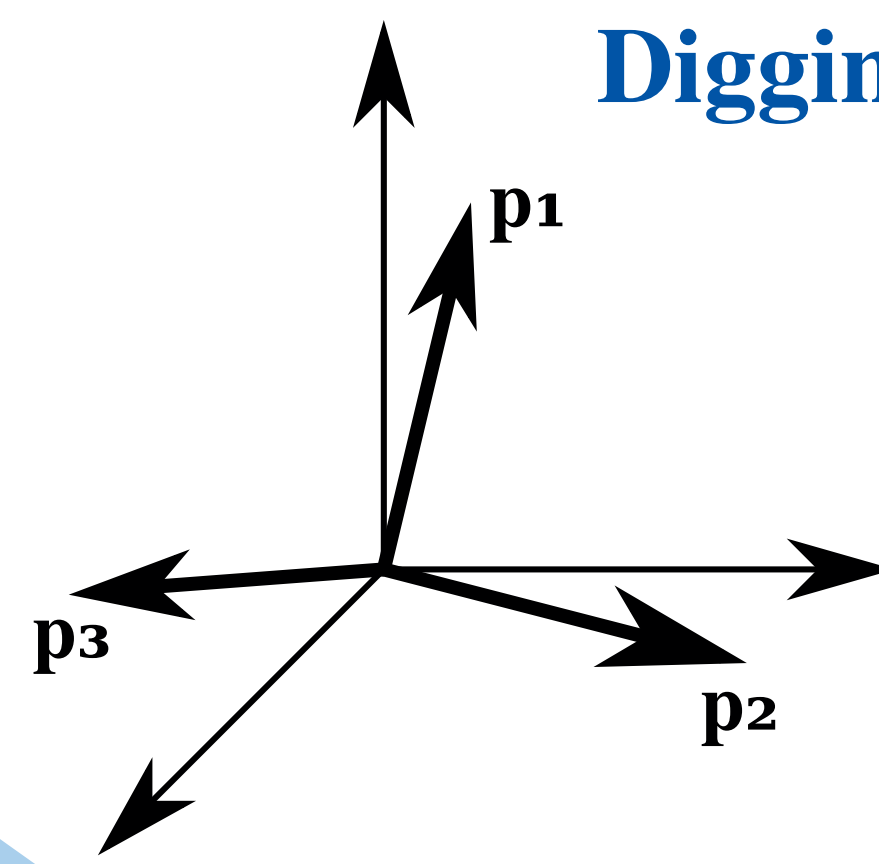


# Semantically Coherent Vector Space Representations



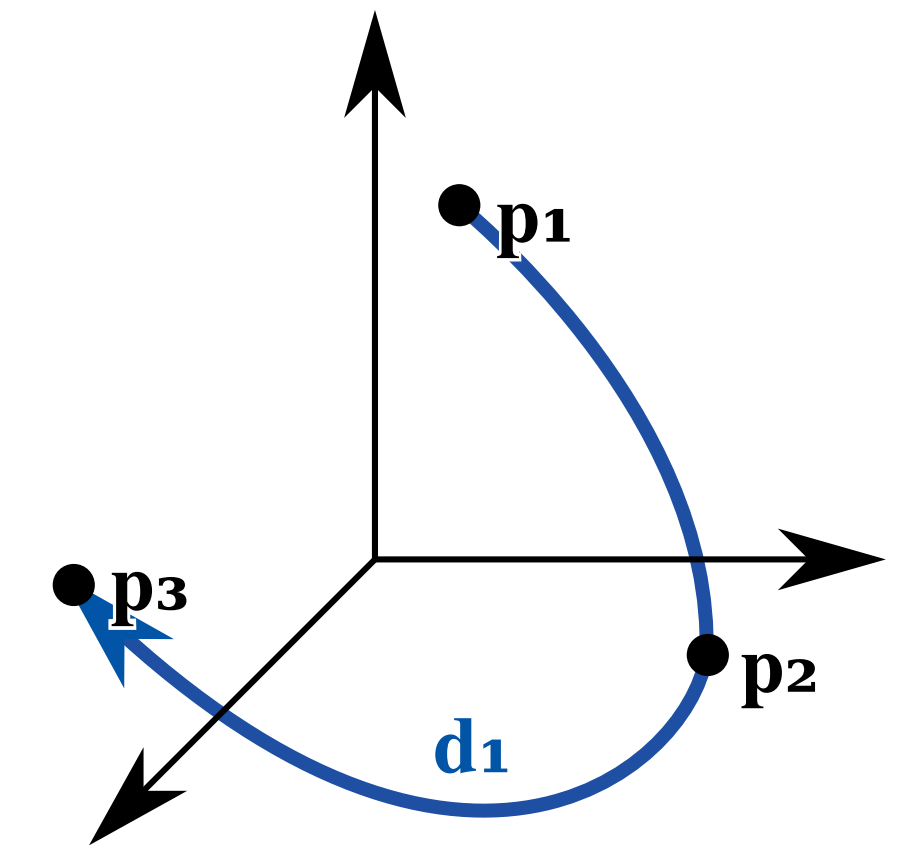
Petr Sojka <sojka@fi.muni.cz>, Vít Novotný <witiko@mail.muni.cz>

MIR research group (mir.fi.muni.cz), Faculty of Informatics, Masaryk University, Brno, Czechia



## Digging for Nuggets of Wisdom in Text [1]

<b>d1</b>	<b>p1</b> Down o'er glaciers, landslips, scars, down the toppling grey moraines,
	<b>p2</b> higher than a big scythe blade and a pale lavender above the dark blue water.
	<b>p3</b> It was marvellous, a feast for the eyes, this complication of coloured tints.



## Trains of Thought: Narrative Similarity [2]

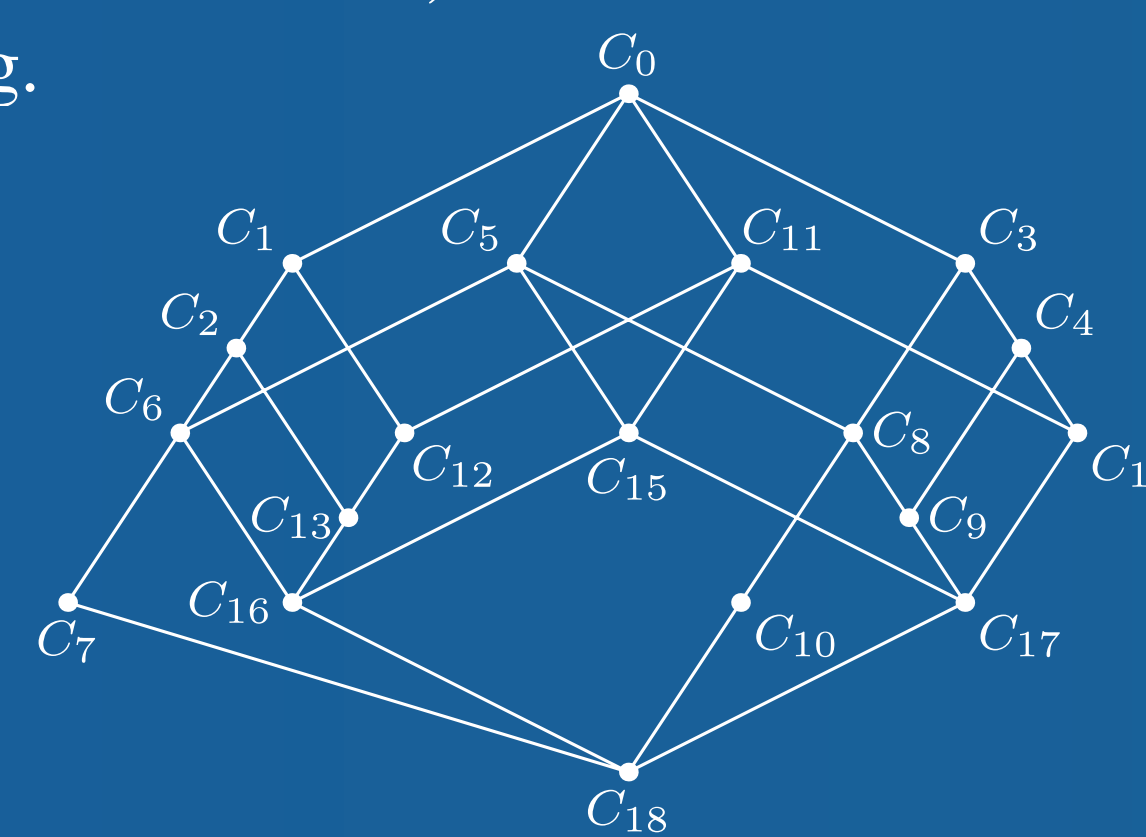
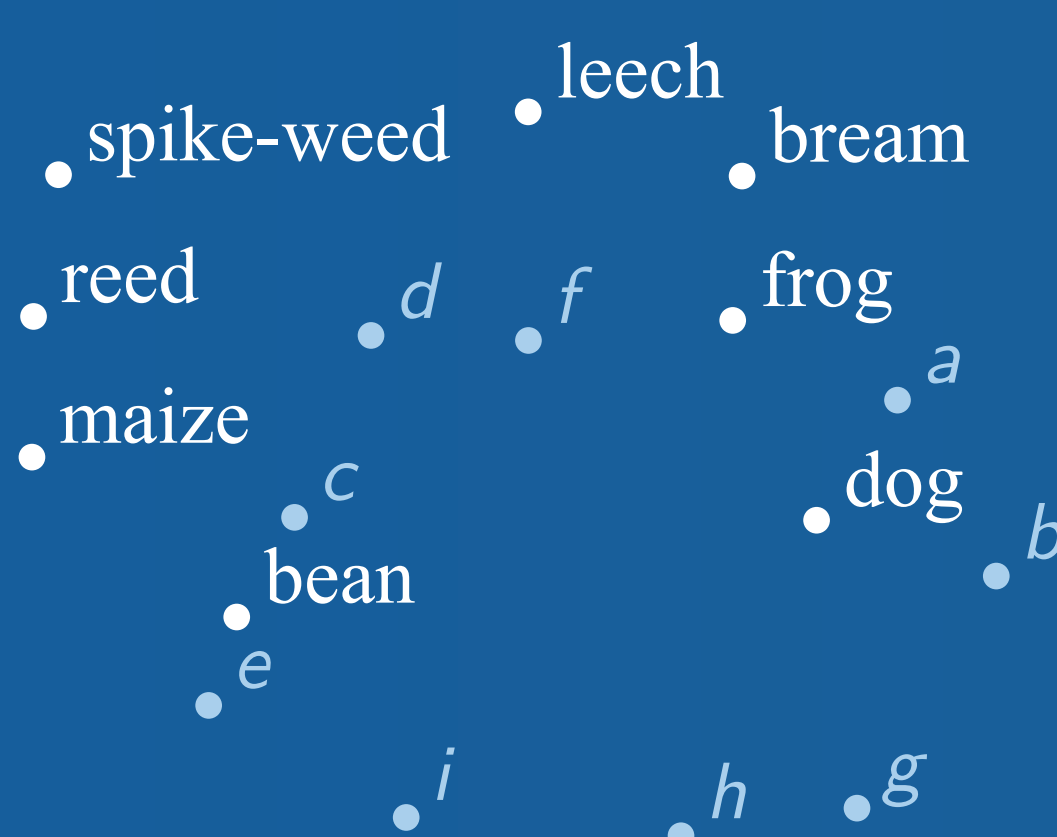
**MARVELLOUS**  
**GLACIERS** **FATHOMS**

## Subword-level Composition Functions for Learning Word Embeddings [3]

## Formal Concept Analysis [4]

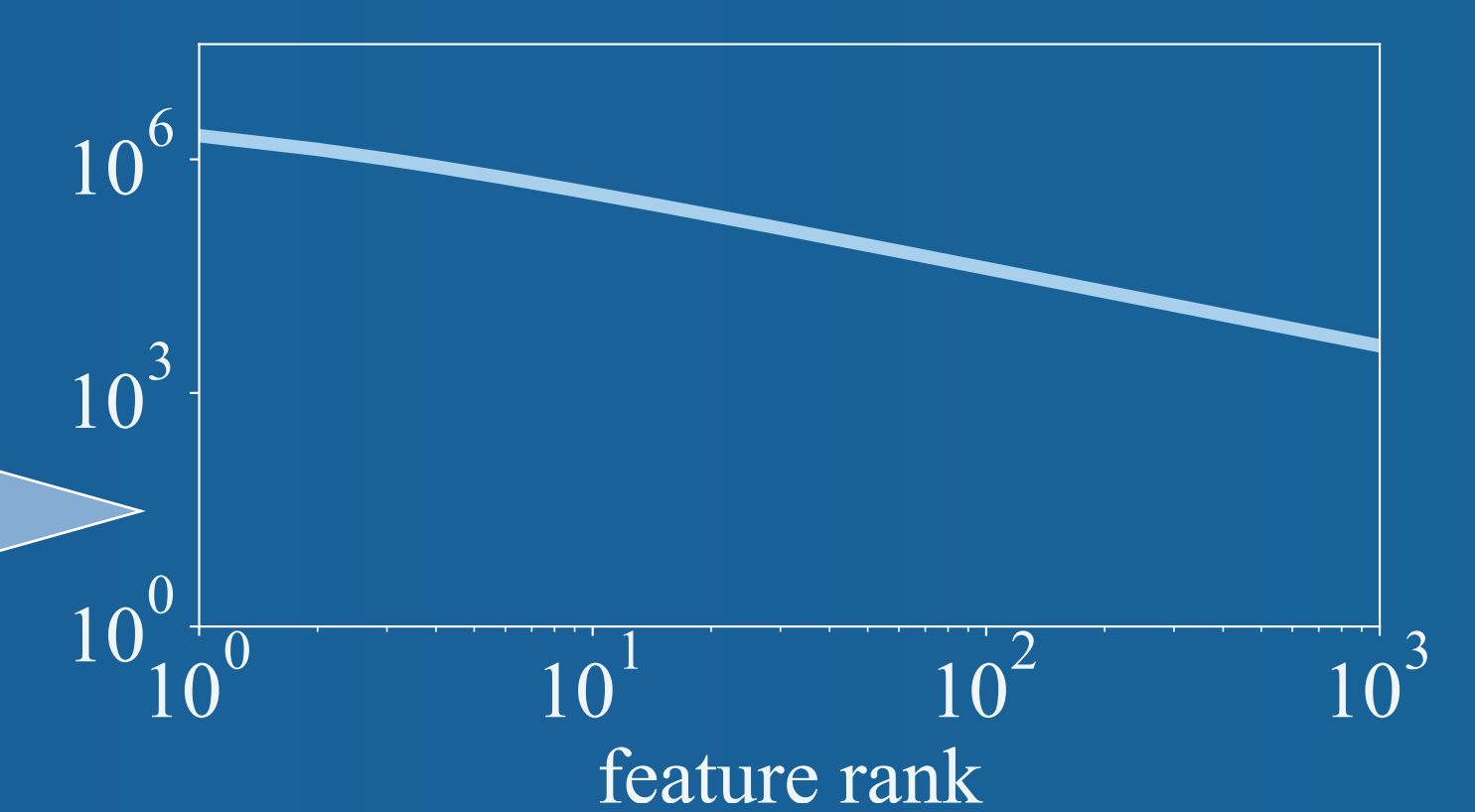
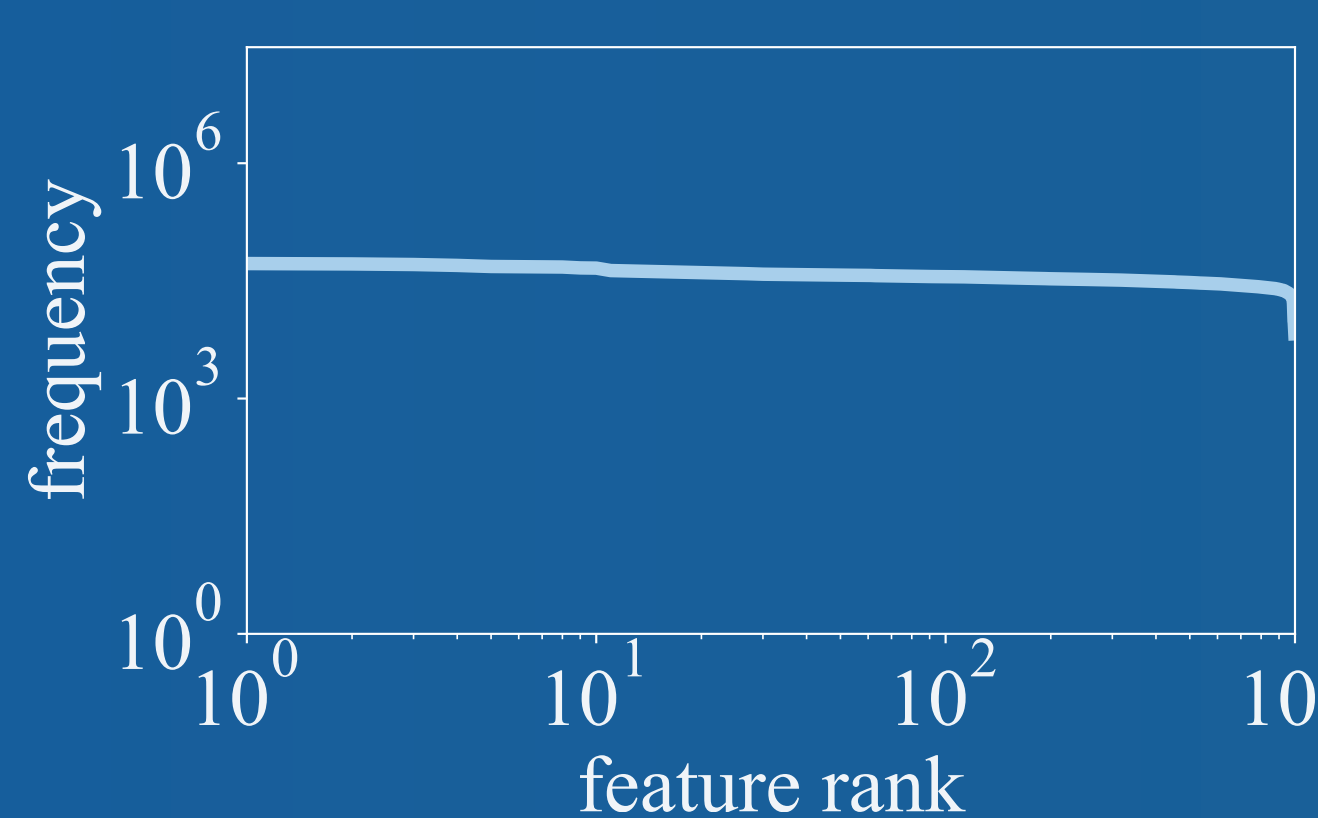
		<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>
leech	1	×	×					×		
bream	2	×	×					×	×	
frog	3	×	×	×				×	×	
dog	4	×		×				×	×	×
spike-weed	5	×	×		×		×			
reed	6	×	×	×	×		×			
bean	7	×		×	×	×				
maize	8	×		×	×		×			

*a*: needs water to live, *b*: lives in water,  
*c*: lives on land, *d*: needs chlorophyll to produce food,  
*e*: two seed leaves, *f*: one seed leaf,  
*g*: can move around, *h*: has limbs,  
*i*: suckles its offspring.



- $C_0 = \langle \{1, 2, 3, 4, 5, 6, 7, 8\}, \{a\} \rangle$ ,
- $C_1 = \langle \{1, 2, 3, 4\}, \{a, g\} \rangle$ ,
- $C_2 = \langle \{2, 3, 4\}, \{a, g, h\} \rangle$ ,
- $C_3 = \langle \{5, 6, 7, 8\}, \{a, d\} \rangle$ ,
- $C_4 = \langle \{5, 6, 8\}, \{a, d, f\} \rangle$ ,
- $C_5 = \langle \{3, 4, 6, 7, 8\}, \{a, c\} \rangle$ ,
- $C_6 = \langle \{3, 4\}, \{a, c, g, h\} \rangle$ ,
- $C_7 = \langle \{4\}, \{a, c, g, h, i\} \rangle$ ,
- $C_8 = \langle \{6, 7, 8\}, \{a, c, d\} \rangle$ ,
- $C_9 = \langle \{6, 8\}, \{a, c, d, f\} \rangle$ ,
- $C_{10} = \langle \{7\}, \{a, c, d, e\} \rangle$ ,
- $C_{11} = \langle \{1, 2, 3, 5, 6\}, \{a, b\} \rangle$ ,
- $C_{12} = \langle \{1, 2, 3\}, \{a, b, g\} \rangle$ ,
- $C_{13} = \langle \{2, 3\}, \{a, b, g, h\} \rangle$ ,
- $C_{14} = \langle \{5, 6\}, \{a, b, d, f\} \rangle$ ,
- $C_{15} = \langle \{3, 6\}, \{a, b, c\} \rangle$ ,
- $C_{16} = \langle \{3\}, \{a, b, c, g, h\} \rangle$ ,
- $C_{17} = \langle \{6\}, \{a, b, c, d, f\} \rangle$ ,
- $C_{18} = \langle \{ \}, \{a, b, c, d, e, f, g, h, i\} \rangle$ .

## Word2Bits, Quantized Word Vectors [5]



## References

- [1] RYGL, Jan, et al. ScaleText: The Design of a Scalable, Adaptable and User-Friendly Document System for Similarity Searches. In: RASLAN. 2016. p. 79–87.
- [2] SOJKA, Petr. Concrete Linguistics. Project proposal (in preparation). 2019.
- [3] LI, Bofang, et al. Subword-level Composition Functions for Learning Word Embeddings. In: Proceedings of the Second Workshop on Subword / Character Level Models. 2018. p. 38–48.
- [4] GANTER, Bernhard; STUMME, Gerd; WILLE, Rudolf (ed.). Formal Concept Analysis: Foundations and Applications. Springer, 2005.
- [5] LAM, Maximilian. Word2Bits – Quantized Word Vectors. arXiv preprint arXiv:1803.05651, 2018.
- [6] KAHNEMAN, Daniel. Thinking, Fast and Slow. Allen Lane, New York, USA, 2011.

Enthusiastically Illustrated by Gabriela Fišerová.