A Lattice Based Algebraic Model for Verb Centered Constructions

Sass Bálint

MAGYAR TUDOMÁNYOS AKADÉMIA
NYELVTUDOMÁNYI INTÉZET

Research Institute for Linguistics, Hungarian Academy of Sciences
sass.balint@nytud.mta.hu

– a new, abstract, mathematical model for *verb centered constructions (VCCs)*
– definition of VCC and *proper VCC (pVCC)*
– *double cube* = simple model for one VCC
– *corpus lattice* = complex model for all the VCCs of a whole corpus
– future work: discovering pVCCs based on the corpus lattice

1 VCC

verb centered construction (VCC) =
verb + other elements around the verb
now: PP/NP dependents (including subject!)

saying	–	the ball is in your court
verbal idiom	–	sweep under the rug
complex predicate	–	take a nap ← **filled place**
prep. phrasal verb	–	believe in ← **free place**
simple transitive verb	–	see
intransitive verb	–	happen

aim: to cover as many types of VCCs as we can, and provide a unified representation.

bottom	**place**	**filler**
↓	↓	↓
sweep	under the	rug

completeness

[sweep + under ⌒ rug] – not complete :(
[sweep + subj + obj + under ⌒ rug] – complete :)

Verbal clauses contain several VCCs which are substructures of each other, one is important:

proper VCC

– complete = contains all necessary elements
  free places constituting complements ✓
  idiomatic/institutionalized fillers ✓
– clean = does not contain any unnec. element
  no free places constituting adjuncts
  no compositional fillers

~ suitable to become a dictionary entry

Example 1 John reads the book.

VCC	proper?
[read + subj + obj]	✓
[read + subj + obj ⌒ book]	–

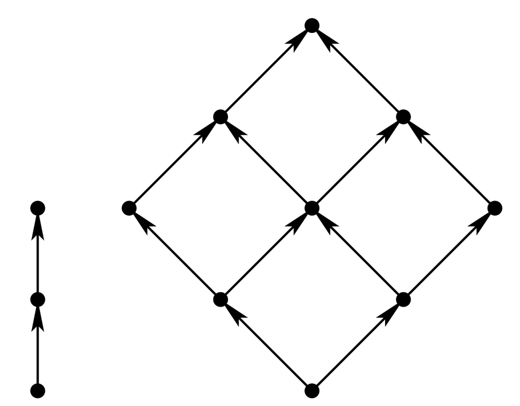
Example 2 John takes part in a demonstration.

VCC	proper?
[take + subj + obj]	–
[take + subj + obj ⌒ part]	–
[take + subj + obj ⌒ part + in]	✓
[take + subj + obj ⌒ part + in ⌒ demo.]	–

2 double cube (dc)

- generalize the cube for n dimensions
- add another cube in every dimension to make a larger cube whose side is twice as long

(double cubes are also called Post-lattices of order 3)



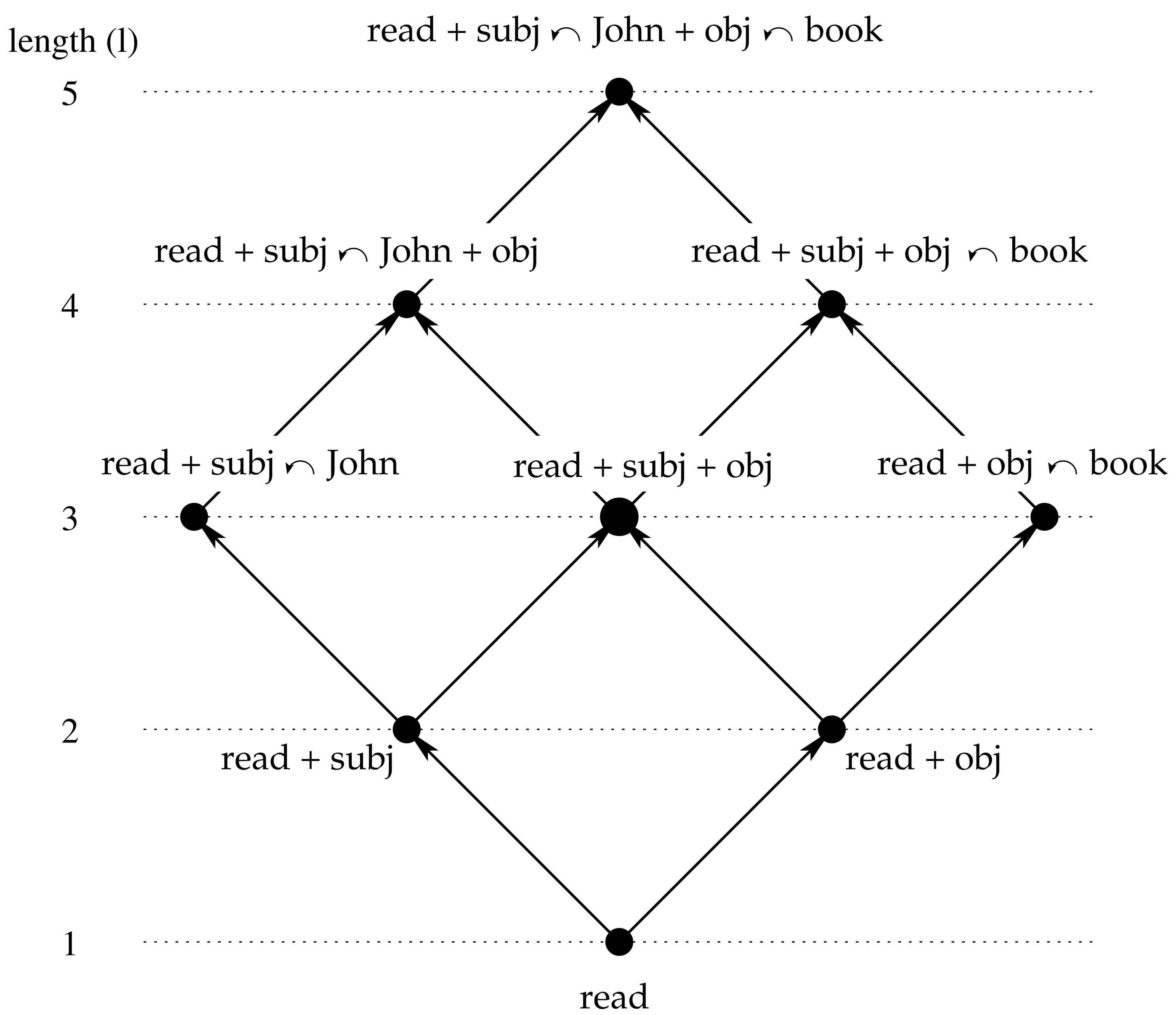
?

1 dimension
2 line segments

2 dimension
4 squares

3 dimension
8 usual cubes

Represent VCCs of a verbal clause as a double cube!



The double cube of John reads the book.

dimension	=	number of places
vertices	→	(sub-)VCCs
edges	→	VCC building operations (+, ⌒)
bottom	=	bare verb
top	=	fully filled VCC (contains all fillers)
proper VCC	=	**one of the vertices** (the center in this case)

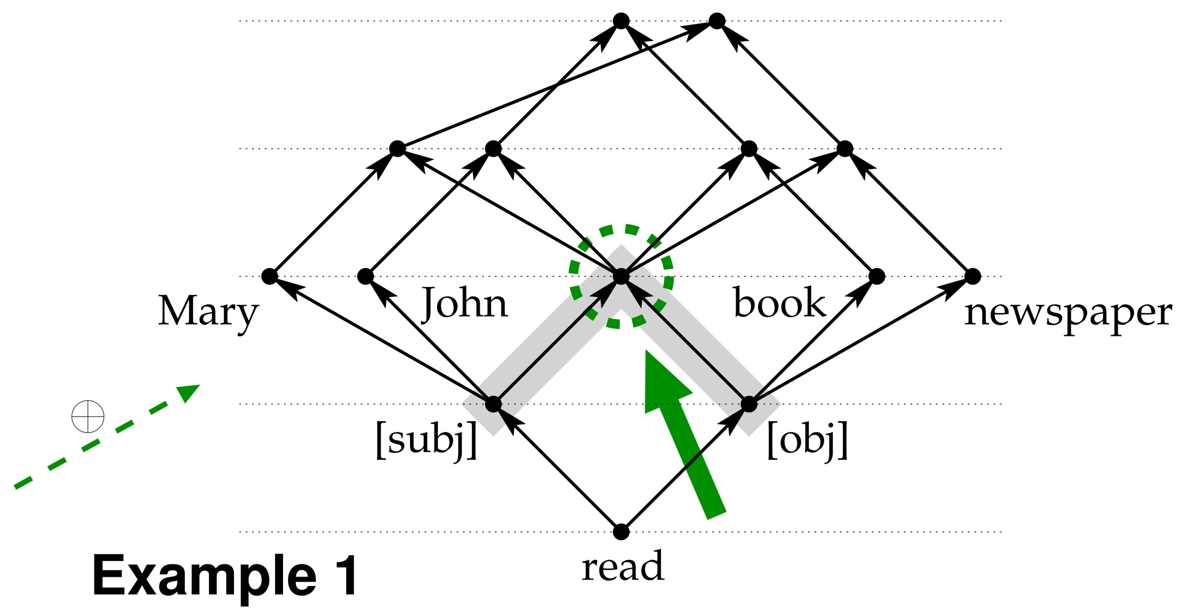
Also: can be viewed as dc of (John) takes part in a demo.
(not depicting the subject dimension is this case)

3 corpus lattice

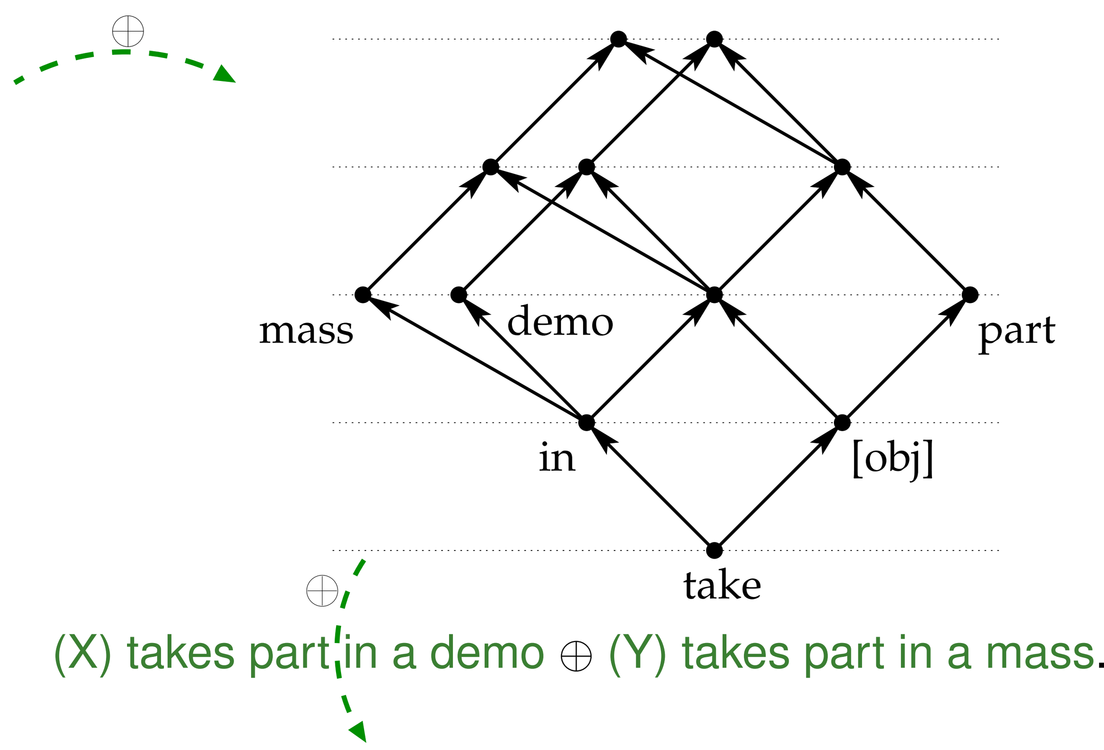
Take a verb. Take all clauses from the corpus containing this verb. Combine the dc of these clauses together. This will be a corpus lattice for this verb.

- *building blocks:* double cubes
- *operation:* lattice combination (⊕)

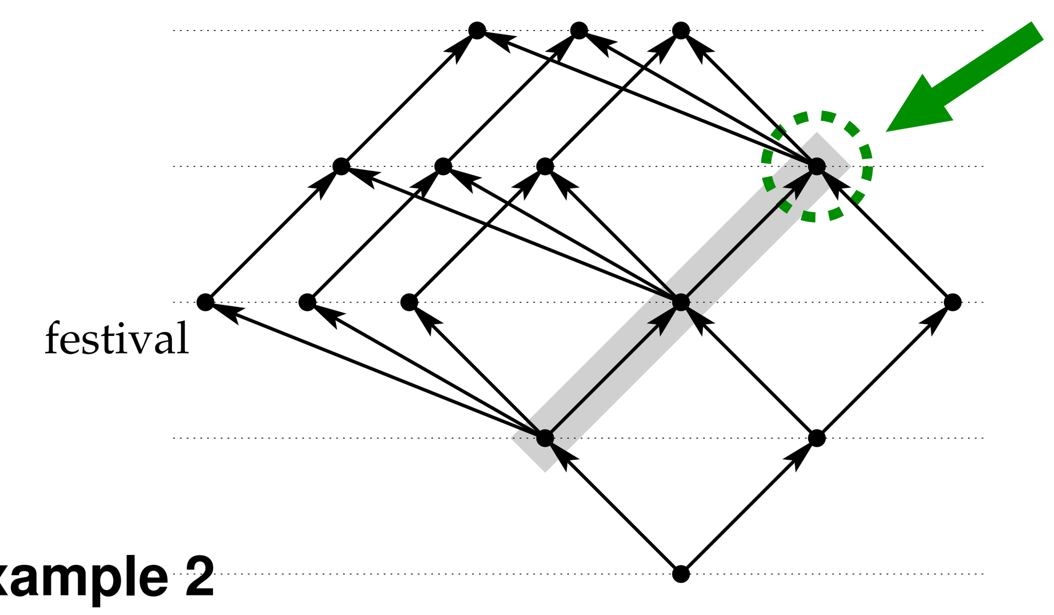
$L_1 \oplus L_2 = K$ means (informally):
where L_1 and L_2 are identical K overlaps,
where L_1 and L_2 are different K splits up.



Example 1
John reads the book ⊕ Mary reads the newspaper.



(X) takes part in a demo ⊕ (Y) takes part in a mass.



Example 2
(X) takes part in a demonstration ⊕ (Y) takes part in a mass ⊕ (Z) takes part in a festival. (subject dim is not shown)

4 FW = discovering pVCCs

Aim is to represent **all VCCs of a corpus together** including their relationships to each other, in order to be able to tackle proper VCCs based on this combined model.

The corpus lattice is the tool which realizes this aim projecting VCCs onto each other in a sense.

As **the corpus lattice** represents the distribution of all free and filled places beside verbs, we think that it is a representation which **can be the basis for discovering proper VCCs** of the corpus.

How can proper VCCs be discovered?

We think that proper VCCs are at some kind of **thickening points** of the corpus lattice.

The prospective algorithm would move through the corpus lattice vertex by vertex, until it reach proper VCCs at such points.

How can thickening points be characterized?

1. As we go top-down, the metric that shows how many clauses (fully filled VCCs) are represented by a given vertex suddenly increases at certain points.
2. Of these vertices, we prefer those which are located higher in the corpus lattice.

Thickening points = **green arrows**.

Conclusion

Example 1 shows that in this case [read + subj + obj] is the pVCC containing two free places, because this is the thickening point.

Example 2 shows that in this case [take + subj + obj ⌒ part + in] is the pVCC containing two free places and a filled place (subject not shown), because this is the thickening point.

In both cases, the thickening point designates the correct pVCC.

Accordingly, it may be worth taking a closer look at thickening points. This is the main direction for future work.

This research was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (case number: BO/00064/17/1; duration: 2017-2020).