

Unifying Models of Cognition

Rens Bod

VICI-Project “Integrating Cognition”

Institute for Logic, Language and Computation
University of Amsterdam

Goals of this Lecture

I'll give you a very brief introduction to the work in my Vici-group ***Integrating Cognition*** (9 people, NWO-funded)

About myself:

Part of the *LaCo* (Language and Computation) group in the ILLC (27 people)

Both affiliated at U. of St Andrews and U. of Amsterdam

Computational linguist and Cognitive Scientist

At the same time, I am writing an overview *History of the Humanities* (will be published in spring 2010)

What do different forms of cognition have in common?

E.g.:

Language:

"List the sales of products in 2003"

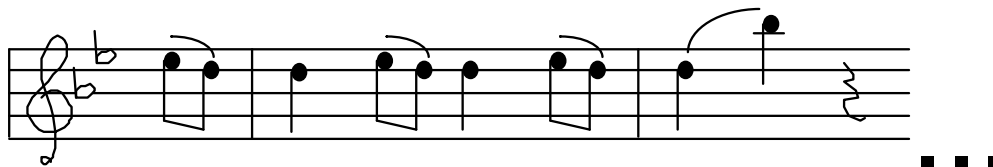
What do different forms of cognition have in common?

E.g.:

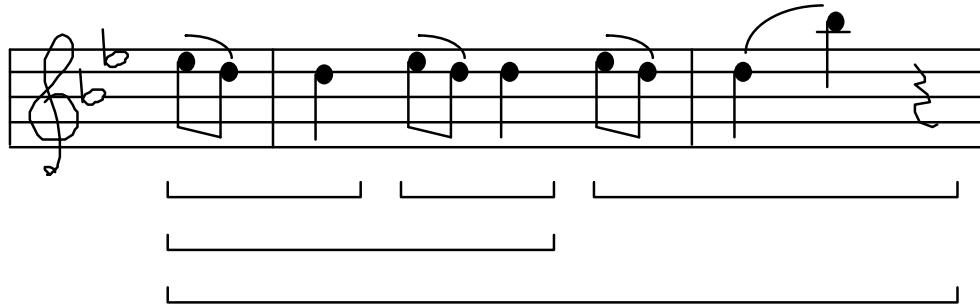
Language:

"List the sales of products in 2003"

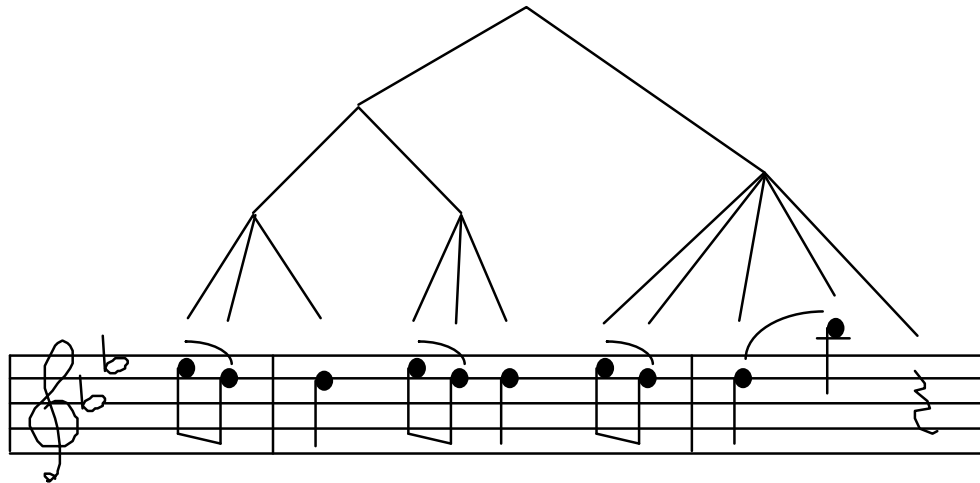
Music:



Grouping Structure = Tree Structure

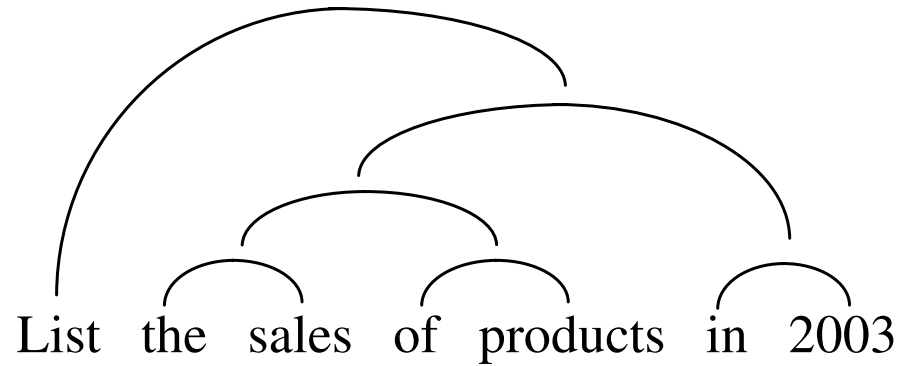


is equivalent (isomorphic) with:



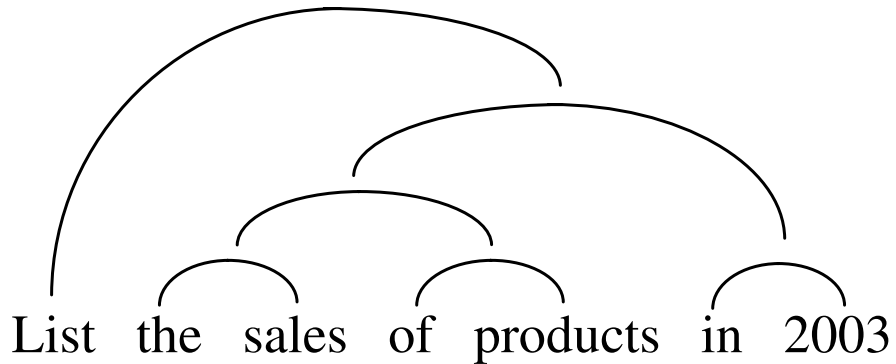
Grouping Structure in Language

Groups in language form a *tree structure* (Wundt 1880):

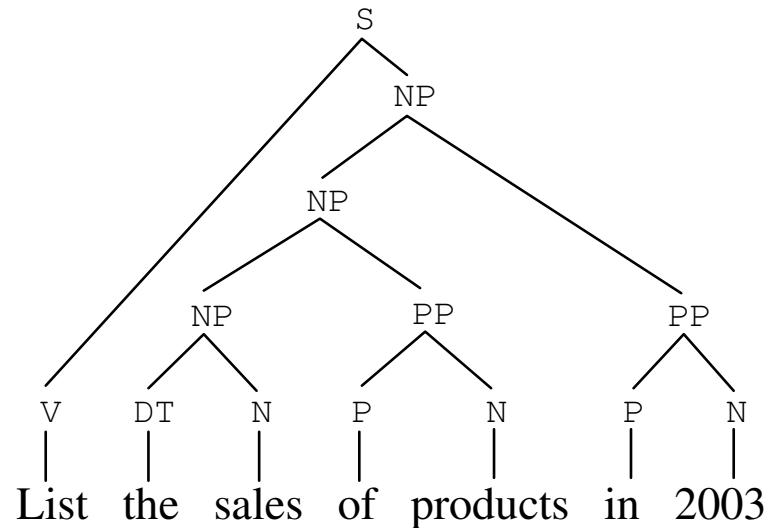
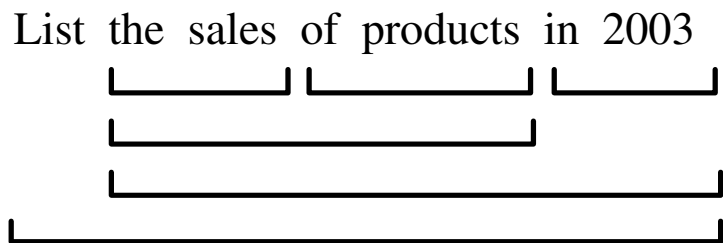


Grouping Structure in Language

Groups in language form a *tree structure* (Wundt 1880):

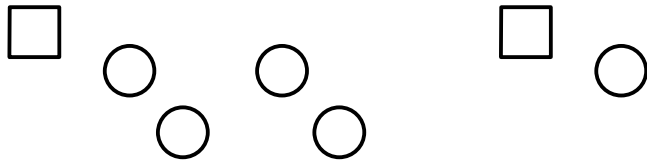


Grouping structure in different representations (Chomsky 1956):

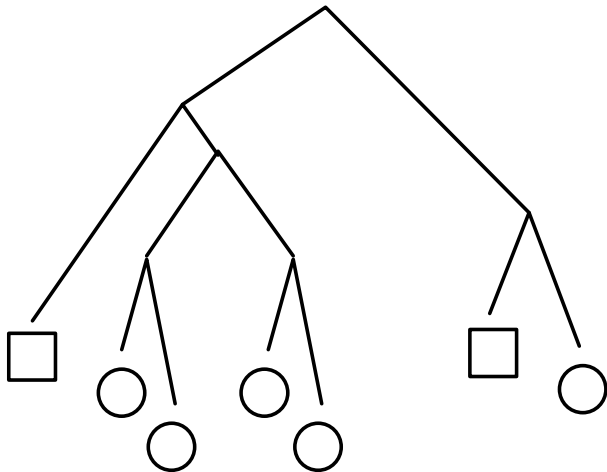


Also Visual Groups form a Tree Structure

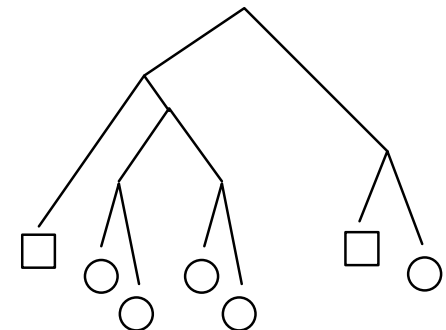
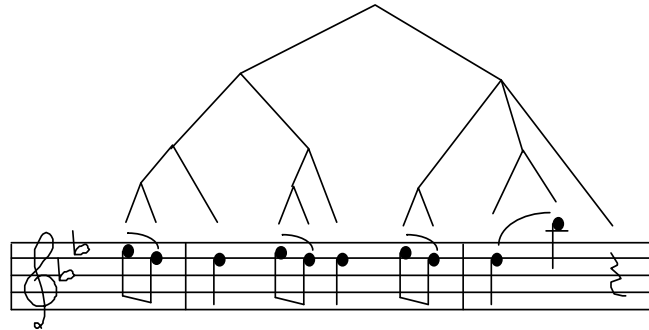
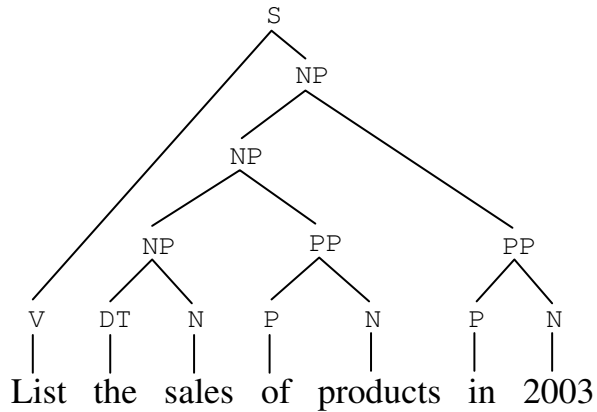
According to Wertheimer (1923) the visual input



is assigned the following structure:



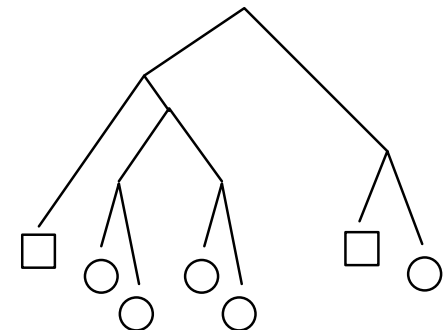
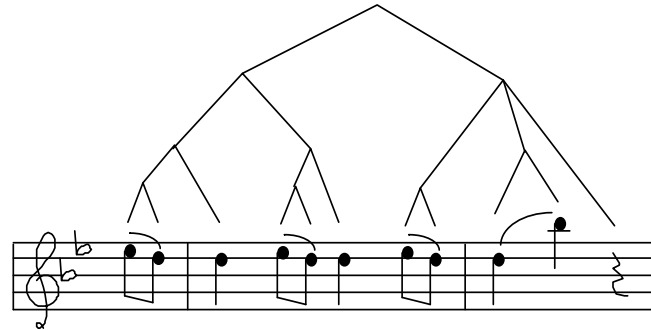
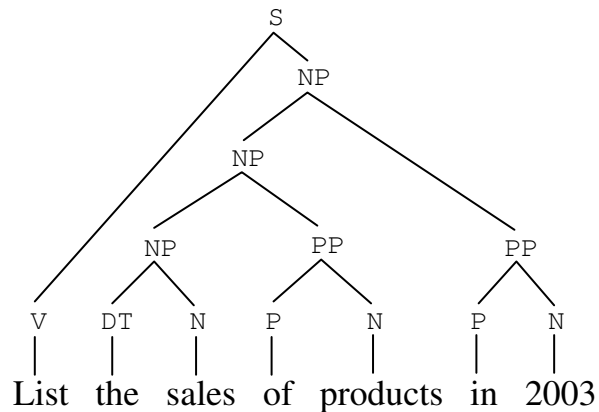
Perceptual Structure = Tree Structure



Relatively Uncontroversial:

There exists *one representation* for structural perception for all modalities

Perceptual Structure = Tree Structure



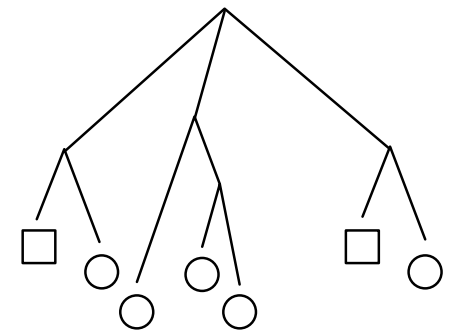
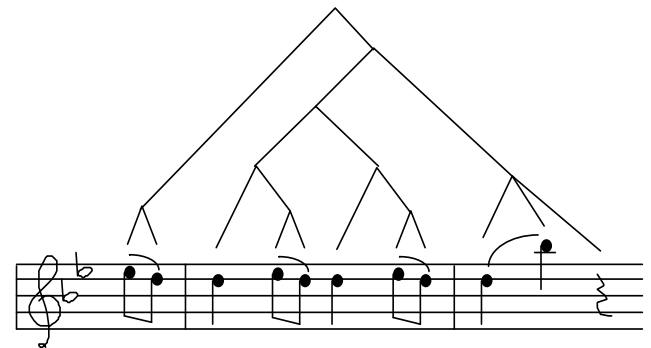
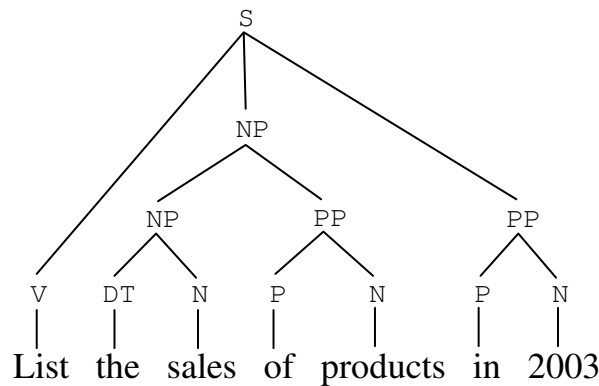
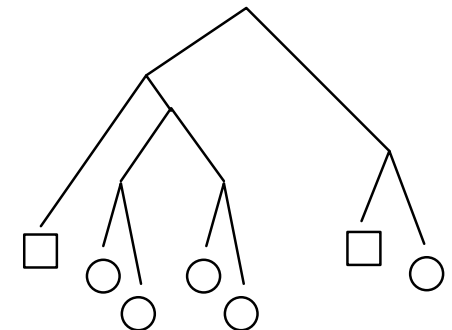
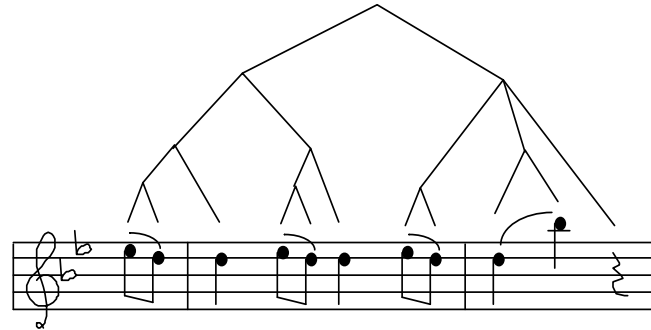
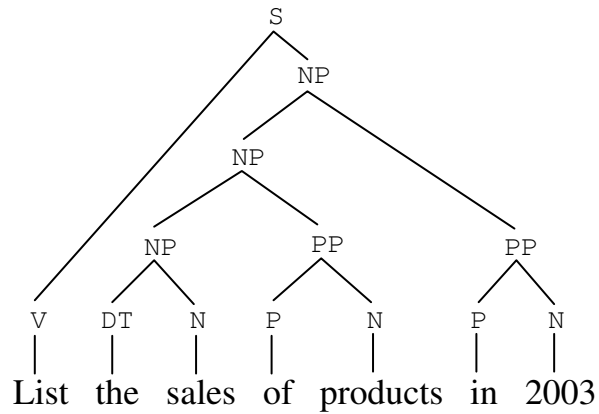
Relatively Uncontroversial:

There exists *one representation* for structural perception for all modalities

Very Controversial:

There exists *one model* that predicts the perceived structure in *language, music, vision* and other modalities... (cf. Newell 1999)

Additional Problem: Perception is Ambiguous



The same input can be assigned several structures: ambiguity

Ambiguity is a major problem

Average sentence from *Wall Street Journal*: more than **one million** different *possible* tree structures (Charniak 1999)

Adding semantics makes the problem even worse!

"Any given sequence of notes is infinitely ambiguous, but this ambiguity is seldom apparent to the listener" (Longuet-Higgins 1987)

Humans perceive mostly just *one* grouping structure

Ambiguity is a major problem

Average sentence from *Wall Street Journal*: more than **one million** different *possible* tree structures (Charniak 1999)

Adding semantics makes the problem even worse!

"Any given sequence of notes is infinitely ambiguous, but this ambiguity is seldom apparent to the listener" (Longuet-Higgins 1987)

Humans perceive mostly just *one* grouping structure

> 96% agreement among subjects (language users)

Language:	<i>Penn Treebank</i>
Music:	<i>Essen Folksong Collection</i>
Vision:	<i>Nijmegen Visual Database</i>

Historically, two competing principles for solving ambiguity

1. **Simplicity Principle** (Wertheimer 1923...Leeuwenberg 2001, Chater 2007)

Preference for the *simplest* structure

2. **Likelihood Principle** (Helmholtz 1910...Suppes 1984, Charniak 2001)

Preference for the *most likely* structure

Can these principles still inspire us?

The Dual Nature of Perception

These principles each play a *different* role in perception:

Simplicity: general preference for "economy", "least effort",
"shortest derivation"

Likelihood: a memory-based bias due to previous experiences

The Dual Nature of Perception

These principles each play a *different* role in perception:

Simplicity: general preference for "economy", "least effort", "shortest derivation"

Likelihood: a memory-based bias due to previous experiences

Hypothesis: perceptual system strives for the *simplest* structure but in doing so it is influenced by the *likelihood* of previous structures

Possible Measures for Simplicity and Likelihood

Simplicity: *number* of "steps" to generate a tree structure

Likelihood: joint *probability* of the steps to generate a tree structure

We can compute this if we have a large, representative collection of tree structures for each modality (a "corpus")

Possible Measures for Simplicity and Likelihood

Simplicity: *number* of "steps" to generate a tree structure

Likelihood: joint *probability* of the steps to generate a tree structure

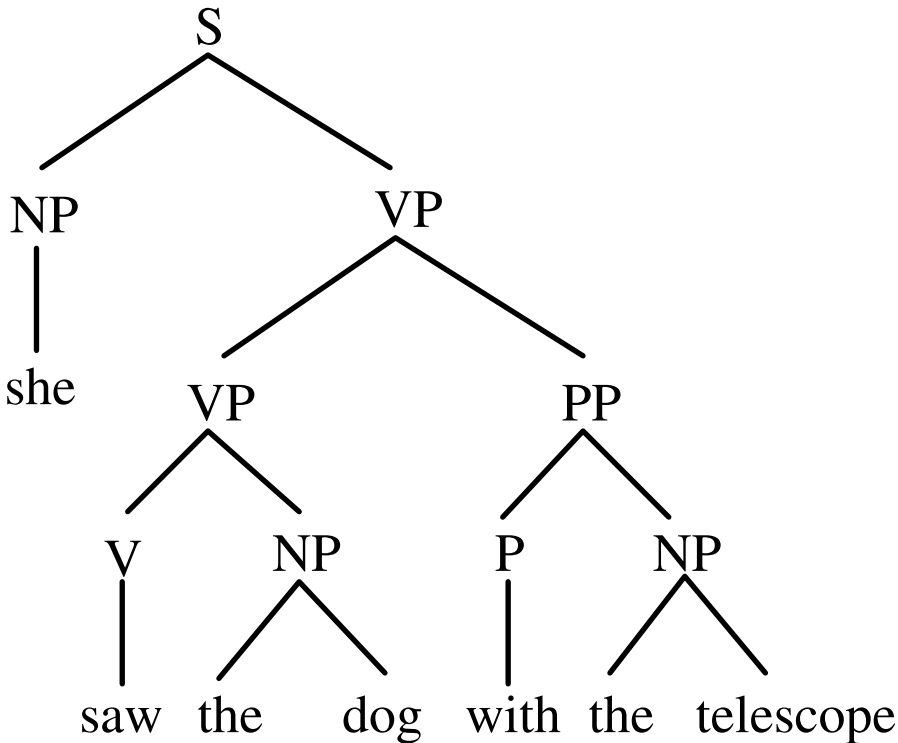
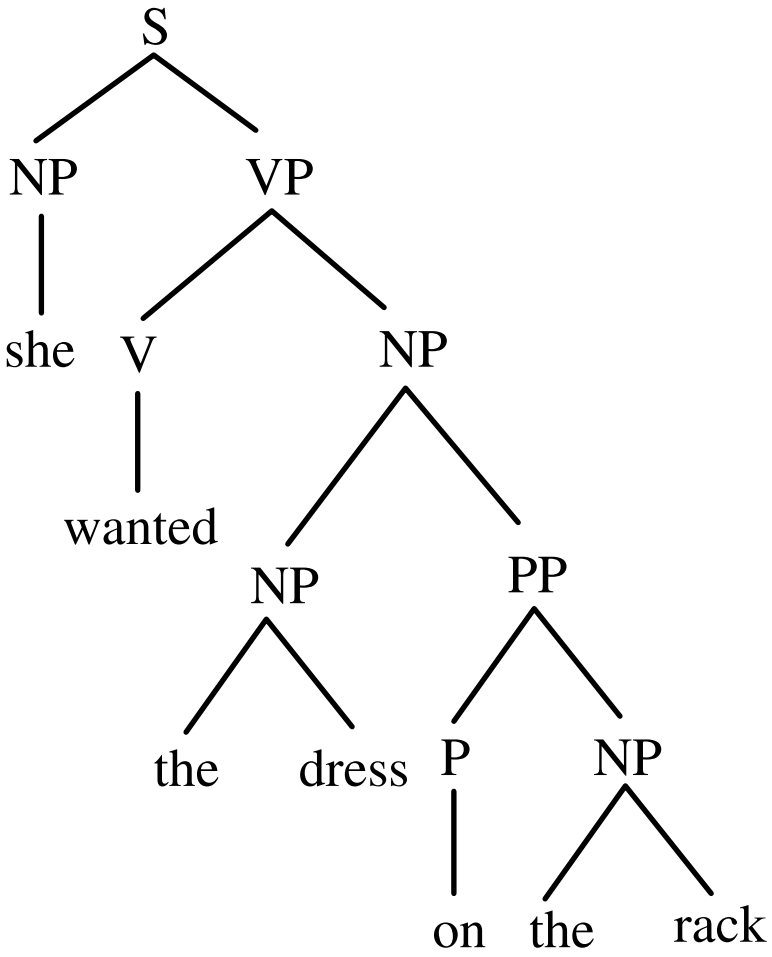
We can compute this if we have a large, representative collection of tree structures for each modality (a "corpus")

Data-Oriented Parsing model (DOP):

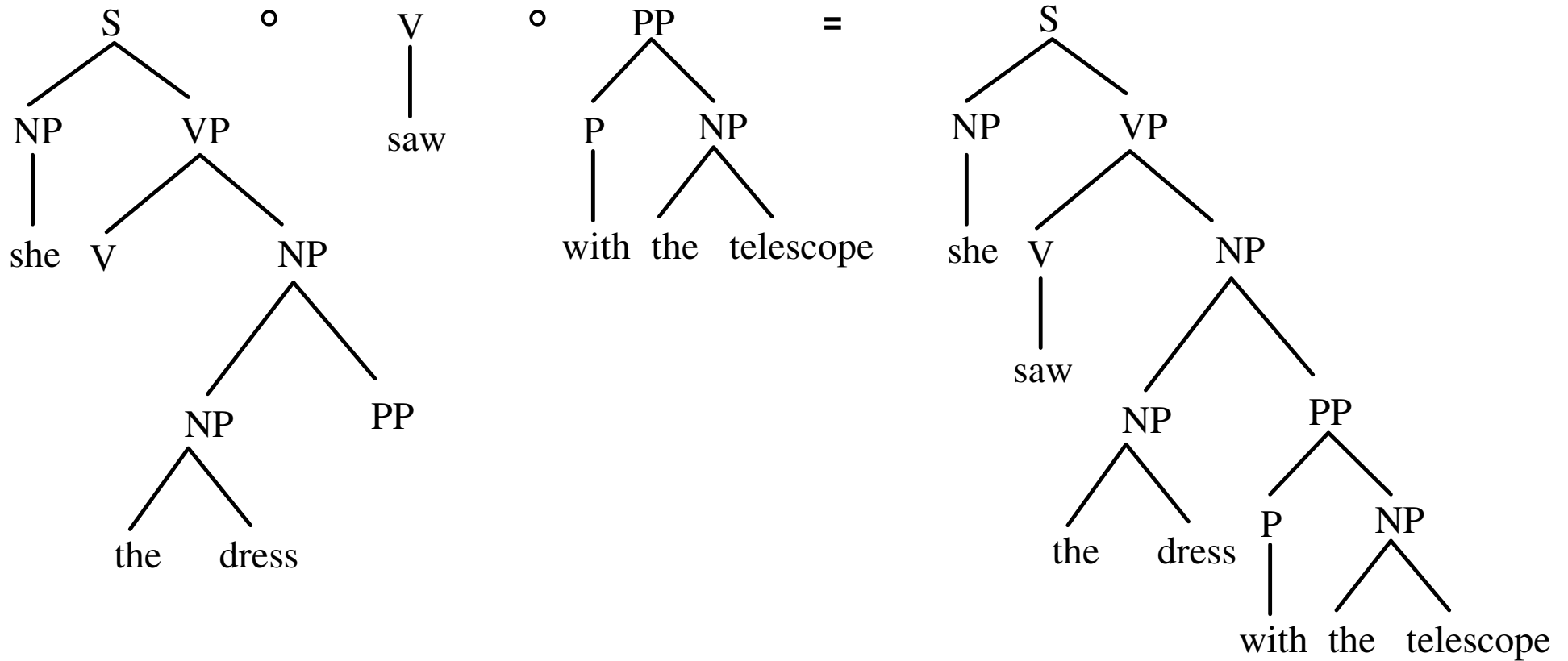
New input is analyzed and interpreted by combining parts of previously perceived input (Scha 1990; Bod 1992, 1998; Sima'an 1995; Kaplan 1996; Goodman 1996; Way 1999; Rajman 1999; Hearne 2003; Post 2009 etc.)

Example of a DOP model for Language

Let's start with an extremely simple corpus:

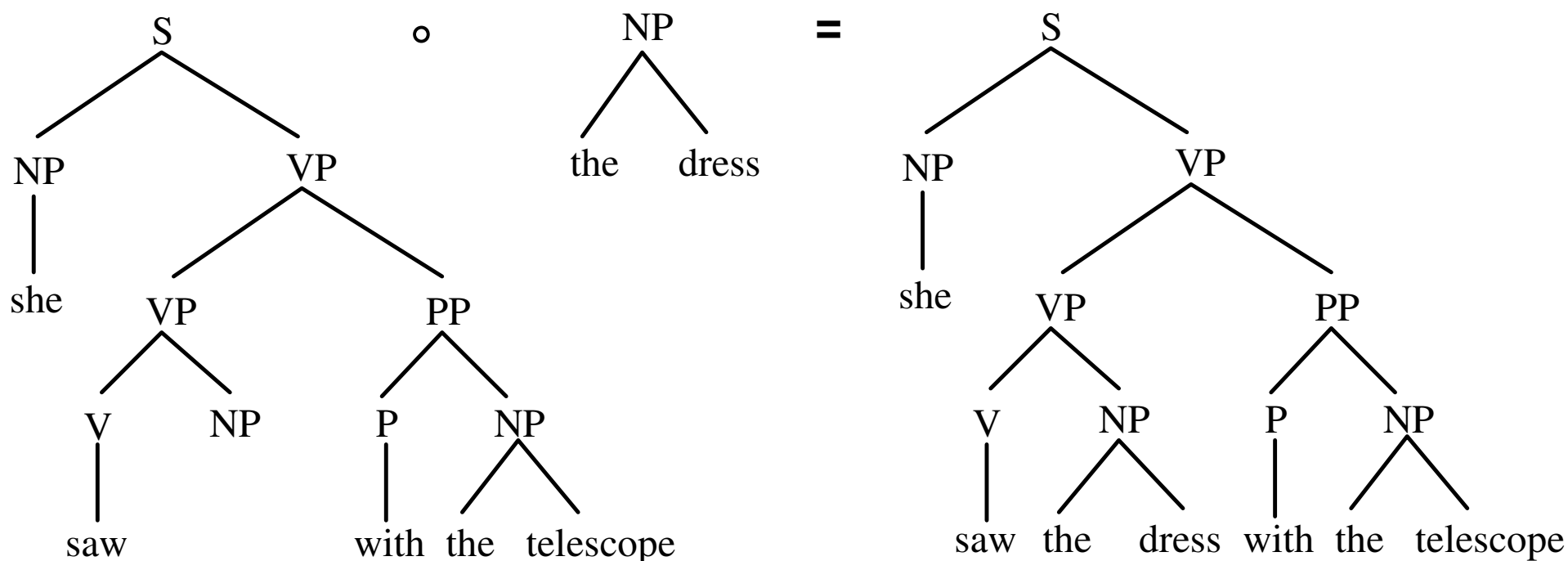


A new sentence such as "*She saw the dress with the telescope*" is analyzed by **combining subtrees from the corpus**



where "o" is left-most node substitution

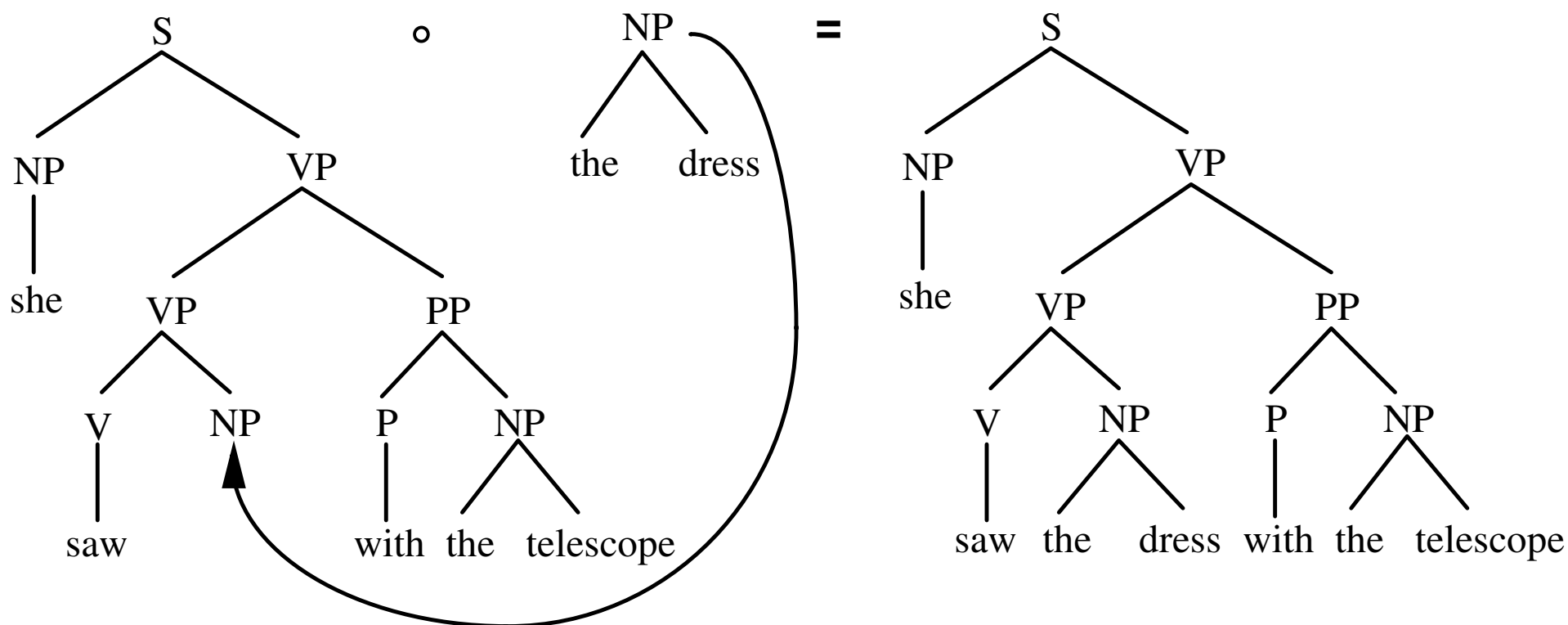
But there is also a "competing" analysis:



This analysis consists of two steps, and is therefore preferred according to the *simplicity principle*: **maximal similarity** with corpus.

But it is **not** preferred according to the *likelihood principle*

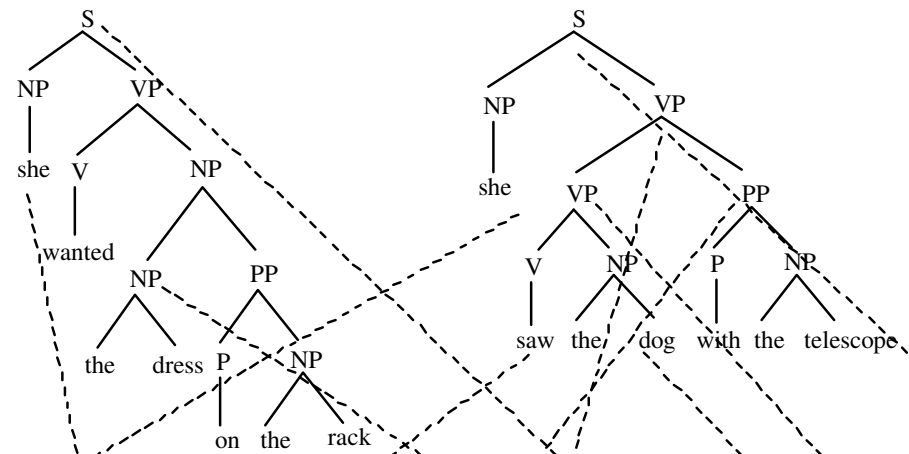
But there is also a "competing" analysis:



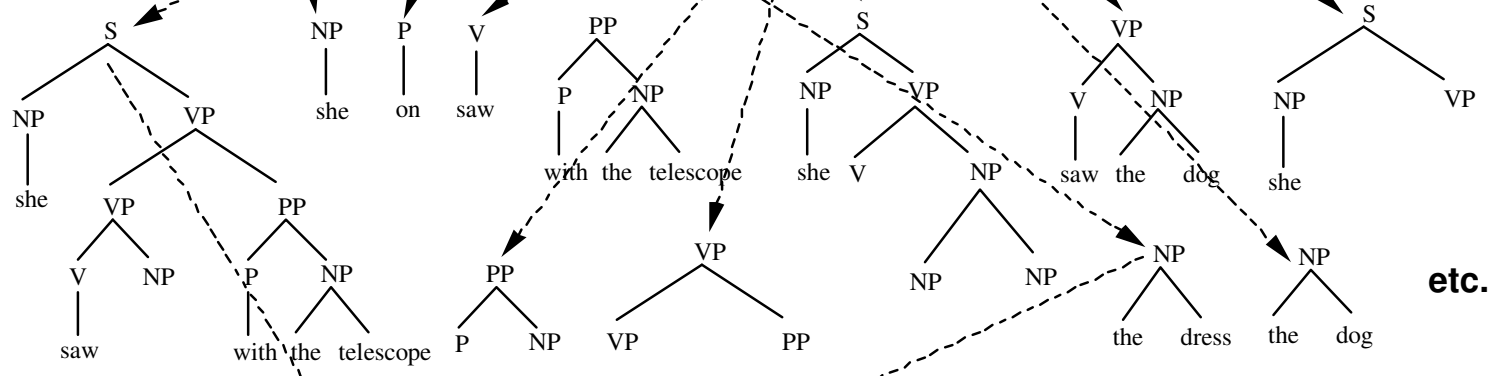
This analysis consists of two steps, and is therefore preferred according to the *simplicity principle*: **maximal similarity** with corpus.

But it is **not** preferred according to the *likelihood principle*

Corpus

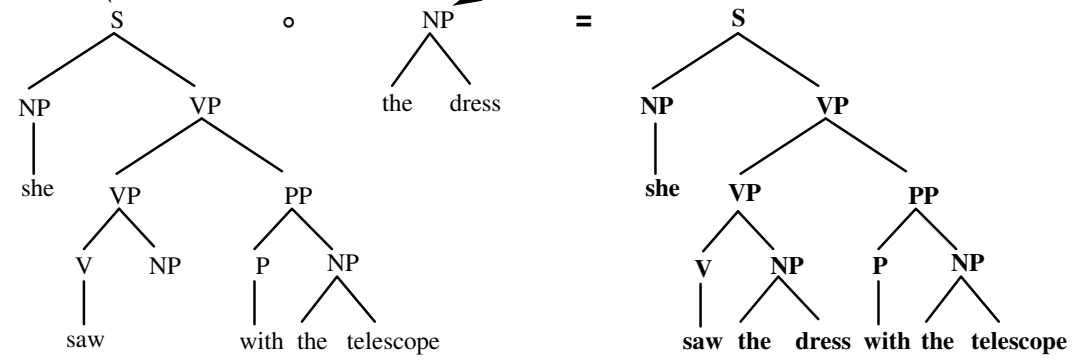


Decompositie



etc.

Recompositie



Definitions of Likelihood & Simplicity

Probability of a subtree:

$$P(t) = \frac{|t|}{\sum_{t': r(t')=r(t)} |t'|}$$

Most probable derivation of a sentence:

$$P(t_1 \circ \dots \circ t_k) = \prod_i P(t_i)$$

Shortest derivation of a sentence:

$$T_{sd} = \operatorname{argmin}_T L(dT)$$

DOP models can be formalized as Stochastic Tree Grammars

DOP is a *tree grammar* where the *tree-units* can be of arbitrary size:

it allows for the possibility that units of *any* size may play a role

By putting constraints on the tree-units, DOP subsumes:

- stochastic context-free grammars
- stochastic head-lexicalized grammars
- stochastic tree-adjoining grammars
- stochastic regular grammars
- ...

Test Domains

- ***Linguistic test domain:***

Penn Treebank Wall Street Journal (WSJ) corpus: **50,000** manually analyzed sentences

- ***Musical test domain:***

Essen Folksong Collection (EFC): **20,150** melodically analyzed western folksongs:

- *Pitches*: numbers from 1 to 7
- *Duration indicators*: underscore (_) or a period (.) *after* the numbers
- *Octave position*: plus and minus signs (+ , -) *before* the numbers
- *Chromatic alterations*: "#" or "b" *after* the numbers
- *Pauses*: 0, possibly followed by duration indicators

Example from Essen Folksong Collection

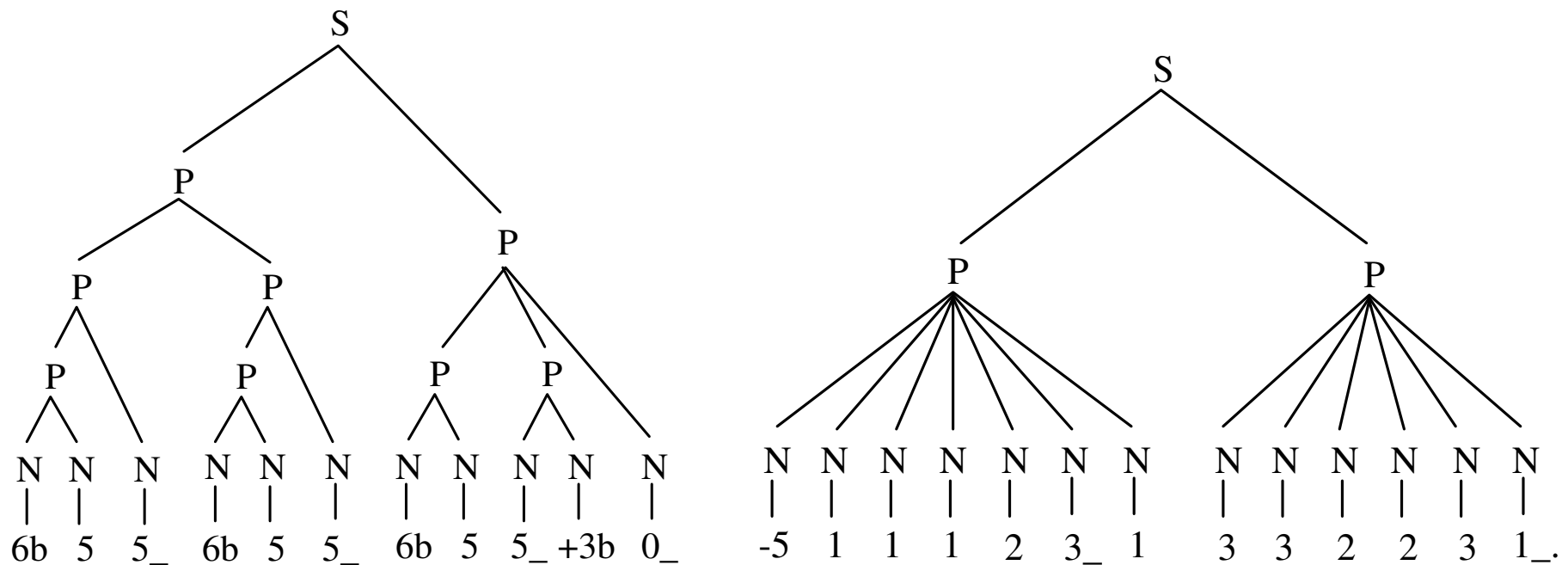
#4551: *Schneckhaus Schneckhaus stecke deine Hörner aus*
(German children song)

5_3_5_3_1234553_1234553_12345_3_12345_3_553_553_553_65432_1_

Grouping structure according to Essen Folksong collection:

((5_3_5_3_) (((1234553_) (1234553_)) ((12345_3_)(12345_3_))) ((553_553_)
(553_65432_1_)))

Examples of musical structures:



We can use the **same DOP model** for both linguistic and musical analysis:

Analysis = decomposition & recomposition

(Example of) Experimental Evaluation

Corpora are randomly divided into 10 *training/test set splits*

Test 1: Simplicity-Likelihood-DOP (SL-DOP)

Selects simplest structure from among n likeliest structures

Test 2: Likelihood-Simplicity-DOP (LS-DOP)

Selects likeliest structure from among n simplest structures

Scores of SL-DOP & LS-DOP

<i>n</i>	SL-DOP		LS-DOP	
	(simplest among <i>n</i> likeliest)		(likeliest among <i>n</i> simplest)	
	Language	Music	Language	Music
1	87.9%	86.0%	85.6%	84.3%
5	89.3%	86.8%	86.1%	85.5%
10	90.2%	87.2%	87.0%	85.7%
11	90.2%	87.3%	87.0%	85.7%
12	90.2%	87.3%	87.0%	85.7%
13	90.2%	87.3%	87.0%	85.7%
14	90.2%	87.2%	87.0%	85.7%
15	90.2%	87.2%	87.0%	85.7%
20	90.0%	86.9%	87.1%	85.7%
50	88.7%	85.6%	87.4%	86.0%
100	86.8%	84.3%	87.9%	86.0%
1,000	85.6%	84.3%	87.9%	86.0%

1. **Same** model obtains **maximal** scores for **both** language and music
2. Perceptual system strives for the **simplest analysis**, but searches only among **the most probable** analyses

Where do initial trees come from?

- How do we learn the *initial* tree structures in language, music and problem solving?
- **Current VICI project:**

Main goal:

The development of a general, *unsupervised* learning model for *different* modalities (language, music, problem-solving, ...)

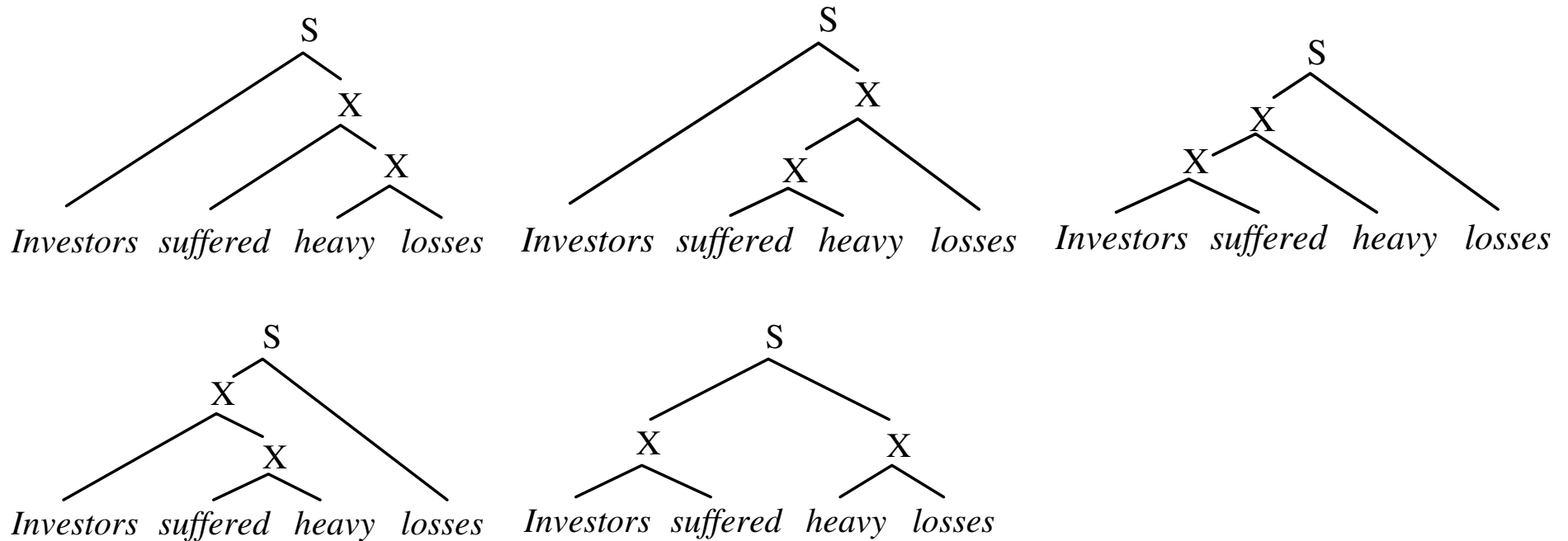
Key-Idea:

If we do not know which trees should be assigned to initial input, assign them all and train them on new data

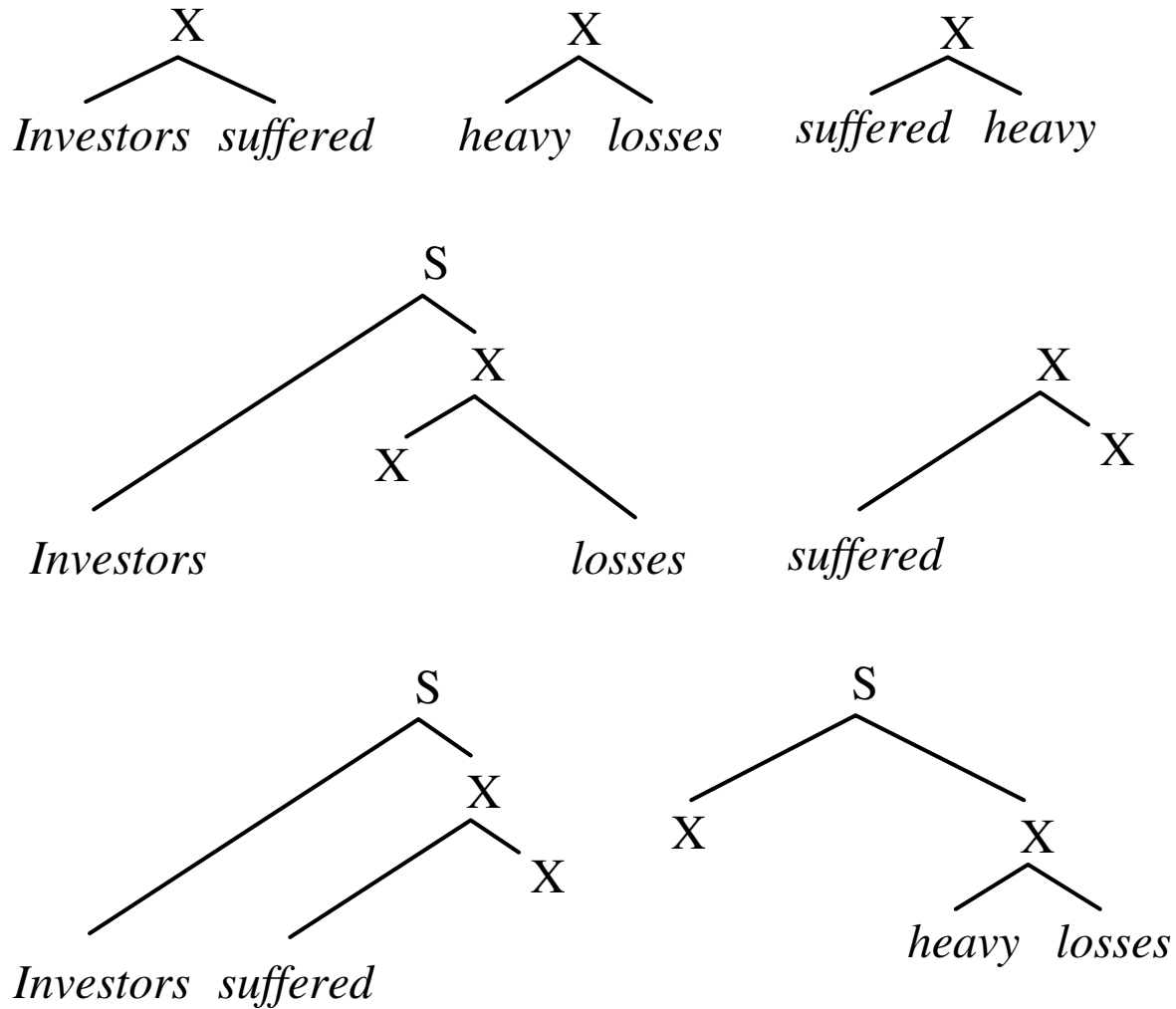
U-DOP: Unsupervised DOP

1. Assign *all* possible binary trees to input (where we leave the internal labels unspecified)

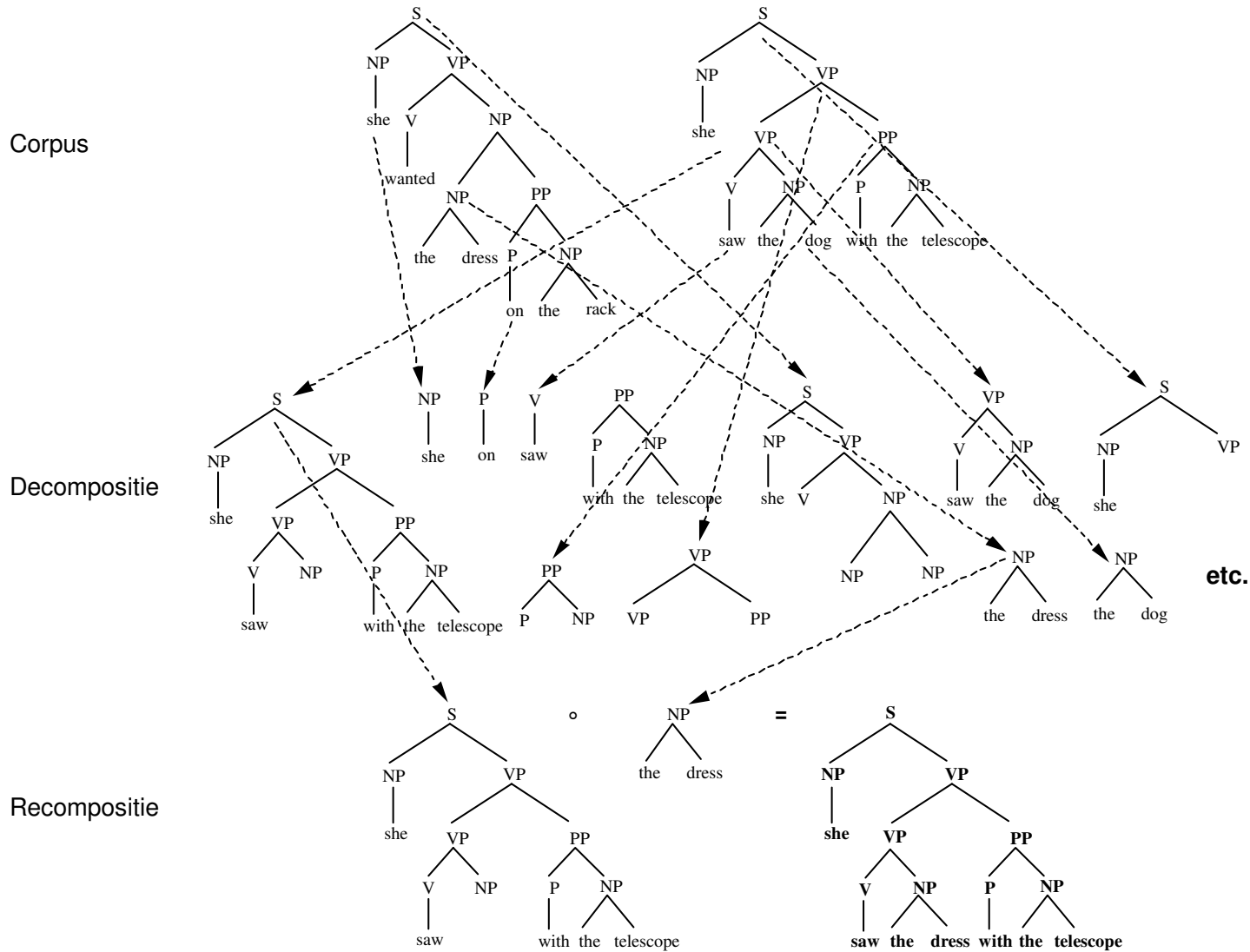
E.g.: *Investors suffered heavy losses:*



2. Divide these trees into all subtrees (& store them in a packed chart)



3. Analyze new input via most probable analysis (as in DOP):



DOP & U-DOP have been used for:

- Modeling language acquisition, surprisals, artificial language learning:

Borensztajn, Zuidema and Bod (2008) (Best Paper Award), Bod (2009), Frank (2009), Ferdinand & Zuidema (2009), Cochran (2009) ...

- Applications: Language models for speech, Machine Translation, Musical Parsing, Linguistic Parsing, Scientific Problem Solving:

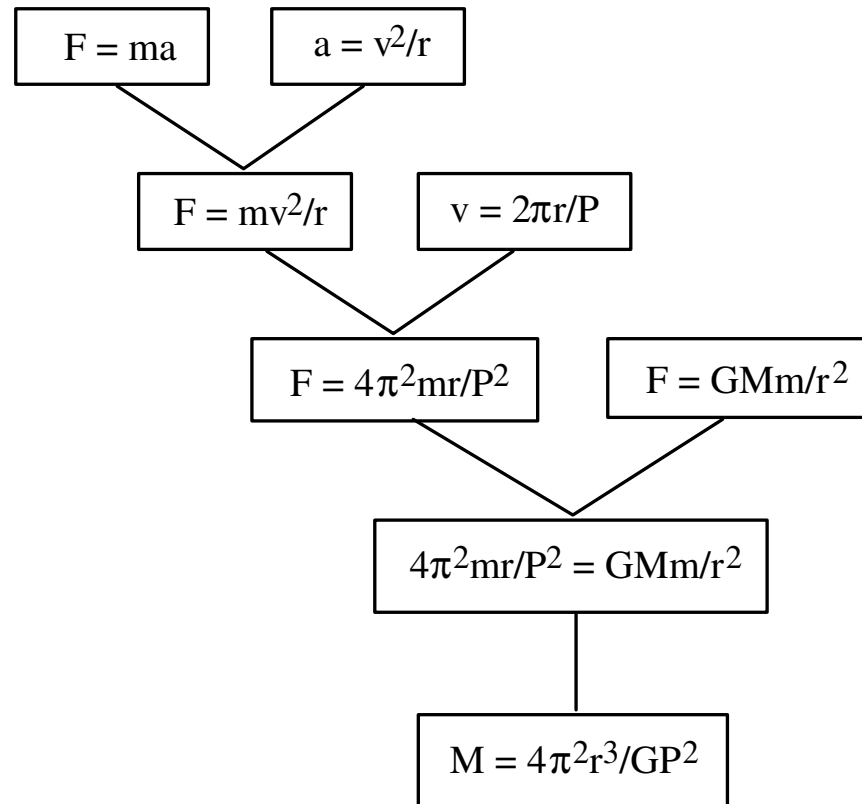
Bod (2007), Sangati & Zuidema (2009), Honingh & Bod (2006), Honingh (2009), Bod et al. (2008)...

How far does DOP stretch?

- **Problem solving** with exemplar-based model such as DOP?
- Exemplar-based reasoning has been proposed in *Case-based reasoning*, *Explanation-based learning*, but without a probabilistic component
- Problem-solutions in physics can be represented by derivation trees

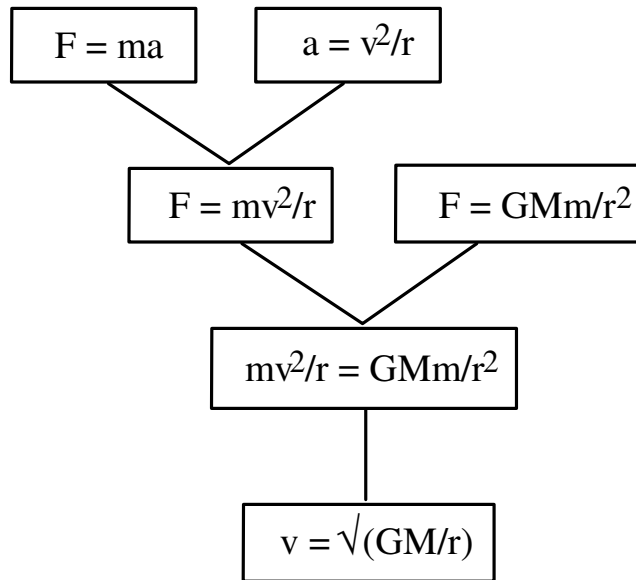
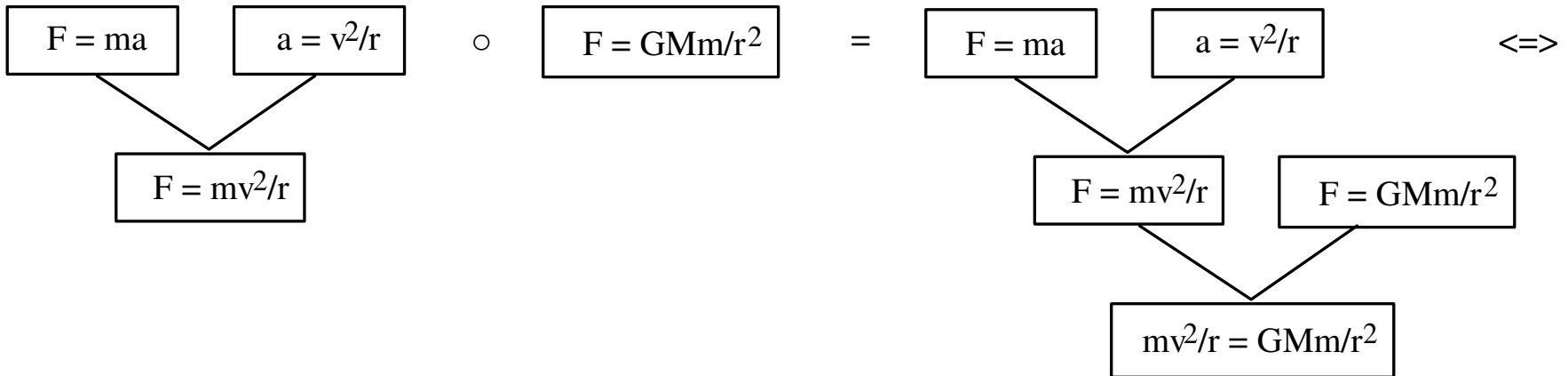
Example of derivation tree in classical mechanics

Derivation of planet's mass using Newton's laws

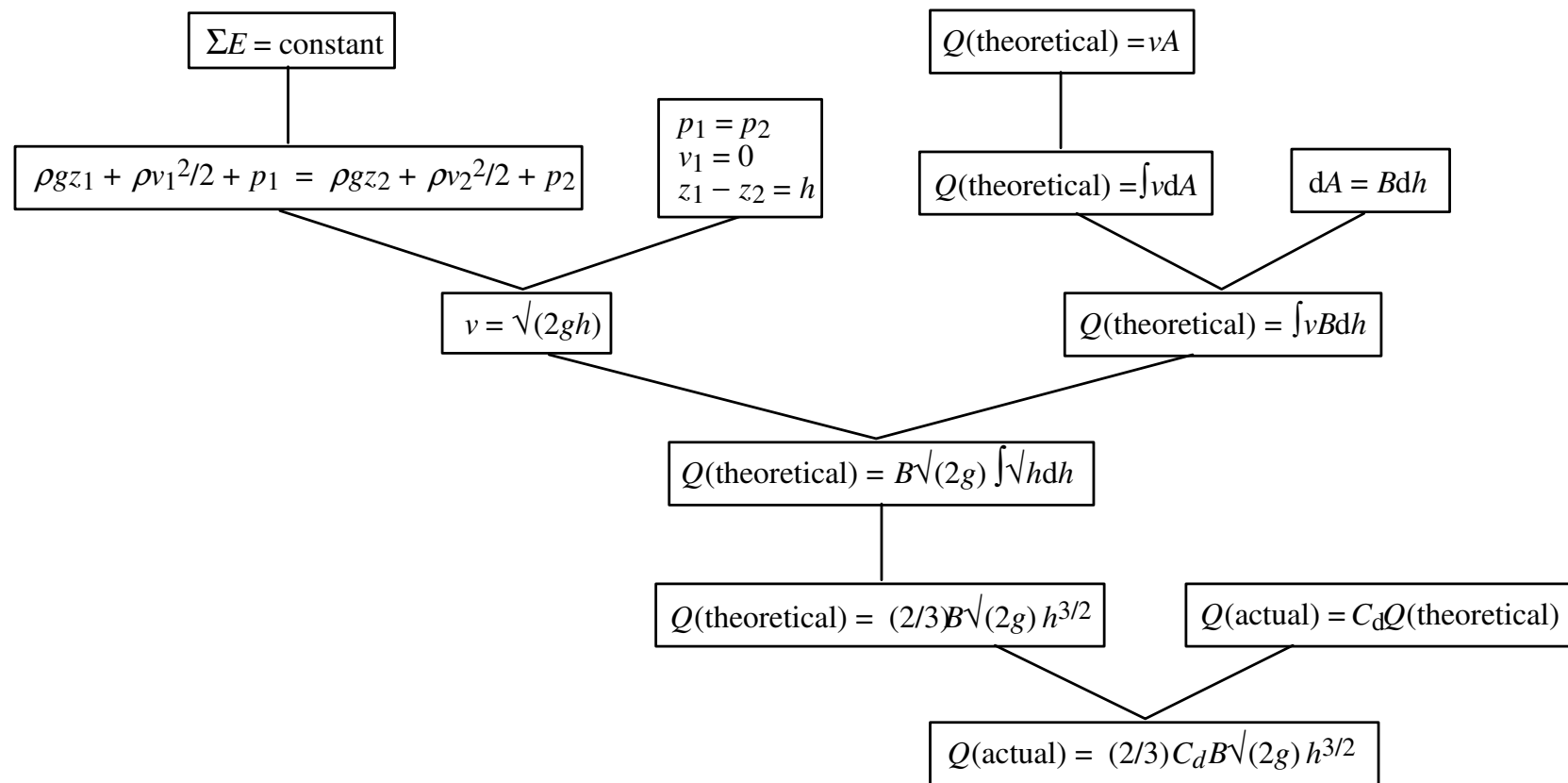


A tree describes the steps from higher-level laws to the solution (formula)

Subtrees can be reused to solve new problems



A more complex derivation tree, from fluid dynamics:



- A similar DOP model for problem solving (enriched with equational reasoning)
- But U-DOP can probably *not* be used for problem solving (but then: while children all learn language from scratch, they do not learn physics from scratch)

Some questions, by way of conclusion

- What is the biological basis of a unified model that understands new input by decomposing previous input?
- What is shared in the mind, and what is special? E.g. can we apply syntactic learning to periodicity (rhythmic) learning?

Some questions, by way of conclusion

- What is the biological basis of a unified model that understands new input by decomposing previous input?
- What is shared in the mind, and what is special? E.g. can we apply syntactic learning to periodicity (rhythmic) learning?

People involved in this VICI project:

Rens Bod

Federico Sangati

Stefan Frank

Aline Honingh

Vanessa Ferdinand

Carlette Jannink

Gideon Borensztajn

David Cochran

Yoav Seginer (till 2008)

Jelle Zuidema (till 2007, now Veni)

Interested to learn more about this?

Then follow one of my second semester courses (electives):

“Cognitive Models of Language and Beyond”

(for linguists, logicians, cognitive scientists)

“Unsupervised Language Learning”

(for computational linguists, computer scientists, AI-ers)