

Recursion and Grammatical Disambiguation

Roland Hausser
Universität Erlangen-Nürnberg (em.)
email: rolandhausser662@gmail.com
©Roland Hausser, July 3, 2022

Abstract

In substitution-driven sign-based Generative Grammar, there are two kinds of ambiguity, *syntactic* and *semantic*. In continuation-driven, agent-based Database Semantics there is a third kind, exemplified by the ‘garden path’ sentence (Bever 1970) and called *continuation* ambiguity.

In combination with unbounded recursion, continuation ambiguity raises the question of whether (i) the ambiguities accumulate or (ii) recursion $n+1$ disambiguates the continuation ambiguity of recursion n . (i) would be an obstacle to successful communication. Also attempts at reconstructing accumulation in a linguistic example are not successful.

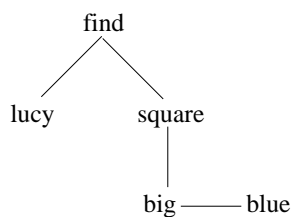
Database Semantics is a linguistic theory of natural language communication which reconstructs a content-surface mapping in the *speak mode* and a surface-content mapping in the *hear mode*.¹ Thereby, the speak mode may have to choose between paraphrases (e.g. active vs. passive) and the hear mode between the readings of an ambiguity (e.g. adnominal vs. adverbial modification).

1 The Speak Mode in Database Semantics

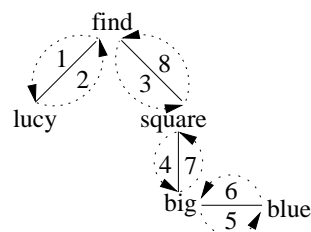
Let us begin with the speak mode as the language variant of an agent’s *action*. The input is a content, the output a language-dependent surface. The drive or motor of the speak mode is a navigation along the semantic relations in the input content:

1.1 GRAPHICAL DBS ANALYSIS OF A SPEAK MODE EXAMPLE

(i) SRG (*semantic relations graph*)



(iii) NAG (*numbered arcs graph*)



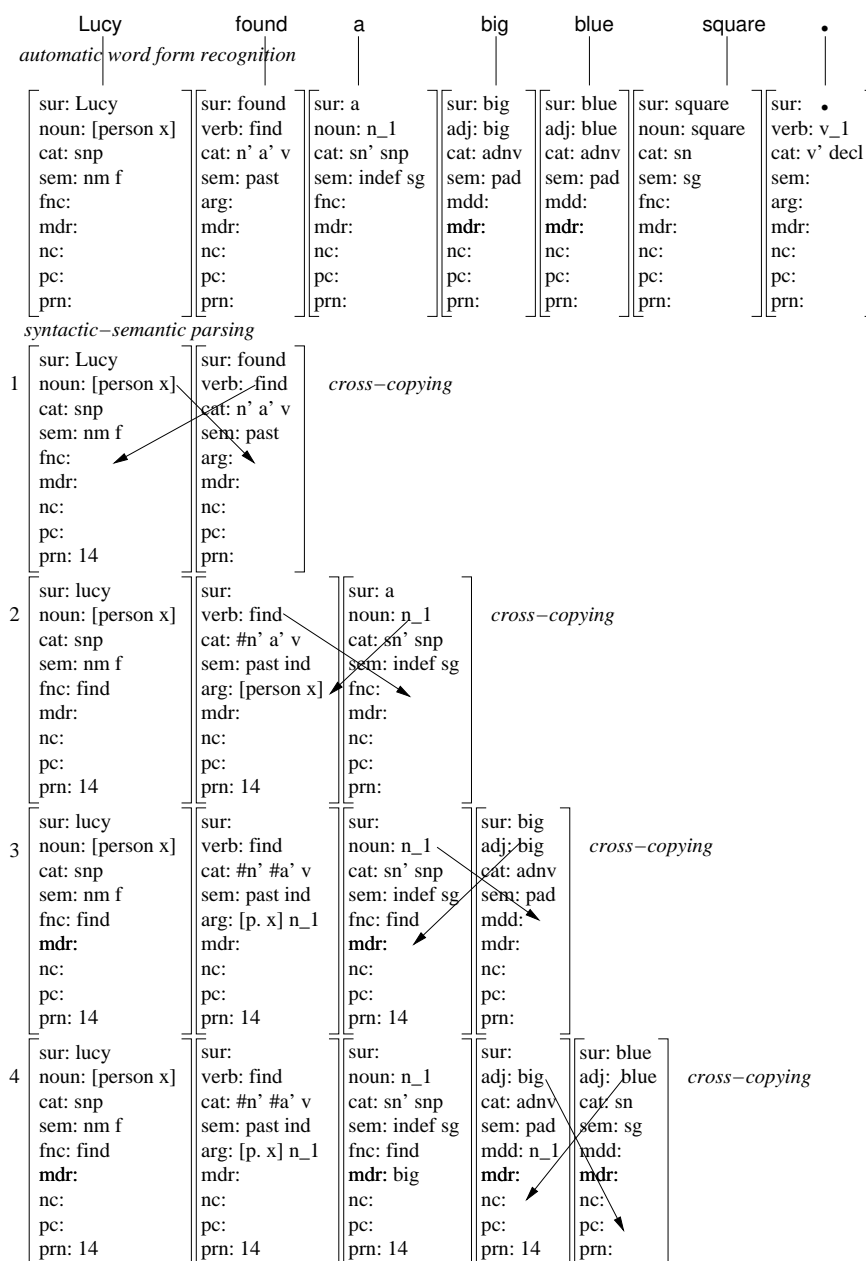
The semantic relations of structure are subject/predicate, object/predicate, modifier/modified, and conjunct–conjunct (Hausser 2022a).

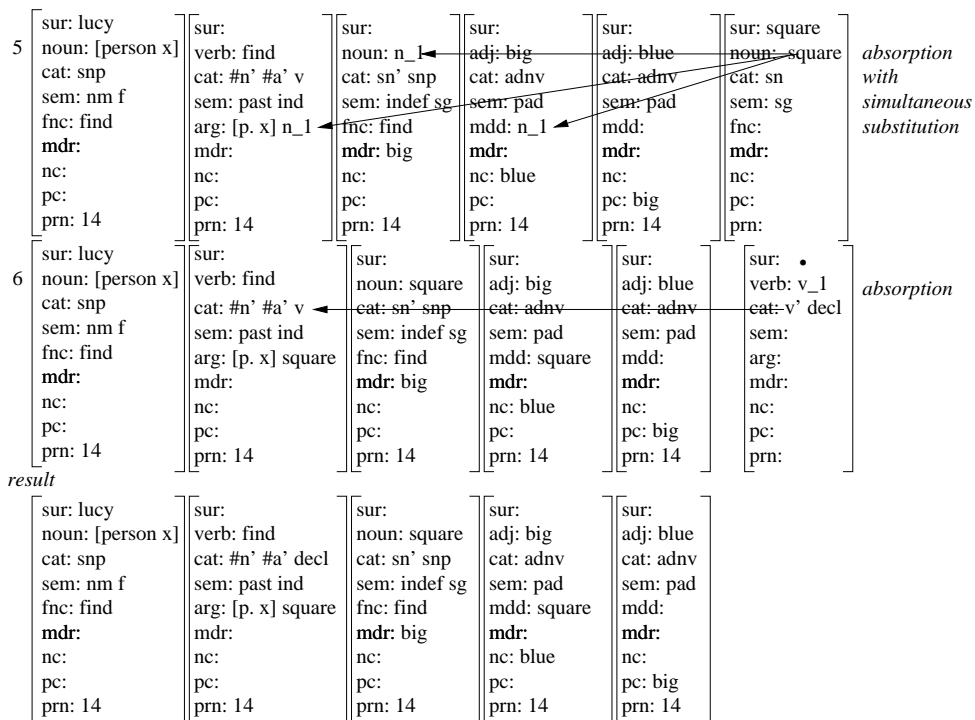
¹This is in contradistinction to Generative Grammar, which aims to randomly generate all well-formed expressions of a natural language from a single abstract input expression, called the S node (Nativism). The ontology of Generative Grammar is sign-based substitution-driven, that of Database Semantics agent-based data-driven.

2 The Hear Mode in Database Semantics

The hear mode is the language variant of an agent's *recognition*. The input is a language-dependent surface and the output a content. The drive or motor is a sequence of word forms in the form of raw data which are recognized by type-token matching in the agent's interface component (Hausser 2021, 12.8.1):

2.1 SURFACE-COMPOSITIONAL TIME-LINEAR HEAR MODE DERIVATION





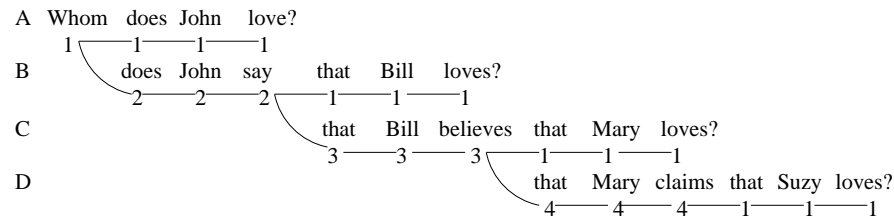
The connectives of the hear mode are \times (cross-copying), \cup (absorption), and \sim (suspension, TExer Sect. 8.3).

The building blocks of a content as input to the speak mode and output of the hear mode are nonrecursive feature structures with ordered attributes,² called *proplets*. By coding the semantic relations in a content by address, the proplets of a content are order-free. This is essential for storage and retrieval in the content-addressable database of DBS³ of artificial cognition (NLC 3.3).

3 Recursion

The following example combines recursion and continuation ambiguity:

3.1 COMBINING RECURSION AND CONTINUATION AMBIGUITY



²This is the opposite to the feature structures popular in Generative Grammar, which are *recursive* with *unordered* attributes (Carpenter 1992). Motivated by a misguided notion of generality, recursive feature structures with unordered attributes are maximally inefficient for computational pattern matching and superfluous in agent-based data-driven DBS.

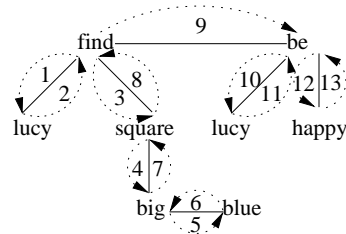
³Called 'A-memory', earlier 'word bank'.

The ambiguities arise in continuing after **John** with a verb taking either a noun or a clause as object. For example, **John** is continuation ambiguous between terminating with **loves Mary** or continuing with **said that**. The ambiguities creating the recursion are strictly local ([-GLOBAL]⁴, FocL 13.3.6).

Because each complete line in 1.1 is unambiguous, the systematic ambiguity originating in this recursion does not affect the linear time complexity of natural language (Hausser 2022b). Furthermore, systematic grammatical disambiguation in recursion holds for natural language in general (universal).

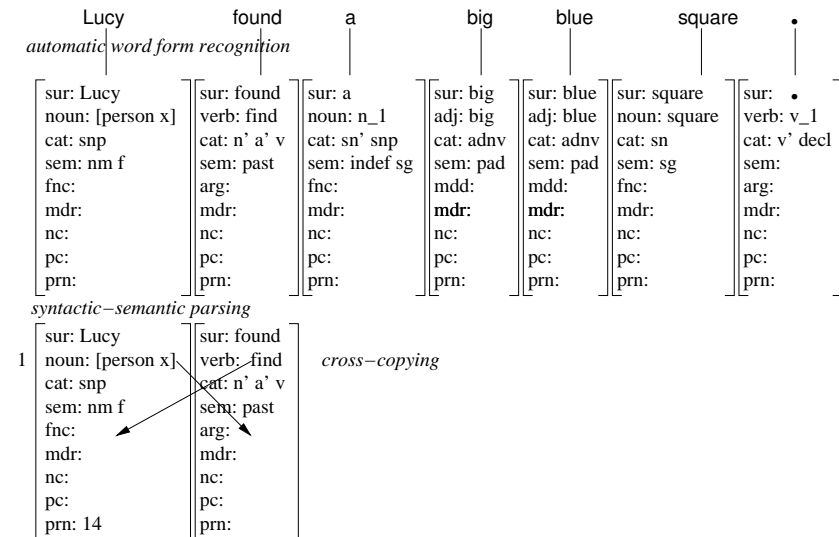
The drive (motor) of the speak mode is a navigation along the semantic relations in a content, called *coactivation*:⁵

3.2 COACTIVATION BY THE TRAVERSAL OF COGNITIVE CONTENT



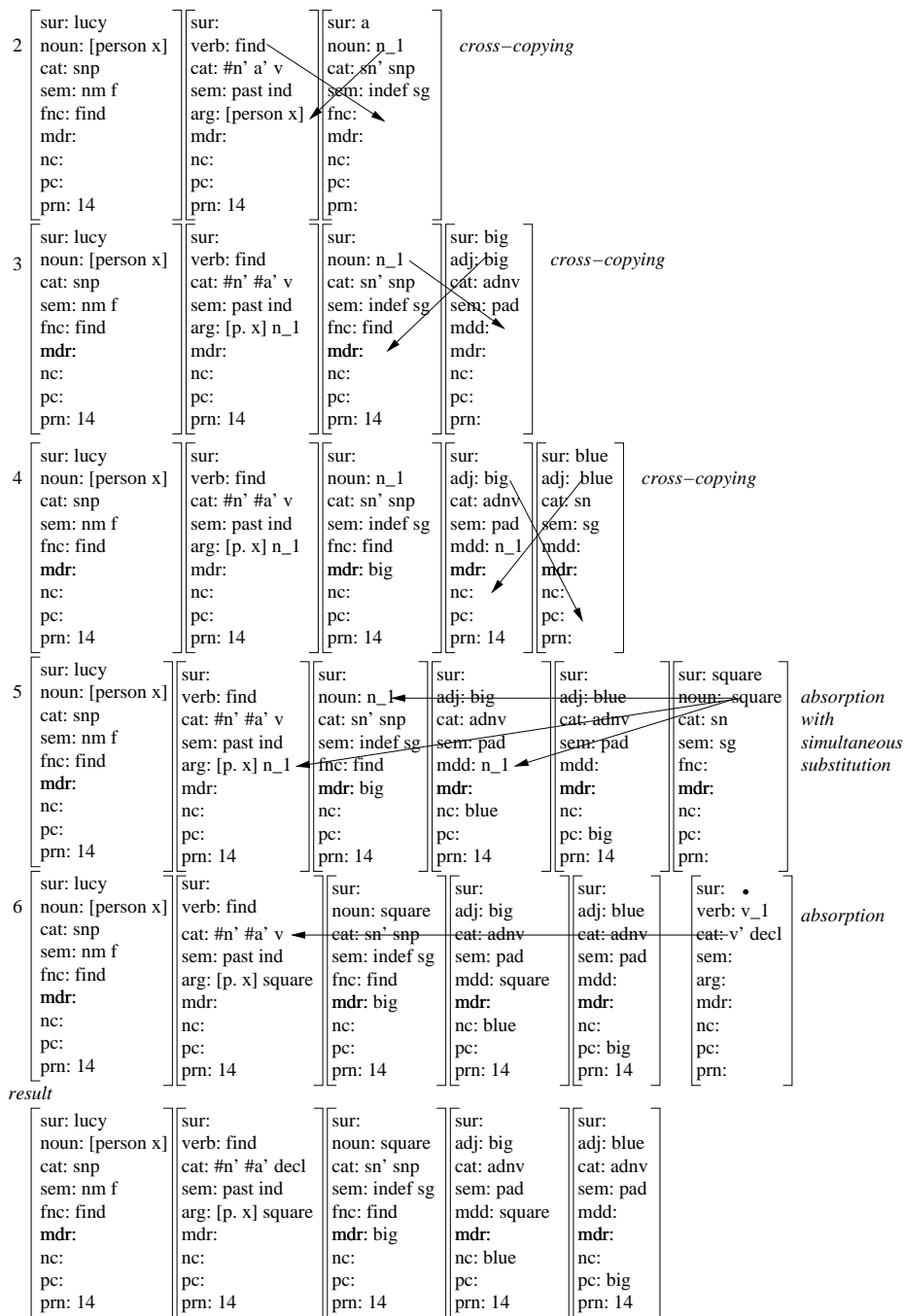
The drive of the hear mode, on the other hand, is the sequence of incoming language-dependent surfaces, as illustrated by the following example:

3.3 TIME-LINEAR SURFACE-COMPOSITIONAL DERIVATION



⁴A classic example of a [-global] ambiguity is the ‘Gardenpath’ sentence “THE HORSE RACED BY THE BARN FELL” by Bever (1970), so-called because the initial interpretation up to **barn** is misleading, as in ‘leading someone down the garden path’. In an era of substitution-driven “Generative Grammar,” Bever’s example is continuation-based, wide awake, and far ahead of its time.

⁵Coactivation in the sense of driving surface production by traversing cognitive content (thought).



The derivation satisfies the two methodological principles of DBS: (i) *surface-compositionality*⁶ and (ii) *time-linearity*⁷.

⁶“The analysis of natural language signs is surface compositional if it uses concrete word forms as building blocks, such that syntactic and semantic properties of a complex expression derive systematically from (i) the lexical properties of the related word forms and (ii) their standard syntactic-semantic composition” (FoCL 4.5.1).

⁷Left-associative derivation order, Aho & Ullman 1977, p.47.

4 What does it take to be a Linguistic Universal?

For a theoretical claim to be a linguistic universal, it must be without a counterexample. Empirically, this is hard to prove. For example, even if a broad-based investigation of ambiguity in unbounded recursion does not come up with a clear example of accumulation, the claim of nonexistence can not be conclusive because there remains the *possibility* of having overlooked such an instance. There remains, however, the functional argument: unbounded ambiguity with unbounded recursion is unlikely because it would be an obstacle to successful communication.

5 Conclusion

In natural language recursion, continuation ambiguities do not accumulate, but are disambiguated by the next recursion step (3.1). This supports successful communication on the side of the hearer.

Bibliography

- Aho, A.V. and J.D. Ullman (1977) *Principles of Compiler Design*, Reading, Mass.: Addison-Wesley
- AIJ'01 = Hausser, R. (2001) "Database Semantics for Natural Language," *Artificial Intelligence Journal*, 130.1:27–74, Elsevier
- Bever, T. (1970) "The cognitive basis for linguistic structures," in R. Hayes (ed.) *Cognition and Language Development*, New York: Wiley & Sons, p. 279–362
- Carpenter, B. (1992) *The Logic of Typed Feature Structures*, Cambridge: CUP
- FoCL = Hausser, R. (1999) *Foundations of Computational Linguistics, Human-Computer Communication in Natural Language*, 3rd ed. 2014, Springer
- Hausser, R. (2021) "12. Language vs. Nonlanguage Cognition," lagrammar.net
- Hausser, R. (2022a) "Outline of Database Semantics," *China Terminology* (3rd issue, 2022), <http://www.term.org.cn>
- Hausser, R. (2022b) "Grammatical Disambiguation, The Natural Language Linear Complexity Hypothesis," *Language and Information*, Korean Society of Linguistics, in print
- NLC = Hausser, R. (2006) *A Computational Model of Natural Language Communication – Interpretation, Inferencing, and Production in Database Semantics*, pp. 360, Springer.
- Peters, S., and R. Ritchie (1973) "On the Generative Power of Transformational Grammar," *Information and Control*, Vol. 18:483–501
- TCS'92 = Hausser, R. (1992) "Complexity in Left-Associative Grammar," *Theoretical Computer Science*, Vol. 106.2:283-308, Elsevier