

VECTOR SEMANTICS: LECTURE 5

András Kornai
SZTAKI Computer Science Research Institute

17 April 2024

DEFINITION SYNTAX

- Definitions have their own syntax: we defined *abash* as 'make (someone) feel embarrassed, disconcerted, or ashamed'
- Aside from a few genre-specific peculiarities, dictionary syntax is not very different from ordinary English syntax
- The peculiarities are mostly driven by the need to express the syntactic or conceptual category of the definiendum: *used to, of or about, to be, someone who, relating to, done as, a way of, according to, to make, something that, a type of, the process of*
- The syntax is recursive: for example the above definition is equivalent to 'make (someone) feel embarrassed or make (someone) feel disconcerted or make (someone) feel ashamed'
- If we want substitutability, the recursive aspect is even more marked: if X is defined as $Y Z$; Y as $A B$; and B as UV , we must have $X = (A(UV))Z$

SYNTAX IN 4LANG

- This suggests string rewriting techniques (derivation trees)
- 4lang uses more complex structures, hypernode graphs rather than trees
- There are two reasons: first, unification brings nodes together, which would naturally lead to DAGs; and second, argument slots can be filled by entire edges, not just nodes
- Example 1: Accusativus cum Infinitivo in Latin (Kálmán and Kornai, 1985)

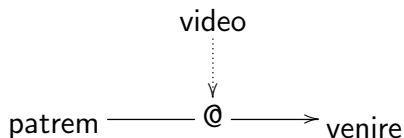
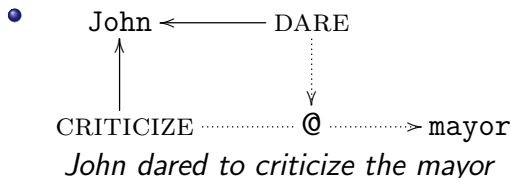


FIGURE: *Video patrem venire*

- How do you know that it's *father's coming* that is being seen, not just *father* (who happens to be coming)?

ACI CON'T

- Many other verbs like *credo* 'believe', *spero* 'hope', *doco/aio* 'say/affirm', *puto* 'think' etc. participate in the same ACI construction, and here the 'I believe/hope/say... father' reading is obviously not available. It must be the entire *father's coming* that is being believed/hoped/...
- In Hungarian this is archaic, but still quite understandable:
Ha zörren egy levél, porozlót jönni vél
- More important, there are a large number of verbs/auxiliaries in many languages that take clausal/sentential objects, and the same logic applies:



MAIN FEATURES OF THE SYNTAX

- 4lang uses RDF-style (Resource Description Framework) linearization. RDF is a widely used standard maintained by the WorldWideWeb Consortium (W3C)
- The key idea is using SVO (subject-verb-object) triples, with recursion permitted only for the subject and object arguments (John DARE (John CRITICIZE mayor))
- This is sufficient for our goals (more on this claim later) and does away with *hyperedges* (edges that touch several nodes) in favor of ordinary edges (between starting node and endnode) while keeping *hypernodes* (subgraphs that act as a single node)
- We need a low-level unification operation that makes sure that the matrix subject John is the same as the John who criticizes the mayor
- Unification is blocked by the keyword *other*, cf
The car overtook the other car

LINKING

- Overall, this syntax is considerably simpler than the one used in many grammatical formalisms, because it uses only subjects and objects, but there are no indirect objects, datives, sources, goals, etc. or distinct thematic roles (other than =agt, =pat), deep cases, kārakas etc.
- It is not that surprising that we have subjects and objects. What is more surprising is that 41ang can manage without extra linkers (Kornai, 2012)
- All the other linkers are handled as relations with their own (prepositional) subject and object
- The linkers are, at least in some languages (Hungarian, Latin, Polish, Japanese), overt morphemes, nominative and accusative case endings, in others like English they are positionally encoded (preverbal is agent, postverbal is patient)
- =agt, =pat are neither vectors nor matrices, they are perturbative terms on the evolving thought matrix

FURTHER SYNTACTIC SUGAR

- The comma ',' stands for conjunction, there is no abbreviation for disjunction or negation
- Angled brackets '<>' denote defaults (optional, but present unless there is evidence otherwise) *sugar* material, sweet, <white>, in food, in drink
- Curly braces '{ }' enclose complex clauses
- Square brackets '[']' used for direct predication: A[B] means A IsA B. We also use B(A) for the same thing
- @ used for external pointers (to WP) *London* city, @London
- Position marker (substitution focus) ..., phonological info (rare) in doublequotes
- 1 and 2 clauses in S-V-O triples
- `def_ply_parser.py` (could be improved to generate dot-style file)

CONCEPTUAL SCHEMAS

- Closely related to Tomkins/Abelson/Schank scripts
- Sequences of *frames*. Restaurant script 1. People go to restaurant. 2. Maitre d' seats them. 3. Waiter brings menu. 4. People order. 5. Food is served. 6. People eat it. 7. They ask for the check 8. Waiter brings check. 9. They pay. 10. They leave
- Temporal sequencing handled by FSA
- Simplest schemes have single frame or two
- Even these may have *before*, *now*, and *after* clauses
- Both the entirety of the frame and the slots are often hard to name
- This makes invoking the frame/schema tricky!

THE “COMMERCIAL EXCHANGE” SCHEMA

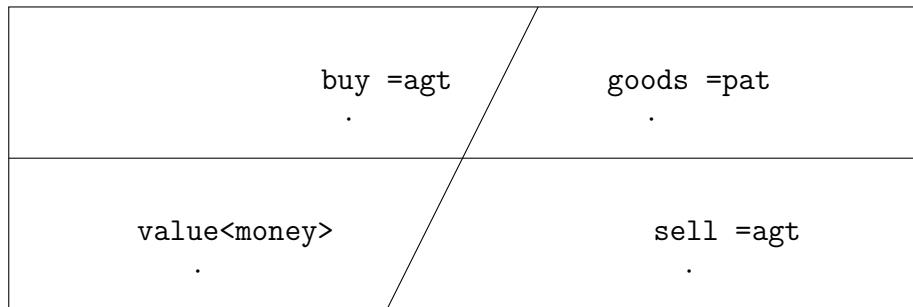





FIGURE: exchange_

GEN 25:29-34

- 29 And Jacob sod pottage: and Esau came from the field, and he was faint:
- 30 And Esau said to Jacob, Feed me, I pray thee, with that same red pottage; for I am faint: therefore was his name called Edom.
- 31 And Jacob said, Sell me this day thy birthright.
- 32 And Esau said, Behold, I am at the point to die: and what profit shall this birthright do to me?
- 33 And Jacob said, Swear to me this day; and he sware unto him: and he sold his birthright unto Jacob.
- 34 Then Jacob gave Esau bread and pottage of lentiles; and he did eat and drink, and rose up, and went his way: thus Esau despised his birthright.

THE LINGUISTIC TAKEWAY

- This whole story makes no sense whatsoever unless we interpret it in the exchange_ frame
- It takes a great deal of work to link up the language with the schema: only in 31, *sell me thy birthright* is the schema triggered
- This already requires sophisticated anaphor resolution: we need to know that Esau will be *seller*, Jacob will be *buyer*.
Morphology for nomen agentis -er/3627 stem_-er is_a =agt, "_ -er" mark_ stem_
- *sell* =agt cause_ buyer has =pat, buyer cause_ =agt has money_, dative_ mark_ buyer
- We need to figure out the other two slots, that the goods are the birthright, and the thing of value is the bowl of lentils.
- The mechanism is **coercive**, it is not that Esau IS a seller, Esau is *the* seller, Esau sell birthright is a valid inference.

-  Kálmán, László and András Kornai (1985). *Pattern matching: a finite state approach to generation and parsing*.
-  Kornai, András (2012). "Eliminating ditransitives". In: *Revised and Selected Papers from the 15th and 16th Formal Grammar Conferences*. Ed. by Ph. de Groote and M-J Nederhof. LNCS 7395. Springer, pp. 243–261. DOI: 10.1007/978-3-642-32024-8_16.
-  — (2023). "Poliszémia politópokkal". In: *Általános Nyelvészeti Tanulmányok* 35. Ed. by Beáta Gyuris, pp. 311–326. ISSN: HU 0569-1338.