Syntax-centered and semantics-centered views of discourse. Can they be reconciled?

Matthias Buch-Kromann and Iørn Korzen

Copenhagen Dependency Treebank Workshop
Copenhagen Business School

23-24 August 2010

The problem: reconciling syntax and discourse

Syntax

Tree structure augmented with word order mechanisms + secondary relations + coreference relations.

Discourse (RST)

Discourse modelled as a tree with nucleus-satellite relations. Expresses centrality wrt. speaker-intentions.

Discourse (PDTB)

Explicit and implicit discourse connectives: functors with clausal arguments (sometimes anaphoric).

Goals

Syntax and discourse both seek to account for linear order and compositional semantics, but at different levels.

Question

What is the relationship between syntax and discourse?

The blurry syntax-discourse boundary

The blurry boundary

- He laughed. That annoyed me.
- He laughed, which annoyed me.
- He annoyed me by laughing.
- His laugh annoyed me.

(Noted by Carlson and Marcu, 2001).

Overlapping units

- Attribution verbs take a complement which may be an entire discourse.
- A long discourse may elaborate on an NP.
- A discourse may function as a sentence-internal parenthetical remark.

Syntax and discourse: differences and overlap

Units and relations

- Syntax: words linked by relations.
- Discourse: clauses linked by relations.

Overlap

- Overlapping annotations in multi-clause sentences.
- Differences not a problem, if discourse trees viewed as semantics, not syntax.
- But similar function. Do they agree? If not, why?

Subordination

Discourse sometimes reverses direction of subordination:

- discourse connectives are often functors (PDTB)
- attribution verbs are satellites (RST)
- some subordinate clauses analyzed as nuclei (RST)

Discourse often assumed to lack crossing relations, multiple dependencies.

PDTB departure from syntax

PDTB example (PDTB manual, (36))

In an invention that drives Verdi purists bananas, Violetta lies dying in bed during the prelude, rising deliriously, **when** then **she remembers the great parties she used to throw.**
Legend: *argument1*, connective, **argument2**

Discourse analysis (eg, PDTB)

- "when" is the functor/head/lexical anchor of one discourse relation
- "then" is the functor of another discourse relation
- the analysis is a non-tree

Syntax analysis

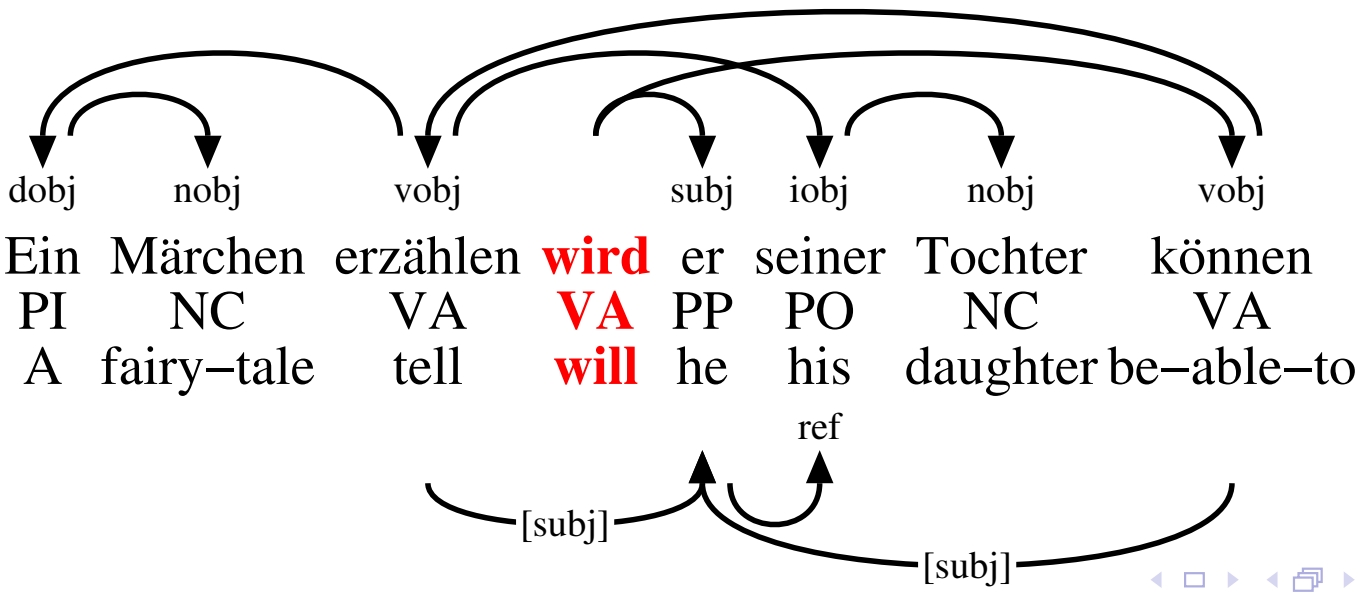
- "when" is a subordinating conjunction
- "then" is an (anaphoric?) adverbial modifying "remembers" VP
- the analysis is a tree

Main mechanisms in syntax

Syntax tree with additional mechanisms for...

- non-projective word order (eg, topicalization)
- multiple heads + gaps (eg, control, relatives, gapping)
- anaphora and binding
- functor-argument structure

Near-agreement about abstract analyses across frameworks (GB, LFG, HPSG, LTAG, dependency, CCG), significant differences in detailed analyses and mechanisms.



Syntax and compositional semantics

Dependency tree

Functor argument tree (Dowty)

Relation

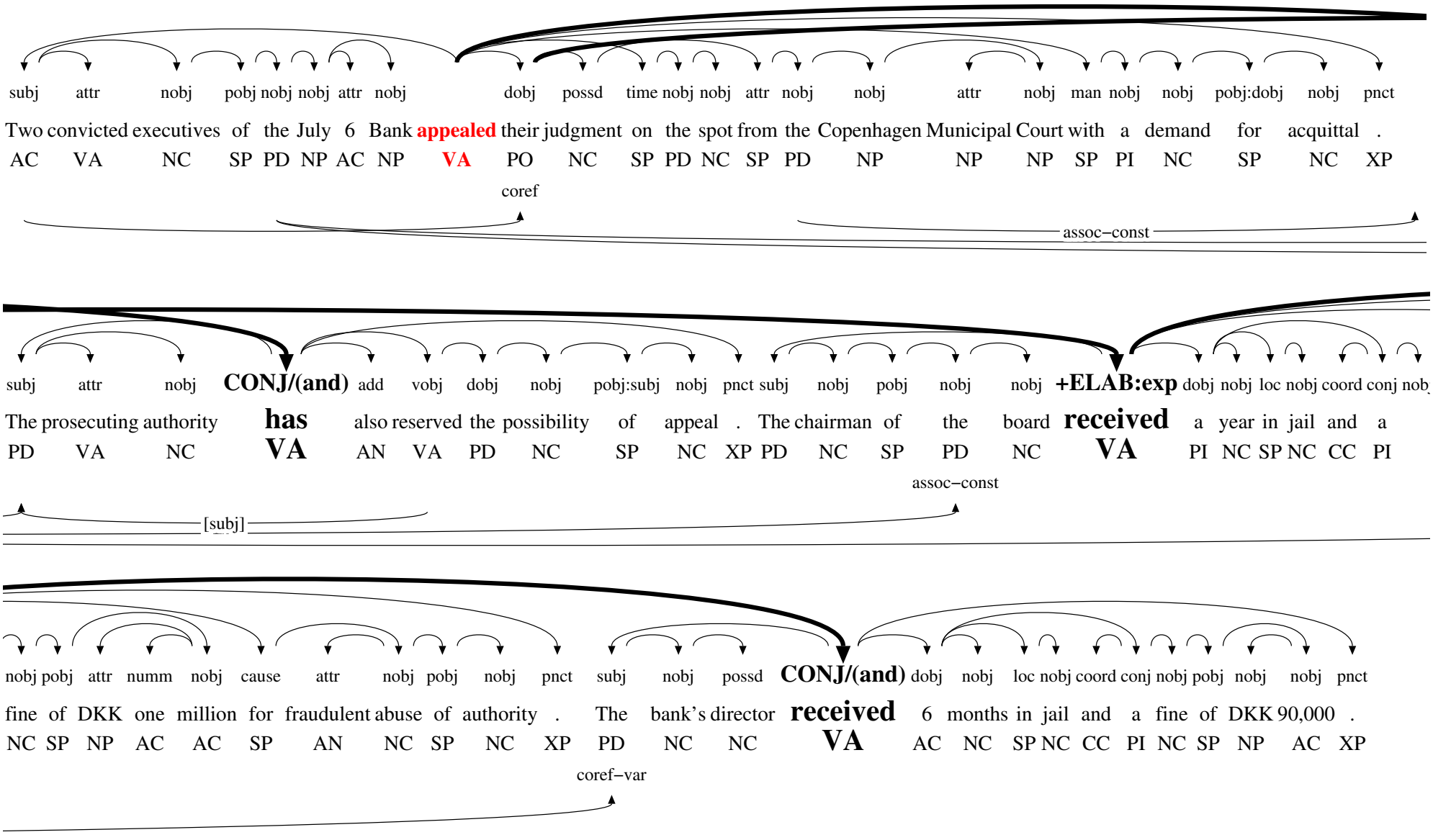
functor-argument tree = dependency tree + scope order

Four syntactic analyses of connectives

Four analyses of connectives (eg, discourse connectives) with essentially identical functor-argument structures. The syntactic head determines the syntactic and semantic type, the semantic head is the highest functor in the functor-argument tree.

D-LTAG/PDTB analyzes discourse connectives as heads (“although”), anaphoric heads (“as a result”), or conjunctions (“and”, “but”).

CDT discourse analysis

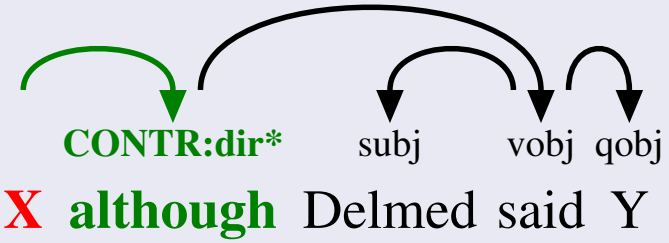


Discourse problems 1

Problems with attribution: Dinesh et al 2005

The current arrangement ends in March 1990, although Delmed said **it will continue to...**

CDT analysis



Effectively relation both between reported events and reporting events.

CDT compositional semantics

- $K_a\phi$: epistemic formula (agent a knows ϕ)
- π: operator with $\pi(K_a\phi) = \phi$.
- $c_{ij}(X', Y') = c(\pi^i(X'), \pi^j(Y'))$ where c is composition function for discourse relation C.
- annotate relation c_{ij} as iCj (using '*' to count i and j).
- multiple composition functions in marker analysis anyway.
- can deal with recursion: Delmed said John believed Mary thought...

Discourse problems 2

Crossing relations (eg, Wolf and Gibson 2005)

Allowed in CDT framework for both syntax and discourse, controlled by surface trees.

PTB-PDTB disambiguation

Dinesh et al (2005) report some PDTB-PTB mismatches caused by different disambiguation choices in the two treebanks.

PTB-PDTB granularity

Lee et al (2006) report some mismatches caused by different granularity (words/clauses).

Discourse problems 3

Shared arguments (Lee et al)

"(X but [Y) so Z]": *Z* not part of argument to "but".

CDT allows compositional semantics to strip off "so Z" from "Y so Z" (a mechanism needed for parentheticals anyway, eg, using a dynamic semantics to model these side-effects).

Other problems (Lee et al)

Properly contained arguments, pure crossings, partially overlapping arguments can be explained similarly by composition mechanisms for parentheticals, coupled with crossing, adverbial anaphors, and syntactic dependencies for missing links.

Discourse problems 4

Speaker intentions + non-crossing (Stede 2008)

Some of Stede's counter-examples are of the form "X Y Z" with "X Y" multinuclear, Z with forced attachment to X or Y. Non-crossing forces head of Z to be the head of "X Y".

Not a problem in CDT because crossing is allowed.

Intentional structure (Stede 2008)

Stede gives examples where syntactic structure conflicts with a tree that encodes the intention of the text (which segment is most central to the writer's purposes).

CDT focuses on syntactic structure, and does not seek to capture intentional structure.

CDT compared to other discourse approaches

Syntax vs. semantics

Seeks to model syntax. Tends to be right-branching. Functor-argument structure induced.

Scope

Modifier scope is underspecified instead of constituency. Functor-argument structure more directly specified in PDTB.

Analysis of connectives

A more syntactic analysis (connectives, markers, anaphoric adverbials) compared to the functor-analyses in D-LTAG.

Mechanisms

Assumes crossing and richer compositional mechanisms for attribution and parentheticals.

Conclusion: Semantics- vs. syntax-centered view

Shared assumptions about discours structure

- controls compositional semantics
- controls linear order
- tree, or close to tree

Semantics-centered view

- clausal units
- unambiguous wrt. argument structure (but non-tree)
- RST: more central to the writer's purposes = high in tree

Syntax-centered view

- lexical units
- links sentence trees into one big discourse tree
- induced functor-argument structure
- no notion of centrality

Why viewing discourse structure as a tree?

Analogy with syntax

- controlling text order
- controlling compositional semantics
- artificial syntax-discourse boundary
- avoiding the blurry syntax-discourse interface
- sticking to the principle of compositionality may provide information about compositional semantics

Analogy with parsing and NLP

- joint parsing and discourse parsing (reusing parsing models)
- cognitive modelling

Other reasons

- model structural relationships between languages
- theory-driven science: our data are not observable, but theoretical constructions