Applicative Structures and Immediate Discourse in the Turkish Discourse Bank

Işın Demirşahin, Adnan Öztürel, Cem Bozşahin and Deniz Zeyrek

Cognitive Science Department, The Informatics Institute, Middle East Technical University (METU/ODTÜ), Ankara

ii.metu.edu.tr

LAW VII & ID, Sofia, ACL 51

Aug 9, 2013

Supported by BAP–Scientific Research Fund of METU Informatics Institute

Nature of Discourse Representation

- Entity chains (Knott et al. 2001)
- A single tree of discourse relations (Rhetorical Structure Theory, RST, Mann and Thompson, 1988)
- Successive trees of varying sizes connected and occasionally intertwined at the peripheries (Hobbs, 1985)
- Adjoined trees (Discourse-Lexicalized Tree Adjoining Grammar, D-LTAG, Webber, 2004)
- Directed acyclic graphs (Lee et al., 2006, 2008)
- Chain graphs (Wolf and Gibson, 2005)
- S/DRT: inferential

- Sentence-level structures require more than context-free power,
  but not to the extent of dealing with general graphs,
  or with strings that grow out of constant control
  Shieber (1985); Joshi (1985)
- Do we need something above context-freeness for discourse?
- Can we decide by looking at ways of combination?
  application, composition and their associated semantics

- How far do discourse structures deviate from applicative semantics?
- Applicative structures: binary operations on data

  e.g. a connective's meaning depending only on two arguments.
- A system is applicative if it makes use of function application only, not graph reduction or general function composition.
- Function application can be linked to applicative structures by currying.
- Do we need that in discourse annotation?
- If more complex structures are found, we must go above applicative semantics, to do function compositions and graph reductions, which are known to require more computational power.

The level of discourse which can be studied "just above" syntax. Joshi (2011)

Before (or without) reaching for higher-level mechanisms such as goals, motives.

Argument-taking, semantics, heads, nature of embedding

- Look for signs of structures in corpus over and above trees
- Tree-violating configurations (Lee et al., 2006, Aktaş et al., 2010)
- in terms of:

  surface configuration

  semantics (application v. composition, connective v. anaphor)

- Turkish Discourse Bank (TDB): 400,000+ words, various genres (Zeyrek and Webber, 2008; Zeyrek et al., 2010)
- Annotated in the style of PDTB (Prasad et al., 2008).
  - CONN: the connective that sets up the discourse relation (*but, and, instead*)
  - Arg2: the clause to which the CONN syntactically belongs.
  - Arg1: the other argument.
- Non-structural relations are annotated as such (discourse adverbials, phrasal expressions)
- They take one argument; the other is inferential.
- Abstract Objects (Asher, 1993)

Anaphoric relation

file name–browser index: 00010111-54&55

Sonra ansızın sesler gelir. Ayak sesleri. Birilerinin ya işi vardır, aceleyle yürürler, ya koşarlar. O ZAMAN kız katılaşır ansızın. Oğlan da katılaşır ve her koşunun gizli bir isteği var.

"And then suddenly there is a sound. Foot steps. Someone has an errand to run, they walk hurriedly or run. THEN the girl stiffens suddenly. The boy stiffens, too; and every run has a hidden wish."

Structural relation

file name–browser index: 00001131-2&3

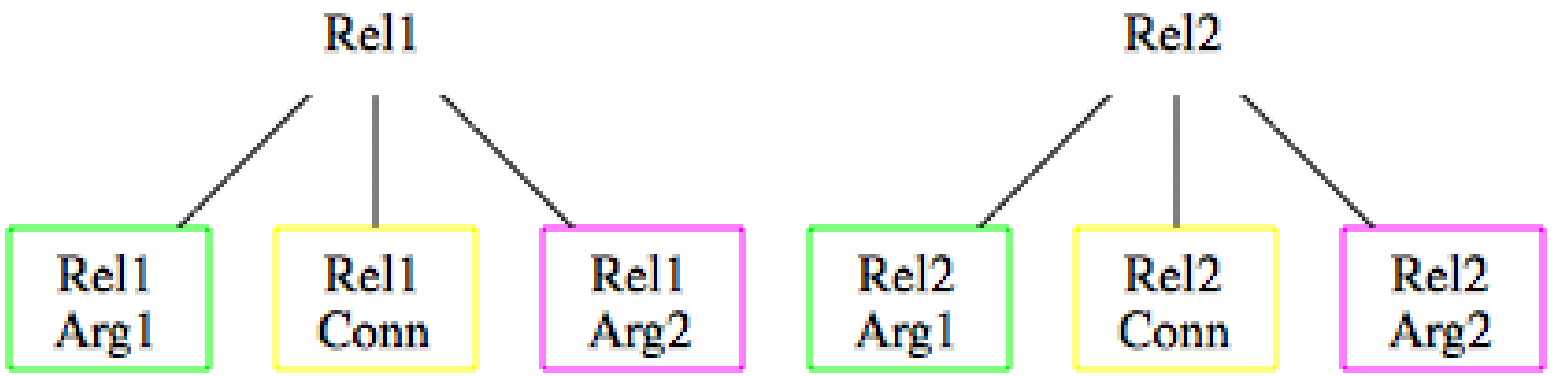
Vazgeçmek kolaydı, ertelemek de. AMA tırmanmaya başlandı mı bitirilmeli! Çünkü her seferinde acımasız bir geriye dönüş vardı.

"It was easy to give up, or to postpone. BUT once you start climbing you have to go all the way! Because there was a cruel comeback everytime."

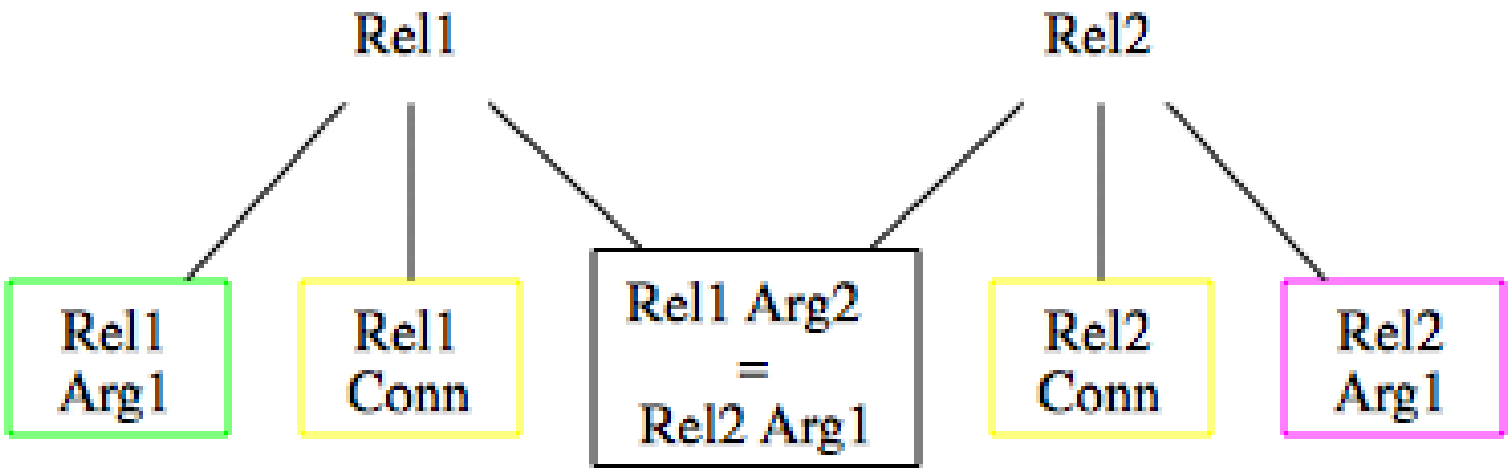
Summary of TDB relations

Configuration	**#**	**%**
Full Embedding	695	27.28
Nested Relations	138	5.42
Total Non-violating Configurations	**833**	**32.69**
Shared Argument	489	19.19
Prop. Cont. Argument	194	7.61
Prop. Cont. Relation	1018	39.95
Pure Crossing	2	0.08
Partial Overlap	12	0.47
Total Violating Configurations	**1715**	**67.31**
Total	**2548**	**100.00**

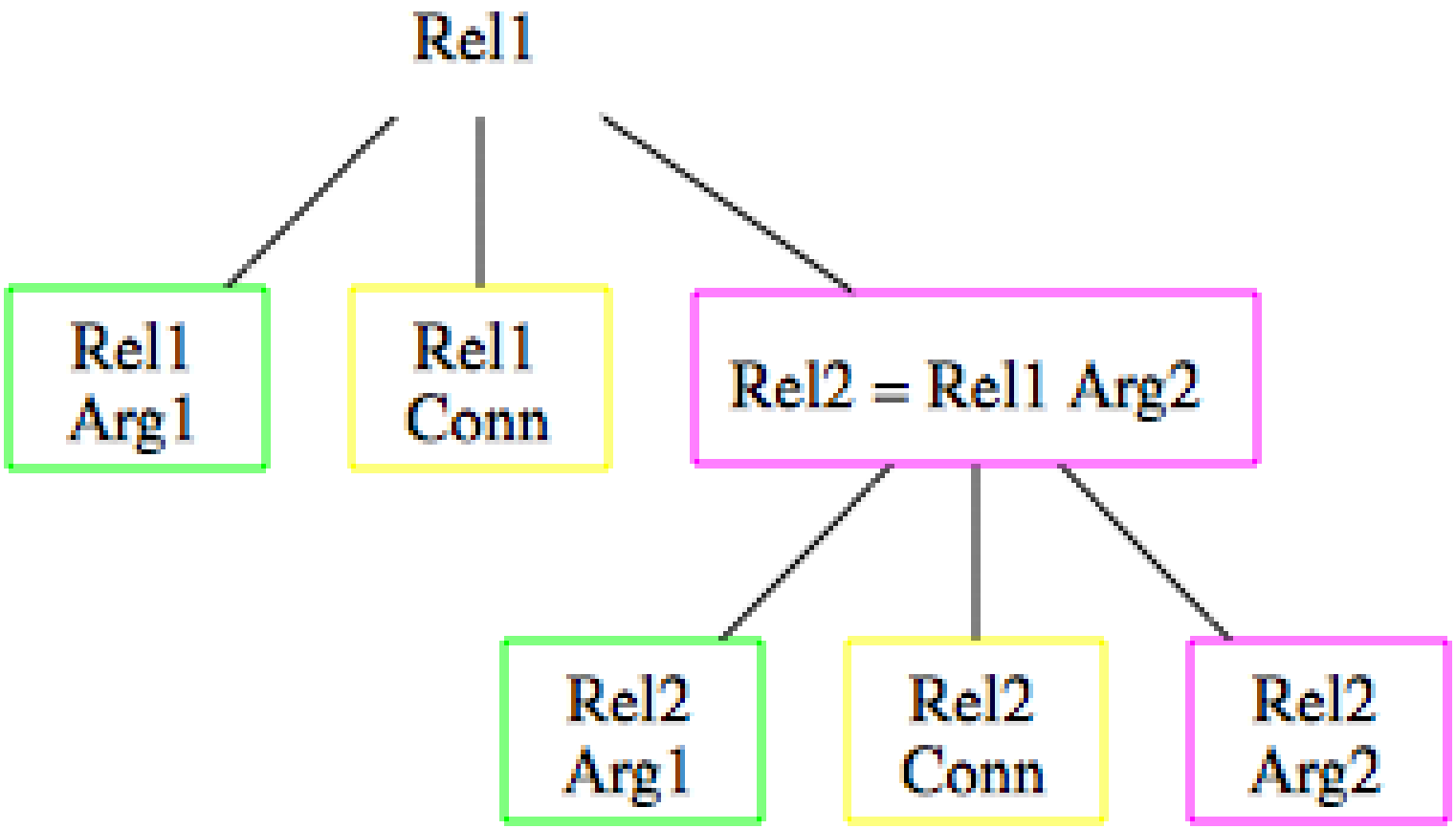
Independent relations



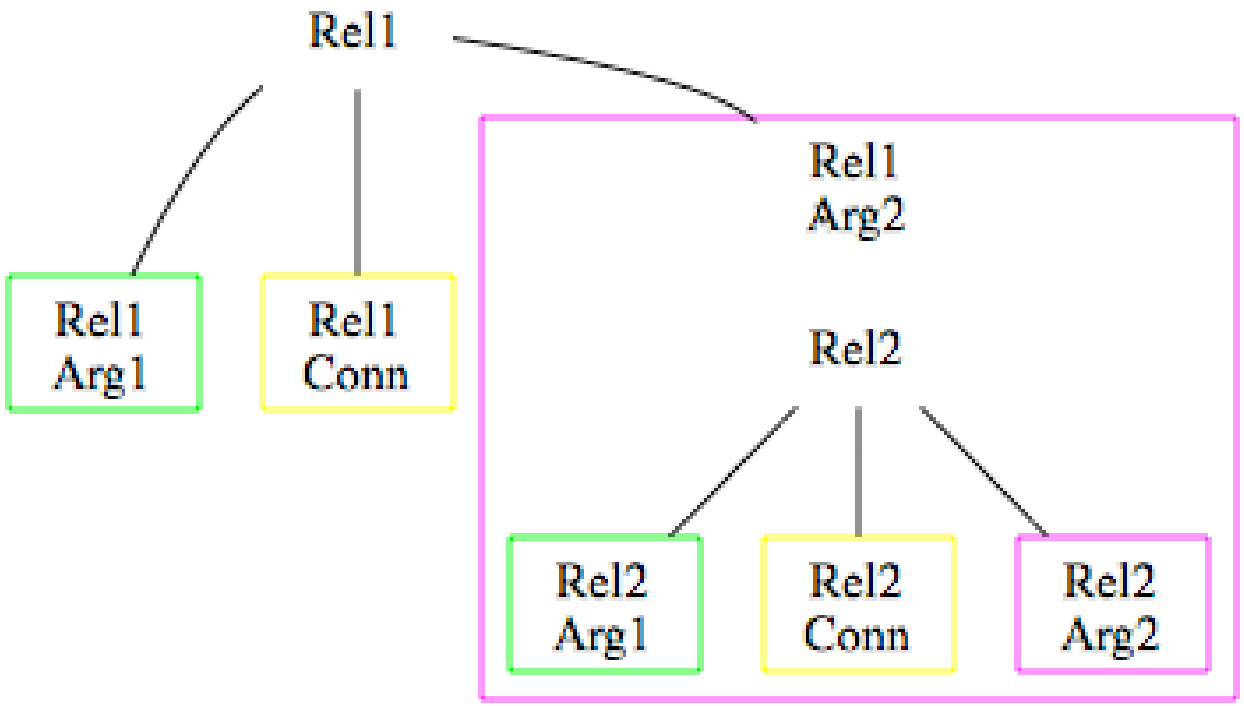
Shared arguments



Full embedding

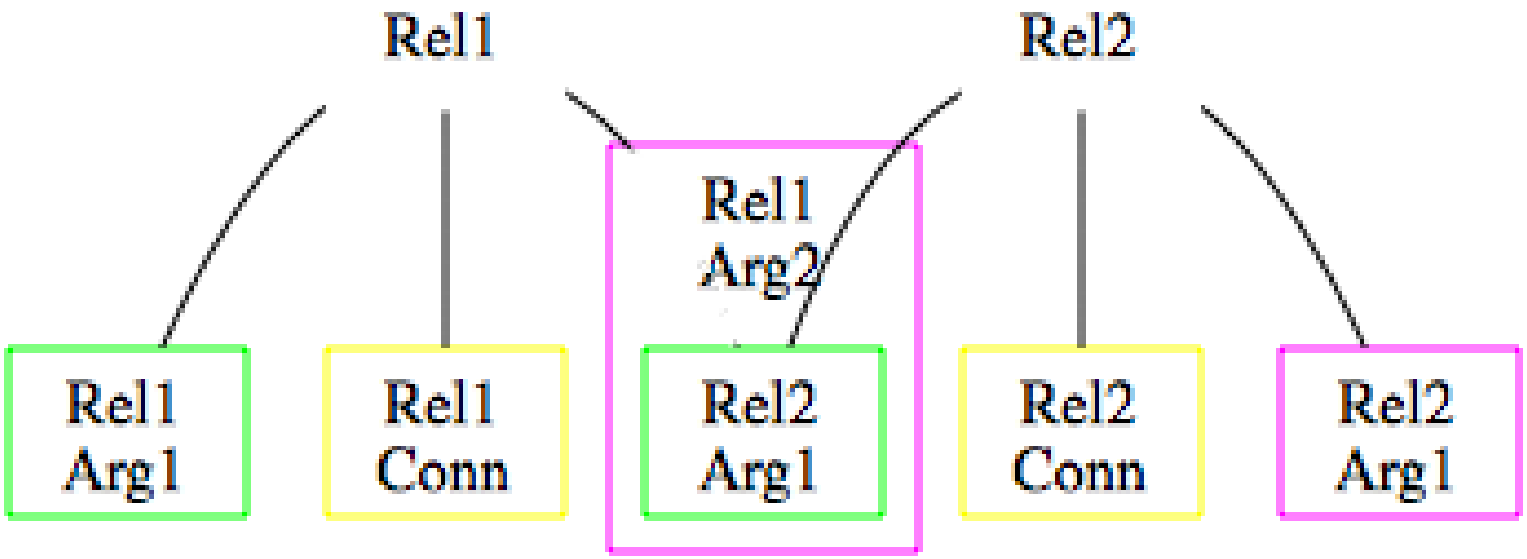


Properly contained relations



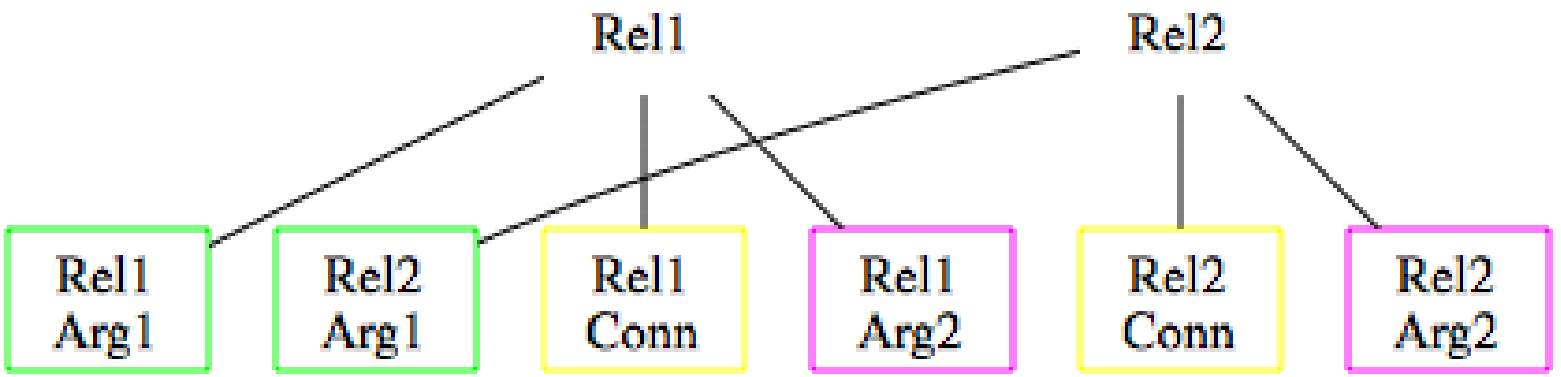
Rel1's Arg2 contains Rel2, and more material

Properly contained argument



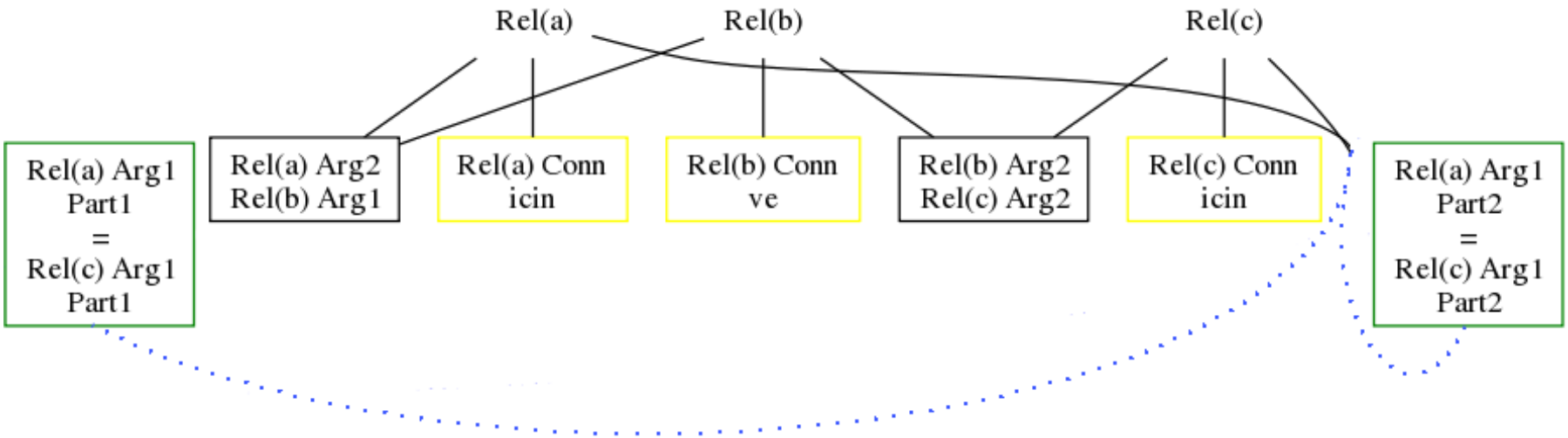
Partial overlap

Crossing


Rel1
Rel2
Rel1 Arg1
Rel2 Arg1
Rel1 Conn
Rel1 Arg2
Rel2 Conn
Rel2 Arg2

Double subordinator

Subordinator (*için* 'because') is repeated. Without repetition, full embedding.



example

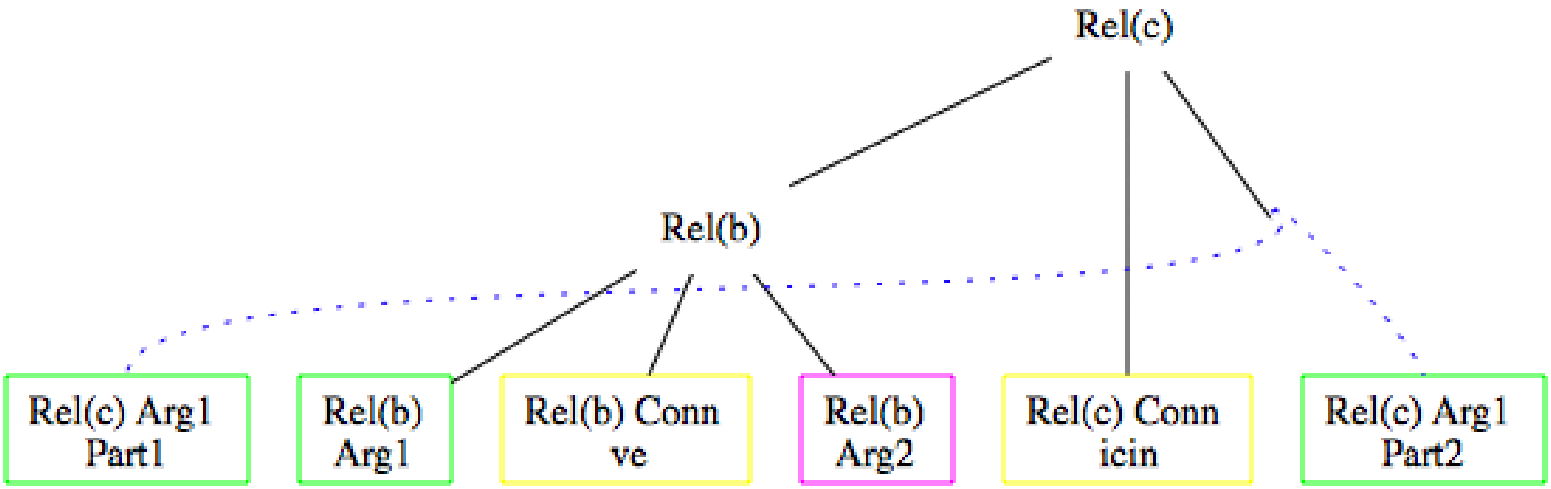
file name–browser index: 20510000-31,32&34

Ceza, Telekom'un iki farklı internet alt yapısı pazarında tekel konumunu kötüye kullandığı IÇIN ve uydu istasyonu işletmeciliği pazarında artık tekel hakkı kalmadığı halde rakiplerinin faaliyetlerini zorlaştırdığı IÇIN verildi.

"The penalty was given BECAUSE Telekom abused its monopoly status in the two different internet infrastructure markets and BECAUSE it caused difficulties with its rivals' activities although it did not have a monopoly status in the satellite management market anymore."

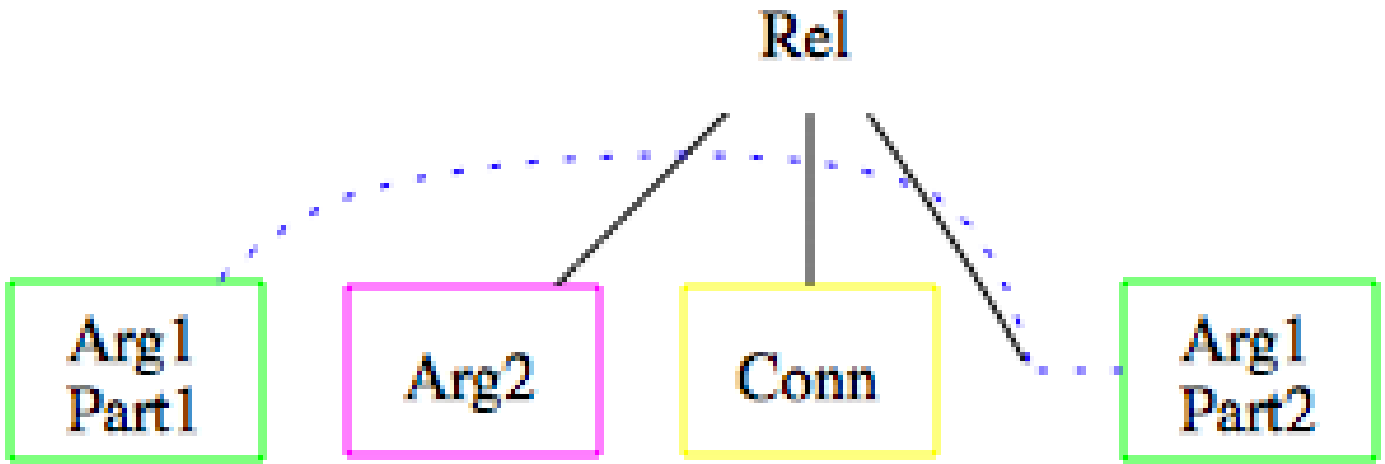
Single subordinator (hypothetical)

Apparent tree-violation if we take one *because* and one *and* relation:

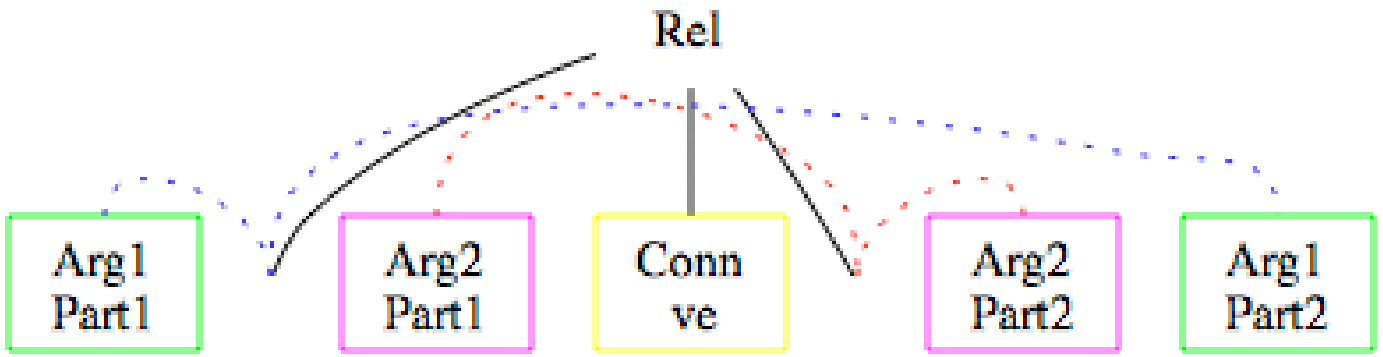


This was not available to annotators. ML can help here.

Wrap



Double wrap



Challenges to surface applicativity

Configuration	Struc-tural	Ana-phoric	Total
Shared Argument	158	331	489
	32.31%	67.69%	100.00%
Prop. Cont. Arg.	65	129	194
	33.51%	66.49%	100.00%
Prop. Cont. Rel.	547	471	1018
	53.73%	46.27%	100.00%
Pure Crossing	1	1	2
	50.00%	50.00%	100.00%
Partial Overlap	9	3	12
	75.00%	25.00%	100.00%
Total	780	935	1715
	45.48%	54.52%	100.00%

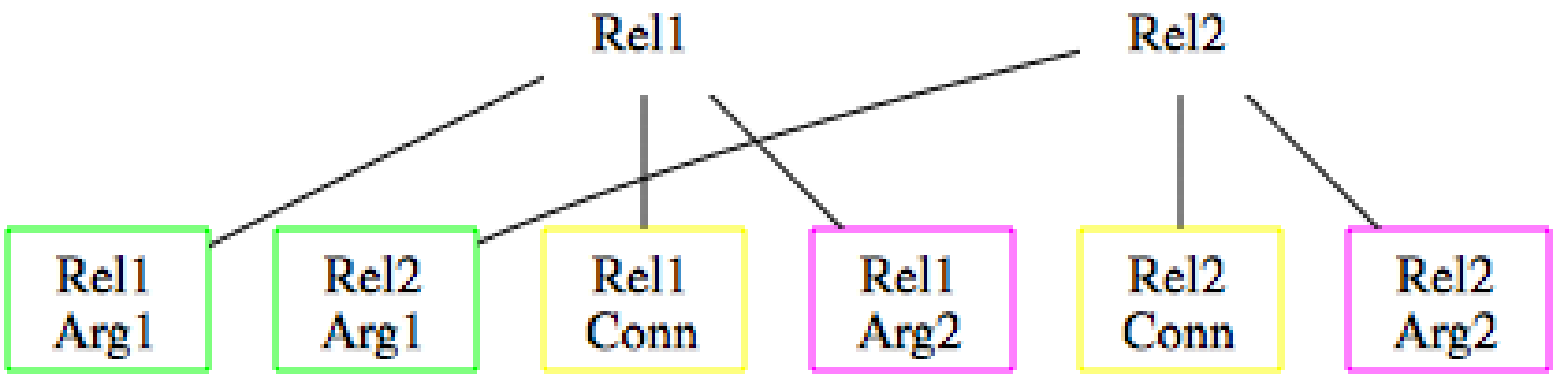
Earlier example has two connectives

file name–browser index: 00010111-54&55

Sonra ansızın sesler gelir. Ayak sesleri. [Birilerinin ya işi vardır, aceleyle yürürler, ya koşarlar.] O ZAMAN kız katılaşır ansızın. Oğlan da katılaşır VE [her koşunun gizli bir isteği var.]

"And then suddenly there is a sound. Foot steps. [Someone has an errand to run, they walk hurriedly or run.] THEN the girl stiffens suddenly. The boy stiffens, too; AND [every run has a hidden wish.]"

Crossing


Rel1
Rel2
Rel1 Arg1
Rel2 Arg1
Rel1 Conn
Rel1 Arg2
Rel2 Conn
Rel2 Arg2

- There is no relation between the two discourse relations.
- No need to compose over trees.
- Violations seem to have applicative semantics (function application)

- Turkish DB: public at http://medid.ii.metu.edu.tr
- Tree violations in TDB
  - annotation differences
  - some can be reanalyzed as wrap, which has applicative semantics
  - anaphoric relations
  - connect at same spot at one end
- They seem to be apparent violations
- We have yet to find genuine cases of tree composition over connectives' meaning

References

Berfin Aktaş, Cem Bozşahin, Deniz Zeyrek. 2010. Discourse Relation Configurations in Turkish and an Annotation Environment. Proc. LAW IV: The Fourth Linguistic Annotation Workshop.

Nicholas Asher. 1993. Reference to Abstract Objects in Discourse. Kluwer Academic Publishers.

Işın Demirşahin. 2008. Connective Position, Argument Order and Information Structure of Discourse Connectives in Written Turkish Texts. Unpublished MS Thesis, Cognitive Science Program, Middle East Technical University.

Işın Demirşahin, Ayışığı Sevdik-Çallı, Hale Ögel Balaban, Ruket Çakıcı and Deniz Zeyrek. 2012. Turkish Discourse Bank: Ongoing Developments. Proc. LREC 2012. The First Turkic Languages Workshop.

Markus Egg, Gisela Redeker. 2010. How Complex is Discourse Structure? Proc. 7th International Conference on Language Resources and Evaluation (LREC 2010) pp. 1619–23.

Michael A. K. Halliday, Ruqaiya Hasan. 1976. Cohesion in English. London: Longman

Jerry R. Hobbs. 1985. On the Coherence and Structure of Discourse. Report CSLI-85-37, Center for Study of Language and Information.

Aravind K. Joshi. 1985. How Much Context-sensitivity is Necessary for Characterizing Structural Descriptions: Tree Adjoining Grammars. In Natural Language Parsing, David Dowty and Lauri Karttunen and Arnold Zwicky (eds.), Cambridge University

Aravind K. Joshi. 2011. Some Aspects of Transition from Sentence to Discourse. Keynote address, In-formatics Science Festival, Middle East Technical University, Ankara, June 9.
Alan Lee, Rashmi Prasad, Aravind K. Joshi, Nikhil Dinesh, Bonnie Webber. 2006. Complexity of de-pendencies in discourse: are dependencies in discourse more complex than in syntax? Proc. 5th Workshop on Treebanks and Linguistic Theory (TLT'06).
Alan Lee, Rashmi Prasad, Aravind K. Joshi, Bon-nie Webber. 2008. Departures from tree structures in discourse. Proc. Workshop on Constraints in Discourse III.
William C. Mann, Sandra A. Thompson. 1988. Rhe-torical structure theory: Toward a functional theory of text organization. Text, 8(3), 243-281.
Alistair Knott, Jon Oberlander, Michael O'Donnel, Chris Mellish. 2001. Beyond elaboration: The interaction of relations and focus in coherent text. In T. Sanders, J. Schilperoord & W. Spooren (Eds.), Text Representation: Linguistic and psycholinguistic aspects (181-196): John Benjamins Publishing.
Jacqueline Kornfilt. 1997. Turkish. New York: Rout-ledge. Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Milt-sakaki, Livio Robaldo, Aravind K. Joshi, Bon-nie Webber. 2008. The Penn Discourse Treebank 2.0. Proc. LREC'08 - The sixth international conference on Language Resources and Evaluation.

Bilge Say, Deniz Zeyrek, Kemal Oazer, and Umut Özge. 2002. Development of a Corpus and a Tree-bank for Present-day Written Turkish. Proc. Eleventh International Conference on Turkish Linguistics (ICTL 2002).

Shieber, Stuart. 1985. Evidence against the Context-Freeness of Natural Language. Linguistics and Philosophy: 8:333-343.
Bonnie Webber. 2004. D-LTAG: Extending Lexicalized TAG to Discourse. Cognitive Science, 28(5), 751-779.
Florian Wolf, Edward Gibson. 2005. Representing discourse coherence: a corpus-based study. Computational Linguistics 31: 249–87.
Deniz Zeyrek, and Bonnie Webber. 2008. A Discourse Resource for Turkish: Annotating Discourse Con-nectives in the METU Turkish Corpus. Proc. 6 th Workshop on Asian Language Resources, The Third International Joint Conference on Natural Language Processing (IJNLP).

Deniz Zeyrek, Işın Demirşahin, Ayışığı Sevdik-Çallı, Hale Ögel Balaban, İhsan Yalçınkaya, Ümit Deniz Turan. 2010. The annotation scheme of Turkish discourse bank and an evaluation of inconsistent annotations. Proc. 4th Linguistic Annota-tion Workshop (LAW IV).