

THE DISCOURSE STRUCTURE OF TURKISH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

IŞIN DEMİRŞAHİN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COGNITIVE SCIENCE

SEPTEMBER 2015

Approval of the thesis:

THE DISCOURSE STRUCTURE OF TURKISH

submitted by **IŞIN DEMİRŞAHIN** in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Cognitive Science, Middle East Technical University by,

Prof. Dr. Nazife Baykal
Director, **Graduate School of Informatics**

Prof. Dr. Cem Bozşahin
Head of Department, **Cognitive Science, METU**

Prof. Dr. Cem Bozşahin
Supervisor, **Cognitive Science, METU**

Examining Committee Members:

Prof. Dr. Deniz Zeyrek Bozşahin
Cognitive Science Department, METU

Prof. Dr. Cem Bozşahin
Cognitive Science Department, METU

Assist. Prof. Dr. Cengiz Acartürk
Cognitive Science Department, METU

Prof. Dr. Varol Akman
Computer Engineering Department, Bilkent University

Prof. Dr. Gülsün Leyla Uzun
Linguistics Department, Ankara University

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: IŞIN DEMİRŞAHİN

Signature :

ABSTRACT

THE DISCOURSE STRUCTURE OF TURKISH

Demirşahin, Işın

Ph.D., Department of Cognitive Science

Supervisor : Prof. Dr. Cem Bozşahin

September 2015, 166 pages

This thesis investigates the structure of immediate discourse in Turkish. The first and foremost question is how discourse is built. Are there components of discourse that constitute a predicate-argument structure, or is discourse realized by underlying non-structural ties that are merely made explicit by these components? If there is structure in discourse, what is the nature of this structure, and what is its complexity?

For this purpose, we analyze the relations annotated in the Turkish Discourse Bank, and their counterparts annotated on the Spoken Turkish Corpus Demo specifically for this study. Through close examination of inter-relational configurations identified in these corpora, we investigate deviations from tree-structure and attempt at eliminating the deviations without compromising the meaning of the text. We show that while some of these deviations can be explained away, some of them stem from the nature of discourse as well as syntactic asymmetries of the components of the discourse relations, and should be accommodated by the discourse theory.

Building upon our findings from the data, we discuss what role discourse connectives play in building the discourse structure. We argue that although discourse relations are best represented as logical predicates, they are fundamentally different from sentence-level predicates. Our conclusion is that the discourse relations anchored by explicit discourse connectives and the inferences represented by implicit discourse connectives are a representation of the structure we perceive in the text, as opposed to sentence-level predicates that build an argument structure and impose linguistic restrictions on their arguments.

Keywords: discourse structure, discourse connectives, turkish discourse bank, spoken turkish corpus, predicate-argument structure

ÖZ

TÜRKÇE'NİN SÖYLEM YAPISI

Demirşahin, Işın

Doktora, Bilişsel Bilimler Programı

Tez Yöneticisi : Prof. Dr. Cem Bozşahin

Eylül 2015 , 166 sayfa

Bu doktora tezi, Türkçe'de anlık söylemin yapısını incelemektedir. Bu bağlamda ilk ve en önemli soru, söylemin nasıl kurulduğudur. Söylemin yapı taşları bir yüklem-üye yapısı mı inşa etmektedirler, yoksa söylem yapı taşları tarafından ortaya çıkarılan, fakat aslında altta yatan bir takım yapısal olmayan bağlar tarafından mı meydana getirilmektedir? Eğer söylemde bir yapı var ise, bu yapının doğası ve karmaşıklığı nedir?

Bu sorulara ışık tutmak için yapılan bu çalışmada, Türkçe Söylem Bankası üzerinde işaretlenmiş olan bağıntılar ve bu bağıntıların Sözlü Türkçe Derlem Demo sürümünde bu çalışmaya özgü olarak işaretlenmiş olan karşılıkları çözümlenmiştir. Söz konusu derlemelerde tespit edilen bağıntılar arası yapılaşmaların incelenmesi yoluyla ağaç yapısından sapmalar tespit edilmiş ve bu sapmaların metnin anlamını bozmadan ortadan kaldırılması amaçlanmıştır. Ağaç yapıdan sapmaların bir kısmının ortadan kaldırılması mümkün olsa da, bir kısmının söylem yapısının doğasından ve bağıntı unsurlarının arasına var olan sözdizimsel eşitsizliklerden kaynaklandığı, ve bu sebeple söylem modelinde yer alması gerektiği görülmüştür.

Bu verilerden yola çıkarak söylem bağlaçlarının söylem yapısındaki rolü tartışılmış, ve her ne kadar söylem bağlaçlarının mantıksal ifadelerde yüklem olarak temsil edilmesi en uygun yaklaşım olarak görülmüşse de, söylem bağlaçlarının sözdizimsel yüklemelerden çok temel ayrılıkları bulunduğu öne sürülmüştür. Açık söylem bağlaçları ile gösterilen söylem bağıntılarının ve örtük söylem bağlaçları ile temsil edilen çıkarımların, söylemi üreten tarafından oluşturulan ya da söylemi okuyan veya dinleyen tarafından algılanan bir yapıyı temsil ettiği, buna karşın, sözdizimsel yüklemeler gibi bir üye yapısı oluşturmadığı ve üyelerine dilbilimsel kısıtlamalar getirmediği sonucuna varılmıştır.

Anahtar Kelimeler: söylem yapısı, söylem bađlacı, türkçe söylem bankası, sözlü türkçe derlemi, eylem-üye yapısı

*To my precious Tofu,
May you always know where your towel is...*

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor Prof. Dr. Cem Bozşahin, from whom I learned how to ask meaningful questions and best practices in answering them, and my project leader Prof. Dr. Deniz Zeyrek for her infinite support and kindness. I would like to thank my jury members Prof. Dr. Gülsün Leyla Uzun, Assist. Prof. Dr. Cengiz Acartürk, and Prof. Dr. Varol Akman for their invaluable comments. I would also like to thank Dr. Ceyhan Temürçü for all his help from the foundations of the knowledge base on which this study was built to brilliant finishing touches; Umut Özge for his support in the very beginning and at the very end of this work; Dr. Ruket Çakıcı for great insights into the inner workings of NLP, academia, and graduate life.

I am grateful to Dr. Ayışığı Başak Sevdik Çallı for so many things from as small as lending me a laser pointer that I did not even know how desperately I needed, to as large as inspiring me to come up with the ideas that make up this thesis. She was first the best colleague and friend, and then the beacon for the light at the end of the tunnel. The red tape of graduation would not resolve as smoothly as it did without her mentorship.

I would like to thank Adnan Öztürel for writing a code that not only works but is also easy to read and modify; Ece K. Takmaz for her help with the final format of this thesis; Hilal Yıldırım for translations; Dr. Ayça Müge Sevinç for solidarity throughout our concurrent PhDs; and everyone involved in the METU Turkish Corpus, Turkish Discourse bank, and the Spoken Turkish Corpus projects. I would also like to acknowledge TÜBİTAK for financially supporting the MEDİD project (107E156).

I offer my gratitude to my sister İnci Demirşahin for answering my every silly question and always encouraging me to write, and her with the rest of my family, Recep Demirşahin, Hasene Demirşahin and Ferah Karter for making me who I am. I thank our princesses Pekmez, Kuki, Bonibon and our one only prince Patates for being the joys of my home and my heart. I also thank Tofu, my imperatrix mundi, whom I miss dearly.

I want to present my thanks to two and a half sisters for being the most fun and supportive cousins at the final sprint of this journey.

I sincerely thank my friends of the Friday nights for comprising such an implausibly comfortable community, and my phorum phriends who keep deceiving me into thinking that I am normal and sane no matter in what medium we find each other.

Many thanks go to Dr. Alp Yürüm, Dr. Meltem Cemre Üstünkaya, and future Dr. Leyla Önal for being crazy, eccentric, depressed, euphoric, sophisticated, intelligent, and silly together.

And last but not the least, I want to express my special gratitude to dear Algan Uskarıcı for being one constant in the hectic tribulations of my mind. Thank you for your endless support of all kinds, for providing me with food, shelter and affection whenever I need, for always being there. And most of all, thank you for bearing with me.

TABLE OF CONTENTS

ABSTRACT	iv
ÖZ	vi
ACKNOWLEDGMENTS	ix
TABLE OF CONTENTS	x
LIST OF TABLES	xiv
LIST OF FIGURES	xvi
LIST OF ABBREVIATIONS	xix
CHAPTERS	
1 INTRODUCTION	1
1.1 The Thesis	2
1.2 Motivation and Challenges	2
1.3 Contribution	3
1.4 Outline	4
2 ELEMENTS OF DISCOURSE	7
2.1 Non-Structural Discourse: Cohesion	7
2.1.1 Reference	8
2.1.2 Substitution	9
2.1.3 Ellipsis	10

2.1.4	Conjunction	10
2.1.5	Lexical Cohesion	11
2.2	Coherence Relations and Structure	12
2.2.1	Tree Structure for Discourse	12
2.2.1.1	Theory of Coherence Relations	12
2.2.1.2	Linguistic Discourse Model	13
2.2.1.3	Rhetorical Structure Theory	14
2.2.1.4	Theory of Tripartite Discourse	16
2.2.1.5	Discourse - Lexicalized Tree Adjoining Grammar (D-LTAG)	17
2.2.1.6	The Penn Discourse Tree Bank (PDTB)	20
2.2.1.7	Discourse Combinatory Categorical Grammar (DCCG)	21
2.2.2	Deviations from Tree Structure	22
2.2.2.1	Complex Interactions Between Trees	22
2.2.2.2	The Segmented Discourse Representation Theory (SDRT)	22
2.2.3	Other Data Structures	24
2.2.3.1	Extended Coherence Relations	24
2.2.3.2	Tree Structure Violations in Penn Discourse Treebank (PDTB)	25
2.2.3.3	Multi-satellite constructions (MSC) in RST	25
2.2.4	Spoken Language	26
3	TURKISH DISCOURSE STRUCTURE	29
3.1	Data	29

3.1.1	Turkish Discourse Bank	29
3.1.2	Spoken Turkish Corpus Demo	32
3.2	Reannotation Methodology	34
3.3	Discourse Relation Dependency Configurations in Written Turkish .	43
3.3.1	Tree Structure	43
3.3.1.1	Independent Relations	43
3.3.1.2	Fully Embedded Relations	45
3.3.1.3	Nested Relations	46
3.3.2	Tree Structure Violations	47
3.3.2.1	Shared Arguments	47
3.3.2.2	Properly Contained Relations	48
3.3.2.3	Properly Contained Arguments	50
3.3.2.4	Partial Overlap	52
3.3.2.5	Pure Crossing	54
3.3.2.6	Distribution of Configurations	59
3.4	A Comparison of Written Discourse vs. Spoken Discourse in Turkish	61
3.4.1	Comparison of the Descriptive Statistics of Discourse Con- nectives in Written vs Spoken Turkish	61
3.4.2	Comparison of the Discourse Relation Configurations in Written vs Spoken Turkish	62
4	EVALUATION AND THE IMPLICATIONS FOR DISCOURSE STRUCTURE	63
4.1	Structure by Explicit Discourse Connectives	63
4.1.1	An analysis of Tree-Structure Deviations	64
4.2	Discourse Structure beyond Explicit Discourse Connectives	68

4.2.1	Implicit Relation	69
4.2.2	AltLex Relation	70
4.2.3	EntRel and NoRel Relations	71
4.3	Variations of a Discourse Relation	71
4.4	Discourse Relations as Predicates	78
5	CONCLUSION	81
5.1	Summary and Conclusions	81
5.2	Limitations	83
5.3	Future Work	84
APPENDICES		
A	DESCRIPTIVES	91
B	A SAMPLE XML FILE FROM TDB	95
C	TOOLS	97
D	LIST OF ALL CONFIGURATIONS	99
	CURRICULUM VITAE	163

LIST OF TABLES

Table 3.1	Connective class breakdown of discourse connectives in the TDB	30
Table 3.2	Breakdown of the unannotated relations in TDB 1.0	40
Table 3.3	Breakdown of the unannotated relations in STC Demo	40
Table 3.4	Distribution of non-independent configurations in TDB	45
Table 3.5	Distribution of fully embedded relations	45
Table 3.6	Distribution of nested relations	46
Table 3.7	Distribution of shared arguments	47
Table 3.8	Reasons for shared argument configurations	48
Table 3.9	Reannotation results for shared argument configurations	48
Table 3.10	Distribution of properly contained relations	50
Table 3.11	Reasons for properly contained relation configurations	50
Table 3.12	Reannotation results for properly contained relation configurations	50
Table 3.13	Distribution of properly contained arguments	52
Table 3.14	Reasons for properly contained argument configurations	52
Table 3.15	Reannotation results for properly contained argument configurations	52
Table 3.16	Distribution of partial overlaps	53
Table 3.17	Reasons for partial overlap configurations	54
Table 3.18	Reannotation results for partial overlap configurations	54

Table 3.19 Distribution of pure crossings	59
Table 3.20 Reasons for pure crossing configurations	59
Table 3.21 Reannotation results for pure crossing configurations	59
Table 3.22 Distribution of non-independent configurations	60
Table 3.23 Distribution of anaphoric relations among tree-violating configurations	61
Table 3.24 Written and spoken uses of <i>ve</i> , <i>için</i> , <i>ama</i> , and <i>sonra</i>	61
Table 3.25 Distribution of non-independent configurations in TDB	62
Table A.1 Number of annotated connectives	91
Table D.1 List of all configurations in the TDB 1.0	99
Table D.2 List of all configurations in the STC Demo	158

LIST OF FIGURES

Figure 2.1 Cohesive ties in (1)	12
Figure 2.2 Typical structure of a conversation from Hobbs (1985) p. 29	13
Figure 2.3 A discourse parse tree from Polanyi (1988) p. 610	14
Figure 2.4 Right frontier constraint from Polanyi (1988) p. 613	15
Figure 2.5 RST schemas from Mann & Thompson (1987) p.7)	16
Figure 2.6 Segmentation and dominance relations for a sample text, Grosz & Sidner (1986), p.183	17
Figure 2.7 Discourse segments, focus spaces and dominance hierarchy, Grosz & Sidner (1986), p.181	18
Figure 2.8 Some elementary trees from Joshi & Schabes (1997) p.7 α trees are initial and the β tree is auxiliary	19
Figure 2.9 Initial tree for the coordinate conjunction <i>so</i> , auxiliary tree for the simple coordinator <i>and</i> from B. Webber et al. (2003) p.31-32	19
Figure 2.10 Violated tree structure for (8)	20
Figure 2.11 The PDTB sense hierarchy (Prasad et al., 2007), p. 27	21
Figure 2.12 Lexical categories for <i>on the one hand</i> and <i>on the other hand</i> , Nakatsu & White (2010), p.21	22
Figure 2.13 A DCCG derivation of nested contrast relations, Nakatsu & White (2010) p.25	22
Figure 2.14 Intersecting and intertwining trees from Hobbs (1985) p. 30	23

Figure 2.15 Modified embedding trees and DR for (9) (Asher, 1993, p. 364)	24
Figure 2.16 Coherence graph from Wolf & Gibson (2005) p. 267	25
Figure 2.17 Non-tree-like dependency structures in PDTB (a) Shared argument; (b) Properly contained argument; (c) Pure crossing; (d) Partially overlapping arguments Lee et al. (2006) p. 84	26
Figure 2.18 RST tree for the same example in 2.17 from Wolf & Gibson (2005) p. 267	27
Figure 3.1 Final structure for (12)	36
Figure 3.2 Final structure for (13)	37
Figure 3.3 Full embedding/shared argument hybrid structure for (14) based on the annotation in (14)((b))i	39
Figure 3.4 Full embedding structure for (14) based on the annotation in (14)((b))ii . . .	39
Figure 3.5 Shared argument configuration for (15)	40
Figure 3.6 Identical relation configuration for (15)	41
Figure 3.7 Shared argument configuration for (18)	42
Figure 3.8 Full embedding configuration for (18). This reading is not available for this item	43
Figure 3.9 Independent relations configuration	44
Figure 3.10 Full embedding configuration	45
Figure 3.11 Nested relations configuration	46
Figure 3.12 Shared argument configuration	47
Figure 3.13 Properly contained relation configuration	49
Figure 3.14 Properly contained argument configuration	51
Figure 3.15 Partial overlap configuration	53

Figure 3.16 Pure crossing configuration	55
Figure 3.17 Double-subordinator analysis for (29) (as-is)	56
Figure 3.18 Single-subordinator analysis for (29) (hypothetical)	56
Figure 3.19 Wrapping	57
Figure 3.20 Double-wrap parenthetical construction for (31)	58
Figure 4.1 Flat tree representation for listing relations	66
Figure 4.2 Shared argument representation for listing relations	67
Figure 4.3 Full embedding representation for listing relations	67
Figure 4.4 D-LTAG derivation and derived trees, B. Webber (2006) p. 352	68
Figure 4.5 The information structure profiles of the connective-argument orders, sorted according to the syntactic type of the connective, from Demirşahin (2008) p. 87	74
Figure 4.6 Possible connective argument orders for non-parallel connectives Demirşahin (2008) p. 40	75
Figure 4.7 Syntactic trees for the connective-argument orders in 4.6	76
Figure 4.8 Simple tree representation for (46)	77
Figure C.1 Discourse Annotation Tool for Turkish	97
Figure C.2 Turkish Discourse Bank Browser	97
Figure C.3 Spoken Turkish Corpus Demo Exmeralda Interface	98
Figure C.4 Flat Spoken Turkish Corpus Transcriptions in Discourse Annotation for Turkish together with the audio on Windows Media Player	98

LIST OF ABBREVIATIONS

AO	Abstract Object
Arg1	The first argument of a discourse connective
Arg2	The second argument of a discourse connective
B	Background
CAO	Connective Argument Order
CCG	Combinatory Categorical Grammar
Conn	Discourse connective
D-LTAG	Lexicalized Tree-Adjoining Grammar for Discourse
DCCG	Discourse Combinatory Categorical Grammar
dcu	Discourse Constituent Unit
DP	Discourse Purpose
DRS	Discourse Representation Structure
DRT	Discourse Representation Theory
DSP	Discourse Segment Purpose
IA	Individual Annotation
L-TAG	Lexicalized Tree-Adjoining Grammar
LDM	Linguistic Discourse Model
MP	Minimality Principle
MSC	Multiple-Satellite Constructions
MTC	Metu Turkish Corpus
NLP	Natural Language Processing
PA	Pair Annotation
PDTB	Penn Discourse TreeBank
PP	Pair Programming
R	Rheme
RST	Rhetorical Structure Theory
SDRT	Segmented Discourse Representation Theory
T	Theme
T-K	Theme-Kontrast
TDB	Turkish Discourse Bank
WSJ	Wall Street Journal

CHAPTER 1

INTRODUCTION

"Let us begin with a fact: discourse has structure"

Hobbs (1985), p. 1

Discourse is characterized by a sense of unity and continuity that random sets of sentences do not have. For example, (1) below is an excerpt from a text, whereas (2) is a random collection of sentences from the same text. The sentences in (2) were taken from the same 2000-word-excerpt as (1), and nevertheless they do not have the unity needed to be a text.

- (1) Sahibi eskiden çöp yuvası olan bu hava aralığını temizlemiş, güzelleştirmişti. Yukarı kadar değil, ama kendi görüş alanına giren bölümü bembeyaz badana etmiş, buraya yeşil çayırlar, masmavi bir gökyüzü çizmiş ve boşluğa açılan pencerenin tam karşısına gelen duvara çiçek saksıları asmıştı. Fazla güneş istemeyen, gölgeyi, rutubeti seven cinsten, koyu yeşil, sarmaşık türü bitkiler... Artur insanlardan sıkıldığı, yalnız kalmak istediği ya da saklanmak zorunda kaldığı zamanlar buraya sığınırdu.

“His owner had cleaned and embellished this air well that used to be a garbage dump. Not all the way up, but he had painted the part in his field of vision in white and painted a blue sky, and he had hung flower pots on the wall that was directly across the window that faced the air well. Plants that do not require much sunlight but like shade and damp, those dark green, ivy-like plants. When he was bored with humans, wanted to be alone, or had to hide, Artur would take shelter here.”

- (2) Pencereden içeri baktı. Daha çok telefonla konuşuyorlar. Yalnızca insanlarla yetine-
mez kediler. Tren hoşuna gitmişti. Birkaç ay sonra tamam! Nina'yla ilk karşılaş-
maları böyle olmuştu. Önceden düşün. Memlekette, onu bu yüzden mi arıyorlar acaba?
Açlığa ve özgürlüğe mahkûm bir zavallı... Bunu sağlayabilmek için kediler ne yap-
malılar? Sepetimde kenarları dantelli kuştüyü yastık bile vardı. Bir başka gün de bun-
ları konuşuruz. Hasta gibiydi. Biliyor musun, bazen sanki kedi değilmişsin gibi bir
duyguya kapılıyorum.

“He looked in through the window. They mostly speak on the phone. Cats cannot be
contented with humans only. He had liked the train. Just a few more months, and then
it's done! His first encounter with Nina was like that. Think beforehand. Are they
looking for him in the homeland because of that? A poor soul confined to hunger and
freedom... What should cats do to ensure this? I even had a laced plume pillow in my
basket. We will talk of these another day. He felt like sick. You know what, sometimes
I get a feeling that you are not a cat.”

The difference between these sequences of sentences stem from a variety of reasons. One reason would be that a text is *structured* through *discourse relations* (or *coherence relations* or *rhetorical relations*), whereas others would argue that the text has *unity* thanks to mostly non-structural *cohesive ties* that are *realized* by the discourse.

1.1 The Thesis

This thesis investigates the structure of immediate discourse in Turkish. The first and foremost question is how the discourse is built. Are there components of discourse that constitute a predicate-argument structure, or is discourse realized by underlying non-structural ties that are merely made explicit by these components? If there is structure in discourse, what is the nature of this structure, and what is its complexity?

For this purpose, we analyze the relations annotated in the Turkish Discourse Bank, and their counterparts annotated on the Spoken Turkish Corpus Demo specifically for this study. Through close examination of inter-relational configurations identified in these corpora, we investigate deviations from tree-structure and attempt at eliminating the deviations without compromising the meaning of the text. We show that while some of these deviations can be explained away, some of them stem from the nature of discourse as well as syntactic asymmetries of the components of the discourse relations, and should be accommodated by the discourse theory.

Building upon our findings from the data, we discuss what role discourse connectives play in building the discourse structure. We argue that although discourse relations are best represented as logical predicates, they are fundamentally different from sentence-level predicates. Our conclusion is that the discourse relations anchored by explicit discourse connectives and the inferences represented by implicit discourse connectives are a representation of the structure we perceive in the text, as opposed to sentence-level predicates that build an argument structure and impose linguistic restrictions on their arguments.

This thesis is concerned with the discourse relations between abstract objects, i.e., propositions, facts, descriptions, situations, or eventualities Asher (1993). *Geldim ve gördüm* ‘I came and I saw’ is within the scope of this thesis whereas *muz ve anans* ‘banana and pineapple’ is out of the scope as there are no abstract object interpretations of *banana* and *pineapple* by default.

In addition, this thesis focuses on the immediate discourse, by which we mean that we are concerned with the local structures built just above clause level. Rhetorical relations such as *coordination*, *contrast*, *cause and effect* are within the scope, as opposed to higher level discourse actions such as *greeting*, *request*, and *apology*.

1.2 Motivation and Challenges

As our opening quote from Hobbs (1985) indicates, for some researchers, it is a fact that discourse has structure; whereas others, such as Halliday & Hasan (1976), argue that discourse is non-structural.

Although most language resources assume some sort of structure, the structural accounts for discourse do not seem to converge on a similar structure. A variety of structures for discourse representation has been proposed, from simplest to most complex: tree structure (Polanyi, 1988), including successive trees of varying sizes connected and occasionally intertwined at the peripheries (Hobbs, 1979, 1985), a single tree structure (Mann & Thompson, 1987, 1988) which may be divided into entity chains (Knott et al., 2001) or may include limited multiparenting (Egg & Redeker, 2010), tree-adjointing grammars (B. Webber & Joshi, 1998; B. Webber et al., 2003; B. Webber, 2004), directed acyclic graphs (Lee et al., 2006, 2008) and chain graphs (Wolf & Gibson, 2004, 2005).

If there is structure in discourse, the complexity of said structure is of interest to linguistics, cognitive science and computer science alike. Is discourse structure more complex or more simple than that of sentence level syntax? 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 (Joshi, 1985; Shieber, 1985). Can discourse, with units much larger than syntax, have more complex structure than sentence? And if such computational power and memory is available for us for linguistic purposes, why don't we use it for sentence level as well?

1.3 Contribution

The contributions of this thesis are the following:

This thesis provides an evaluation of historical and current approaches to discourse representation and discourse annotation from the perspective of structure in discourse and computational complexity. We introduce exemplary theories for each step of complexity from the simplest tree structure to the most complex chain graphs. We initially suspected that discourse may need more complex structures than simple trees (Demirşahin, 2012), but further investigations presented in this thesis showed that discourse seem to have a much simple structure than sentence-level syntax.

The annotations on the Spoken Turkish Corpus Demo version in the style of the Penn Discourse Treebank and the Turkish Discourse Bank is the first of its kind on spoken Turkish data (Demirşahin & Zeyrek, 2014). By carrying this approach to another medium in Turkish, we discovered that it is possible for phrasal expressions to take both their arguments from the distant previous discourse anaphorically. Although in our example one of the anaphoric elements is included in the phrasal expression, the clitic nature of the Turkish question particle may allow even the structural connectives to take arguments in a similar manner.

This thesis offers a complete account of the structure expressed by the explicit connectives in Turkish Discourse Bank. We provided quantitative data for the inter-relational configurations first identified by Aktaş et al. (2010), i.e., tree-conforming *independent relations*, *full embedding*, and *nested relations*, and tree-violating configurations *shared argument*, *properly contained argument*, *properly contained relation*, *partially overlapping arguments*, and *pure crossing* (Demirşahin et al., 2013). In addition we analyzed the reasons for the tree-violating configurations, and reannotated some of them to provide alternative, tree-conforming structures.

In order to investigate whether the tree-structure violations are structural or anaphorical, we

annotated the syntactic class of all explicit discourse connectives annotated in the TDB 1.0. This annotation, along with the complementary annotations of the morphological features of the arguments of subordinating conjunctions, the anaphoric component of phrasal expressions, and the parallel status of the connectives will be included in the further releases of the Turkish Discourse Bank (Demirşahin, Sevdik-Çallı, et al., 2012).

To the best of our knowledge, this thesis provides the first whole-corpus structure analysis in PDTB style. The previous studies were either focused on a single connective (Lee et al., 2006), or were exploratory in nature and were not quantitative (Aktaş et al., 2010). Our study covers all explicit connectives annotated in the TDB 1.0, and all instances of the corresponding search tokens in the STC Demo.

The investigations on the tree-structure violations in the TDB 1.0 resulted in the discovery of the previously undescribed phenomenon of wrapping at discourse level. We found out that one of the reasons for the apparent surface crossings is an information structurally motivated strategy in Turkish, namely bringing the constituent to be focused to the preverbal position, which results in whole arguments of discourse connectives to move the said focus position, due to the free word order of Turkish and the adverbial characteristics of the Turkish subordinate clauses. The matrix clause, which is the other argument of the discourse connective ends up wrapped around the discourse connective and the argument that hosts it.

During the annotations of the Turkish Discourse Bank, we came up with the novel annotation methodology Pair Annotation, named after Pair Programming, which is a collaborative programming paradigm where two programmers work on an algorithm or a piece of code as a unit, assuming equal responsibility and credit for the work done. The Pair Annotation method reduces the possibility of physical errors, increases the inter-annotator agreement, and provides the annotators with the opportunity to discuss hard cases during annotation. By including at least one individual annotator, we preserved the principles of independent and blind annotation (Demirşahin, Yalçınkaya, & Zeyrek, 2012; Demirşahin & Zeyrek, in press).

1.4 Outline

In Chapter 2 we review the previous works that are concerned with the structure of discourse, or lack thereof. We present various approaches to discourse structure, varying in complexity from the simplest tree structure to the most complex chain graphs.

Then in Chapter 3 we analyze the annotations in the first large-scale and public language resource annotated with discourse-level phenomena in Turkish. We take a look at the structures that arise as a result of the annotation of discourse connectives in Turkish Discourse Bank (TDB) 1.0, and quantitatively investigate the computational power required for these structures. We also provide a similar analysis for discourse annotations on the demo release of the Spoken Turkish Corpus (STC) conducted specifically for this study. We try to disentangle structures that arise from the particular approach that was used for the annotation of the TDB 1.0 and the STC demo, and those that are inherent to the discourse.

In Chapter 4 we delve further into the causes for more complex structures that require more computational power than sentence-level complexity. We investigate the structural complexity of the discourse as anchored by explicit discourse connectives, and discuss the possible impact of the annotation of implicit connectives. Then we look into the relation between the discourse

connectives and the semantics they denote, and question their status as predicates.

Finally in Chapter 5 we summarize our findings and discussions. We discuss the limitations of the study that arises from the nature of corpus studies in general, corpus-driven and connective-based approaches to discourse, and the time and budget constraints of this study in particular. We also present the ideas for future work for which this thesis offers a starting point.

CHAPTER 2

ELEMENTS OF DISCOURSE

For the native speaker, the difference between the two sequences of sentences in (1) and (2) is obvious. (1) is *coherent*, whereas (2) is not. However, the exact reason for the coherence and the incoherence of a particular sequence of sentences is somewhat elusive. Hobbs (1979) explains that the mere quality of being about the same entities does not yield coherence. Our examples confirm his intuition: both examples are concerned with the cat Artur and his owner, but one is coherent and the other is incoherent. Also as in Hobbs' examples, when confronted with the challenge of an incoherent sequence, the reader tries to attribute coherence to the piece by imposing certain inferences and assumed backgrounds. For example, although the text provides no antecedent for the pronoun *they*, one can imagine that upon looking through the window, Artur sees some people, who happen to be the antecedent for *they*, who mostly talk on the phone. This alternative reading would account for the next sentence where the cats cannot be contended with humans only, since the humans are spending their time on the phone rather than tending to their cats. Out of boredom of humans, cats would need entertaining activities, such as the train ride Artur likes in the following sentence. Similar stretches of imagination can almost make up for the lack of coherence in the sequence. However, without such determination to impose coherence, the sequence reads more like a stream of consciousness, which as a style is allowed to be somewhat incoherent.

Hobbs interprets this type of accommodation of incoherence as a need for coherence on the part of the reader, and defines coherence as an independent structure which is not caused by being about the same entity; on the contrary, the feeling that a sequence of sentences are about the same thing is a byproduct of coherence. He further argues that while coherence and anaphora resolution are related; coherence is the dominant one of the two.

2.1 Non-Structural Discourse: Cohesion

Hobbs' position is almost the exact opposite of that of Halliday & Hasan (1976). Whereas Hobbs takes it as a fact that discourse has structure as it defining property, Halliday & Hasan claim that the essential property of text is *cohesion*, a mostly non-structural property that unifies a sequence of sentences and gives it texture. According to Halliday & Hasan, cohesion is based on *reference*, *substitution*, *ellipsis*, *conjunction*, and *lexical cohesion*. Of these five bases, the first three are all concerned with different facets of the same process, a concrete or abstract entity is anaphorically retrieved by either a pronoun, a substitute, or by omission. They make a point of emphasizing that the cohesive ties do not form syntactic structures. They argue that a text is a semantic unit of *realization* and not that of *constituency*, and while

structure implies texture, texture does not necessarily imply structure.

2.1.1 Reference

Reference is a very broad term concerning proper nouns, definite noun phrases, and indexicals. For the purposes of this section, we will restrict our definition to reference as discussed in Halliday & Hasan (1976).

Halliday & Hasan (1976) distinguish two broad types of reference. *Exophoric* (situational) referential items stand for things in the world outside of the text. For example the demonstrative *bu*, when used to point at an object, refers to a real object and not a linguistic object. Ostensive references and many deictic expressions such as *today* as referring to the actual day of the utterance or *here* as in the physical place that the utterance is taking place are all considered exophoric. *Endophoric* (textual) referential items, on the other hand, refer to entities, or linguistic objects, that are already mentioned in the text. Halliday & Hasan (1976) consider only endophoric reference to be cohesive. Endophoric ties can either be anaphoric, meaning that the resolution of the referential item takes place in the preceding discourse, or cataphoric, meaning that the resolution is to be found in the following discourse.

Reference is semantically definite, as in it invokes a specific antecedent, meaning that something that was previously mentioned has reentered the discourse, or in the case of cataphora, the item will again enter the discourse in the near future. This continuity of reference results in cohesion. Personal pronouns, demonstrative pronouns and comparatives can form cohesive ties.

Personal reference ties are realized by personal pronouns. The category person is used liberally here. Personal reference can refer to roles in discourse as in the speaker and the addressee, and other people, but it is not restricted to human entities only. It also applies to non-human entities, objects, and passages of text. In English, *I, you, he, she, it, we, they* and the generalized *one*, and their accusative and possessive counterparts refer to persons. In Turkish, the personal pronouns *ben, sen, o, biz, siz, onlar* and the reflexive *kendi* and their inflected forms perform similar functions. In (3), the underlined phrases all refer to the same entity, the girl who read Kierkegaard on Lange Leidsewards Straat. These ongoing chains of reference realize cohesive ties.

- (3) Lange Leidsewards Straat'da Kierkegaard okuyan kıza, kendisiyle yeniden görüşmekten sevinç duyacağımı söylemiş, ertesi gün öğleye doğru, onun oturduğu sokağın başındaki o güzel, iki katlı kahveye çağırmıştım onu.

"I told the girl who was reading Kierkegaard on Lange Leidsewards Straat that I would be very happy to see her again on the next day towards noon, I invited her to the beautiful, two-story cafe at the end of the street she was living in."

Demonstrative reference items are essentially ostensive determiners are pronouns. When used to point to an object in the text, they realize cohesive ties. In English, *this, these, here* and *now* are demonstratives that are used to point to close objects and places, whereas *that, those, there* and *then* are used to point to distant objects and places. Turkish also has close (*bu, bunlar, bura*) and distant (*o, onlar, ora*) as well as a middle, or moderately distant, set of

demonstratives *şu, şunlar, şura*. Just as they are used to point objects in varying distances in the world, there items can be used to point to object in varying distances in the text, too.

Halliday & Hasan state that the singular form of object reference in English, *it*, can also refer to a passage of text. In Turkish, *o*, can also refer to a passage of text, however, our intuition is that it is not a personal reference, but a demonstrative reference that is employed when referring to passages of texts. None of the other personal reference items refer to passages of text, whereas almost all demonstrative reference items frequently refer to passages of text. Note that the distant demonstrative reference item root is *o*, same as the third person singular.

When referring to a text passage, *o* is anaphoric, i.e., *o* refers to a passage of text in the preceding discourse. On the other hand, *şu* is cataphoric, i.e., *şu* refers to a passage of text in the following discourse. *Bu* is usually anaphoric, but there are cases it can be cataphoric too. In (4) *bu* anaphrically refers to the previous sentence.

(4) Sen beni iyice işletiyorsun. Dur bakalım bunun sonu nereye varacak?

“You’re having me on. Let’s wait and see where this will end up.”

Comparatives realize cohesive ties through identity, similarity, and difference. By definition, a comparative presupposes an existing entity, one which is being compared to another entity. The comparison adjectives and adverbs such as *same, identical, similar, additional, other, different, else, identically, similarly, likewise, so, such, differently, otherwise*, and particular comparison adjectives and adverbs such as *better, more*, and comparative forms of other adjectives form comparative reference ties, too. Turkish comparative reference items include but are not limited to: *aynı, benzer, farklı, başka, değişik*.

2.1.2 Substitution

During substitution a word takes the place of another word in the text. The resulting cohesive relation, according to Halliday & Hasan, is between words. Unlike reference, which is a semantic cohesive relation, Halliday & Hasan take substitution, including ellipsis, to be grammatical. Therefore, reference can point to anywhere in and out of the text, but substitution is confined to the text. Even in the rare case of exophoric substitution, Halliday & Hasan expect to find an assumption or implication that something has been said.

Substitution has three types: nominal, verbal and clausal (Halliday & Hasan, 1976). Nominal substitution occurs when a word takes the place of the head of a nominal group. In English, *one, ones* and *same* can substitute nominal heads. Though Turkish can employ *biri* for nominal substitution as English employs *one*, the use of definitive morphology seems more common for this job. Where the English native speaker would use *the red one* to refer to a red dress, the Turkish native speaker would prefer *kırmızıyı* ‘red-DEF.ACC’ or *kırmızı olanı* ‘red be-REL-DEF.ACC’ both meaning ‘the red one’ without substitution. The Turkish counterpart of *same* is *aynısı*. This word carries a possessive marker, morphologically indicating the cohesive relation.

Verbal substitution occurs when a word takes the place of a lexical verb, acting as the head of a verbal group. The English word for verbal substitution is *do*. Its Turkish equivalent is *yap*, and *yap* can be used as a verbal substitution item.

In the case of clausal substitution, a word does not take the place of another word or word group, but a whole clause. In English *so* and *not* are used for clausal substitution. In Turkish the clausal substitution can be conveyed by *öyle*. In negative situations, *öyle* is used with the appropriate negative form.

Substitution items can also be taken as complements by discourse connectives. They can even form discourse adverbials as *öyleyse* has done through lexicalization from an inflected form with *-se*, a subordinator-type discourse connective.

2.1.3 Ellipsis

When the discourse connective is defined by taking arguments that are abstract objects (B. Webber, 2004), and when the notion of abstract object depends on being a proposition, fact, description, situation, or eventuality (Asher, 1993), it becomes exceptionally important to understand the nature of ellipsis. A group of words that seem to be grouped together without an obvious predicate may constitute a proposition, fact, description, situation or eventuality, thus may be an abstract object: a valid argument for a discourse connective.

Ellipsis is not very different from substitution from a viewpoint of cohesion. In fact, Halliday & Hasan, take ellipsis to be “substitution by zero” (p.142). Ellipsis is the case when something is not said, but is still understood.

Like substitution, ellipsis has three types: nominal ellipsis, verbal ellipsis and clausal ellipsis. Nominal ellipsis occurs within a nominal group, i.e., some part of a nominal group is missing from the utterance.

Verbal ellipsis means something in the verbal group is left unsaid. The unsaid material may be the lexical verb in the verbal group, in which case Halliday & Hasan call it a lexical ellipsis, or it may be other materials, subjects, modals, etc., in which case it is called operator ellipsis.

2.1.4 Conjunction

Conjunction is another type of cohesive link, and in some ways different from the others (Halliday & Hasan, 1976). Reference, substitution and ellipsis instruct the reader or hearer to search for an element, most of the time in the preceding or following text. Conjunction, on the other hand, instructs the addressee how to bring two parts of text together. The meaning of the conjunctive item itself is not dependent on what is presupposed.

A relation can be expressed in many ways in natural languages. Two events, A and B, in a relation can be expressed by grammatical predication, as in “A caused B”, by minor predication as in “B happened because of A”, by means of a subordinator as in “Because A happened, B happened”, by means of an adverbial expression relating two separate sentences as in “A happened. As a result B happened.” This adverbial expression is called a conjunctive adjunct or a discourse adjunct by Halliday & Hasan (1976) and a discourse adverbial by B. Webber (2004).

Halliday & Hasan draw a line between coordination and conjunction. They state that *and* and *or* relations in their very basic logical sense are structural and not cohesive. One of their

arguments against coordination being a cohesive relation is that coordinated items form a single complex element, which behaves as simple elements behave.

They define four major types of conjunctive relations: *additive*, *adversative*, *causal* and *temporal*. These types are further specified according to too detailed criteria to mention here. The conjunctive relations can be external or internal. Halliday & Hasan propose these terms to express functional dichotomy that might be called objective/subjective or experiential/interpersonal. The external relations exist simply between two events, or rather situations. Internal relations occur in the communication process. This dichotomy is most explicit in temporal relations. For example, in a text after this might refer to after something already mentioned in the text (external, in “thesis time”) or after the time the text is being realized (internal, in “thesis time”).

The indication of such a division also exists in the Penn Discourse Tree Bank (PDTB) sense list in their annotation manual (Prasad et al., 2007). In this relatively theory independent treebank’s sense hierarchy, there are four major semantic classes: *temporal*, *comparison*, *contingency* and *expansion*. These classes are further divided into types and subtypes, where some senses have ‘pragmatic’ subtypes. Pragmatic senses involve the interpretation of an argument rather than simply compositional meanings, or involve evaluation of speech acts.

One major difference between the two approaches is that Halliday & Hasan put conjunctives under certain types, for example, *thus* is put under *additive*, *internal*, *apposition*, *exemplificatory* in their table. In PDTB annotations, on the other hand, the exact sense of a particular instance of *thus* would be clear only when the annotators put that particular *thus* into context.

2.1.5 Lexical Cohesion

Lexical cohesion occurs when semantically close words are used repetitively in a text.

Halliday & Hasan propose that lexical cohesion occurs in two ways, reiteration and collocation. Reiteration, as the name implies, is repetition of the same referent but this is not restricted to the repetition of the same word. In fact, repetition of the same word is only one of the ways reiteration can take place. Other ways are use of synonyms like ascent-climb, near-synonyms such as sword- brand, superordinates such as Jaguar-car (Halliday & Hasan, 1976, 278), and use of general words such as people, thing, place, etc.

In reiteration, all the words used refer back to the same referent even though the words themselves are not the same. In collocation, on the other hand, the referents are not the same, they even may be opposites, but the words are still cohesive. Such semantically close words often come from complementary sets as in boy-girl, or antonyms such as like-hate, members of the same ordered series, for example, Tuesday-Thursday, members of unordered lexical sets like red-green, words in a part-whole relation such as box-lid, or part-part relation as in mouth-chin, as well as words which are not easy to put under a systematic semantic class, but are related nevertheless, for instance, comb-curl.

Though Halliday & Hasan prefer to keep cohesion distinct from discourse structure, lexical cohesion stands close to some relations in discourse structure theories. What discourse structure theories name elaboration (Mann & Thompson, 1987, 1988) or entity relation (EntRel) (Prasad et al., 2007; B. Webber et al., 2006) are relations where two discourse units are re-

lated by means of providing more information about the same thing or even just being about the same thing. Unlike lexical cohesion ties, which can exist between any items in the text, both of these relations are restricted to adjacent text spans, elaboration by virtue of being an Rhetorical Structure Theory (RST) relation and EntRel by virtue of being an implicit relation which is defined at sentence boundaries. The status of elaboration as a discourse relation has been questioned (Knott et al., 2001).

Even a small piece of text can be abundant with the cohesive ties proposed by Halliday & Hasan. Figure 2.1 displays some of the cohesive ties in (1).

Yukarı kadar değil, **ama** kendi görüş alanına
giren bölümü **bembeyaz** badana etmiş,
buraya yeşil çayınlar, **masmavi** bir gökyüzü
çizmiş **ve** boşluğa açılan pencerenin tam
karşısına gelen duvara **çiçek saksıları**
asmıştı. Fazla **güneş** istemeyen, **gölgeyi**,
rutubeti seven cinsten, **koyu yeşil**, **sarmaşık**
türü **bitkiler...**

Figure 2.1: Cohesive ties in (1)

2.2 Coherence Relations and Structure

If there is structure in discourse, the complexity of the said structure is of interest to linguistics, cognitive science and computer science alike. Is discourse structure more complex or more simple than that of sentence level syntax? How and to what degree is that structure constrained? In order to answer questions along these lines, researchers explore the possible data structures for discourse in natural language resources.

2.2.1 Tree Structure for Discourse

2.2.1.1 Theory of Coherence Relations

Hobbs (1985) takes it as a fact that discourse has structure. Building upon the “combinations of predications” Longacre (1976) that denote conjunction, contrast, comparison, alternation, temporal overlap and succession, implication and “rhetorical predicates” in Grimes (1975) that denote alternation, specification, equivalence, attribution, and explanation, he calls the relations that build the discourse structure *coherence relations*. He claims that unlike previous work that only formally define these relations or relate the structure of coherence relations to memory, his theory of coherence relations are integrated into a knowledge-based discourse interpretation theory.

For this purpose, the knowledge base, i.e., all knowledge accessible to the speaker and the

audience, and the sentences in a text are translated into a logical form. A deductive mechanism interprets and manipulates the axioms that make up the knowledge base and the logical forms of the sentences. Discourse operations specify the possible interpretations and select the ones relevant to the current text. In the final step, “the best interpretation” for the sentence is specified from the possible interpretations by taking into account to *internal coherence* of the sentence and the *local coherence*, i.e. the relation in which the sentence stands with its surrounding text.

Hobbs identifies nine coherence relations: occasion, evaluation, background, explanation, parallel, elaboration, exemplification, contrast, violated expectation. Through these coherence relations, clauses, which are basic segments of discourse, are linked together and *constitute* a single segment of discourse. Parallel and elaboration are coordinating relations, whereas background, explanation, exemplification and generalization, contrast, and violated expectation are subordinating relations. In coordinating relations, a common proposition is the assertion of the composed segment. In subordinating relations, one of the segments is subordinated to the other, dominant segment and the assertion of the composed segment is the assertion of the dominant segment. Hobbs (1985) is undecided about the status of the occasion relation.

According to Hobbs, well planned discourses can be composed to a single segment. However, tangents happen, and the discourse is fragmented to a series of trees connected by smaller trees that combine or intertwine at the edges as in 2.2.



Figure 2.2: Typical structure of a conversation from Hobbs (1985) p. 29

2.2.1.2 Linguistic Discourse Model

Polanyi (1988) proposes a formal model for discourse, the Linguistic Discourse Model (LDM). LDM is an incremental discourse parser that builds a Discourse Parse Tree.

In LDM, the basic unit of discourse is the *discourse constituent unit (dcu)*, of which the most elementary one is the clause. The four types of dcus are the *sequence*, a string of similar dcus, the *expansion*, a clause that is expanded by a semantically subordinated dcu, the *binary structures*, structures that are formed by linking dcus with explicit logical operators such as *and, because, or, if, then.,* and the *interruption*.

In addition to the dcus, there are *discourse operators* that modify the dcus. Discourse operators include affirmative and negative particles, discourse markers, discourse connectives, interjections, vocatives. Interjections such as *hello*, *goodbye* and vocative proper nouns are *assigners*, discourse connectives such as *and*, *because*, *therefore* are *connectors* discourse markers such as *well*, *so* and *anyway* are discourse PUSH/POP markers.

Dcus and discourse operators compose *Discourse Genre Units* such as stories and plans, and *Discourse Adjacency units* such as question & answer pairs. The *Discourse Units (DUs)* make up the context for each dcu. The LDM parser processes the text left-to-right, clause by clause. All clauses, including digressions and interruptions, are processed in the same manner, resulting in a *Discourse Parse Tree* as in 2.3.

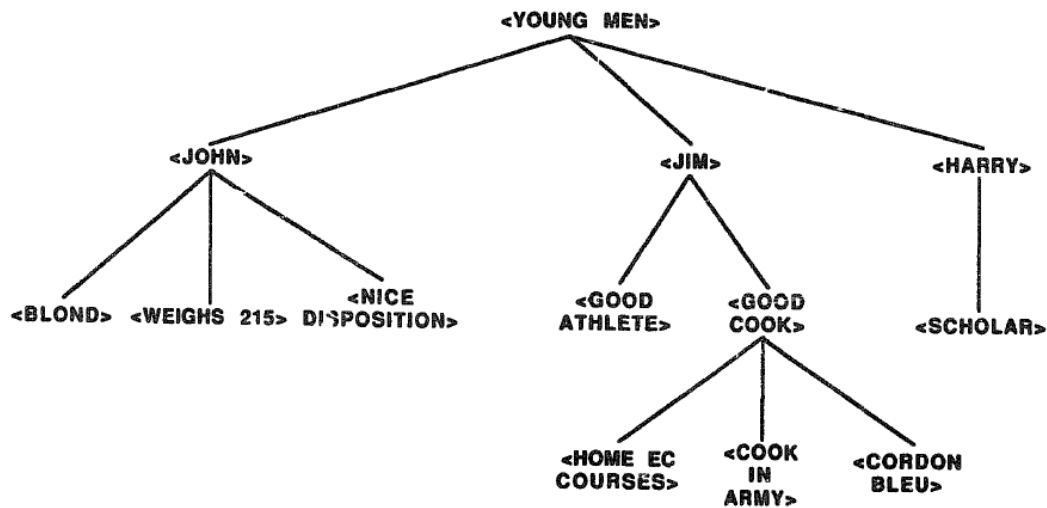


Figure 2.3: A discourse parse tree from Polanyi (1988) p. 610

LDM also introduces the *Right Frontier Constraint*, which means that each discourse constituent unit can only attach the rightmost open nodes at various levels of the tree, thus formalizing the accessibility of previous discourse constituent units to new discourse operations, and ensuring the resulting structure is indeed a tree.

Polanyi (1988) admits that the LDM makes a very strong claim in terms of the possible structure of the discourse. They maintain that although it is possible to go back to the subject of a closed note, it will only be possible by intonational repair or initiation signals, and will be added as a new unit rather than continuing an older one.

2.2.1.3 Rhetorical Structure Theory

(Mann & Thompson, 1987, 1988) proposes that a text can be analyzed as a single tree structure by means of predefined rhetorical relations. Rhetorical relations hold between adjacent constituents either asymmetrically between a nucleus and a satellite, or symmetrically between two nuclei, in which case, the relation is said to be multinuclear. The notion of nuclearity allows the units to connect to previous smaller units that are already embedded in a larger tree structure, because a relation is assumed to be shared by the nuclei of non-atomic constituents. In other words, a relation to a complex discourse unit can be interpreted as either between the

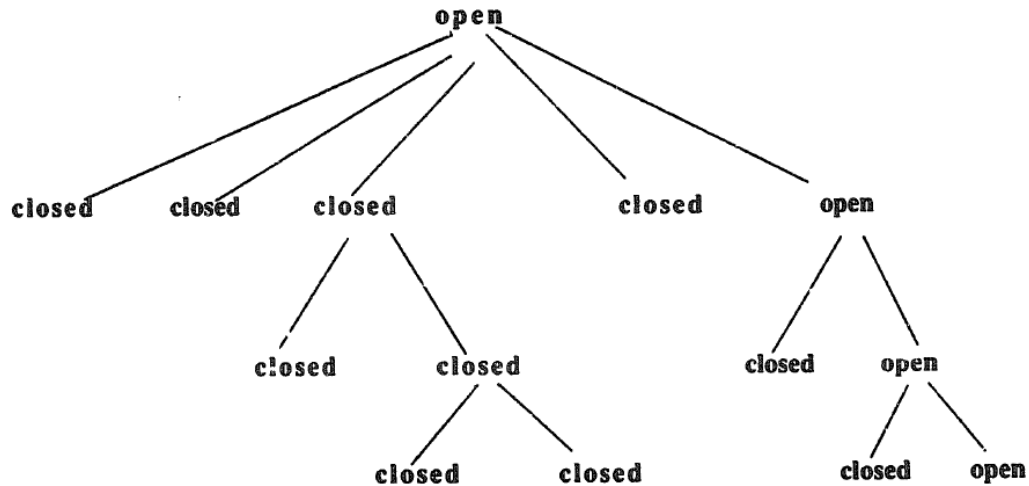


Figure 2.4: Right frontier constraint from Polanyi (1988) p. 613

adjacent unit and the whole of the complex unit, or between the adjacent unit and a nucleus of the complex unit.

RST assumes that coherence occurs when every part of a text is one way or another connected to another part in the text and these connections between parts of text can be represented by functions, i.e., plausible reasons for the presence of particular parts in the text.

RST proposes a hierarchical structure for text. Relations among clauses are analyzed independent from any lexical cue. A relation in RST consists of constraints on the nucleus, constraints on the satellite, constraints on the combination of the two and the effect, i.e., what the writer intended to achieve, or how this relation changes the reader's ideas. For example an EVIDENCE relation exists between a nucleus satisfying the constraint "R might not believe N to a degree satisfactory to W" and a satellite satisfying the constraint "The reader believes S or will find it credible". The constraint on the combination of these two is "R's comprehending S increases R's belief on N" and the effect of this relation is that "R's belief of N is increased"

(Mann & Thompson, 1987) Though these features seem plausible, the analyst has to guess what the writer intended in order to determine the nature of relation. Writers do not always write what they intend to. The task of analyzing low level semantic relations between parts of text is more or less mechanical, whereas the task of identifying intentions requires a deeper understanding of the text, the context and the author. What is more, one relation may be used with different intentions in different situations.

RST schemas define how spans of text can interact with each other. The schemas apply recursively, i.e., a text span resulting from the application of a schema can be, or rather, is expected to be the nucleus or satellite of another relation higher in the hierarchy.

The RST schemas are applied in a way to satisfy four constraints. *Completeness* requires that the application of schemas to the entire text results in one schema application. *Connectedness* requires that all text spans in the text are either a minimal unit or take part in another schema application in the analysis. *Uniqueness* requires that schema applications are on different sets of text spans, and *Adjacency* requires that the text spans of a schema application result in another text span (Mann & Thompson, 1987) . The schema application constraints are well

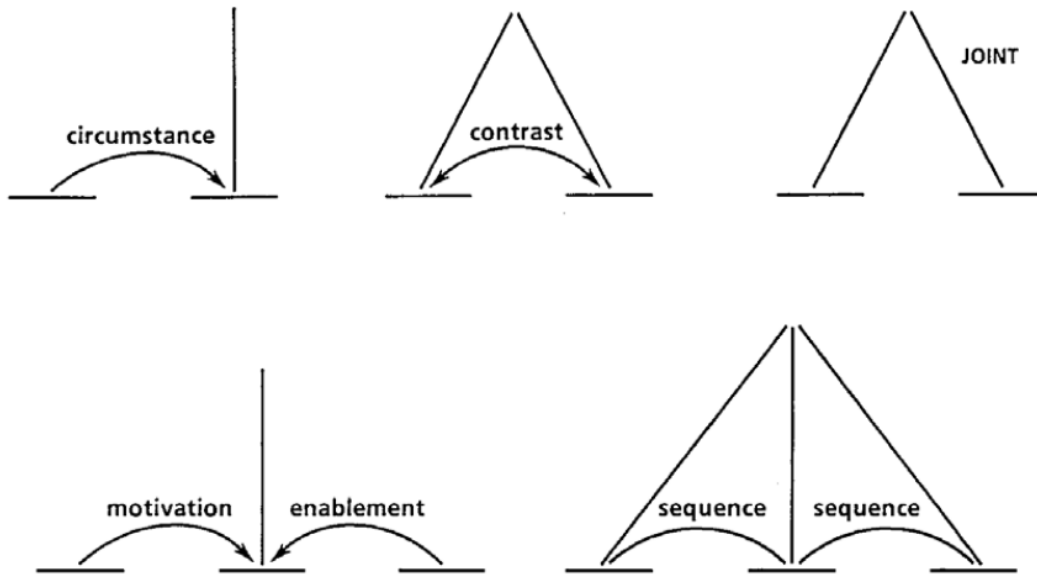


Figure 2.5: RST schemas from Mann & Thompson (1987) p.7)

defined and they are at the same time quite strict. Such strict restrictions are bound to result in consistent analyses between analysts; however, they are also likely to interfere with the analyst when determining the features of a relation.

One of the rhetorical structures in RST, elaboration is criticized by Knott et al. (2001) who propose an elaboration-less coherence structure, where the global focus defines linearly organized entity chains, which can contain multiple atomic or non-atomic RS trees, and which are linked via non-rhetorical resumptions.

2.2.1.4 Theory of Tripartite Discourse

Grosz & Sidner (1986) propose a theory of tripartite discourse. They claim that discourse includes three separate components which interact with each other. The first component is the linguistic structure, which consists of a sequence of utterances. Segments of utterances are not necessarily continuous. This discourse segment structure interacts with the utterances that make up the segment. Some expressions in these utterances, i.e., *cue phrases*, express information about the discourse structure, and are among the primary indicators of segment boundaries. In return, the generation and interpretation of these expressions are constrained by the discourse.

The second component is the intentional structure. It concerns the purpose of the discourse. Grosz & Sidner (1986) differentiate the purpose essential to the discourse from private purposes. The discourse purpose (DP) explains why that particular discourse is happening and why it is happening the way it does. Each discourse segment has a discourse segment purpose (DSP). DSPs make up the DP and each individual DSP indicates how the discourse segment contributes to the discourse. DSPs are structurally related by dominance and satisfaction-precedence. A DSP dominates another when the latter contributes to the satisfaction of the

dominant DSP. Satisfaction-precedence relation occurs when one DSP needs to be satisfied before another DSP. Their analyses show that one DSP can dominate several DSPs, whereas no DSP is dominated by multiple DSPs, resulting in a tree structure.

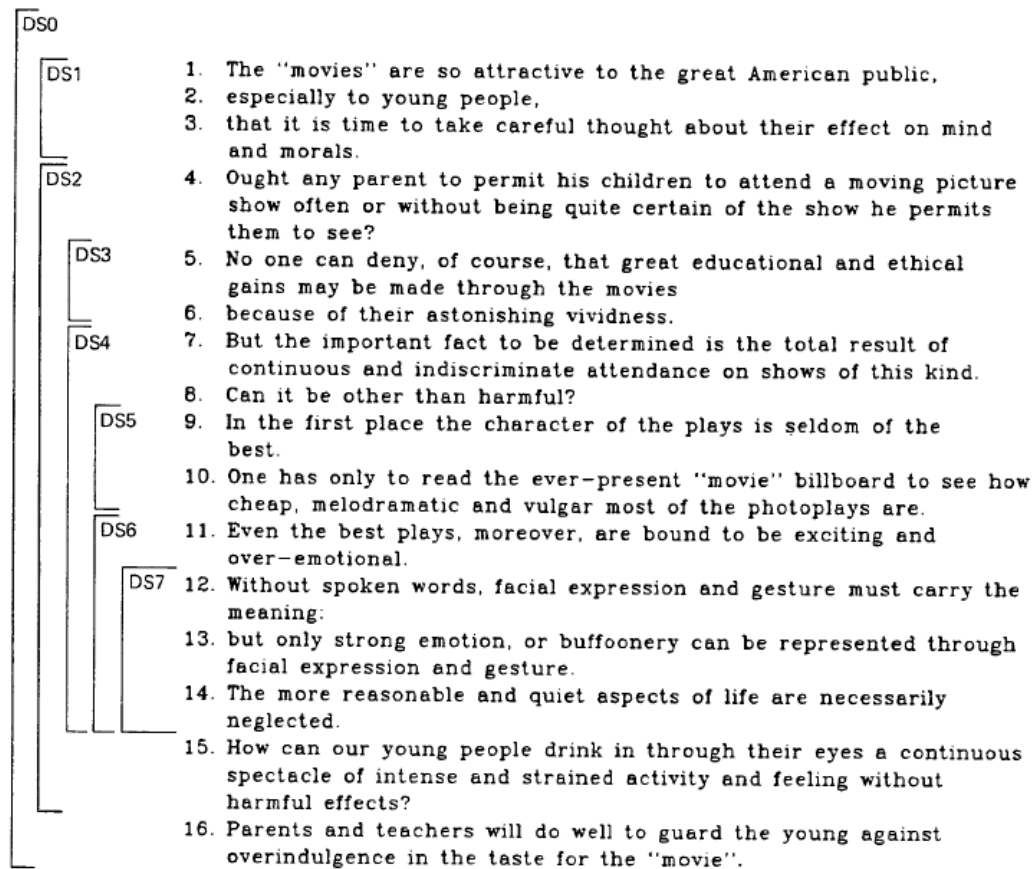


Figure 2.6: Segmentation and dominance relations for a sample text, Grosz & Sidner (1986), p.183

The third component is the attentional state, which concerns the focus of attention. The attentional state is represented by a focus space which defines the salient entities at that point of discourse. Naturally, the focus space is updated as the discourse progresses. A focus space, in a way, includes both (parts of) the discourse segment and the DSP, so that it represents that the conversational participants are aware of what is being discussed and why it is being discussed (Grosz & Sidner, 1986). Although Grosz & Sidner propose a two-stack alternative to handle flashbacks in discourse, they do not expect this mechanism to be necessary precisely because of its added complexity. The focus state is mostly handled by a single-stack mechanism, confirming that the complexity is within tree-structure-level.

2.2.1.5 Discourse - Lexicalized Tree Adjoining Grammar (D-LTAG)

Discourse - Lexicalized Tree Adjoining Grammar (D-LTAG) (B. Webber, 2004) is an extension of the sentence-level Tree Adjoining Grammar (Joshi, 1987) to discourse level.

Discourse connectives act as discourse level predicates that connect two spans of text with

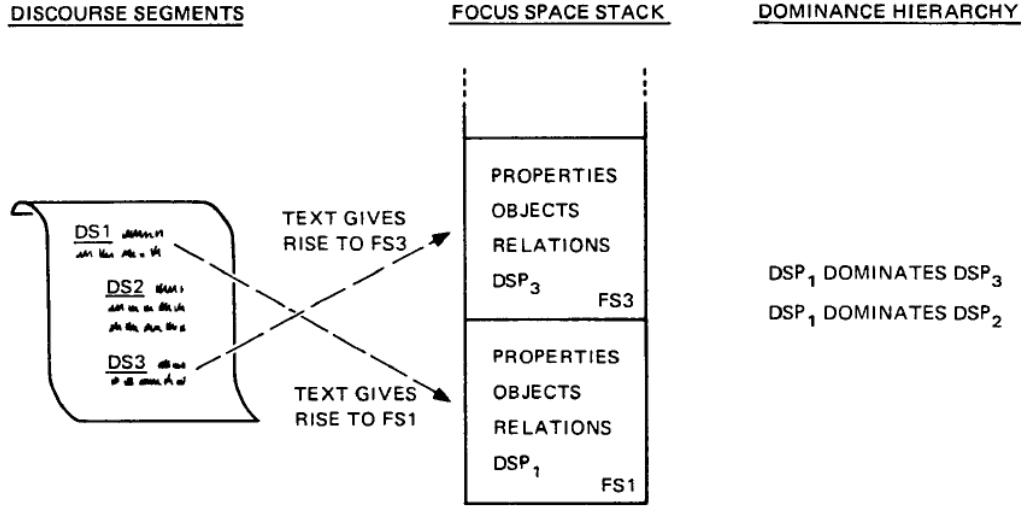


Figure 2.7: Discourse segments, focus spaces and dominance hierarchy, Grosz & Sidner (1986), p.181

abstract object (Asher, 1993) interpretations. Coordinating and subordinating conjunctions such as *fakat* ‘but’ (5) and *rağmen* ‘although’ (6), take their host clauses by substitution and the other argument either by substitution or by adjoining; whereas discourse adverbials such as (7) take the host argument by adjoining, and the other argument anaphorically.¹

(5) 00013212-3

*Araştırma Merkezi aşağı yukarı bitmiş durumda, **fakat iç ve dış donanımı eksik.***

*“The Research Center is more or less complete **but its internal and external equipments are missing.**”*

(6) **Benim için çok utandırıcı bir durum olmasına rağmen** *oralı olmuyordum.*

*“**Although it was a very embarrassing situation for me,** I didn’t pay much heed.”*

(7) *İlgisizliğim seni şaşırtabilir; ama üvey babamı görmek istemediğim için yıllardır o eve gitmiyorum. **Anneme çok bağlı olduğumu da söyleyemem ayrıca.***

*“My indifference might surprise you, but since I do not want to see my stepfather, I have not been to that house for years. **In addition, I cannot say I am attached to my mom much.**”*

As in sentence level syntax, the anaphoric relations are not part of the structure; as a result, the discourse adverbials can access their first arguments anywhere in the text without violating non-crossing constraint of tree structure. When a structural connective such as *ve* ‘and’ and a discourse adverbial such as *bundan ötürü* ‘therefore’ are used together as in (8), an argument

¹ In the examples from TDB the first line indicates the file name and the browser index of the connectives involved in the example. The first arguments (Arg1) of the connectives are in italic, the second arguments (Arg2) are in bold. Shared arguments, i.e., spans that are interpreted as belonging to both arguments are both in boldface and italic. The connectives are in boldface and underlined. Modifiers of the connectives are underlined but not in boldface. For the sake of simplicity, the supplementary materials to the arguments are left out unless critical to the example in discussion.

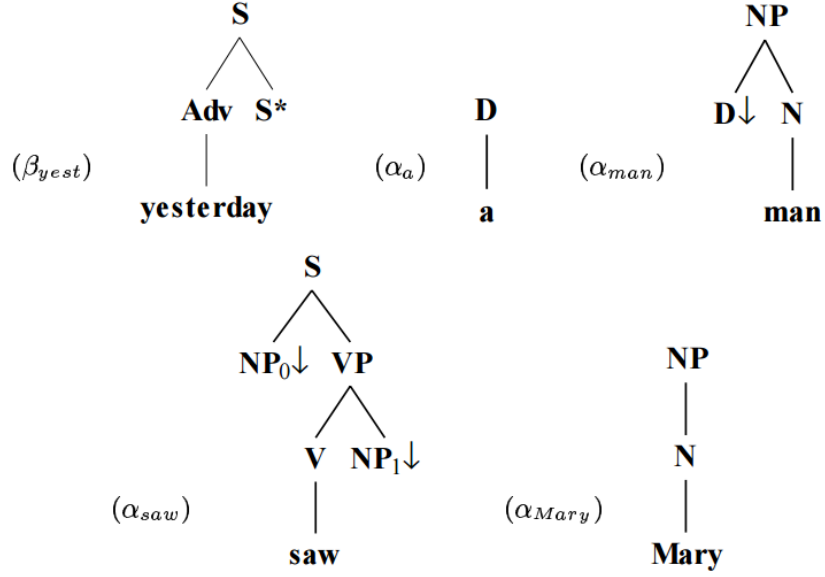


Figure 2.8: Some elementary trees from Joshi & Schabes (1997) p.7 α trees are initial and the β tree is auxiliary

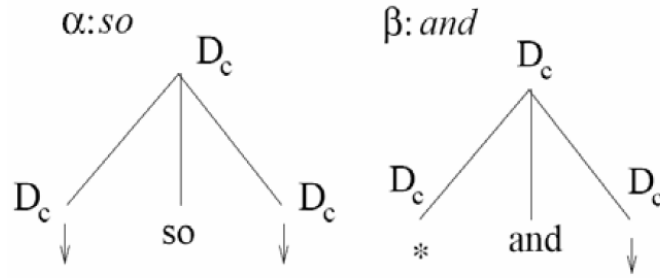


Figure 2.9: Initial tree for the coordinate conjunction *so*, auxiliary tree for the simple coordinator *and* from B. Webber et al. (2003) p.31-32

may have multiple parents violating one of the constraints of the tree structure; but since the discourse adverbial takes the other argument anaphorically, the non-crossing constraint is not violated.

- (8) (a) ***Dedektif romanı içinden çıkılmaz gibi görünen esrarlı bir cinayetin çözümünü sunduğu için***, her şeyden önce mantığa güveni ve inancı dile getiren bir anlatı türüdür ve bundan ötürü de burjuva rasyonelliğinin edebiyattaki özü haline gelmiştir.

Because it unravels the solution to a seemingly intricate murder mystery, the detective novel is a narrative genre which primarily gives voice to the faith and trust in reason and therefore, it has become the epitome of bourgeois rationality in the literature.

- (b) ***Dedektif romanı içinden çıkılmaz gibi görünen esrarlı bir cinayetin çözümünü sunduğu için***, her şeyden önce mantığa güveni ve inancı dile getiren bir anlatı türüdür **ve** bundan ötürü de **burjuva rasyonelliğinin edebiyattaki özü haline gelmiştir**.

Because it unravels the solution to a seemingly intricate murder mystery, the detective novel is a narrative genre which primarily gives voice to the faith and trust in reason **and** therefore, it has become the epitome of bourgeois rationality in the literature.

- (c) *Dedektif romanı* içinden çıkılmaz gibi görünen esrarlı bir cinayetin çözümünü sunduğu için, her şeyden önce mantığa güveni ve inancı dile getiren bir anlatı türüdür ve **bundan ötürü** de burjuva rasyonelliğinin edebiyattaki özü haline gelmiştir.

Because it unravels the solution to a seemingly intricate murder mystery, the detective novel is a narrative genre which primarily gives voice to the faith and trust in reason and **therefore**, it has become the epitome of bourgeois rationality in the literature.

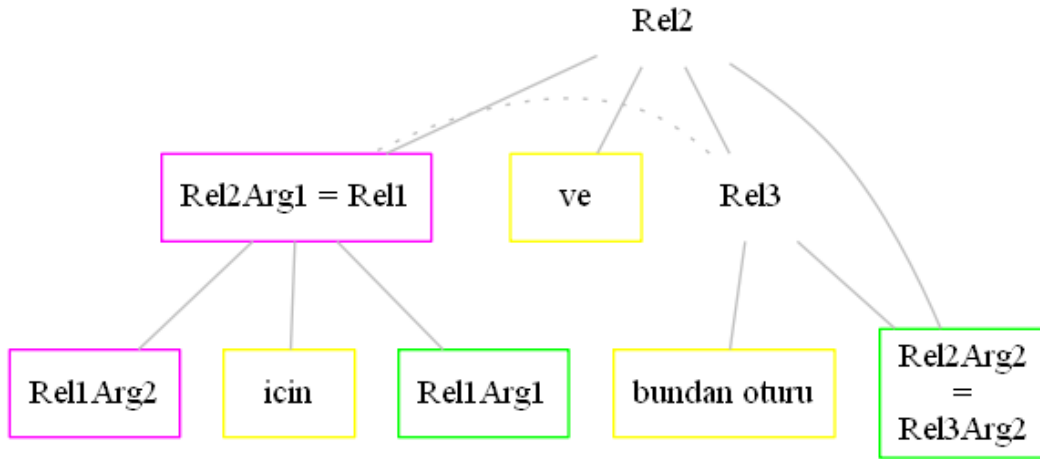


Figure 2.10: Violated tree structure for (8)

Bundan ötürü ‘therefore’ takes one argument anaphorically, shown as a dotted line in this representation. Since the anaphora is non structural, there is no crossing in (8). However, tree structure is still violated because *Rel2* and *Rel3* share an argument, resulting in multiple-parent structure.

Implicit connectives always link two adjacent spans structurally, the host span by substitution and the other by adjoining. Since after adjunction the initial immediate dominance configurations are not preserved, the semantic composition is defined on the derivation tree rather than the derived tree (Forbes et al., 2003; Forbes-Riley et al., 2006).

2.2.1.6 The Penn Discourse Tree Bank (PDTB)

The Penn Discourse Treebank (PDTB) (Prasad et al., 2008), although intended as a theory-neutral language resource, is loosely based on D-LTAG: the discourse connectives are annotated as discourse level predicates with two arguments; but the focus is no longer on the global structure of discourse but on individual relations.

Explicit connective in the PDTB is annotated for their connective span and two argument

spans, as well as the modifier span if available. Implicit connectives are either inserted, or selected from a predefined list of AltLex, EntRel, and NoRel.

All connectives are annotated for sense and attribution. The sense of connective is selected from the PDTB sense hierarchy 2.11. Connectives are allowed multiple senses. Attribution annotation includes the attribution span, the source and the type of attribution, and the scope and the determinacy of the attribution. Attribution is annotated as a feature of the relation and not as a structural constituent.

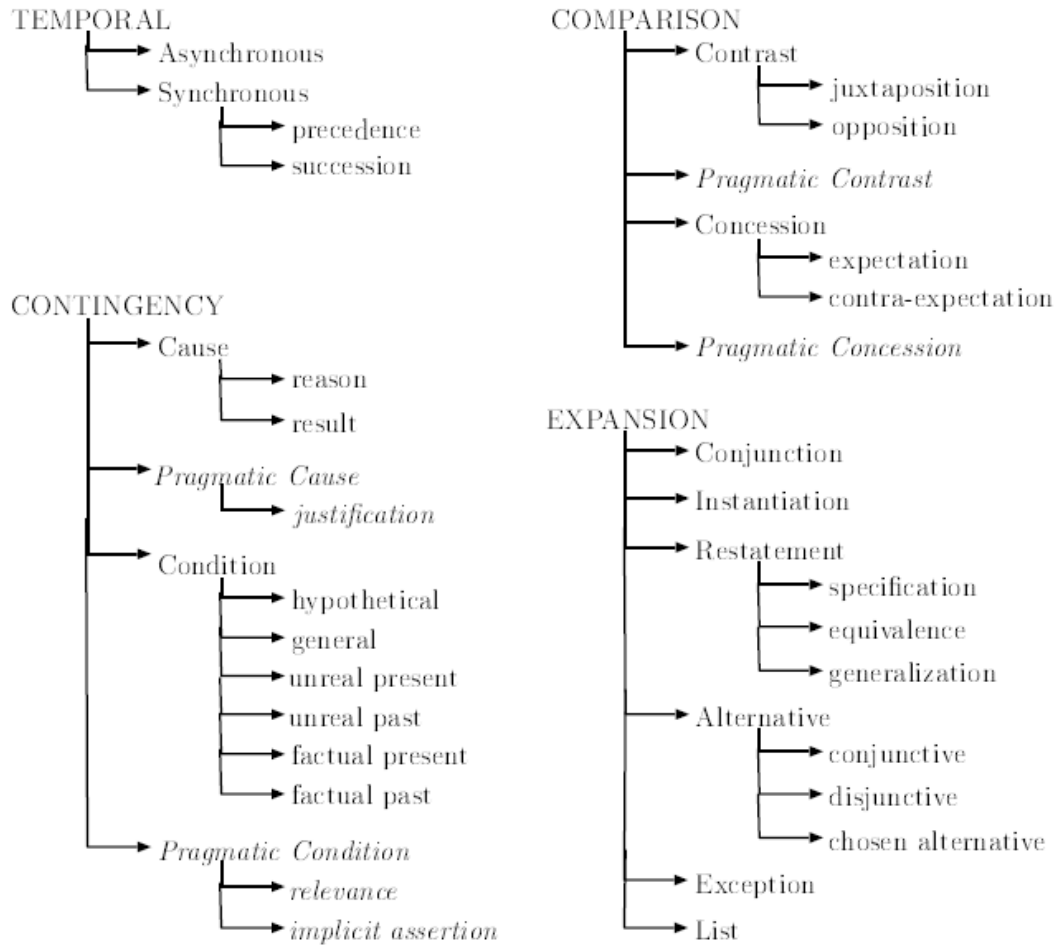


Figure 2.11: The PDTB sense hierarchy (Prasad et al., 2007), p. 27

2.2.1.7 Discourse Combinatory Categorical Grammar (DCCG)

Just as D-LTAG is the extension of Lexicalized Tree Adjoining Grammar to discourse, Discourse Combinatory Categorical Grammar (DCCG) is the extension of Combinatory Categorical Grammar (CCG) to discourse Nakatsu & White (2010). Like DLTAG, the DCCG focuses on connectives, and recognizes structural and adverbial connectives, the latter taking one of their arguments anaphorically.

Unlike DLTAG, which provides a second, distinct layer of syntactic structure for discourse, DCCG is truly an extension of the CCG. Discourse connectives are lexical items that take

sentential arguments to produce sentential outputs (15).

$$\begin{aligned} \textit{on the one hand} &\vdash s_{\text{ot1h}/\diamond} s_{\text{nil}/*} \textit{punc}, \\ \textit{on the other hand} &\vdash s_{\text{otoh}/\diamond} s_{\text{nil}/*} \textit{punc}, \end{aligned}$$

Figure 2.12: Lexical categories for *on the one hand* and *on the other hand*, Nakatsu & White (2010), p.21

Although CCG has mildly context sensitive power and can go beyond simple tree-structure, the nature of discourse connectives as simple binary predicates is likely to result in clean tree structures for structural connectives. An example of nested contrastive relations is given in 2.13. If DCCG adopts the somewhat circular criterion of discourse adverbials as discourse connectives that enter more complex relations, the anaphoric nature of the first arguments of the discourse adverbials is likely to eliminate any violation of tree structure.

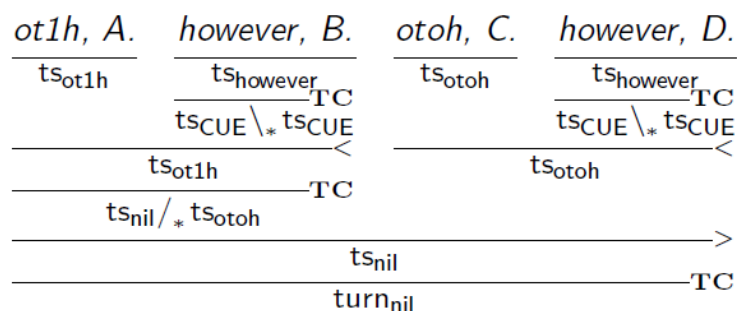


Figure 2.13: A DCCG derivation of nested contrast relations, Nakatsu & White (2010) p.25

Nakatsu & White (2010) propose employing Hybrid Logic Dependency Semantics (HLDS) (Kruijff, 2001; Baldrige & Kruijff, 2002) for DCCG. The sense of the connective is introduced in its HLDS representation. For example the semantics for *on the one hand* in 2.13 would be $@_e(\textit{contrast} - \textit{rel} \wedge \langle \textit{Arg1} \rangle e_1 \wedge \langle \textit{Arg2} \rangle e_2)$, introducing the sense *contrast-rel*.

2.2.2 Deviations from Tree Structure

2.2.2.1 Complex Interactions Between Trees

The trees proposed by Hobbs (1985) can connect or intertwine at the peripheries. This means that there is both multiparenting and crossing at boundaries. Although inner nodes of the trees are not available for these interactions, computationally the structure could be as complex as chain graphs in order to accommodate these interactions - unless the peripheries are handled non-structurally.

2.2.2.2 The Segmented Discourse Representation Theory (SDRT)

The Segmented Discourse Representation Theory (SDRT) (Asher, 1993) expands the basic Discourse Representation Theory (DRT) proposed by Kamp (1981) by introducing a constituent structure for DRT, a dynamic semantic representation, in an attempt to extend the

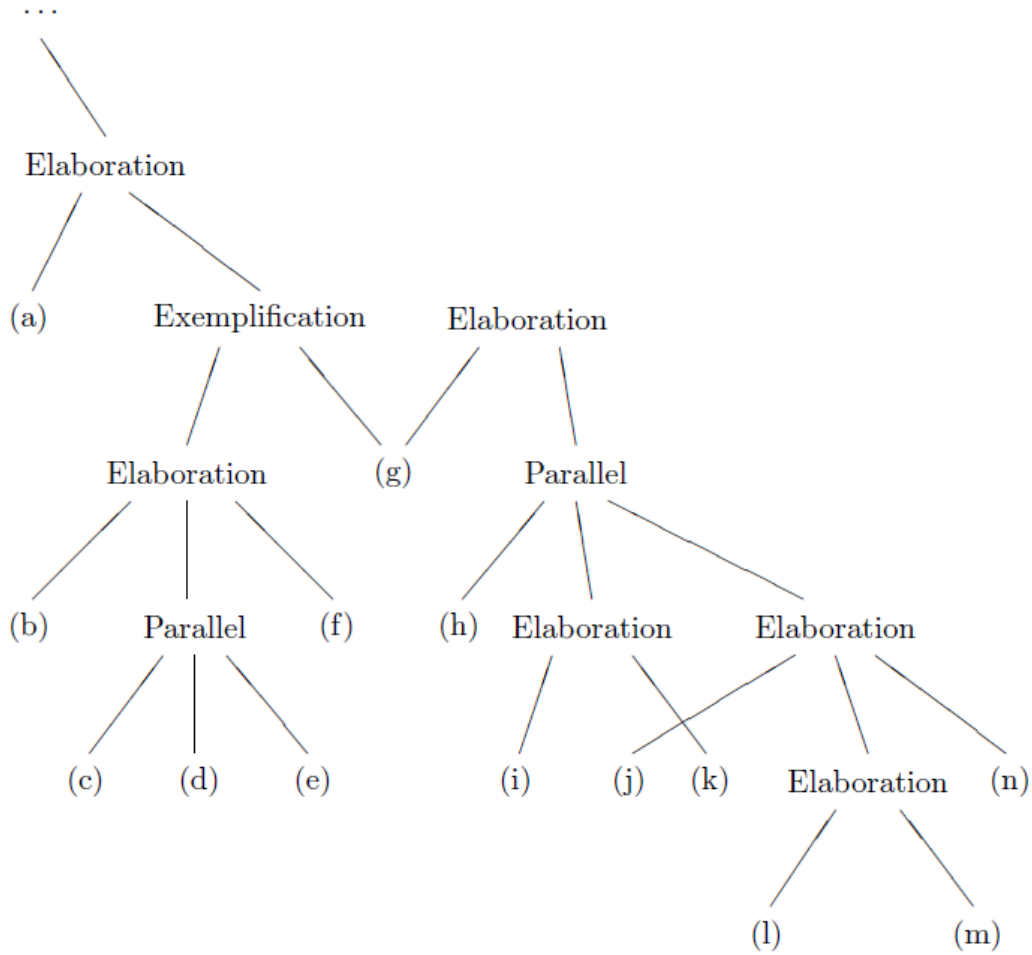


Figure 2.14: Intersecting and intertwining trees from Hobbs (1985) p. 30

theory to cover a wider range of anaphoric phenomena including reference to abstract objects. The constituent graphs are trees, but they are overlaid with arrows that donate tree isomorphisms. Tree isomorphism representations are used for revision of the trees as they are dynamically built. However, the final constituent graphs may include tree isomorphisms as in 2.15, the DRS and modified embedding trees for (9).

- (9) Every Swiss farmer who owns a donkey beats it. But if Austrian farmer does, he doesn't.

Since all discourse relations are considered to be inferential in SDRT, the formal distinction between tree-forming relations and isomorphism-depicting relations, and therefore the computational complexity of the constituent trees, are unclear.

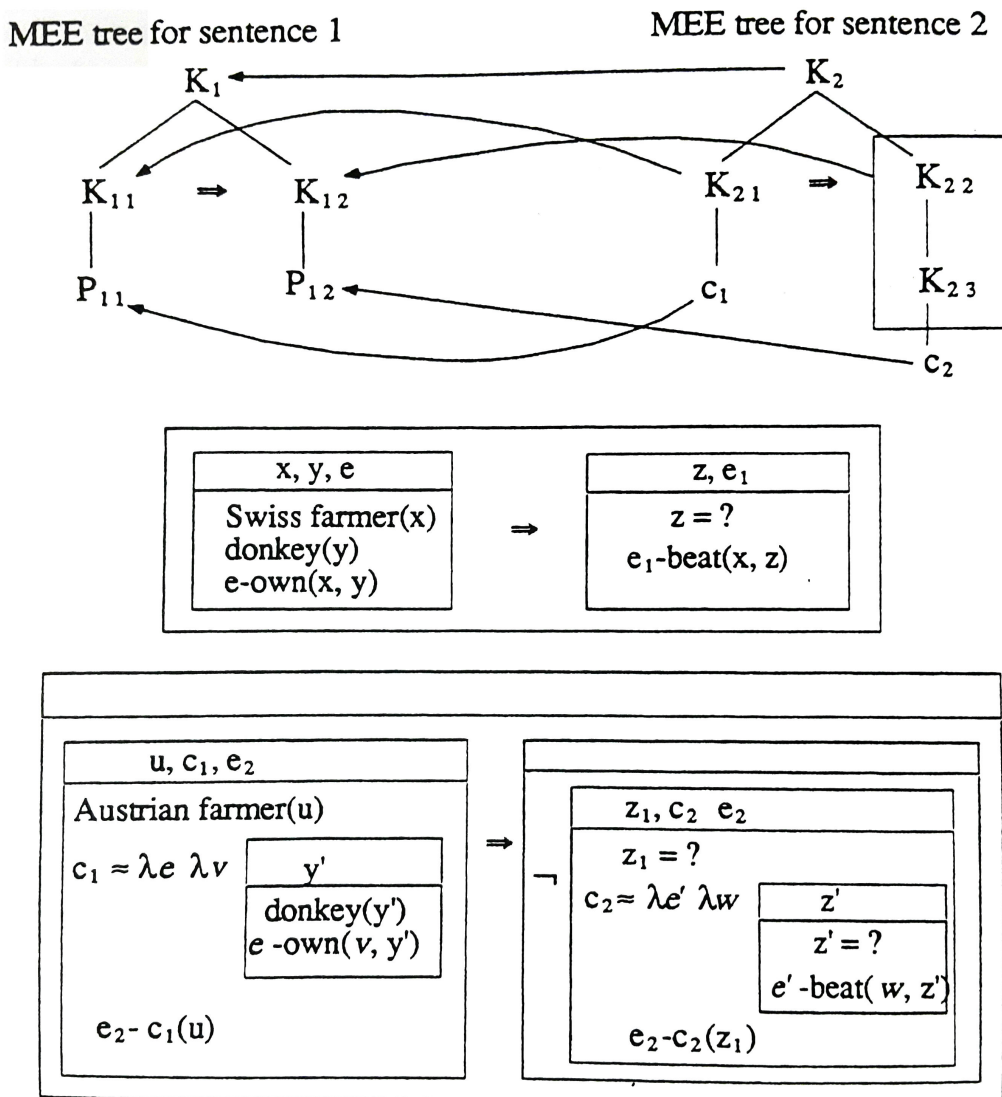


Figure 2.15: Modified embedding trees and DR for (9) (Asher, 1993, p. 364)

2.2.3 Other Data Structures

2.2.3.1 Extended Coherence Relations

Wolf & Gibson (2004, 2005), judging from a corpus annotated for a set of relations that is based on Hobbs (1985), argue that the global discourse structure cannot be represented by a tree structure. They point out that the definition for the anaphoric connectives in D-LTAG seems to be circular, since they are defined by their anaphoric arguments which can be involved in crossing dependencies, and in turn they are defined as anaphoric and thus outside the structural constraints. They propose a chain graph-based annotations scheme, which they claim express the discourse relations more accurately than RST, because the relations can access embedded, non-nuclear constituents that would be inaccessible in an RST tree.

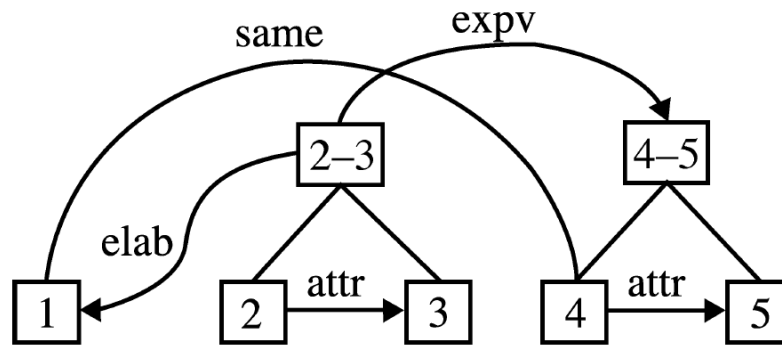


Figure 2.16: Coherence graph from Wolf & Gibson (2005) p. 267

2.2.3.2 Tree Structure Violations in Penn Discourse Treebank (PDTB)

Since Wolf & Gibson use attribution and same relations, which are not considered discourse relations in D-LTAG or the PDTB, a direct comparison of chain graph annotations and the PDTB does not seem possible at this point; but violations of tree structure are also attested in the PDTB.

Lee et al. (2006, 2008) investigate the PDTB and identify dependencies that are compatible with tree structure, *independent relations* and *full embedding*; as well as incompatible dependencies, *shared argument*, *properly contained argument*, *partially overlapping arguments*, and *pure crossing*. They claim that only shared arguments (same text span taken as argument by two distinct discourse connectives) and properly contained arguments (a text span that is the argument of one connective properly contains a smaller text span that is the argument of another connective) should be considered as contributing to the complexity of discourse structure; the reason being that the in-stances of partially overlapping arguments and pure crossing can be explained away by anaphora and attribution, both of which are non-structural phenomena. The presence of shared arguments carries the discourse structure from tree to directed acyclic graphs (B. Webber et al., 2012).

Aktaş et al. (2010) have identified similar tree structure violations in the Turkish Discourse Bank (TDB) (Zeyrek et al., 2010). In addition to the dependencies in Lee et al. (2006), Aktaş et al. have identified *properly contained relations* and *nested relations*. A quantitative analysis of the tree structure violations will be presented in 3

2.2.3.3 Multi-satellite constructions (MSC) in RST

Egg & Redeker (2008, 2010) argue that tree structure violations can be overcome by applying an underspecification formalism to discourse representation. They adopt a weak interpretation of nuclearity, where although the relation between an atomic constituent and a complex constituent is understood to hold between the atomic constituent and the nucleus of the complex constituent, structurally the relation does not access the nucleus of the complex, and therefore does not result in multiple parenting. This approach is not directly applicable to PDTB-style relations, because of the minimality principle, which constrains the annotators to select the smallest text span possible that is necessary to interpret the discourse relation when

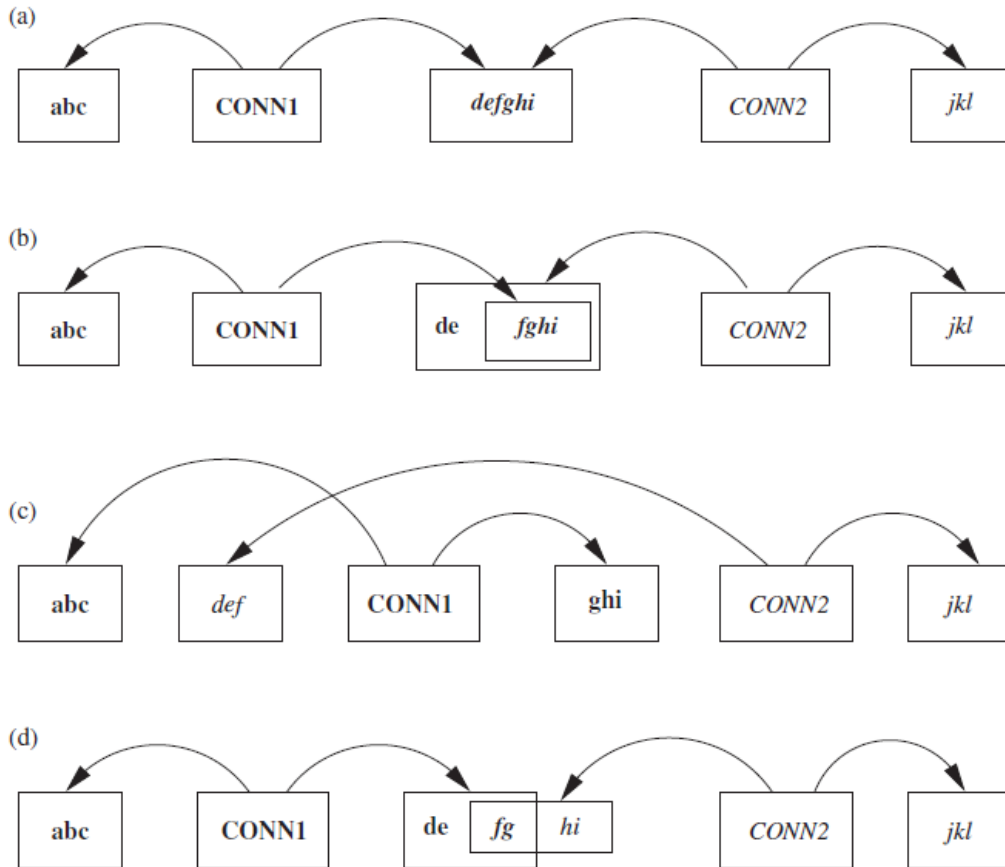


Figure 2.17: Non-tree-like dependency structures in PDTB (a) Shared argument; (b) Properly contained argument; (c) Pure crossing; (d) Partially overlapping arguments Lee et al. (2006) p. 84

annotating the arguments of a discourse connective.

Egg & Redeker also argue that most of the crossing dependencies in Wolf & Gibson (2005) involve anaphora, which is considered non-structural in discourse as well as in syntax.

However, they admit that multi-satellite constructions (MSC) in RST, where one constituent can enter into multiple rhetorical relations as long as it is the nucleus of all relations, seems to violate tree structure. They state that only some of the MSCs can be expressed as atomic-to-complex relations, but they also state that those the MSCs that cannot be expressed so seems to be genre specific. The fact that both Egg & Redeker (2008) and Lee et al. (2008) cannot refute the presence of multiple parenting in discourse structure is striking.

2.2.4 Spoken Language

All studies cited above investigate discourse structure in written texts. There are spoken corpora annotated for RST such as Stent (2000) and SDRT Baldrige & Lascarides (2005), but the only PDTB-style spoken discourse structure annotation within the author's knowledge is part of the LUNA corpus in Italian (Tonelli et al., 2010).

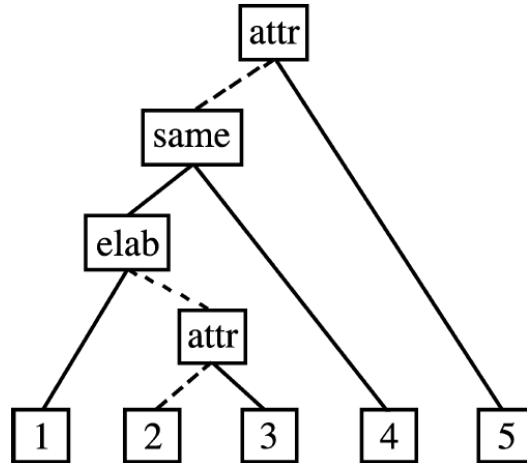


Figure 2.18: RST tree for the same example in 2.17 from Wolf & Gibson (2005) p. 267

The most striking change Tonelli et al. made in the PDTB annotation scheme when annotating spoken dialogues is to allow for implicit relations between non-adjacent text spans due to higher fragmentation in spoken language. They also added an interruption label for when a single argument of a speaker was interrupted. Some changes to the PDTB Sense Hierarchy was necessary including the addition of the GOAL type under CONTINGENCY class, fine tuning of PRAGMATIC sub-types, exclusion of LIST type from EXPANSION class and merging of syntactically distinguished REASON and RESULT subtypes into a semantically defined CAUSE type.

No structural analysis of Tonelli et al.'s data is available for the time being.

Whether tree structure is sufficient to represent discourse relations is an open question that will benefit from diverse studies in multiple languages and modalities. Here we have presented some of the arguments for and against tree structure in discourse. The current study aims to reveal the constraints in simultaneous spoken Turkish discourse structure. The proposed framework for dis-course structure analysis is based on PDTB-style, with adjustments for Turkish and spoken language. The adjustments will be based on the existing PDTB-style studies in Turkish conversational speech, although they are likely to evolve further as research progresses. The methodology for the study is to search for possible tree-violations, and try to apply the explanations in the literature to explain them away. The violations that cannot be plausibly explained away by non-structural mechanisms should be accommodated by the final discourse model.

CHAPTER 3

TURKISH DISCOURSE STRUCTURE

3.1 Data

3.1.1 Turkish Discourse Bank

Turkish Discourse Bank (TDB) is the first large-scale publicly available language resource with discourse level annotations for Turkish built on an approximately 400,000-word sub-corpus of METU Turkish Corpus (MTC) (Say et al., 2002)(Say et al., 2002), annotated in the style of Penn Discourse Tree Bank (PDTB) (Prasad et al., 2008). Connectives are annotated together with their modifiers and arguments, and with supplementary materials for the arguments (Zeyrek et al., 2013).¹

Penn Discourse Tree Bank (PDTB) takes inspiration from D-LTAG as the framework for annotation. Theoretically, D-LTAG treats discourse connectives as discourse level predicates that take as argument two text spans that can be interpreted as abstract objects (facts, events, situations, propositions, etc.) Asher (1993); B. Webber (2004). The fundamental components of the PDTB annotation framework are explicit and implicit connectives, their two arguments, and their senses. The PDTB also annotates the material that semantically supplements the first or the second argument, as well as attribution. The TDB 1.0 includes explicit discourse connectives, their two arguments, modifiers, supplementary materials and the shared elements amounting to 197 files and 8483 relations.

As in PDTB, the connectives in TDB come from a variety of syntactic classes (Zeyrek et al., 2008). The coordinating and subordinating conjunctions such as *ve* ‘and’ and *için* ‘for’ and ‘in order to’, respectively, are considered structural connectives, meaning that they take both arguments structurally. Discourse adverbials and phrasal expressions that are built by combining a discourse-anaphoric element with a subordinating conjunction are considered to be anaphoric connectives, meaning that they only take the argument that is syntactically related, and the other argument is interpreted anaphorically. In PDTB and TDB style, the syntactically related argument is called the second argument (Arg2), and the other argument is called the first argument (Arg1), for both structural and anaphoric connectives (Zeyrek et al., 2013).

The TDB 1.0 annotations were created manually with three different annotation procedures: independent annotation (IA), group annotation (GA) and pair annotation (PA). Regardless

¹ The first release of TDB is freely available to researchers at <http://medid.ii.metu.edu.tr/>

Table 3.1: Connective class breakdown of discourse connectives in the TDB

Syntactic Class	No. of relations in TDB	% of relations in TDB
Coordinators	4477	52.78 %
Subordinators	2287	26.96 %
Discourse Adverbials	1225	14.44 %
Phrasal Expressions	494	5.82 %
Total	8483	100 %

of the annotation procedure, the annotators are asked to obey the minimality principle, i.e. they have to select as arguments the minimal textual span necessary to interpret the discourse relation (Prasad et al., 2008). The minimality principle ensures that the annotators focus on the local text while annotating a particular discourse connective without having to remember the global structure of the text. All the annotations are adjudicated in periodical agreement meetings with the leadership of at least one of the research team members. The leader helps the annotators to resolve the differences (if any) and the team produces an agreed version of the annotations unanimously.

In the IA procedure, the data is triply-annotated blindly; i.e. three annotators annotate the data without seeing the others' annotations, and the other search tokens previously annotated on the file. In the GA procedure, the annotators gather to produce a single set of annotations for a search token, noting any disagreements to be discussed in a subsequent agreement meeting. In the PA procedure, a pair of annotators produces a single set of annotations, which is blind to a third annotator's annotations.

The PA process, inspired by Pair Programming, is a novel annotation approach developed during the TDB project. Section 4.0 below explains this procedure in more detail. Of the total 8483 relations in the TDB 1.0, 3804 (44.84%) discourse relations were annotated by the IA procedure, 3985 (46.98%) by PA, and 694 (8.18%) were annotated by GA (Zeyrek et al., 2013).

When the inter-annotator reliability among three (independent) annotators stabilized, a new procedure was proposed, namely the use of a pair of annotators to carry out the task together. We call the procedure Pair Annotation after the pair programming (PP) procedure in software engineering (Demirşahin, Yalçınkaya, & Zeyrek, 2012).

PP is a collaborative programming paradigm where two programmers work on an algorithm or a piece of code as a unit, assuming equal responsibility and credit for the work done (Williams et al., 2000). The unit is composed of two roles, the driver and the navigator. The driver is the one who is physically creating the code or algorithm, whereas the navigator is the one who monitors the driver. The monitoring is an active process: the navigator is expected to be involved in the creation of the code at all times by watching for errors, suggesting alternatives and supplementing the driver with additional resources when necessary. The pair periodically switches the roles of the driver and the navigator. Maintaining active involvement of the navigator and changing roles regularly ensures that the pieces of code created via PP does not only belong to the programmer who was the driver at the time, but the pair as a unit; i.e. the result is a joint ownership.

The PA annotation procedure emerged out of the need to accelerate the annotation process. It was proposed by two of the annotators quite independently of PP, and its principles emerged

in a short time on their own accord. In quite a spontaneous way, one of the annotators came to annotate the data while the other annotator checked, corrected otherwise simply agreed with the first annotator's annotation. Therefore, the roles of the driver and the navigator used in the PP literature arose. The PA, then, is the procedure where one of the annotators assumes the driver role physically handling the keyboard and the mouse with the other annotator sitting next to her, looking at the screen and working together with her as a navigator as in PP. The driver and navigator roles are occasionally switched between the annotators, as in PP. To assess the reliability of pair-annotations, we always compare them with the annotations produced by a third, independent annotator.

Demirşahin & Zeyrek (in press) observed that in the PA procedure, physical errors, e.g. erroneously leaving a few letters of a word unmarked, or selecting spaces at the peripheries of the arguments are more easily noticed and corrected: the navigator readily sees such mistakes and warns the driver who then corrects them immediately. A related benefit is that the annotation of ambiguous cases can be handled more efficiently because the pair can easily resolve the ambiguity by discussing the options among them. The end result of this collaborative task is fewer disagreements in the annotations.

Demirşahin & Zeyrek also noticed that the annotators have a higher motivation during the PA procedure, as mentioned in the PP literature. During PA, the annotators are quite focused on the task and can easily resist being sidetracked since they do not want to waste each other's time. In our case, annotating numerous instances of the same connective is often monotonous. The pair of annotators uses the advantage of having a partner to collaborate, discuss, and occasionally joke to lighten up the mood. Thus, the task that is tiresome when carried out alone becomes interactive and pleasant when carried out with a partner.

Thirdly, the PA can be time saving because the pair is well prepared for the discussion of the hard cases in the agreement meetings. The pair annotators share the results of their discussions with the research team (through the notes field of the annotation tool) and offer their solution resulting from in-depth discussions and careful thinking. In hard cases, the pair annotators were particularly careful in recording their first intuitions and their reasoning process in producing the joint annotation; sometimes they even declared an unresolved difference of opinion. These comments were highly beneficial for the research team as they provided more insight about the reasoning behind the annotation itself, thus accelerating the agreement meetings.

One of the most prominent objections against PP is the increased man-hours. In the IA procedure, three annotators produce three sets of annotations, whereas in the PA procedure, three annotators produce two sets of annotations; it is as if PA increases the cost of a set of annotations by 50%. Yet, the benefits are high because the PA procedure increases the annotation pace of the pair and increases the quality of the annotations.

Another concern is the possibility of losing the input of one of the annotators, most likely those of the navigator. This can take place in several ways. For example, the navigator may lose interest and watch passively as the driver annotates, or the driver may take control over the whole annotation and ignore the input from the navigator. The TDB team was an already well-established research group before the inception of PA, and the annotators had intrinsic and extrinsic motivations to produce a high quality corpus in a limited time; hence these issues did not arise. In other projects where annotators are not a part of the research team or their involvement is limited to annotations only, they might be inclined to overlook the principles

of PA. If such cases arise, it would be advisable to incorporate peer evaluation to get periodic feedback and ensure that the procedure is working as intended.

These concerns are common to PP and PA, but issues specific to annotation projects may also arise. In annotation projects it may be desirable to involve several annotators to annotate the same text files so as to capture the intuitions of many native speakers. PA may appear as if a limited range of native speaker intuitions is captured. It may also be argued that the constant interaction between the pair may contaminate their own intuitions. To avoid both criticisms, we have effectively utilized the notes field in the DATT to record the annotators' initial intuitions in cases when one of them felt that the pair annotation did not reflect her intuitions. Thus in the agreement meetings, the intuitions of each annotator were taken into consideration to ensure that the input from one of the annotator was not lost.

Demirşahin & Zeyrek do not claim that PA is the solution to all problems in annotation, or that it offers the perfect annotation procedure. That is why we suggest keeping an independent individual annotator in the process. As such, this procedure is akin to having two independent annotators, where one of the annotators is like a composite consisting of two individuals thinking independently but producing a single set of annotations collaboratively. Similar to the joint ownership of PP, neither annotator claims the annotation as her own. It is treated as a single set of annotations both during the agreement meetings and in calculating the agreement statistics.

3.1.2 Spoken Turkish Corpus Demo

The Spoken Turkish Corpus demo version is an approximately 20,000-word resource of spoken Turkish². The demo version contains 23 recordings amounting to 2 hours 27 minutes. Twenty of the recordings include casual conversations and encounters, comprising 2 hours 1 minutes of the total, the 3 remaining recordings are broadcasts lasting a total of 26 minutes. The casual conversations include a variety of situations such as conversations among families, relatives and friends, and service encounters. The broadcasts are news commentaries. The topics of conversation range from daily activities such as infant care and naming babies to biology e.g. the endocrine system, to politics such as European Union membership process or the clearing of the mine fields on Syrian border. Such wide range of topics provide for a wide coverage of possible uses of discourse connectives even in such a relatively small corpus.

The STC Demo was annotated using the Discourse Annotation Tool for Turkish (DATT) (Aktaş et al., 2010). We used the transcription texts included in the STC Demo version as the DATT input and provided the annotators with separate audio files.

This approach was a trade-off: the annotators could not make use of the rich features of the time-aligned annotation of the STC; but by importing text transcripts directly into an existing specialized annotation tool we did not have to go through any software development and/or integration stage. The annotators reported only slight discomfort in matching the text and the audio file during annotation, but stated that it was manageable as few of the files are long enough to get lost between the two environments.

² The STC Demo is available to researchers for free at <http://std.metu.edu.tr/en/>. At the time of the completion of this thesis, a revised version of the STC Demo was released; however, the study could not be reconducted for the revised version due to time constraints.

Some of the challenges of annotating discourse connectives we have already observed in written language transfer to the spoken modality. For example, in written discourse it is possible for an expression to be ambiguous between a discourse and non-discourse use, as the anaphoric elements can refer to both abstract objects and non-abstract entities. This applies to spoken language as well.

- (10) SER000062: Şey Glomerulus o yuvarlak topun adı mıydı (bu)? Ordan şey oluyor ...
AFI000061: hı-hı hı-hı
AFI000061: Süzülme ondan sonra oluyor ama. Şu Henle kulpu falan var ya. Şöyle geri.
“SER000062: Um Glomerulus was (this) the name of that round ball? Stuff happens there ...
AFI000061: Yes, yes.
AFI000061: Filtration occurs after that, though. That Loop of Henle and such. Reverse like this.”

In (10) *ondan sonra* ‘after that’ could be interpreted as resolving to the clause ‘Stuff happens there’, which is an abstract object although a vague one. The pronoun can also refer to *the glomerulus*, which is an NP. This was exactly the case during the annotation of this specific example: one annotator interpreted it as a temporal discourse connective that indicates the order of two sub-processes of kidney function, whereas the other annotator interpreted that ‘that refers to the NP and did not annotate this instance of *ondan sonra*. As a TDB principle, if an expression has at least one discourse connective meaning, it is annotated. As a result, this example was annotated as per the first annotator’s annotation.

- (11) (a) AFI000061: [_{Sup1}Tiroksin. Ha bak. Metabolizma hızımı artırıyor.]
...
(b) SER000062: *Tiroit bezinden tiroksin salgılanıyor.*
(c) AFI000061: Hmm salgılanıyor dedin sen. Tamam. Doğru.
(d) SER000062: Tamam.
(e) SER000062: Hatta tiroit şey olan... Emm **tiroidinde sorun olanlar çok ee şey olur ya aktif olur ya.**
(f) AFI000061: Hmm?
(g) SER000062: Çok hareketli olurlar. Evet.
(h) AFI000061: Onun için mi?
(a) “AFI000061:[_{Sup1}Thyroxin. Oh look. It speeds up the metabolism.]
...
(b) SER000062: *Thyroxin is secreted by the thyroid gland.*
(c) AFI000061: Hmm you said secreted. Ok. Right.
(d) SER000062: Ok.
(e) SER000062: Actually thyroid is the one that... Emm **you know, those who have problems with thyroid are ee they tend to be very active.**

- (f) AFI000061: Hmm?
- (g) SER000062: They tend to be very energetic. Yes.
- (h) AFI000061: Is (it) because of that?”

In spoken language, particularly spontaneous casual dialogue, phrasal expressions can take their first arguments from anywhere in the previous discourse. This is very much like discourse adverbials. For example, *için* in (11) displays an unattested use in TDB, as it appears distant from both its arguments, allowing the participant to question the discourse relation between two previous text spans. Given the supplemental material thyroxin increases the metabolism in line (a) by speaker AFI, speaker SER provides two propositions, thyroxin is secreted by the thyroid gland in line (b) and people with overactive thyroids tend to be hyperactive in line (e). In line (h), AFI offers a discourse connective because in order to show her understanding of the preceding discourse, i.e., something like ‘(so they tend to be very active) because of that?’, where the material in parentheses are elided. One can argue that this connective builds a new discourse relation with one anaphoric and one elliptic argument. Nevertheless, we kept the annotations as shown in the example, because (a) it was the most intuitive annotation according to the annotators and (b) the DATT does not allow annotation of ellipsis as arguments for now.

Another problem with spoken corpus is that some elements may be missing. There are many examples that could not be annotated as discourse connectives, because the speakers were interrupted before they could complete, or at times even start, the latter argument of a possible discourse relation. In other examples, the argument may be there but not recorded clearly, or may be completely inaudible even though they were uttered because of background noise or overlapping arguments.

3.2 Reannotation Methodology

The quantitative analysis in this study is two-fold. In the first stage, we analyzed the explicit connectives annotated on the TDB and the STC Demo. Following the structural analysis PDTB Lee et al. (2006) has done on the annotations of *however*, we have analyzed all annotations of explicit connectives on both corpora, we have determined the distributions of the inter-relational configurations that confirm to or deviate from tree-structure.

There are 2547 inter-relational interactions in the TDB and 164 in the STC Demo. Our first analysis shows that 1715 (67.31%) of those in the TDB and 81 (60.45%) of those in the STC Demo violates tree-structure constraints. In the second part of the study, we analyze the reasons for these violations in an attempt to pinpoint which tree-structure deviations should indeed be accommodated by the final discourse model.

First of all, we should keep in mind that the TDB 1.0 does not claim completeness. The TDB 1.0 contains annotations for explicit connectives only, and the annotation of implicit connectives are in progress. In addition, the discursive use of particles and the simplex subordinators, i.e., the subordinators that are composed of only suffixes and not postpositions were not annotated in TDB 1.0. Due to the lack of morphological analysis and part-of-speech tagging in the source data, the disambiguation of these highly polysemous morphemes were out of the scope of the initial project. In order to produce comparable data, the STD data was annotated only for the explicit connectives that were annotated in the TDB 1.0.

(12) 00001131 56&57

- (a) *Üzerine gittikçe sinirleniyor **ve** bir daha asla kapımı çalmayacağını düşünerek gitmeden önce bana öldürücü bir darbe vurup intikam almaya hazırlanıyordu.*
“She was getting angrier as she was pushed around **and** thinking that she won’t knock on my door anymore, she was getting ready to get revenge by giving me a fatal blow before leaving.”
- (b) *Üzerine gittikçe sinirleniyor ve bir daha asla kapımı çalmayacağını düşünerek gitmeden önce bana öldürücü bir darbe vurup intikam almaya hazırlanıyordu.*
“She was getting angrier as she was pushed around and **thinking** that she won’t knock on my door anymore, she was getting ready to get revenge by giving me a fatal blow before leaving.”
- (c) *Üzerine gittikçe sinirleniyor ve bir daha asla kapımı çalmayacağını düşünerek gitmeden önce bana öldürücü bir darbe vurup intikam almaya hazırlanıyordu.*
“She was getting angrier as she was pushed around and thinking that she won’t knock on my door anymore, she was getting ready to get revenge by giving me a fatal blow **before** leaving.”
- (d) *Üzerine gittikçe sinirleniyor ve bir daha asla kapımı çalmayacağını düşünerek gitmeden önce bana öldürücü bir darbe vurup intikam almaya hazırlanıyordu.*
“She was getting angrier as she was pushed around and thinking that she won’t knock on my door anymore, she was getting ready to get revenge **by giving me a fatal blow before leaving.**”

(12) illustrates how simplex subordinators take part in Turkish discourse relations, and how their annotation will change the structure of the annotated discourse. This sentence includes four explicit connectives. *Ve* ‘and’ is a coordinating conjunction and *(-mAdAn)*³ *önce* ‘before’ is a complex subordinator. Both connectives are annotated in TDB 1.0 as in (a) and (c), respectively. Without the annotation of simplex subordinator *-ArAk* ‘by’, the annotations in (a) and (c) result in a properly contained relation configuration, as the *önce* relation is completely contained in the *ver* relation, and the *-ArAk* clause is left out. The annotation of the relation expressed by *-ArAk* as in (b) will get rid of the tree-violation and result in a full embedding configuration instead of a properly contained relation configuration.

Notice that the annotation of the simplex subordinators do not necessarily change the distribution of discourse relation configurations in favor of tree-structure. The currently unannotated relation expressed by the other simplex subordinator in the sentence, *-Hp* ‘by, after’ as in (d) results in another properly contained relation configuration, as the relation as a whole is the complement of the verb *hazırlanıyordu* ‘was preparing’.

(13) 00006231 32&33

³ The vowels of the suffixes in Turkish harmonize with the final vowel of the stem, and the suffix-initial consonants may devoice due to assimilation. We use capital letters to represent the following sets of letters, to which they will realize in the surface form:

A = { a, e }

H = { ı, i, u, ü }

D = { d, t }

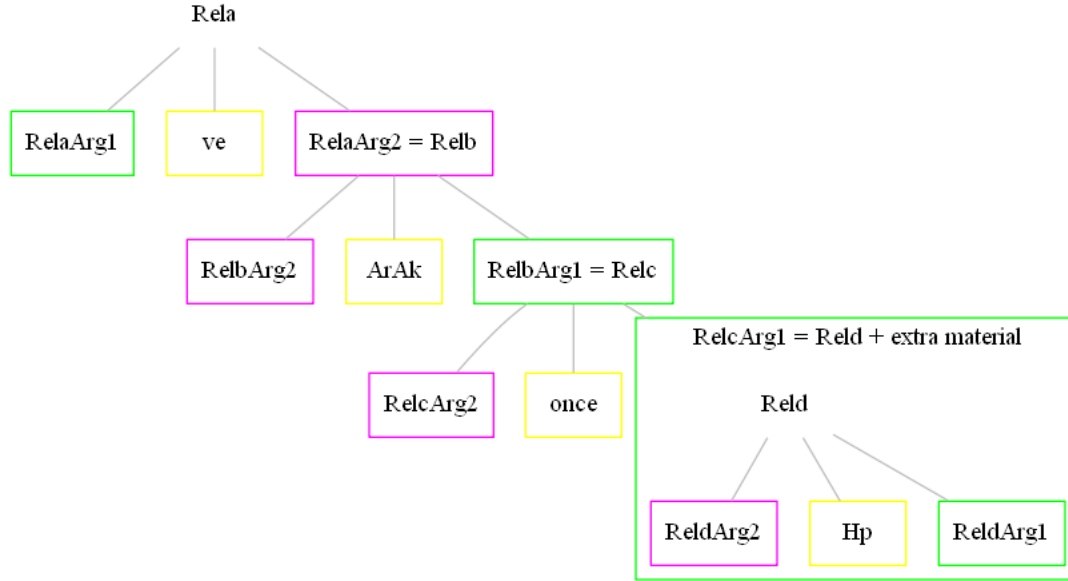


Figure 3.1: Final structure for (12)

- (a) *Hiçbir zaman birbirine uygun düşmeyecekti bu iki şey. (Implicit = ve) Uygun düştüğü sanıldığı zaman da hemen birbirlerinin üzerinden kayıp gideceklerdi.* Bu yüzden yaşam, baştan sona kaygı, acı çekme ve bunaltıydı.
“Those two things would never ever fit together. (Implicit = and) When they were thought to fit together, they would slip over each other. This is why life, from the beginning to the end, was worry, agony, and anxiety.”
- (b) *Hiçbir zaman birbirine uygun düşmeyecekti bu iki şey. Uygun düştüğü sanıldığı zaman da hemen birbirlerinin üzerinden kayıp gideceklerdi.* Bu yüzden yaşam, baştan sona kaygı, acı çekme ve bunaltıydı.
“Those two things would never ever fit together. When they were thought to fit together, they would slip over each other. This is why life, from the beginning to the end, was worry, agony, and anxiety.”
- (c) *Hiçbir zaman birbirine uygun düşmeyecekti bu iki şey. Uygun düştüğü sanıldığı zaman da hemen birbirlerinin üzerinden kayıp gideceklerdi. Bu yüzden yaşam, baştan sona kaygı, acı çekme ve bunaltıydı.*
“Those two things would never ever fit together. When they were thought to fit together, they would slip over each other. This is why life -from the beginning to the end- was worry, agony, and anxiety.”

(a) is an example of inter-sentential implicit connective, the only kind of implicit connectives annotated in the PDTB. (13) contains two explicit connectives *zaman* ‘when’ and *bu yüzden* ‘this is why’ which are annotated in TDB 1.0. Notice that in the PDTB *bu yüzden* ‘this is why’ would be considered an AltLex, i.e., an implicit connective. Here we remain loyal to the annotations in TDB 1.0 and treat it as an explicit connective of the phrasal expression type.

The two explicit connectives result in a properly contained relation configuration, as the first sentence has no explicit connections to the relation expressed by *zaman*, but is contained

in the relation expressed by *bu yüzden*. The insertion of an explicit connective *ve* ‘and’ or any other connective that expresses a simple expansion relation results in a full embedding configuration.

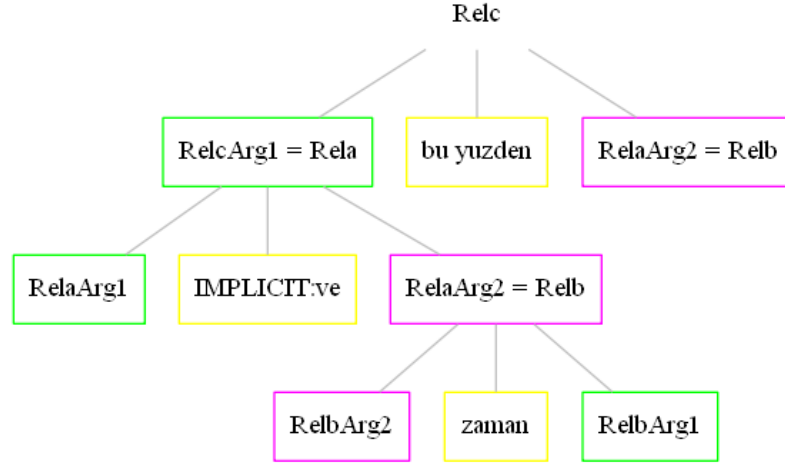


Figure 3.2: Final structure for (13)

Another important type of missing annotations in TDB 1.0 is intra-sentential implicit connectives which are not annotated in PDTB. However, consecutive clauses separated by commas within the same sentence is a common occurrence in Turkish, and they should be taken into account for a complete description of Turkish discourse structure.

(14) 00014113 14&15

- (a) **Ortaçağın kapanmasından sonra insanlığın gelişimi hızlanmış**, gelişim 18. yüzyılda en yüksek noktasına ulaşmış, süreç bu yüzyılda en klasik formuna erişmiştir. Bundan dolayı, 18. yüzyıla Aydınlanma Çağı denir
 “**After the end of the Medieval period the progress of mankind accelerated**, the progress peaked in the 18th century, the process reached its most classic form in this century. This is why, the 18th century is called the Age of Enlightenment.”
- (b) i. **Ortaçağın kapanmasından sonra insanlığın gelişimi hızlanmış**, (**Implicit = sonra**) **gelişim 18. yüzyılda en yüksek noktasına ulaşmış**, süreç bu yüzyılda en klasik formuna erişmiştir. Bundan dolayı, 18. yüzyıla Aydınlanma Çağı denir
 “*After the end of the Medieval period the progress of mankind accelerated*, (**Implicit = then**) **the progress peaked in the 18th century**, the process reached its most classic form in this century. This is why, the 18th century is called the Age of Enlightenment.”
- ii. **Ortaçağın kapanmasından sonra insanlığın gelişimi hızlanmış**, (**Implicit = sonra**) **gelişim 18. yüzyılda en yüksek noktasına ulaşmış**, **süreç bu yüzyılda en klasik formuna erişmiştir**. Bundan dolayı, 18. yüzyıla Aydınlanma Çağı denir
 “*After the end of the Medieval period the progress of mankind accelerated*, (**Implicit = and then**) **the progress peaked in the 18th century, the process reached its most classic form in this century**. This is why, the 18th century is called the Age of Enlightenment.”

- (c) Ortaçağın kapanmasından sonra insanlığın gelişimi hızlanmış, *gelişim 18. yüzyılda en yüksek noktasına ulaşmış, (Implicit = ve) süreç bu yüzyılda en klasik formuna erişmiştir.* Bundan dolayı, 18. yüzyıla Aydınlanma Çağı denir
 “After the end of the Medieval period the progress of mankind accelerated, *the progress peaked in the 18th century, (Implicit = and) the process reached its most classic form in this century.* This is why, the 18th century is called the Age of Enlightenment.”
- (d) *Ortaçağın indan sonra insanlığın kapanmasgelişimi hızlanmış, gelişim 18. yüzyılda en yüksek noktasına ulaşmış, süreç bu yüzyılda en klasik formuna erişmiştir. Bundan dolayı, 18. yüzyıla Aydınlanma Çağı denir*
 “After the end of the Medieval period the progress of mankind accelerated, *the progress peaked in the 18th century, the process reached its most classic form in this century. This is why, the 18th century is called the Age of Enlightenment.*”

(14) contains two explicit connectives, *sonra* ‘then’ and *bundan dolayı* ‘this is why’, which are annotated in TDB 1.0 as in (a) and (d). It also contains two intra-sentential implicit relations, as displayed in (b) and (c). (14)((b))i and (14)((b))ii are alternatives for the scope of the implicit temporal succession and/or expansion relation. Note that the explicit *sonra* is a complex subordinator meaning ‘after’, whereas the implicit *sonra* is a structural implicit connective which in meaning is akin to the discourse adverbial *sonra*, meaning ‘and then’.

Without the implicit relations, the structure appears to be another properly contained relation configuration. With the implicit connectives included, it results in either a full embedding configuration, or a full embedding/shared argument hybrid configuration.

Our analysis shows that these missing annotations, namely the lack of inter-sentential and intra-sentential implicit connectives, simplex subordinators, and the particles in the data is the direct cause of 308 (17.9 %) of the tree-structure violations in the TDB 1.0 and 31 (18.90%) in the STD Demo. The breakdown of the missing relations for the TDB 1.0 and the STD can be found below.

The ongoing annotation of implicit connectives and the planned annotation of simplex subordinators is likely to eliminate almost one-fifth of the tree-structure violations in the corpora, although as figure 4.1 and figure 4.2 demonstrate, they might possibly result in some additional non-crossing tree-violations.

Secondly, there are errors and inconsistencies in the annotations that create false tree-violations. In some relations a space, punctuation, or interjection that should have been left out were included in an argument. As a result, configurations that should be full embedding or shared argument showed up in the results as properly contained arguments or relations. 148 such errors were identified in the annotations and correcting these errors will result in eliminating 143 (8.34%) tree-violations in the TDB 1.0. 4 (4.94%) of tree-violations in the STC Demo were also eliminated by correcting such errors.

The annotation guidelines in the TDB 1.0 causes a small number of apparent tree-violations, too. When an argument contains the connective that anchors another discourse relation at its periphery, the connective is left out as a principle. Since that connective is part of another relation, it shows up as partially contained argument or relation in the inter-relational configuration. Apparent violations due to the guideline conventions make up only 19 (1.1 %) of

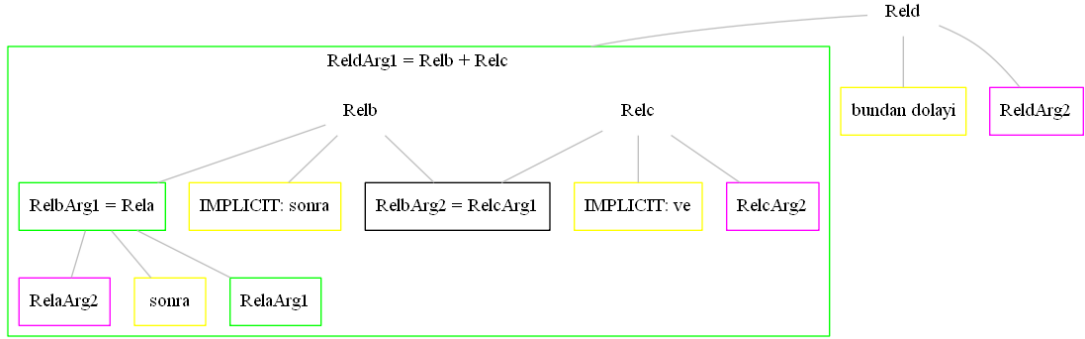


Figure 3.3: Full embedding/shared argument hybrid structure for (14) based on the annotation in (14)((b))i

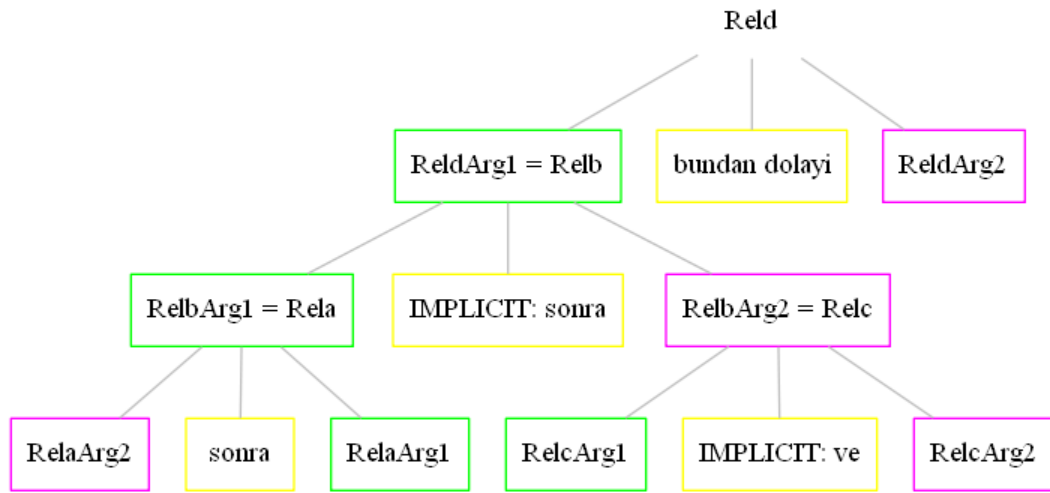


Figure 3.4: Full embedding structure for (14) based on the annotation in (14)((b))ii

the tree-violations in the TDB 1.0, and no such violations were attested in the STC Demo annotations.

Also, there is an artifact of the annotation style of the TDB 1.0 when it comes to multiple connectives denoting a single discourse relation. The TDB 1.0 was annotated connective by connective. On each pass, all instances of one search token was annotated. As a result, when multiple connectives denote a single relation, these connectives were annotated separately, each one on their own pass. In our analyses, these relations showed up as shared argument configurations as both the whole first argument and the whole second argument belonged to both connectives. We believe that these multiple connectives do not represent two distinct relations, thus we dubbed such cases *identical relation*.

(15) *Henüz çok iyi öğrenememiştim New York metrosunu ama gene de her gece gideceğim yere varabiliyordum.*

“I hadn’t learned the New York subway very well yet but still every night I could get to wherever I was going.”

In the TDB 1.0 137 identical relations make up 7.99% of the tree-violations, and in STD

Table 3.2: Breakdown of the unannotated relations in TDB 1.0

Unannotated relation	# of instances	% of unannotated	% of tree-violations
Inter-sentential implicit	145	47.08	8.45
Intra-sentential implicit	72	23.38	4.20
Simplex subordinator	89	28.90	5.19
Discourse particle	2	0.65	0.12
Total	308	100.00	17.96

Table 3.3: Breakdown of the unannotated relations in STC Demo

Unannotated relation	# of instances	% of unannotated	% of tree-violations
Inter-sentential implicit	26	83.87	15.85
Intra-sentential implicit	3	9.68	1.83
Simplex subordinator	1	3.23	0.61
Discourse particle	1	3.23	0.61
Total	31	100.00	18.90

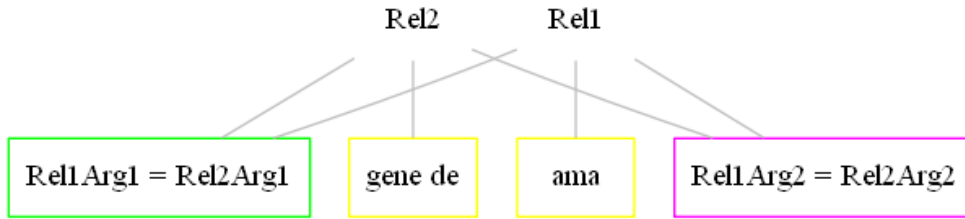


Figure 3.5: Shared argument configuration for (15)

Demo, 5 identical relations make up 6.17% of the tree-violations.

While selecting the boundaries of the spans that are connected by the discourse connectives, the PDTB/TDB approach applies the minimality principle which states that the annotators should select the minimal text span that is necessary for the interpretation of the connective. The minimality principle is an essential guideline that increases both the annotation speed and the inter-rater agreement, because it enables the annotators to discard the non-essential pieces of text that does not directly contribute to the core meaning of the connective. Such loosely related pieces of texts were considered to be more likely to be interpreted differently by different annotators, thus decreasing the inter-annotator agreement and increasing the noise in the data Zeyrek et al. (2010). For a connective-oriented annotation approach that aims to explore the linguistic aspects of the connectives or train NLP applications with data with as little noise as possible, this is a sound approach.

However, there is a downside to the minimality principle. It encourages the annotators to converge on the shortest span possible that is enough to get the core meaning of the connective, but it does not necessarily point to the whole spans of text that particular instance of the connective connects in the context of the current text.

- (16) (a) *Ali sinemaya gitmeyi seviyor. **Oysa** Ayşe tiyatroyu tercih ediyor.* Dahası, resim sergilerinden de hoşlanıyor.
*“Ali likes to go the movies. **But** Ayşe prefers plays.* Moreover, she enjoys art

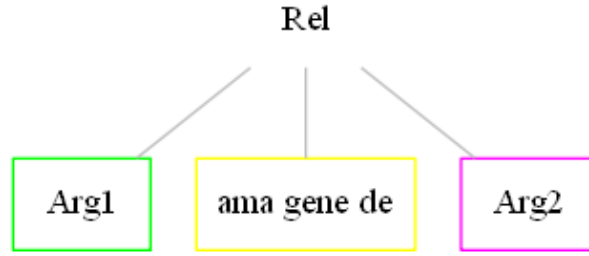


Figure 3.6: Identical relation configuration for (15)

exhibitions, too.”

- (b) Ali sinemaya gitmeyi seviyor. Oysa Ayşe tiyatroyu tercih ediyor. **Dahası, resim sergilerinden de hoşlanıyor.**

Ali likes to go the movies. But Ayşe prefers plays. **Moreover, she enjoys art exhibitions, too.**

For the constructed example (16), in the TDB/PDTB scheme the annotators are likely to select the first and the second sentences as arguments of *oysa* ‘but, however’ because these are the minimum spans that are necessary to interpret the connective. However, in this context, it is possible to extend the second argument of *oysa* to include the third sentence so as to contrast the things Ali likes and the things Ayşe likes. The minimality principle here serves to limit the possibilities for the annotators so as to make the annotation task as reliable as possible in terms of inter-annotator agreement, as well as making annotation easier, as hard cases increase the noise in the data and make machine learning more difficult Calhoun et al. (2010); however, it does not necessarily reflect the true structure in the text. *Dahası* ‘moreover’ takes the second and the third sentences as its arguments as it connects the things Ayşe likes. It is not possible to extend its first argument to the first sentence. The resulting structure is a *shared argument configuration*, which results in violation of tree-constraints since multiparenting is not allowed in trees. Without the minimality principle, it would be possible to extend the second argument of *oysa* to the third sentence, resulting in a *full embedding configuration*, which confirms to tree structure.

- (17) (a) Ali sinemaya gitmeyi seviyor. **Oysa Ayşe tiyatroyu tercih ediyor. Dahası, resim sergilerinden de hoşlanıyor.**

“Ali likes to go the movies. **But Ayşe prefers plays. Moreover, she enjoys art exhibitions, too.”**

- (b) Ali sinemaya gitmeyi seviyor. Oysa Ayşe tiyatroyu tercih ediyor. **Dahası, resim sergilerinden de hoşlanıyor.**

Ali likes to go the movies. But Ayşe prefers plays. **Moreover, she enjoys art exhibitions, too.**

In our analysis, we reinterpreted the relations in the non-independent relations in the corpora. Instead of looking for the minimal span necessary for the interpretation of the connective a la PDTB, or instead of imposing a predefined structure to the text a la RST, we loosened the minimality principle to see if this changes the particular configuration the relation participates.

This approach sometimes resulted in direct violation of the TDB guidelines, for example by including elaborations, examples, and explanations in the arguments, which were explicitly excluded from the arguments in order to comply with the minimality principle. However, if the adjacent spans were not extended simply for sake of expanding them. The guiding principle was the semantic integrity of the relation, if adding a span conflicted with the meaning conveyed by the connective or even changed it dramatically, that particular span was not included in the argument. For example:

- (18) (a) *Ağır ekonomik koşullar durgunluk yaratıyor. Sıfır hatta eksi kalkınma yaşıyor. Milli gelir dağılımındaki adaletsizlik sürüyor. **Ama, uygulanan ekonomik program yavaş yavaş ekonomiyi rayına oturtmak üzeredir.** Ancak, reçetedeki ilaçların acı tadı henüz halkın damağından silinmemiştir.*

*“Hard economic conditions create stagnation. Development rate falls to zero, even below zero. The injustice of the distribution of the national income persists. **But the economic program in progress is slowly putting the economy back on its track.** However, the bitter taste of the medications on the prescription has not been wiped away from the mouths of the people yet.”*

- (b) *Ağır ekonomik koşullar durgunluk yaratıyor. Sıfır hatta eksi kalkınma yaşıyor. Milli gelir dağılımındaki adaletsizlik sürüyor. Ama, uygulanan ekonomik program yavaş yavaş ekonomiyi rayına oturtmak üzeredir. **Ancak, reçetedeki ilaçların acı tadı henüz halkın damağından silinmemiştir.***

*“Hard economic conditions create stagnation. Development rate falls to zero, even below zero. The injustice of the distribution of the national income persists. **But the economic program in progress is slowly putting the economy back on its track.** However, the bitter taste of the medications on the prescription has not been wiped away from the mouths of the people yet.”*

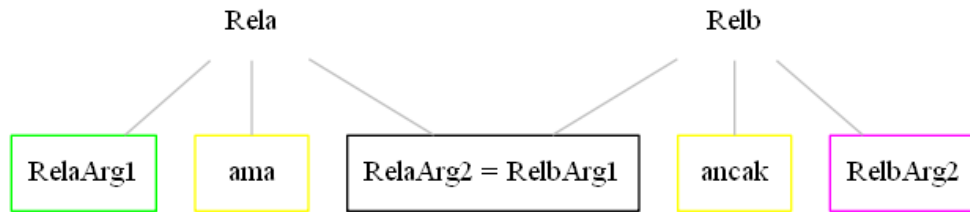


Figure 3.7: Shared argument configuration for (18)

In (18), the list of the negative conditions contrast with the expected recovery through the new program, which in turn contrasts with the ongoing unrest of the people. We cannot include the first argument of the first relation in the second relation, nor can we include the second argument of the second relation into the first relation without conflicting with the meaning of the anchoring connective. Unlike structure-oriented approaches that impose the presumed structure onto the text no matter what, we refrained from extending such relations in order to achieve tree-structure. As a result of this annotation exercise, we concluded that 480 cases could be reinterpreted, and of these reinterpretations 474 would result in tree structure. Notice that what we did was not trying to come up with the exact scope of the connective in its particular context, as this proves highly subjective in most cases. What we did was more akin to applying another principle, almost the exact opposite of the minimality principle, in

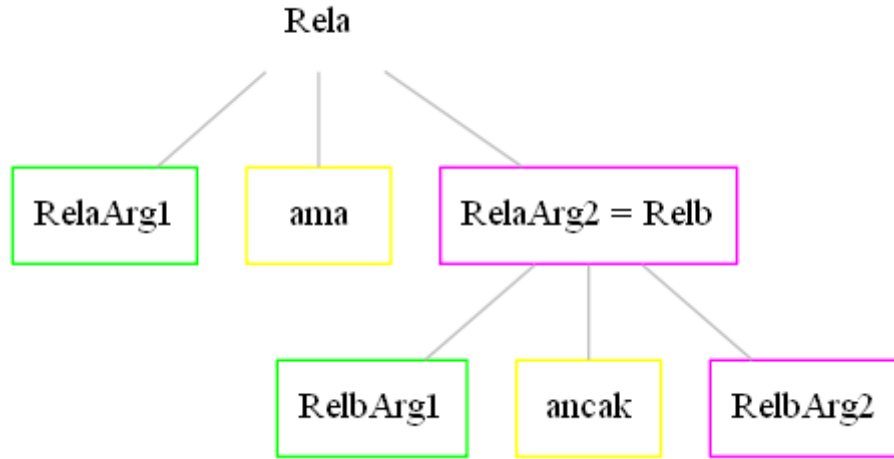


Figure 3.8: Full embedding configuration for (18). This reading is not available for this item

order to look for simpler inter-relation configurations. As a result, we saw that we could get rid of 474 (27.64%) of tree-violations through reinterpretation . Similarly, 38 configurations were reinterpreted in the STC Demo and as a result we eliminated 36 (44.44%) of the tree violations.

Missing annotations, false violations due to errors and leftout material due to the annotation guidelines, and reinterpretation can explain away a total of 1081 (63.03 %) tree-violations in the TDB 1.0 and 78 (96.3 %) tree-violations in the STC Demo. The remaining tree violations can not be reannotated in our current annotation scheme.

3.3 Discourse Relation Dependency Configurations in Written Turkish

3.3.1 Tree Structure

As mentioned in 2, Lee et al. (2006, 2008) identified *independent relations* and *fully embedded relations* as conforming to the tree structure, and *shared arguments*, *properly contained arguments*, *pure crossing*, and *partially overlapping arguments* as departures from the tree structure in PDTB. Although most departures from the tree structure can be accounted for by non-structural explanations, such as anaphora and attribution, Lee et al. state that shared arguments may have to be accepted in discourse structure. Aktaş et al. (2010) identified similar structures in TDB, adding *nested relations* that do not violate tree structure constraints, as well as *properly contained relations* that introduce further deviations from trees. Following their terminology, we will reserve the word *relation* to discourse relations, or coherence relations, and use the term *configuration* to refer to relations between discourse relations.

3.3.1.1 Independent Relations

The first release of TDB consists of 8,483 explicit relations. The argument spans of some discourse connectives do not overlap with those of any other connectives in the corpus. We call

them *independent relations*. All others are called *non-independent relations*. (19) includes two relations that are not part of a configuration anchored by explicit discourse connectives. The possibility of configurations with unannotated simplex subordinators, implicit relations and alternative lexicalizations will be discussed in ch. 4.

(19) 00001131- 7 & 8

- (a) *Sen de haberdar değildin **ve** ben hayatımda ilk kez yıkmaya değil aşmaya çalışıyordum.* İzin vermiyor, engeller koyuyordun. Dikenli tellerle çeviriyordun bu duvarı. Yaralanıyordum tırmanırken, kanyordum. Kırılıyordum, acıyordum, ama bırakmıyordum.

*“You weren’t aware of it either **and** for the first time in my life I was trying not to take down something but to go over it.* You weren’t allowing me and you were creating obstacles. You were surrounding this wall with barbed wires. I was getting hurt while climbing, I was bleeding. I was falling to pieces, hurting but I wasn’t giving up.”

- (b) *Sen de haberdar değildin ve ben hayatımda ilk kez yıkmaya değil aşmaya çalışıyordum.* İzin vermiyor, engeller koyuyordun. Dikenli tellerle çeviriyordun bu duvarı. *Yaralanıyordum tırmanırken, kanyordum. Kırılıyordum, acıyordum, **ama** bırakmıyordum.*

“You weren’t aware of it either and for the first time in my life I was trying not to take down something but to go over it. You weren’t allowing me and you were creating obstacles. You were surrounding this wall with barbed wires. *I was getting hurt while climbing, I was bleeding. I was falling to pieces, hurting **but** I wasn’t giving up.”*

Figure 3.9 represents the independent relations configuration.



Figure 3.9: Independent relations configuration

We have identified 2,548 non-independent configurations consisting of 3,474 unique relations, meaning that 5,010 relations (59.05%) are independent in the TDB 1.0.

A total of 419 relation were annotated on the STC Demo. 151 unique relations take part in non-independent relations, meaning that 268 relations only take part in independent relations.

After the reannotation, the number of independent annotations in the TDB 1.0 increased to 5148 (60.69%) and in the STC Demo to 273 (65.15%) as seen in 3.4.

Table 3.4: Distribution of non-independent configurations in TDB

	Annotation		Reannotation	
	#	%	#	%
TDB 1.0	5010	59.05	5148	60.69
STC Demo	268	63.69	273	65.15

3.3.1.2 Fully Embedded Relations

Fully embedded relations conform to tree structure. In (20), the relation in (b), anchored by *önce* ‘before’, is fully embedded in the relation in (a), anchored by *ve* ‘and’.

(20) 00001131- 32 & 33

- (a) *Gün ağarana dek uğraşiyor **ve** kadın terasa çıkmadan **önce** kaçıyordu.*
 “He would try until the morning dawned **and** he would ran away **before** the woman went out to the terrace.”
- (b) *Gün ağarana dek uğraşiyor ve kadın terasa çıkmadan **önce** kaçıyordu.*
 “He would try until the morning dawned and he would ran away **before** the woman went out to the terrace.”

Figure 3.10 represents the fully embedded relations configuration.

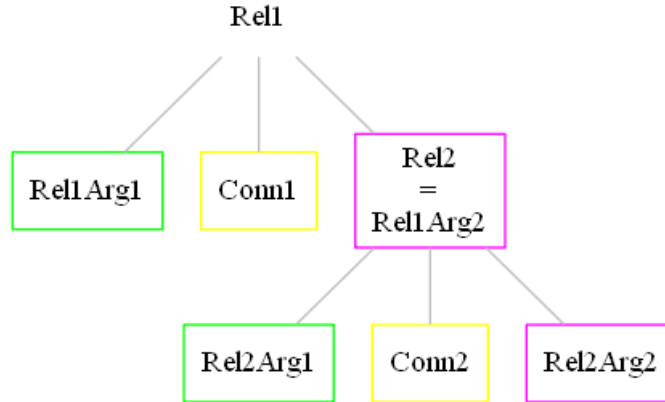


Figure 3.10: Full embedding configuration

Table 3.5 shows the distribution of fully embedded relations in the TDB 1.0 and the STC Demo before and after reannotation.

Table 3.5: Distribution of fully embedded relations

	Annotation		Reannotation	
	#	%	#	%
TDB 1.0	743	29.17	1631	64.04
STC Demo	23	17.16	106	64.63

3.3.1.3 Nested Relations

Nested relations also conform to tree structure. The relation in (a) is nested within the relation in (b). Neither relation contains any part of the other relation, yet they are not independent either. All arguments of the relation in (a) are located between arguments of the relation in (b) without any connections or crossing dependencies.

(21) 00002213- 23 & 24

- (a) Bir süre kapısında bir köpek gibi süründüm. Benden sonra âşık olduğu adamı gece gündüz izledim. İçim kıskançlık, acı, kin ve nefretle doluydu. Anlatması güç duygular bunlar. Adam onu dövüyordu. *Bazı geceler kulağımı kapısına dayar, dayak yerken attığı çığlıkları dinlerdim.* **Sonra barışıldardı.** Ne tuhaf bir şeydi bu! Sonra da bu parka düştüm işte.
- (b) *Bir süre kapısında bir köpek gibi süründüm. Benden sonra âşık olduğu adamı gece gündüz izledim.* İçim kıskançlık, acı, kin ve nefretle doluydu. Anlatması güç duygular bunlar. Adam onu dövüyordu. Bazı geceler kulağımı kapısına dayar, dayak yerken attığı çığlıkları dinlerdim. **Sonra da bu parka düştüm işte.**

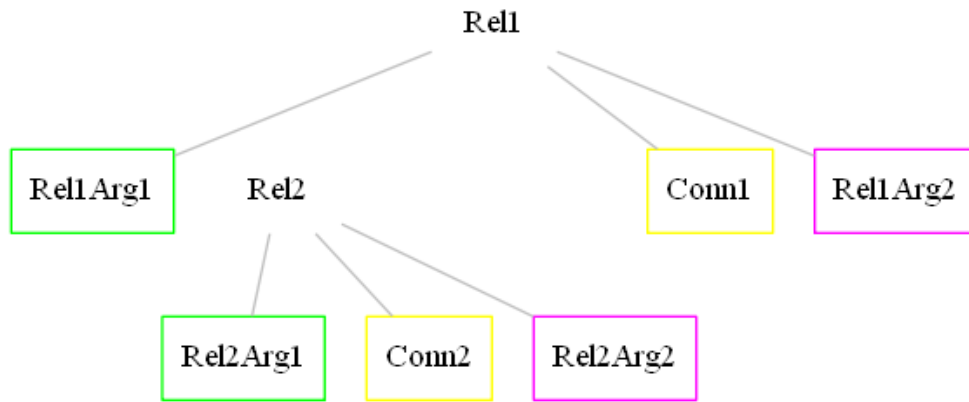


Figure 3.11: Nested relations configuration

Table 3.6 shows the distribution of nested relations in the TDB 1.0 and the STC Demo before and after reannotation.

Table 3.6: Distribution of nested relations

	Annotation		Reannotation	
	#	%	#	%
TDB 1.0	138	5.42	140	5.5
STC Demo	30	22.39	32	19.51

3.3.2 Tree Structure Violations

3.3.2.1 Shared Arguments

Lee et al. (2006, 2008) state that shared argument is one of the configurations that cannot be explained away, and should be accommodated by discourse structure. Similarly, Egg & Redeker (2008) admit that even in a corpus annotated within RST Framework, which enforces tree structure by annotation guidelines, there is a genre-specific structure that is similar to the shared arguments in Lee et al. (2006).

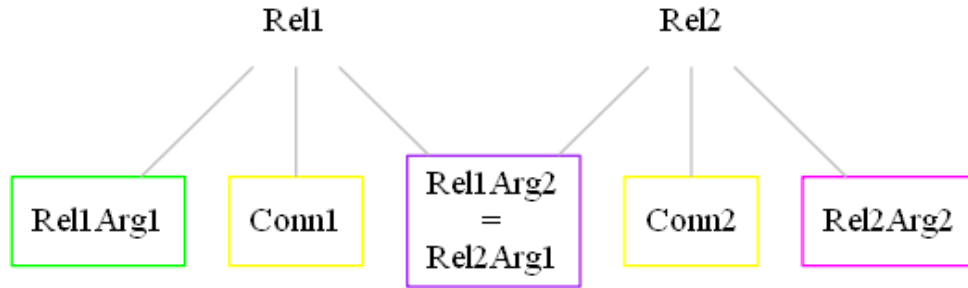


Figure 3.12: Shared argument configuration

(22) 00001131- 2 & 3

- (a) *Vazgeçmek kolaydı, ertelemek de. **Ama** tırmanmaya başlandı mı bitirilmeli! Çünkü her seferinde acımasız bir geriye dönüş vardı.*
*It was easy to give up, so was to postpone. **But once you start climbing you have to go all the way!** Because there was a cruel comeback everytime.*
- (b) *Vazgeçmek kolaydı, ertelemek de. Ama tırmanmaya başlandı mı bitirilmeli! **Çünkü** her seferinde acımasız bir geriye dönüş vardı.*
*It was easy to give up, so was to postpone. But *once you start climbing you have to go all the way!* **Because there was a cruel comeback everytime.***

In (22), the first argument of *ama* ‘but’ annotated in (a) completely overlaps with the first argument of *çünkü* ‘because’, annotated in (b) on the same text for comparison. The result is a shared argument configuration.

Table 3.7 shows the distribution of shared argument configurations in the TDB 1.0 and the STC Demo before and after reannotation.

Table 3.7: Distribution of shared arguments

	Annotation		Reannotation	
	#	%	#	%
TDB 1.0	488	19.16	79	3.1
STC Demo	35	26.12	7	4.27

Table 3.8 lists the reasons for the shared argument configurations identified during reannotation, and table 3.9 shows how the shared argument configurations were reannotated.

Table 3.8: Reasons for shared argument configurations

	TDB 1.0	STC Demo
Missing annotation	44	16
Multiple connectives	117	2
Leftout material	3	1
Annotation error	9	-
MP Reinterpretation	251	15
Syntactic asymmetry	-	-
Semantic tree violation	61	4

Table 3.9: Reannotation results for shared argument configurations

	TDB 1.0	STC Demo
Independent relations	5	-
Identical relations	128	2
Full embedding	290	32
Nested relations	1	-
Shared argument	61	4
Properly contained relation	2	-
Properly contained argument	1	-
Partial overlap	-	-
Pure crossing	-	-

3.3.2.2 Properly Contained Relations

Properly contained relations where anaphoric connectives are not involved can be caused by attribution, complement clauses, and relative clauses. (23) is a relation within a relative clause (a), which is part of another relation in the matrix clause (b). The result is a properly contained relation.

(23) 00001131-27&28

- (a) Sabah çok erken saatte **bir önceki akşam gün batmadan hemen önce** astığı çamaşırları toplamaya çıkıyordu ve doğal olarak da gün batmadan o günkü çamaşırları asmak için geliyordu.

She used to go out to gather the clean laundry *she had hung to dry* **right before the sun went down** the previous evening, and naturally she came before sunset to hang the laundry of the day.

- (b) Sabah çok erken saatte *bir önceki akşam gün batmadan hemen önce* astığı çamaşırları toplamaya çıkıyordu **ve doğal olarak da gün batmadan o günkü çamaşırları asmak için geliyordu.**

She used to go out to gather the clean laundry *she had hung to dry the previous evening right before the sun went down,* **and naturally she came before sunset to hang the laundry of the day.**

Sometimes a verb of attribution is the only element that causes proper containment. Lee et al. (2006) argue that since the relation between the verb of attribution and the owner of the

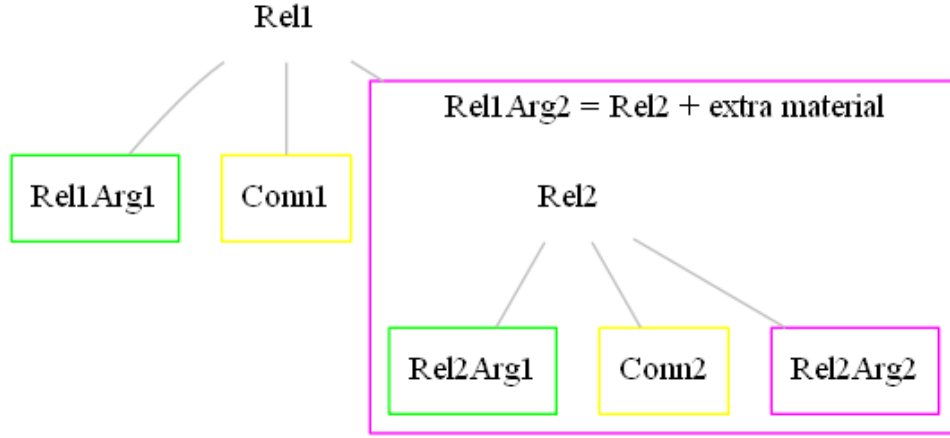


Figure 3.13: Properly contained relation configuration

attribution is between an abstract object and an entity, and not between two abstract objects, it is not a relation on the discourse level. Therefore, those stranded verbs of attribution should not be regarded as tree-structure violations. In (24) the properly contained relations occur in a quote, but the intervening materials are more than just verbs of attribution. Because the intervening materials in (24) are whole sentences that participate in complex discourse structures, we believe that (24) is different than the case proposed by Lee et al. (2006) and should be considered a genuine case of properly contained relation.

(24) 00003121-10, 11&13

- (a) Evet, küçük amcamdı o, nur içinde yatsın, yetmişlik bir rakıyı devirip ipi sek sek geçmeye kalkmış; kaptan olan amcam ise kocaman bir gemiyi sulara gömdü. Aylardan kasımdı, ben çocuktum, çok iyi anımsıyorum, fırtınalı bir gecede, Karadeniz'in batısında batmışlardı. *Kaptandı, **ama yüzmeye bilmezdi amcam.*** Bir namaz tahtasına sarılmış olarak kıyıya vurduğunda kollarını zor açmışlar, yarı yarıya donmuş. *Belki de o anda Tanrı'ya yakarıp yardım istiyordu, **çünkü çok dindar bir adamdı.*** Ama artık değil; küp gibi içip meyhanelerde keman çalıyor. Sonra da Nesli'nin ilgiyle çatılmış alnına bakıp gülüyor: Çok istavritsin!

Yes, he was my younger uncle, may he rest in peace, he tried to hop on the tightrope after quaffing down a bottle of raki; my other uncle who was a captain, on the other hand, sank a whole ship. It was October, I was a child, I remember it vividly, in a stormy night, they sank by the west of the Black Sea. *He was a captain, **but he couldn't swim, my uncle.*** When he washed ashore holding onto a piece of driftwood, they pried open his arms with great difficulty, he was half frozen. *Maybe at that moment he was begging God for help, **because he was a very religious man.*** But not anymore, now he hits the bottle and plays the violin in taverns. Then he sees Nesli's interested frown and laughs: You're so gullible!

- (b) *Evet, [...] Ama artık değil; küp gibi içip meyhanelerde keman çalıyor. **Sonra da Nesli'nin ilgiyle çatılmış alnına bakıp gülüyor: Çok istavritsin!***

*Yes, [...] But not anymore, now he hits the bottle and plays the violin in taverns. **Then he sees Nesli's interested frown and laughs: You're so gullible!***

Whereas attribution can be discarded as a nondiscourse relation, a discourse model based on discourse connectives should be able to accommodate partially contained relations resulting from relations within complements of verbs and relative clauses.

Table 3.10 shows the distribution of properly contained relation configurations in the TDB 1.0 and the STC Demo before and after reannotation.

Table 3.10: Distribution of properly contained relations

	Annotation		Reannotation	
	#	%	#	%
TDB 1.0	975	38.28	532	20.89
STC Demo	32	19.51	14	8.54

Table 3.11 lists the reasons for the properly contained relation configurations identified during reannotation, and table 3.12 shows how the shared argument configurations were reannotated.

Table 3.11: Reasons for properly contained relation configurations

	TDB 1.0	STC Demo
Missing annotation	267	9
Multiple connectives	1	-
Leftout material	25	1
Annotation error	2	-
MP Reinterpretation	158	8
Syntactic asymmetry	522	14
Semantic tree violation	-	-

Table 3.12: Reannotation results for properly contained relation configurations

	TDB 1.0	STC Demo
Independent relations	4	-
Identical relations	1	-
Full embedding	446	18
Nested relations	1	-
Shared argument	3	-
Properly contained relation	519	14
Properly contained argument	1	-
Partial overlap	-	-
Pure crossing	-	-

3.3.2.3 Properly Contained Arguments

As in properly contained relations, properly contained arguments may arise when an abstract object that is external to a quote is in a relation with an abstract object in a quote. Likewise, a discourse relation within the complement of a verb or a relative clause can cause properly contained arguments.

- (a) **Bakan Türker**, IMF ile görüşmelerde bazı konuları açık bir şekilde masaya getirmelerinin IMF tarafından olumlu karşılandığını söyledi **ve şöyle devam etti**: "Örneğin bu ay sonuna kadar işten çıkarılması gereken işçileri çıkartmayacağımızı söyledim. Emeklilik sistemi içinde hazırana kadar daha fazla adam çıkacağını, eğer devlet adam çıkarırsa çift tazminat ödeyeceğimizi ve iç talepte lüzumsuz bir daralmaya ve işsizliğe neden olacağımızı anlattığımız zaman çok olumlu karşıladılar."

"Minister Türker said that the IMF reacted positively to the fact that they talked over some issues explicitly during the conference with the IMF and added that: "For example, I have told that we are not going to dismiss the employees who are to be dismissed till the end of this month. They have reacted very positively when we have told them more people will quit until June in pension regime, and if the government fires people, we will pay double indemnity and we will give cause for an unnecessary shrinkage in domestic demand and unemployment."

- (b) Bakan Türker, IMF ile görüşmelerde bazı konuları açık bir şekilde masaya getirmelerinin IMF tarafından olumlu karşılandığını söyledi ve şöyle devam etti: "**Örneğin** bu ay sonuna kadar işten çıkarılması gereken işçileri çıkartmayacağımızı söyledim. Emeklilik sistemi içinde hazırana kadar daha fazla adam çıkacağını, eğer devlet adam çıkarırsa çift tazminat ödeyeceğimizi ve iç talepte lüzumsuz bir daralmaya ve işsizliğe neden olacağımızı anlattığımız zaman çok olumlu karşıladılar."

"Minister Türker said that the IMF reacted positively to the fact that they talked over some issues explicitly during the conference with the IMF and added that: "For example, I have told that we are not going to dismiss the employees who are to be dismissed till the end of this month. They have reacted very positively when we have told them more people will quit until June in pension regime, and if the government fires people, we will pay double indemnity and we will give cause for an unnecessary shrinkage in domestic demand and unemployment."

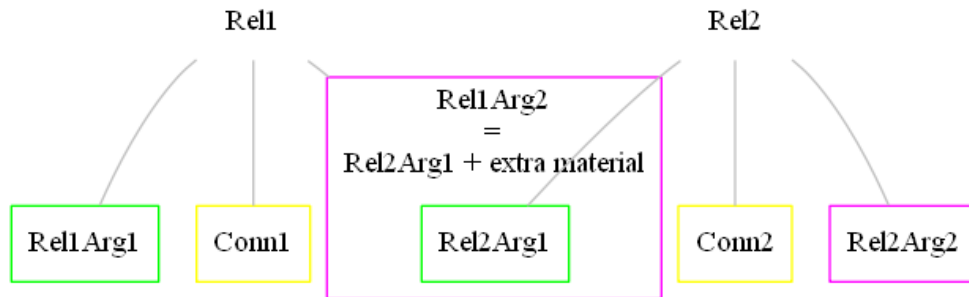


Figure 3.14: Properly contained argument configuration

Table 3.13 shows the distribution of properly contained argument configurations in the TDB 1.0 and the STC Demo before and after reannotation.

Table 3.14 lists the reasons for the properly contained argument configurations identified during reannotation, and table 3.15 shows how the shared argument configurations were reannotated.

Table 3.13: Distribution of properly contained arguments

	Annotation		Reannotation	
	#	%	#	%
TDB 1.0	189	7.42	7	0.27
STC Demo	30	18.29	0	0

Table 3.14: Reasons for properly contained argument configurations

	TDB 1.0	STC Demo
Missing annotation	19	9
Multiple connectives	8	3
Leftout material	4	-
Annotation error	1	-
MP Reinterpretation	141	18
Syntactic asymmetry	16	-
Semantic tree violation	-	-

Table 3.15: Reannotation results for properly contained argument configurations

	TDB 1.0	STC Demo
Independent relations	7	-
Identical relations	10	3
Full embedding	144	24
Nested relations	-	-
Shared argument	15	3
Properly contained relation	8	-
Properly contained argument	5	-
Partial overlap	-	-
Pure crossing	-	-

3.3.2.4 Partial Overlap

In (26), the argument span of *amacıyla* ‘in order to’ partially overlaps with the argument span of *için* ‘for’, resulting in a partial overlap of the arguments of two structural connectives. The first argument of relation (26) (a) properly contains the first argument of (26) (b), whereas the second argument of (b) properly contains the second argument of (a). This double containment results in a complicated structure that will be analyzed in detail in 3.3.2.5.

(26) 20630000-44&45

- (a) *Hükümetin, 1998’de kapatılan kumarhaneleri, kaynak sorununa çözüm bulmak amacıyla yeniden açmak için harekete geçmesi, tartışma yarattı.*

The fact that *the government took action for reopening the casinos that were closed down in 1998 in order to come up with a solution to the resource problem* caused arguments.

- (b) *Hükümetin, 1998’de kapatılan kumarhaneleri, kaynak sorununa çözüm bulmak amacıyla yeniden açmak için harekete geçmesi, tartışma yarattı.*

The fact that *the government took action* **for** **reopen the casinos that were closed down in 1998 in order to come up with a solution to the resource problem** caused arguments.

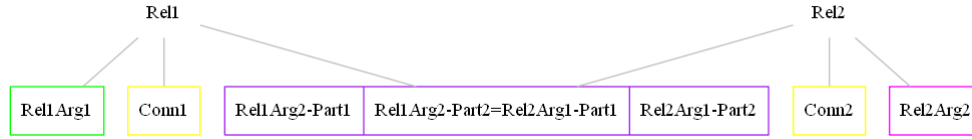


Figure 3.15: Partial overlap configuration

In (27) the second argument of but (relation (27) (a)) contains only one of the two conjoined clauses, whereas the first argument of after (relation (27) (b)) contains both of them. The most probable cause for this difference in annotations is the combination of "blind annotation" with the "minimality principle". This principle guides the participants to annotate the minimum text span required to interpret the relation. Since the annotators cannot see previous annotations, they have to assess the minimum span of an argument all over again when they annotate the second relation. Sometimes the minimal span for one relation is annotated differently than the minimal span required for the other, resulting in partial overlaps.

(27) 00001131-42&43

- (a) *Yine istediği kişiyi bir türlü görememişti, **ama** aylarca sabrettikten sonra gözeltlediği bir kadın soluğunu daralttı*, tüyleri diken diken oldu.

*Once again he couldn't see the person he wanted to see, **but** after waiting patiently for months, a woman he peeped at took his breath away*, gave him goose bumps.

- (b) *Yine istediği kişiyi bir türlü görememişti, ama aylarca sabrettikten sonra gözeltlediği bir kadın soluğunu daralttı, tüyleri diken diken oldu.*

*Once again he couldn't see the person he wanted to see, but **after** waiting patiently for months, a woman he peeped at took his breath away, gave him goose bumps.*

Table 3.16 shows that all partially overlapping argument configurations in the TDB 1.0 and the STC Demo were eliminated during reannotation.

Table 3.16: Distribution of partial overlaps

	Annotation		Reannotation	
	#	%	#	%
TDB 1.0	12	0.47	0	0
STC Demo	2	1.22	0	0

Table 3.17 lists the reasons for the partially overlapping argument configurations identified during reannotation, and table 3.18 shows how the partial overlaps were reannotated.

Table 3.17: Reasons for partial overlap configurations

	TDB 1.0	STC Demo
Missing annotation	-	1
Multiple connectives	-	-
Leftout material	-	-
Annotation error	-	-
MP Reinterpretation	9	1
Syntactic asymmetry	3	-
Semantic tree violation	-	-

Table 3.18: Reannotation results for partial overlap configurations

	TDB 1.0	STC Demo
Independent relations	2	-
Identical relations	-	-
Full embedding	7	2
Nested relations	-	-
Shared argument	-	-
Properly contained relation	3	-
Properly contained argument	-	-
Partial overlap	-	-
Pure crossing	-	-

3.3.2.5 Pure Crossing

There are only two pure crossing examples in the current release of TDB, a number so small that it is tempting to treat them as negligible. However, the inclusion of pure crossing would result in the most dramatic change in discourse structure, raising the complexity level to chain graph and making discourse structure markedly more complex than sentence level grammar. Therefore, we would like to discuss both examples in detail.

(28) 00010111-54&55

- (a) *Sonra ansızın sesler gelir. Ayak sesleri. Birilerinin ya işi vardır, aceleyle yürürler, ya koşarlar. **O zaman kız katılaştır ansızın.** Oğlan da katılaştır ve her koşunun gizli bir isteği var.*

And then *suddenly there is a sound. Footsteps. 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.*

- (b) *Sonra ansızın sesler gelir. Ayak sesleri. Birilerinin ya işi vardır, aceleyle yürürler, ya koşarlar. O zaman kız katılaştır ansızın. Oğlan da katılaştır **ve her koşunun gizli bir isteği var.***

And then *suddenly there is a sound. Footsteps. **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.***

In (28), the discourse relation encoded by then is not only anaphoric -and therefore not determinant in terms of discourse structure- but also the crossing annotation does not necessarily arise from the coherence relation of the connective's arguments. It is more likely imposed by lexical cohesive elements (Halliday & Hasan, 1976), as the annotators apparently made use of the repetitions of *ansızın* 'suddenly' and [*koş*] 'run' in the text when they could not interpret the intended meaning.

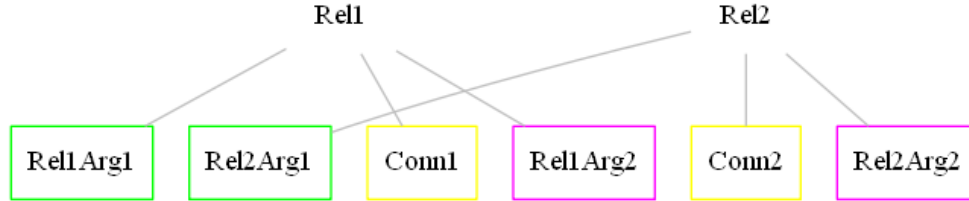


Figure 3.16: Pure crossing configuration

The other example, (29), is not anaphoric. It is more interesting as it points to a peculiar structure similar to (26) in 3.3.2.4, a surface crossing which is frequent in the subordinating conjunctions of Turkish.

(29) 20510000-31,32&34

- (a) *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.

- (b) *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.

- (c) *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.

A closer inspection reveals that the pure crossings in (29) are caused by two distinct reasons.

The first reason is the repetition of the subordinator *için* 'because'. Had there been only the rightmost subordinator, the relation would be a simple case of Full Embedding, where *ve*

‘and’ in (a) connects the two reasons for the penalty, while the rightmost subordinator connects the combined reasons to the matrix clause (see 3.17). However, since both subordinators were present, they were annotated separately. They share their first arguments, and take different spans as their second arguments, which are also connected by *ve* ‘and’, resulting in an apparent pure crossing.

Our alternative analysis is that *ve* ‘and’ actually takes the subordinators *için* ‘because’ in its scope, and it should be analyzed similar to an assumed single-subordinator case. This kind of annotation was not available in TDB because the annotation guidelines state that the discourse connectives at the peripheries of the arguments should be left out as in figure 3.18.

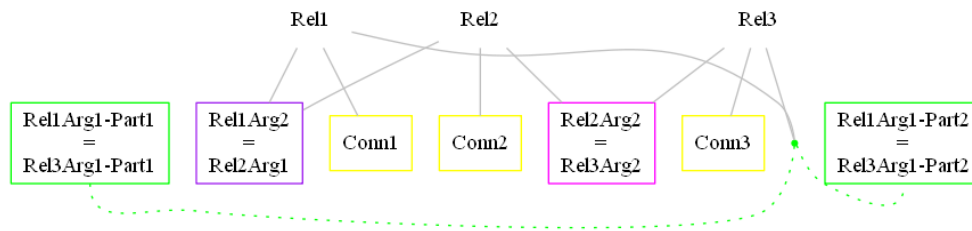


Figure 3.17: Double-subordinator analysis for (29) (as-is)

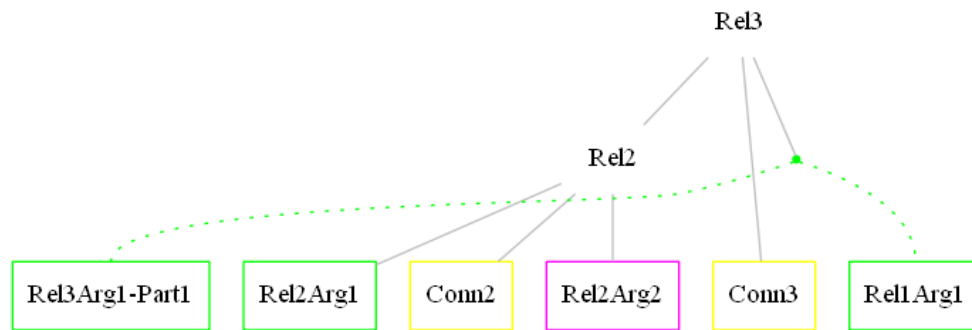


Figure 3.18: Single-subordinator analysis for (29) (hypothetical)

The second reason for crossing is the wrapping of the first arguments of (a) and (c) around the subordinate clause. This crossing is in fact not a configuration-level dependency, but a relation-level surface phenomenon confined within the relation anchored by *için* because, without underlying complex discourse semantics. Example (30) is a simpler case where the surface crossing within the relation can be observed.

(30) 10380000-3 1882’de İstanbul Ticaret Odası, **bir zahire ve ticaret borsası kurulması için girişimde bulunuyor** ama sonuç alamıyor.

*In 1882, İstanbul Chamber of Commerce makes an attempt **for founding a Provisions and Commodity Exchange Market** but cannot obtain a result.*

Subordinators in Turkish form adverbial clauses (Kornfilt, 2013), so they can occupy any position that is legitimate for a sentential adverb. Wrapping in discourse seems to be motivated information-structurally. In the unmarked position, the subordinate clause comes before the matrix clause and introduces a theme. However, the discourse constituents can occupy different positions or carry non-neutral prosodic features to express different information structures

Demirşahin (2008). In (29), wrapping takes *ceza* ‘penalty’ away from the rheme and makes it part of the theme, at the same time bringing the causal discourse relation into the rheme.

As is clear from the gloss in (29) and its stringset, this is function application, where *ceza verildi* ‘penalty was given’ wraps in the first argument as a whole. Double occurrence of the connective within the wrapped-in argument is causing the apparent crossing, but there is in fact one discourse relation.

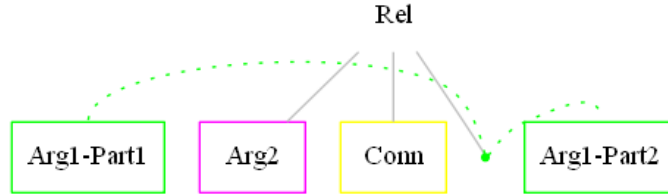


Figure 3.19: Wrapping

Wrapping in discourse is almost exclusive to subordinating conjunctions, possibly due to their adverbial freedom in sentence-level syntax. The subordinators make up 468 of the total of 479 wrapping cases identified in TDB. However, there are also four cases of coordinating conjunctions with wrapping. Two of them result in surface crossing as in (30), and the other two build a nested-like structure, as in (31) and (32). The latter two are both parentheticals.

(31) 10690000-32

Bezirci'nin sonradan elimize geçen ve 1985'lerde yaptığı antoloji hazırlığında [...]

In the preparation for an anthology which Bezirci made during 1985's and which came into our possession later [...]

In (31) *ve* ‘and’ links two relative clauses, one of which seems to be embedded in the other. It should be noted that the first part of Arg1 (*Bezirci-nin*) has an ambiguous suffix. The suffix could be the agreement marker of the relative clause, as reflected in the annotation, or it could be the genitive marked complement of the genitive-possessive construction *Bezirci'nin antoloji hazırlığı* ‘Bezirci’s anthology preparation’. The latter analysis does not cause wrapping.

(32) 00003121-26

Biz yasalar karşısında evli sayılacak, ama *gerçekte evli iki insan gibi değil de (evlilikler sıradanlaşıyordu çünkü, tekdüze ve sıkıcıydı; biz farklı olacaktık), aynı evi paylaşan iki öğrenci gibi yaşayacaktık.*

We would be married under the law, but *in reality we would live like two students sharing the same house rather than two married people (because marriages were getting ordinary, (they were) monotonous and boring; we would be different).*

(33) 00008113-10

Masa ya da duvar saatleri bulunmayan, ezan seslerini her zaman duyamayıp zamanı öğrenmek için *erkeklerin (evde oldukları zaman, tabii) cep saatiyle doğanın ışık saatine ve kendi içgüdüleriyle tahminlerine bel bağlayan* birçok aile, yaşamlarını bu top sesine göre ayarlarlardı.

Lots of families who didn't have a table clock or a wall clock and couldn't always hear the prayer calls, who *relied upon the men's pocket watch (when they were home, of course) and their instincts and guesses* to learn the time adjusted their lives according to this cannon shot.

Both (32) and (33) are parentheticals, resulting in a double-wrapping construction (figure 3.20). However, parentheticals move freely in the clause and occupy various positions, so we believe that this construction should be taken as a peculiarity of the parenthetical, rather than the structural connectives involved in the relation.

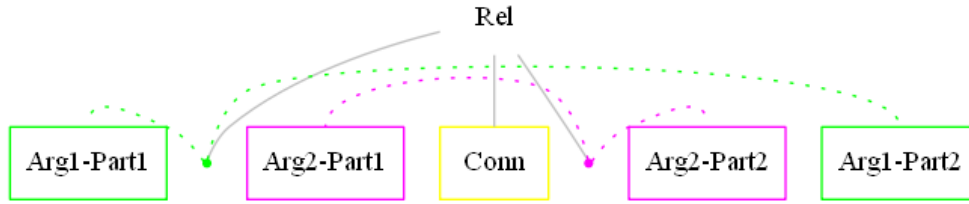


Figure 3.20: Double-wrap parenthetical construction for (31)

In STC Demo, only one pure crossing configuration was attested.

- (34) (a) HAL000098: Üşürüm ama ya. Hmm nice. İçine ne giyeceksin?
 ONU000099: Bilmiyorum işte!
 HAL000098: *John Travolta gibi olursun. Beyaz tişört giy.*
 ONU000099: Yani **mesela otuz sene önceki hali gibi di mi?**
 HAL000098: Tabii ki! Şimdiki hali değil. Sen filinta gibisin. Adam şimdi yaşlı ve şişman ... Ya da uzun kollu o siyah söylediğim şeyi giysene.
- (a) HAL000098: Üşürüm ama ya. Hmm nice. İçine ne giyeceksin?
 ONU000099: Bilmiyorum işte!
 HAL000098: *John Travolta gibi olursun. Beyaz tişört giy.*
 ONU000099: Yani mesela otuz sene önceki hali gibi di mi?
 HAL000098: Tabii ki! Şimdiki hali değil. Sen filinta gibisin. Adam şimdi yaşlı ve şişman ... **Ya da uzun kollu o siyah söylediğim şeyi giysene.**

In (a) the relation anchored by *mesela* ‘for example’, which is a discourse adverbial. Since it takes the first argument anaphorically, it does not increase the computational complexity of the configurations in the STC Demo.

In addition, *mesela* exist together with *yani* ‘i.e, in other words, namely, that is to say’, a connective that was not annotated in either TDB or the STC Demo. *Yani* introduces parantheticals (Ruhi, 2009). Just like in (32) and (33), we believe this crossing dependency may be caused by the paranthetical nature of the text span introduced by *yani*.

Table 3.19 shows that one of the pure crossing configurations in the TDB 1.0 was eliminated during reannotation. One pure crossing in the TDB 1.0 and the only one in the STC Demo remain as semantic tree violations. Note that both remaining pure crossing configurations include at least one anaphoric connective.

Table 3.19: Distribution of pure crossings

	Annotation		Reannotation	
	#	%	#	%
TDB 1.0	2	0.08	1	0.04
STC Demo	1	0.61	1	0.61

Table 3.20 lists the reasons for the pure crossing configurations identified during reannotation, and table 3.21 shows how the pure crossing configurations were reannotated.

Table 3.20: Reasons for pure crossing configurations

	TDB 1.0	STC Demo
Missing annotation	-	-
Multiple connectives	-	-
Leftout material	1	-
Annotation error	-	-
MP Reinterpretation	-	-
Syntactic asymmetry	-	-
Semantic tree violation	1	1

Table 3.21: Reannotation results for pure crossing configurations

	TDB 1.0	STC Demo
Independent relations	-	-
Identical relations	-	-
Full embedding	1	-
Nested relations	-	-
Shared argument	-	-
Properly contained relation	-	-
Properly contained argument	-	-
Partial overlap	-	-
Pure crossing	1	1

3.3.2.6 Distribution of Configurations

In addition to the *shared arguments* that were accepted in discourse structure by Lee et al., we have also identified *partially contained arguments* and *partially contained relations* in the Turkish data. These configurations arise not only from attribution as argued in the PDTB study, but also from verbal complements and relative clauses. These structures can be treated differently in other frameworks; for instance in RST, they are treated as discourse constituents taking part in coherence relations. However, for the connective-based approach adopted in this study, they need to be accommodated as deviations from tree structure. What is more interesting for our study is that these proper containments were always due to some sort of syntactic asymmetry. We are yet to find any proper containments due to a semantic tree violation.

The few partial overlaps we have encountered were all explained away by reinterpretation or

Table 3.22: Distribution of non-independent configurations

Configuration	TDB Before		TDB After		STC Before		STC After	
	#	%	#	%	#	%	#	%
Full Embedding	744	29.2	1632	64.51	29	17.68	106	64.63
Nested Relations	138	5.42	140	5.53	32	19.51	32	19.51
Identical relation	-	-	139	5.49	-	-	5	3.05
Total Non-violating	882	34.62	1910	75.49	61	37.2	143	87.2
Shared Argument	488	19.15	79	3.12	38	23.17	7	4.27
Properly Cont. Rel.	975	38.27	532	21.03	32	19.51	14	8.54
Properly Cont. Arg.	189	7.42	7	0.28	30	18.29	-	-
Partial overlap	12	0.47	-	-	2	1.22	-	-
Pure crossing	2	0.08	1	0.04	1	0.61	-	-
Total tree-violating	1666	65.38	759	30	103	62.8	21	12.8
Total	2548	100	2530	100	164	100	164	100

syntactic asymmetry, and were reannotated as other configurations.

Table 3.22 shows the distribution of all non-independent configurations in the TDB 1.0 and the STC Demo before and after reannotation.

The single pure crossing example we identified in the STC Demo includes an anaphoric connective. Of the two pure crossing examples we have found in TDB 1.0, one was anaphoric, whereas the other could be explained in terms of information structurally motivated relation-level surface crossing, i.e. wrapping. Recall that wrapping has applicative semantics. If we leave the processing of information structure to other processes, the need for more elaborate annotation disappears. In Joshi (2011)’s terminology, immediate discourse in the TDB 1.0 and the STC Demo appears to be an applicative structure, which, unlike syntax, seems to be in no need of currying.

As a result, we can state that structural pure crossing (i.e. crossing of the arguments of structural connectives) is not genuinely attested in the TDB 1.0 and the STC Demo. The annotation scheme need not be enriched to allow more complex algorithms to deal with unlimited use of crossing. There seems to be a reason in every contested case to go back to the annotation, and revise it in ways to keep the applicative semantics, without losing the connective’s meaning.

Overall, about half of the tree-violating configurations can be accounted for by anaphoric relations, i.e. they are not structural tree violations. Note that if one of the relations in a configuration is anaphoric, we treat the configuration as anaphoric.

Table 3.23 shows the distribution of anaphoric and structural tree violations in all non-independent configurations in the TDB 1.0 and the STC Demo after reannotation.

Table 3.23: Distribution of anaphoric relations among tree-violating configurations

Configuration	TDB 1.0			STC Demo		
	Anaphoric	Structural	Total	Anaphoric	Structural	Total
Prop. Cont. Arg.	6	1	7	-	-	-
%	85.71	14.29	100	-	-	-
Prop. Cont. Rel.	210	322	532	6	8	14
%	39.47	60.53	100	42.86	57.14	100
Pure Crossing	1	-	1	1	-	1
%	100	0	100	100	0	100
Shared Arg.	55	24	79	4	3	7
%	69.62	30.38	100	57.14	42.86	100
Total	272	347	619	11	11	22
%	43.94	56.06	100	50.00	50.00	100

3.4 A Comparison of Written Discourse vs. Spoken Discourse in Turkish

3.4.1 Comparison of the Descriptive Statistics of Discourse Connectives in Written vs Spoken Turkish

Because of the large difference in size between the two corpora, we converted the raw numbers to frequencies. We used number/1000 words as the frequency unit in 3.24.

The top five most frequent connectives in the TDB in descending order are *ve* ‘and’, *için* ‘for’, *ama* ‘but’, *sonra* ‘later’ and *ancak* ‘however’ and the top five most frequent connectives in the STC are *ama* ‘but’, *ve* ‘and’, *mesela* ‘for example’, *sonra* ‘later’ and *için* ‘for’. Here we compare the four most frequent connectives, namely, *ve*, *için*, *ama* and *sonra*, which make up 4951 (58.3%) of the total 8484 annotations in TDB and 217 (52.2%) of the total 416 relations annotated in the STC.

Conn	TDB						STC Demo					
	Discourse Conn			Total			Discourse Conn			Total		
	#	<i>f</i>	%	#	<i>f</i>	%	#	<i>f</i>	%	#	<i>f</i>	%
<i>ve</i> ‘and’	2112	5.31	28.2	7501	18.86	100	50	2.40	48.1	104	5.00	100
<i>için</i> ‘for’	1102	2.77	50.9	2165	5.44	100	32	1.54	61.5	52	2.50	100
<i>ama</i> ‘but’	1024	2.57	90.6	1130	2.84	100	96	4.61	80.7	119	5.72	100
<i>sonra</i> ‘later’	713	1.79	56.7	1257	3.16	100	39	1.87	72.2	54	2.60	100

Table 3.24: Written and spoken uses of *ve*, *için*, *ama*, and *sonra*

Although both the frequency of the total occurrences of the connectives and their discourse uses seem to be lower in the spoken corpus, chi square tests show that the differences are not statically significant ($p > 0.5$). The percentage of the use of tokens as discourse connectives across modalities is not significant either ($p > 0.5$). The preliminary results indicate that the distribution of these five connectives and their uses as discourse connective are similar in

written and spoken language.

The similarity is expected, as the MTC and the subcorpus that the TDB is built on are multi-genre corpora. Specifically, the TDB includes novels and stories, which in turn include dialogues. Also, there are interviews in news excerpts, which are basically transcriptions of spoken language. As a result, the TDB texts reflect some aspects of spoken language. In addition, 3 of the 23 files of the STC Demo are news broadcasts and interviews, which are probably scripted and/or prepared. Thus they may not necessarily reflect all aspects of spontaneous spoken language.

3.4.2 Comparison of the Discourse Relation Configurations in Written vs Spoken Turkish

Configuration	TDB		STC Demo	
	#	%	#	%
Full Embedding	695	27.28	23	17.16
Nested Relations	138	5.42	30	22.39
Total Non-Violating Configurations	833	32.69	53	39.55
Shared Argument	489	19.19	35	26.12
Properly Contained Argument	194	7.61	28	20.90
Properly Contained Relation	1018	39.95	17	12.69
Pure Crossing	2	0.08	1	0.75
Partial Overlap	12	0.47	0	0
Total Tree-Violating Configurations	1715	67.31	81	60.45
Total	2548	100.00	134	100.00

Table 3.25: Distribution of non-independent configurations in TDB

The distribution of the tree-violating and non-tree violating configurations are similar; however, the distribution of individual configurations (such as nested relations, properly contained relations, properly contained arguments, and partially overlapping arguments) change across modalities. The difference could be across genres rather than across modalities. Since the STC Demo is significantly smaller than TDB, more spoken data is needed to achieve more meaningful statistical data.

CHAPTER 4

EVALUATION AND THE IMPLICATIONS FOR DISCOURSE STRUCTURE

4.1 Structure by Explicit Discourse Connectives

We observed that the discourse structure that is expressed by explicit connectives in written and spoken Turkish includes tree-conforming configurations such as *independent relations*, *full embedding* and *nested relations*, as well as tree violating configurations such as *shared argument*, *properly contained argument*, and *properly contained relation*. *Partially overlapping arguments* were attested in the TDB 1.0 and the STC Demo, but they were few in numbers and could be completely eliminated by reannotation.

Only a handful of *pure crossing* configurations were attested in both TDB and STC Demo. All pure crossing examples were accounted for by surface crossing due to wrapping, anaphoric discourse relations, and parentheticals. We conclude that structural pure crossing was not attested in either TDB or STC Demo.

Neither PDTB, nor TDB and STC Demo approaches claim that all discourse relations are anchored by explicit discourse connectives. PDTB tries to capture the remaining discourse relations by annotating implicit connectives. There are four types of implicit connective tags: Implicit relations, Alternative Lexicalizations (AltLex), Entity Relations (EntRel), and No relation (NoRel). In PDTB all implicit connectives take adjacent arguments. The TDB 1.0, and by extension the STC Demo do not include implicit connectives.

Note that neither TDB 1.0 nor STC Demo annotations include annotation of simplex subordinators i.e. subordinators that are simple suffixes or suffix groups that are not immediately connected to postpositions, or implicit connectives. Although the annotation of these discourse relation anchors is expected to have an impact on the distribution of number of different types of configurations, we do not expect them to increase the computational complexity. Both simplex subordinators and implicit connectives are likely to take adjacent first arguments. In the few cases simplex subordinators may have elliptic arguments as in (11). We propose that elliptic arguments should be handled as anaphoric. An elliptical argument is anaphoric as in a demonstrative pronoun is anaphoric; therefore, structural discourse connectives can take elliptic arguments by substitution, rather than taking them by adjunction like discourse adverbials.

Pure crossing relations require distant arguments. As a result, further annotations should not change the computational complexity of the discourse structure as far as they are anchored by

discourse connectives.

In summary, our preliminary analysis shows that discourse structure may have to accommodate partial containment and wrap in addition to shared arguments. Both TDB and STC Demo have an applicative structure, and the discourse structures that are constructed by discourse connectives do not need chain-graph-level computational power.

4.1.1 An analysis of Tree-Structure Deviations

Tree-violations due to syntactic asymmetry occurs when a relation or the argument of a relation is in a complement clause, such as the complement of an attribution (35), (36) or a relative clause (37), or when an argument is the subject or the nominalized predicate of a clause. Since the relations or the arguments of the relations are in syntactically asymmetrical positions, they result in properly contained arguments or relations. All 15 (18.52%) of the remaining tree-violations in the STC Demo and 538 (31.37%) of the remaining tree violations in the TDB 1.0 result from a syntactic asymmetry between the arguments and/or relations.

(35) 10380000 15 & 16

- (a) *Osmanlı'da ilk matbaanın 1727'de açıldığı söylenir **fakat** nedense **15 yıl sonra kapandığı söylenmez...***

*“It is said that the first printing house in the Ottoman Empire was founded in 1727 **but** for some reason it is not mentioned that it was closed **15 years later.**”*

- (b) *Osmanlı'da ilk matbaanın 1727'de açıldığı söylenir fakat nedense 15 yıl sonra kapandığı söylenmez...*

*“It is said that the first printing house in the Ottoman Empire was founded in 1727 but for some reason it is not mentioned that **it was closed 15 years later.**”*

(36) 00008113 12 & 13

- (a) ***Eskenazi, Manisalı bir Yahudi, sonradan Amerika'ya gidip doktor oluyor ve öldüğü zaman mirasıyla doğum yerinde bir hastane kurulmasını, naaşının yakılmasını, küllerinin o hastaneye götürülmesini vasiyet ediyor.***

*“Eskenazi, a Hebrew from Manisa, later goes to the States, becomes a doctor **and** wishes that when he dies a hospital will be established where he was born, he will be cremated, his ashes will be brought to that hospital.”*

- (b) *Eskenazi, Manisalı bir Yahudi, sonradan Amerika'ya gidip doktor oluyor ve **öldüğü zaman** mirasıyla doğum yerinde bir hastane kurulmasını, naaşının yakılmasını, küllerinin o hastaneye götürülmesini vasiyet ediyor.*

*“Eskenazi, a Hebrew from Manisa, goes to the States later, becomes a doctor and wishes that **when he dies** a hospital will be established where he was born, he will be cremated, his ashes will be brought to that hospital.”*

(37) 00013112 5&6

- (a) Prof. Dr. Ufuk Esin ile Aşıklı Höyük Kazısı ve buluntuları üzerine söyleştik. Yine Sayın Esin'in bir makalesinden Neolitik Dönemi tanımlayan kısa bir alıntı yaptık. **Ayrıca antropolog Prof. Dr. Metin Özbek'in Aşıklı Höyük'te bulunan beyin ameliyatı geçirmiş bir kafatası üzerindeki incelemeleriyle ilgili bir makalesi ile Dr. Henk Woldring'in Aşıklı Höyük'te yerleşmenin o zamanki bitki örtüsünü belirlemek amacıyla yaptığı polen analizini konu alan makalesinden birer bölüme yer verdik.**

“We had a chat with Professor Doctor Ufuk Esin about Aşıklı Mound Dig and the findings. One again we quoted a brief definition of the Neolithic Period from one of Mr. Esin's articles which. **Besides, we covered one of anthropologist Professor Doctor Metin Özbek's articles about the research on a skull that underwent a brain operation which was found in Aşıklı Mound and one of Dr. Henk Woldring's articles about a polen analysis which was conducted in order to determine the flora of the settlement at Aşıklı Mound in those times.**”

- (b) Prof. Dr. Ufuk Esin ile Aşıklı Höyük Kazısı ve buluntuları üzerine söyleştik. Yine Sayın Esin'in bir makalesinden Neolitik Dönemi tanımlayan kısa bir alıntı yaptık. Ayrıca antropolog Prof. Dr. Metin Özbek'in Aşıklı Höyük'te bulunan beyin ameliyatı geçirmiş bir kafatası üzerindeki incelemeleriyle ilgili bir makalesi ile Dr. Henk Woldring'in *Aşıklı Höyük'te yerleşmenin o zamanki bitki örtüsünü belirlemek* **amacıyla** yaptığı polen analizini konu alan makalesinden birer bölüme yer verdik.

“We had a chat with Professor Doctor Ufuk Esin about Aşıklı Mound Dig and the findings. One again we quoted a brief definition of the Neolithic Period from one of Mr. Esin's articles which. Besides, we covered one of anthropologist Professor Doctor Metin Özbek's articles about the research on a skull that underwent a brain operation which was found in Aşıklı Mound and one of Dr. Henk Woldring's articles about a polen analysis *he conducted* **in order to determine the flora of the settlement at Aşıklı Mound in those times.**”

In (37), the relative clause contains a relation, and is incidentally contained within the argument of another relation. The relative clause modifies a non-abstract object in the span of another relation, and the semantics of neither relation is dependent on the other.

Another type of syntactic asymmetry, not between relations, but the between the arguments of the same relation can be observed in (38).

(38) 10520000 39

Bazı sürtüşmeler yaşadığı tiyatroyu sinema ve dizi filmlerle aldattığını söyleyen Özyağcılar, “tiyatro yârine çok sadık bir sevgili olmadıgı” itirafında bulunuyor ardından.

“Özyağcılar, *who says that he has cheated with cinema and TV series on theatre with which he had some quarrels,* **then makes the confession that “he wasn't able to be quite a faithful lover for his beloved theatre”.**”

The last 67 (3.91%) of the tree-violations in the TDB are genuine, discourse-level tree-violations that cannot be explained away by missing annotations, errors, guideline restrictions and minimality principle, nor can they be traced back to a syntactic asymmetry. One non-reinterpretable relation is the single pure crossing instance that was discussed in 3.1.2.8. All

other tree-violations are Shared Argument configurations. 46 of these configurations include at least one anaphoric connective, i.e., either a discourse adverbial or a phrasal expression. None of the remaining 20 Shared Arguments can be explained away by any of the criteria in our analysis. Although they are few in number and make up only 1.17% of all tree-violations and 0.79% of all inter-relational configurations, our final discourse model has to account for the Shared Argument configuration.

The simplest structure proposed for the discourse structure is a tree, which treats discourse structure simpler than sentence-level syntax. The most complex representation, chain graphs that allow for crossing dependencies and other tree-violations, treats discourse as more complex than sentence level. Sentence level syntax lay between context-free and context sensitive (Shieber, 1985; Joshi, 1985), more complex than trees but not as complex as general graphs.

Discourse relations are usually defined as either between two discourse units, or a listing type of relation between an unbound number of units, which are best described as recursive binary relations.

(39) 20360000 15

Daha çok 35 yaş altındaki internet kullanıcılarının yüzde 50.8’i bekâr, yüzde 40.1’i evli, diğerleri ise ya [birlikte yaşıyor], ya [boşanmış] ya da [dul]...

“Of the internet users who are mostly below 35 years old, 50.8 percent are single, 40.1 percent are married, the others on the other hand **either** [live together], **or** [(are) divorced], **or** [(are) widows].”

(40) 00002113 8

Simsiyah saçlı, orta boylu, siyah deri yelekli, boynunda kırmızı fular olan bir adam bir kızla delice dans ediyordu. [Kızı sırtüstü yatırıyor], [birden kendine doğru çekiyor], [bacağına bir çimdik atıyor],[yere bırakıveriyor], [derken havaya kaldırıyor], **sonra** [ona sımsıkı sarılıyordu].

“A middle sized man with jet-black hair, leather vest, and a red foulard on his neck was dancing with a girl like crazy. He was [laying her down], [pulling her suddenly], [pinching her leg], [letting her drop], [lifting her up], **then** [finally hugging her tightly].”

(39) and (40) illustrate listing discourse relations with syntactic and adverbial connectives, respectively. These relations can be represented in various ways.

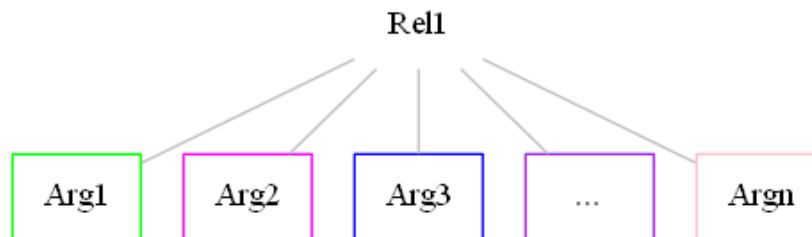


Figure 4.1: Flat tree representation for listing relations

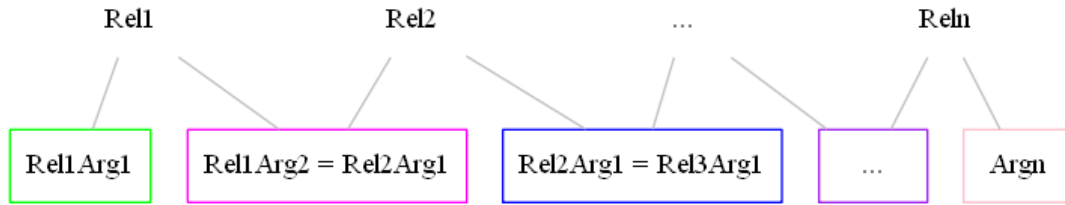


Figure 4.2: Shared argument representation for listing relations

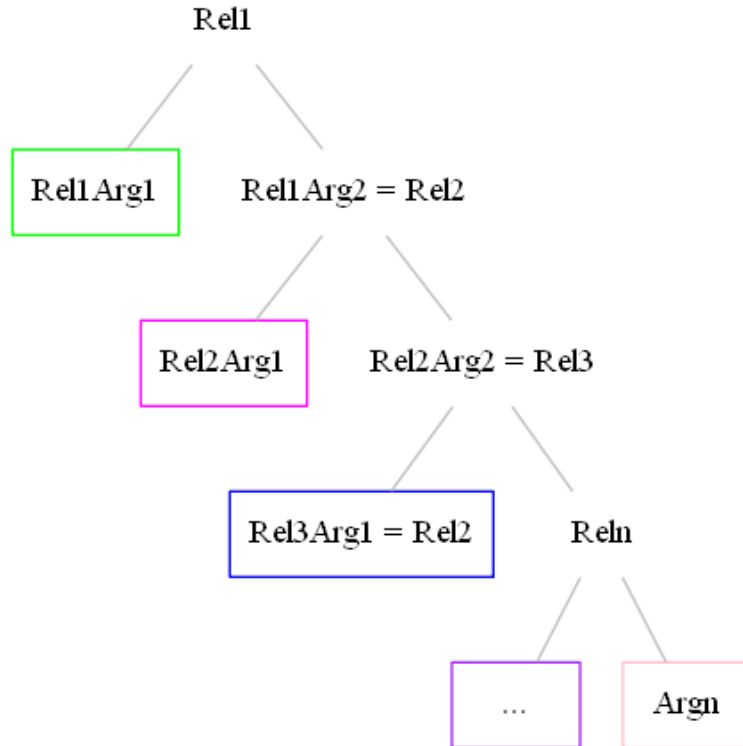


Figure 4.3: Full embedding representation for listing relations

The problem with the single predicate, flat tree representation in 4.1 is that since listing relations have an arbitrary number of items, it is not possible to pinpoint the arity of any connective that takes part in listing relations. It would also imply that the *ya* ‘or’ in a two-alternative-relation, three-alternative-relation and n-alternative-relation are all distinct lexical entries with different numbers of arguments. The representations in 4.2 and 4.3 have superior explanatory power as in they account for an arbitrary number of arguments with a single lexical entry for *ya*.

The resulting embedding structure in 4.3 implies there is asymmetry, a command or domination relation among the arguments, which is not true for discourse. Both SDRT and the derived trees of D-LTAG exhibit this structure. In order to avoid this interpretation, the semantic structure in D-LTAG is computed over the derivation trees, rather than the derived trees Forbes-Riley et al. (2006) 4.4. Shared argument reflects that all arguments are at equal level, but violates the tree structure constraints. Note that, however, applicative semantics are still adequate due to the fact that no function-composition is necessary to compute the semantics

of the resulting discourse structure.

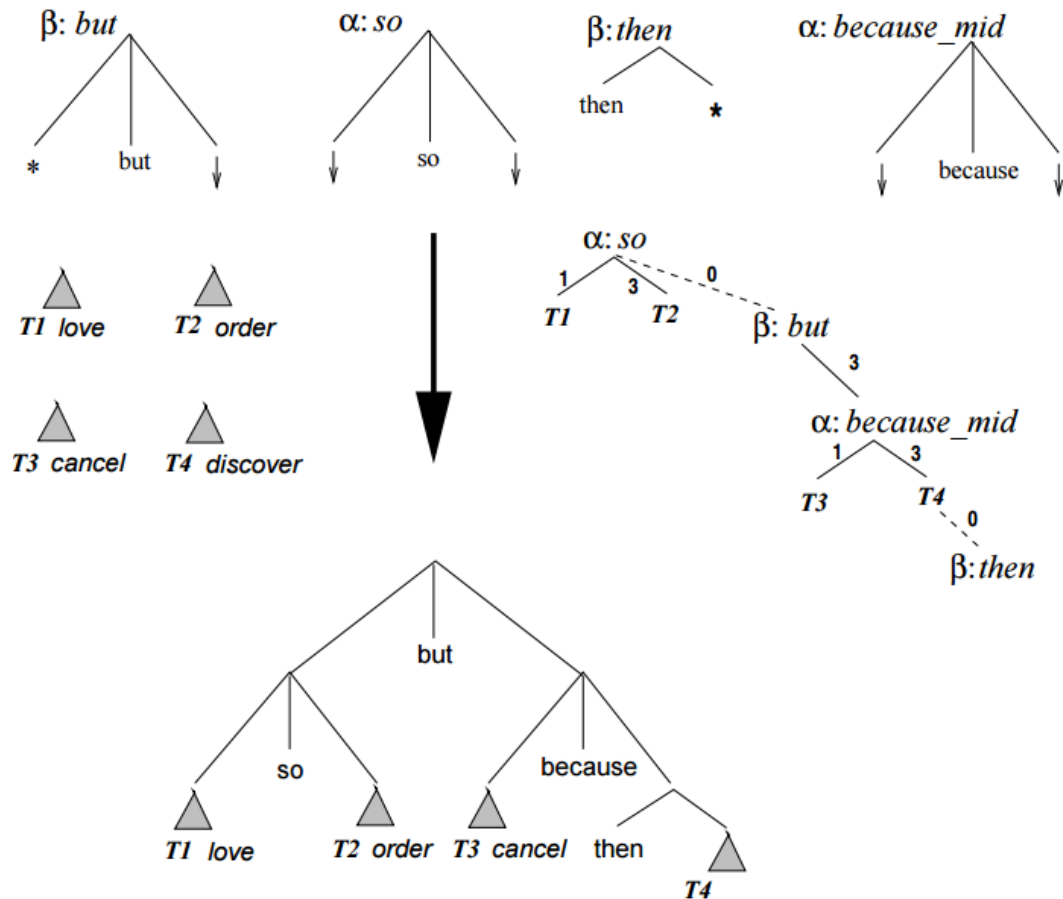


Figure 4.4: D-LTAG derivation and derived trees, B. Webber (2006) p. 352

If all we need is binary trees, the discourse-level relations can be accounted for by applicative structures, i.e. binary function application, without resorting to more complex operations such as function composition or graph reduction.

4.2 Discourse Structure beyond Explicit Discourse Connectives

In the PDTB/TDB scheme, there are four kinds of implicit connectives. The first type is the inserted **Implicit** connectives, the other tree are non-insertable implicit connectives, namely **AltLex**, **EntRel** and **NoRel**. All implicit relations in the PDTB scheme is between adjacent sentences. Since they are always adjacent and take whole sentences as arguments, they can not result in *pure crossing* configurations. In addition, presupposition is considered non-structural and the term presuppositional is used interchangeably with anaphorical as the complementary of structural (eg. in B. L. Webber (1988) and Zeyrek et al. (2008)).

4.2.1 Implicit Relation

Inserted **Implicit** connectives are annotated by representing the discourse relations between two adjacent sentences by inserting the corresponding explicit connectives inferred by the annotators. A similar example in Turkish would be the *Implicit = ve* ‘Implicit = and’ relation in (13).

The fact that some discourse relations can be inferred without an explicit head is somewhat problematic for a purely syntactic discourse representation model that tries to unify discourse structure with sentence structure, or treats discourse as merely the extension of sentence-level syntax. Sentence-level syntax is incremental and compositional, where each lexical item contributes to the sentence and the literal meaning of the complete sentence is completely dependent on its constituents.

Inference, on the other hand, is a semantic process which depends on a variety of sentence-external components including the textual context, the backgrounds of the speaker/author and the audience, as well as general world knowledge. Unlike entailment, another semantic process that is objective and necessary, inference is subjective: both its presence and the precise content may change depending on the context. As a result, the inserted **Implicit** connective represents a possible inference. It may not be necessarily intended by the author/speaker, nor inferred exactly the same by the rest of the audience. For example in (41), each reader may infer a different discourse relation. It is in fact possible to infer completely opposite inferences depending on the expectation of the reader from the author.

(41) Çok yorgundum. Dört saat uyumuşum.

“I had been very tired. (Apparently) I had slept for four hours.”

One of the possible interpretations for (41) is *Implicit = çünkü* ‘Implicit: because’. In this reading, the utterer is tired, because four hours is considerably less than the average nighttime sleep, which can be considered seven to eight hours for the purposes of this sentence. In this case, the second sentence is the reason for the first sentence.

Another reading would completely invert the direction of causality. If we assume that the utterer did not intend for a full night’s sleep because the event occurs during daytime, or if we were told before that the utterer intended for only a short nap, the inferred relation becomes one of *Implicit = dolayısıyla* ‘Implicit = so’. In this reading, the first sentence is the reason for the second sentence.

Still another available reading invokes a concession meaning. In this case, the utterer was very tired before going to bed, and despite being very tired slept only for four hours. With the discourse relation *Implicit = yine de* ‘Implicit = still’, the first sentence raises the expectation that the utterer should get at least an average night’s sleep if not more, and the second sentence counters this expectation by revealing that they slept about half of the expected duration.

In this constructed example we tried to make the sentences as unmarked as possible. One can still argue that the tenses and the aspects of the predicates favor one reading or the other. In addition, in a real life situation, the context or the prosody of the utterance can easily select one interpretation among the possible set of inferences. However, that is exactly the point we are presenting. An inferred relation does not compositionally contribute to the meaning

of the text, but is realized by the text. This case of inferred **Implicit** connectives seems to support Halliday & Hasan (1976)'s strictly non-structural case of cohesion in text, which is one of realization rather than constitution, although it does not exactly fit into the five ways cohesion is realized.

On the other hand, the relations realized by the text do give rise to some sort of structure. Binary relations between spans of text can be identified with reasonable accuracy.

The implicit relations are annotated by inserting an explicit connective that represents the inferred relation between two adjacent spans. When there are inferred relations between two spans that are already connected with an explicit discourse connective, no implicit connectives are inserted even when the explicit and implicit connectives express different senses. This approach means that there are unannotated senses, in other words discourse relations, between two spans that are already arguments of a connective. The implication is that there may be multiple discourse relations between two spans, and only some of them are expressed by explicit connectives.

In addition, intra-sentential, across-paragraph, and non adjacent implicit relations are not annotated. The reasons behind this decision are likely practical. Defining guidelines and creating consistent annotations for implicit relations are already a difficult task when they are restricted to adjacent clauses. Still, the lack of these annotations mean that not all discourse relations are covered by this annotation scheme.

4.2.2 AltLex Relation

AltLex label is used when there is an explicit expression in the text that expresses a discourse relation, and thus makes the insertion of an **Implicit** connective redundant.; but the expression does not fit the expectations from discourse connectives, i.e., it is not easily recognizable as the lexical head of a discourse relation. In PDTB, AltLex expressions include, but are not limited to, phrases like *because of that* and *despite this*. In TDB 1.0 and the STC Demo, the corresponding phrasal expressions built by a subordinating conjunction and an anaphoric expression are annotated as explicit discourse connectives similar to discourse adverbials.

The case for Turkish phrasal expressions as discourse adverbial-like connectives, subordinating discourse connectives with anaphoric expressions, or implicit **AltLex** relations was one of practical choice rather than a theoretical implication. Many Turkish discourse adverbials are anaphoric because they include a possessive morpheme, eg. *ıdolayısıyla* 'so', *aksine* 'on the contrary' etc. Annotating the phrasal expressions as adverbial-like connectives result in a unified treatment of the more lexicalised adverbials that have dropped the genitive counterpart of the possessive morphemes they carry and the phrasal expressions that include the genitive or bare anaphoric component.

TDB 1.0 and the STC Demo annotations do not include annotations for any other type of alternative lexicalisations, but PDTB uses AltLex to annotate other ways to express discourse relations such as causative *make* to express causality. In Turkish, AltLex tag would be useful for a variety of constructions that express discourse relations, for instance, the repetition of positive and negative aorist *-A/Hr ... -mAz* on the same root *gel* 'come' to express TEMPO-RAL:immediate succession relation 'as soon as' in (42).

(42) **Eve gelir gelmez peyniri yedim.**

“I ate the cheese **as soon as I came home.**”

The need for AltLex tag seems to be largely pragmatic, as in it is used for low frequency and highly productive under a single tag, instead of counting them all as different discourse connectives. However, their placement in the implicit category seems to be somewhat problematic, as these expressions are clearly explicit in the text. It is possibly the case that the PDTB group wished to reserve the *explicit connective* label for fixed expressions that would likely be the predicate of a discourse relations, following D-LTAG, and the highly productive nature of the AltLex expressions may make it counterproductive in such a system. However, in the interest of creating a theory neutral language resource, we propose either renaming the implicit/explicit convention, or moving the AltLex category to the explicit category.

4.2.3 EntRel and NoRel Relations

EntRel tag is used to annotate two adjacent spans that are not connected by a discourse relation, but they are about the same entity. This corresponds to the elaboration relation in DRT that was criticized by Knott et al. (2001) for not being a true discourse relation. In a way, the EntRel tags in PDTB represents the entity chains proposed by Knott et al.. Neither TDB 1.0 nor our annotations on the STC Demo include EntRel relations.

Finally, the NoRel tag is used for the sake of completeness. It is used to annotated adjacent spans that are not connected by any explicit or implicit discourse connective, and also are not about the same entity. As the name implies, this so called implicit connective shows that there are no relations between that particular set of adjacent sentences. The TDB 1.0 and the STC demo do not include NoRel annotations. Moreover, we believe that NoRel relations should be excluded form any study that investigates the structure in discourse, as they obviously do not denote any semantic relation.

4.3 Variations of a Discourse Relation

(43) demonstrates some of the ways a very simple causal relation between being hungry and eating the cheese can be expressed.

(43) (a) *Peyniri yedim **çünkü** açtım.*

“I ate the chesse **because I was hungry.**”

(b) *Peyniri yedim **zira** açtım.*

“I ate the chesse **because I was hungry.**”¹

(c) ***Aç olduğumdan** peyniri yedim.*

“**Because I was hungry,** I ate the chesse.”

¹ We provided a single translation for items that are so close semantically that we cannot provide distinct counterparts in English. For example, (a) *Peyniri yedim çünkü açtım.* and (b) *Peyniri yedim zira açtım.* are both translated as ‘I ate the cheese because I was hungry.’

- (d) **Aç olduğum için** peyniri yedim.
“**Because I was hungry**, I ate the chesse.”
- (e) **Aç olduğumdan dolayı** peyniri yedim.
“**Because I was hungry**, I ate the chesse.”
- (f) **Aç olduğumdan ötürü** peyniri yedim.
“**Because I was hungry**, I ate the chesse.”
- (g) **Aç olmam dolayısıyla** peyniri yedim.
“**Due to me being hungry**, I ate the chesse.”
- (h) **Aç olmam sebebiyle** peyniri yedim.
“**Due to me being hungry**, I ate the chesse.”
- (i) **Aç olmam nedeniyle** peyniri yedim.
“**Due to me being hungry**, I ate the chesse.”
- (j) **Aç olmam sayesinde** peyniri yedim.
“**(Unfortunately) due to me being hungry**, I ate the chesse.”
- (k) **Aç olmam yüzünden** peyniri yedim.
“**(Fortunately) due to me being hungry**, I ate the chesse.”
- (l) **Aç olmam sonucunda** peyniri yedim.
“**Resulting from me being hungry**, I ate the chesse.”
- (m) **Açtım, bu yüzden** peyniri yedim.
“I was hungry, **because of this I ate the chese.**”
- (n) **Açtım, bu sebeple** peyniri yedim.
“I was hungry, **because of this I ate the chese.**”
- (o) **Açtım, bu nedenle** peyniri yedim.
“I was hungry, **because of this I ate the chese.**”
- (p) **Açtım, bu sayede** peyniri yedim.
“I was hungry, **(fortunately) because of this I ate the chese.**”
- (q) **Açtım, bunun sonucunda** peyniri yedim.
“I was hungry, **as a result I ate the chese.**”
- (r) **Açtım, dolayısıyla** peyniri yedim.
“I was hungry, **as a result I ate the chese.**”
- (s) **Açtım, sonuç olarak** peyniri yedim.
“I was hungry, **as a result I ate the chese.**”
- (t) **Aç olmam peyniri yememle sonuçlandı.**
“My being hungry, **resulted in my eating the cheese.**”
- (u) **Aç olmam peyniri yememin sebebiydi.**
“My being hungry, was **the reason of my eating the cheese.**”
- (v) **Aç olmam peyniri yememin nedeniydi.**
“My being hungry, was **the reason of my eating the cheese.**”
- (w) **Peyniri yememin sebebi aç olmamdı.**
“**The reason that I ate the cheese was that I was hungry.**”

- (x) **Peyniri yememin nedeni aç olmamdı.**
 “**The reason that I ate the cheese** was *that I was hungry.*”
- (y) **Açtm. (Implicit = Bu yüzden) peyniri yedim.**
 “*I was hungry, (Implicit = because of this) I ate the chese.*”
- (z) **Peyniri yedim. (Implicit = Çünkü) açtm.**
 “*I ate the chesse (Implicit = because) I was hungry.*”

Admittedly, the variations in (43) are neither the same, nor can they be used interchangeably. In this section we will try to pinpoint what are the defining differences between these variations.

First of all, there are the obvious syntactic differences. The connectives in (a) and (b) are coordinating conjunctions, the *-dHğHndAn* ‘ablative factive’ in (c) is a simplex subordinator, the connectives in (d)-(l) are all complex subordinators, (m)-(q) include phrasal expressions, (r) and (s) include discourse adverbials and the relations in (t)-(x) are expressed through other types of alternative lexicalisations. Notice that the PDTB would not annotate (t)-(x) since they only annotate inter-sentential implicit connectives, but we included these examples here for the sake of completeness. In the PDTB, alternative lexicalisations are not annotated like the TDB 1.0 phrasal expressions. In PDTB the first sentence in the relation is annotated as the first argument and the second sentence is annotated as the second argument. The predefined Implicit = AltLex connective is inserted, and the alternative lexicalisation span is not explicitly marked. In this example, we annotated the span of the alternative lexicalisation as a phrasal expression, selecting the syntactically closer argument as its second argument, thus trying for a more unified approach for representing the spans that express discourse relations explicitly in the text. Finally in (y)-(z), there are no explicit connectives and the discourse relations are inferred, rather than expressed.

The syntactic differences are not limited to the syntactic type of the connective. With the syntactic type of the connective, the finiteness of the clauses change, too. In addition, the linear order of being hungry and eating switch depending on the syntactic construction, though the temporal order is preserved. These changes are in close relation with the information structure of the sentence. In English, subordinate clauses predominantly express theme, i.e., content that is already known and links the new information to be introduced to the previous discourse. Even when the subordinate clause introduce new content, it is presented as if old information (Quirk et al., 1985). Turkish subordinate clauses are not restricted in this manner. Demirşahin (2008) analyzed the information structure of the discourse connectives and their arguments in Turkish. Whereas discourse adverbials are the most permitting class in terms of word order in English, subordinate clauses are the most flexible both in terms of word order and information structure in Turkish. In 4.5, T stands for theme, T-K stands for theme kontrast, R stands for rheme and B stands for backgrounded information. CAO stands for connective argument order. 4.6 explains all possible connective argument orders for non-parallel connectives, i.e. connectives whose components are not distributed to each argument as in English *either...or* and *neither...nor* and their Turkish counterparts *ya...ya* ‘either...or’ and *ne...ne* ‘either...or’.

In (43), in their default positions, (a)-(b) and (w)-(x) are more likely to present *peyniri yemek* ‘eating the cheese’ as the known and *aç olmak* ‘being hungry’ as the new information. Note that with prosodic changes, one can either select *peyniri yemek* among possible alternative causes by employing a theme-kontrast tune, or present *peyniri yemek* as the new information by employing a rheme tune, and thus put *aç olmak* in a backgrounded position, post-rheme

		Arg1-Arg2			Arg2-Arg1			Arg1-Arg2-Arg1			
		CAO-1	CAO-2	CAO-3	CAO-4	CAO-5	CAO-6	CAO-7	CAO-8	CAO-9	
S T R U C T U R A L	Coordinating	<i>çünkü</i>	T, T-K	B	B				T, T-K	B	
	Conjunctions	<i>zira</i>	T	T-K,B	B				T-K	B	
	Subordinating Conjunctions	<i>p-için</i>		B	B		T-K	T-K, R		T, T-K	T, T-K,R
		<i>r-için</i>		B	B, R		T-K	T, R		T-K	T-K, R
<i>s-dolayı</i>				B, R			T-K, R			R	
		<i>s-dolayısıyla</i>			B		T-K, R			R	
E N D O P H O R I C	Anaphoric	<i>a-için</i>	T, T-K, R	T, R	B						
		<i>a-dolayı</i>	T-K, R	T-K, R	B						
		<i>a-dolayısıyla</i>	T-K, R	T-K, R	B						
C A T A P H O R I C	Cataphoric	<i>c-için</i>				R	R	R			
		<i>c-dolayı</i>				R	R	R			
			Arg-i	Arg-m	Arg-f	Arg-i	Arg-m	Arg-f	Arg-i	Arg-m	Arg-f

Figure 4.5: The information structure profiles of the connective-argument orders, sorted according to the syntactic type of the connective, from Demirşahin (2008) p. 87

positions are prosodically restricted to a flat background tune in Turkish Özge (2003); Özge & Bozsahin (2010). Items (c)-(v), on the other hand, are more likely to present *aç olmak* as the known information and *peyniri yemek* as the new information, together with the prosodic variations. However, prosody is not the only way subordinator clauses can take the rheme role. Because of the aforementioned prosodic restrictions, employing the rheme tune to a sentence-initial subordinate clause leaves no positions for a theme rune in the sentence. In order to present a subordinate clause as rheme, together with another theme in the sentence, the Turkish subordinators, and the subordinate clauses they occur in, can take on the rheme role by means of the wrapping process as demonstrated in 3.3.2.5. When both clauses introduce new information, the subordinate clauses can even fragment into independent incomplete sentences, providing space for two rhemes in two different information structures (Demirşahin, 2008). (44) demonstrates these variations for *için* ‘because, for’ in (43)(d).

- (44) (a) **Aç olduğum için** peyniri yedim.
“**Because I was hungry**, I ate the chesse.”
- (b) Peyniri **aç olduğum için** yedim.
“I ate the chesse **because I was hungry**.”
- (c) Peyniri yedim. **Aç olduğum için**.

Arg1-Arg2	CAO-1	Arg1	Conn-	Arg2	i		
	CAO-2	Arg1	Arg2	-Conn-	Arg2	m	
	CAO-3	Arg1	Arg2	-Conn	f		
Arg2-Arg1	CAO-4	Conn-	Arg2	Arg1	i		
	CAO-5	Arg2	-Conn-	Arg2	Arg1	m	
	CAO-6	Arg2	-Conn	Arg1	f		
Arg1-Arg2-Arg1	CAO-7	Arg1	Conn-	Arg2	Arg1	i	
	CAO-8	Arg1	Arg2	-Conn-	Arg2	Arg1	m
	CAO-9	Arg1	Arg2	-Conn	Arg1	f	
Arg2-Arg1-Arg2	CAO-10	Conn-	Arg2	Arg1	Arg2	i	
	CAO-11	Arg2	-Conn-	Arg1	Arg2	m	
	CAO-12	Arg2	Arg1	-Conn-	Arg2	m	
	CAO-13	Arg2	Arg1	Arg2	-Conn	f	

Figure 4.6: Possible connective argument orders for non-parallel connectives Demirşahin (2008) p. 40

*“I ate the chesse. **Because I was hungry.**”*

Whereas the variations in the information structure of the subordinate clauses arise from moving arguments in the sentence, other information structure varieties can be expressed by moving coordinating conjunctions, discourse adverbials, phrasal expressions and possibly other alternative lexicalisations within the second argument. These connectives can be focused in a preverbal slot or backgrounded by moving to the end of the argument, alone or together with other backgrounded constituents. In order to provide more slots for connectives, (45) provides examples enriched with adjuncts.

- (45) (a) *Eve gelir gelmez peyniri yedim, **çünkü** sabahtan beri açtım.*
*“As soon as I came home, I ate the cheese, **because I was hungry since morning.**”*
- (b) *Eve gelir gelmez peyniri yedim, sabahtan beri açtım **çünkü.***
*“As soon as I came home, I ate the cheese, **because I was hungry since morning.**”*
- (c) *Sabahtan beri açtım, **bu yüzden** eve gelir gelmez peyniri yedim.*
*“I was hungry since morning, **this is why** as soon as I came home, I ate the cheese.”*
- (d) *Sabahtan beri açtım, eve gelir gelmez peyniri yedim **bu yüzden.***
*“I was hungry since morning, **this is why** as soon as I came home, I ate the cheese.”*
- (e) *Sabahtan beri açtım, eve gelir gelmez **bu yüzden** peyniri yedim.*
*“I was hungry since morning, **this is why** as soon as I came home, I ate the cheese.”*

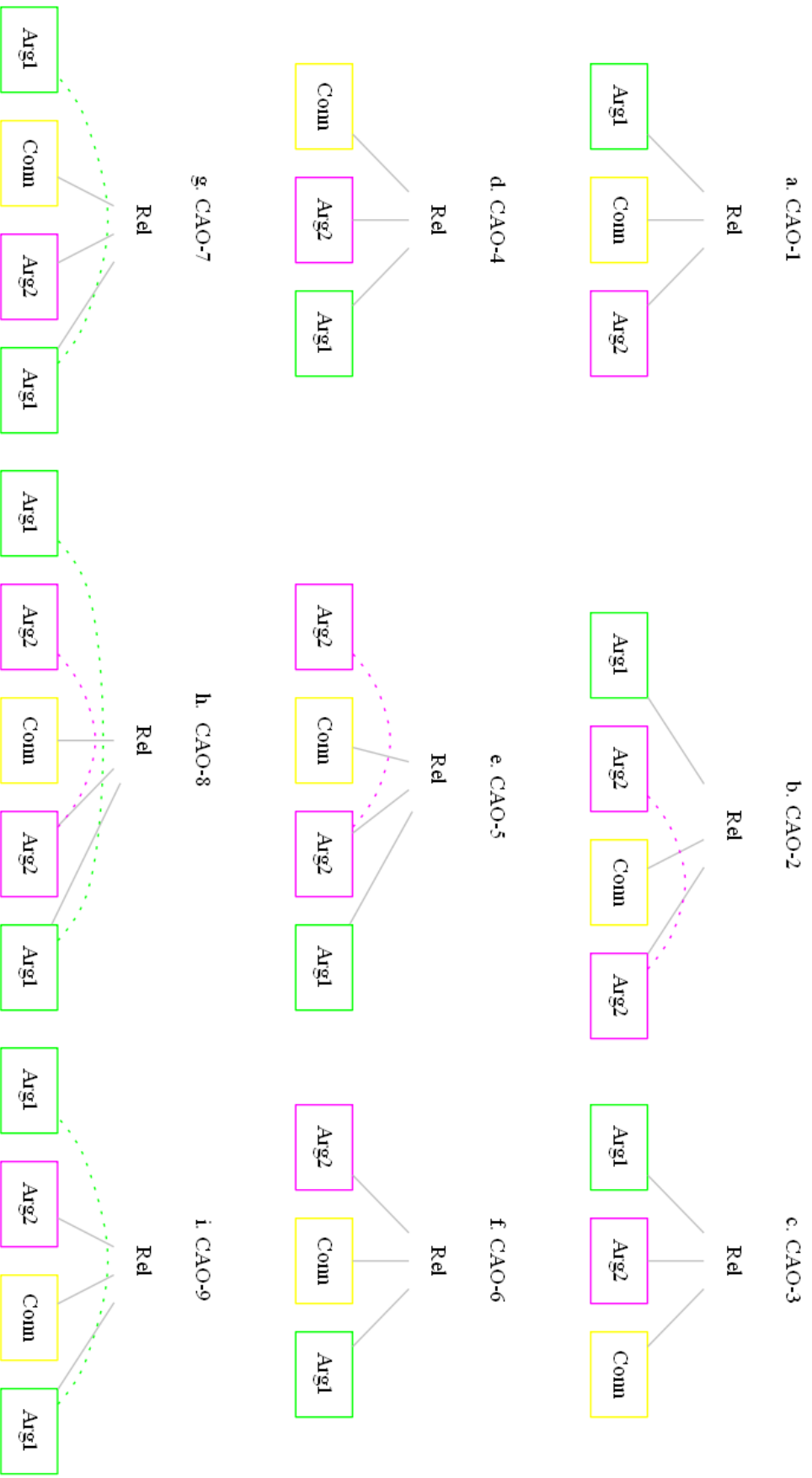


Figure 4.7: Syntactic trees for the connective-argument orders in 4.6

These information structure-motivated variations introduce further connective-argument order variations, resulting in more discourse-level syntactic variation. The discourse-level syntactic trees, constructed in a D-LTAG-like fashion, are presented in 4.7.

These variations are a direct result of the syntactic class of the discourse connectives and their arguments, as well as the information structure. However, neither syntactic type, nor the information structure seem to affect the semantic representation directly. A purely semantic representation of the variations in (44) seems to be the same. It is possible to represent all variations with a very simple and theory neutral proposition in (46) and 4.8.

(46) CAUSE(HUNGRY(speaker), EAT(speaker,cheese)).

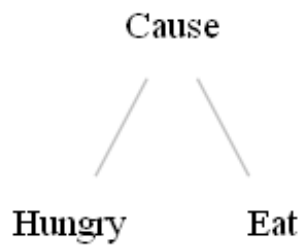


Figure 4.8: Simple tree representation for (46)

Semantically, the temporal relation between the hunger state and the eating event, as well as the direction of the causality is preserved. However, there are slight to moderate differences of meaning among these variations. One can argue that some variations in (43) are REASON relations whereas others are RESULT relations, both relations being a specification of CAUSALITY or CONTINGENCY. In (43)(a), (b),(w), (x), and (z) the effect, namely *eating the cheese* precedes the cause, namely *being hungry*. These variations may be analyzed as having the REASON relation, as opposed to the other items, where the cause precedes the result following the natural order of the eventualities, leading to the RESULT relation. One can argue that this distinction is a pragmatic one; by distinguishing the REASON and RESULT, we do not make a logical distinction between the underlying eventualities, but we mark the point of view of the utterer. In none of the variations can the act of eating be the cause for the state of hunger. However, it is possible for the statement of the act of eating to be cause for the statement of the state of hunger which at this point pragmatically becomes an explanation or justification in addition to semantically being the cause.

In addition to the linear order of the arguments or the statements ² variations (43)(j), (k) and (p) introduce another pragmatic distinction, namely the sentiment of the utterer concerning the turn of events. *Saye*, ‘shadow, protection’ in Persian, has a positive connotation in Turkish, which adds the meaning of *thanks to* or *with the help of* meaning to the cause. *Yüz* ‘face’, on the other hand, has a negative connotation as a subordinator, and introduces an accusatory meaning. Note that the phrasal expression constructed with *yüz* in (m) is largely neutral, and does not necessarily have a negative meaning.

² In the PDTB/TDB scheme, the order of the arguments and the statements do not correspond directly, as the order of the arguments are reversed between subordinating conjunctions and coordinating conjunctions in the default word-order of Turkish. The argument order of the discourse adverbials and the implicit connectives follow that of coordinating conjunctions.

4.4 Discourse Relations as Predicates

The logical representation in (46) CAUSE is a predicate. To this there is not much objection in the discourse literature, as CAUSE is taken to be a predicate in formal logic as well (eg. by McCarthy (1963)). However, other discourse relations, such as simple conjunction, simple disjunction, and implication, are traditionally logical connectives, operators rather than predicates. This distinction is evident in more semantically oriented approaches such as DRT and its followers like SDRT Asher (1993). The syntactically oriented D-LTAG takes all discourse connectives to be predicates (Forbes-Riley et al., 2006).

Although it is possible to rewrite all operators as predicates, the distinction between an operator and a predicate can be of theoretical interest. Syntactic predicates typically assign theta roles to their arguments, which largely correspond to their semantic thematic assignments; whereas the syntactic counterpart of the logical conjunction, the simple coordinator *and* does not. The coordinated items are interchangeable because of the lack of thematic assignment.

It is not a simple task to decide whether the discursive use of *ve* ‘and’ is just a logical operator or a discourse predicate, and it becomes mostly a matter of practical application in corpora annotation.

When *ve* coordinates finite clauses, usually it is not possible to use the coordinated clauses interchangeably. However, it is not easy to entangle the source of this prevention. If the discursive *ve* is predicative at the discourse level, the thematic assignment of the arguments may put a syntactic constraint on the arguments. On the other hand, the order of the eventualities, often marked by tense or constrained by states of affairs in the world, also prevents the arguments from interchanging freely. For example in (47) *ve* coordinates two finite clauses: the butterfly takes off and starts to fly. The arguments in this coordination are not interchangeable, but it is not clear if the constraint is imposed by the connective, the temporal order of the events as marked by tense, or the logical order of the take off and flight.

(47) Derken *kelebek havalandı ve sokağın öbür ucuna doğru uçmaya başladı.*

“Just then *the butterfly took off and started to fly towards the end of the street.*”

(48) Altı ay önce bitirdiği bir resmi *uzun süre dayanması ve renklerini koruması için verniklediği bir gece ansızın bir tekme savurarak üst kata çıktı.*

“During a night at which he was varnishing a picture he finished 6 months ago for it to last longer **and keep its colors**, he suddenly kicked it and went upstairs.”

(48) includes coordinated nonfinite clauses. More specifically, two nonfinite clauses are coordinated and the resulting coordinate structure is the argument of the subordinator *için* ‘for, in order to’. In this example, the coordinated items can switch places, but there is a subtle change in the meaning. In the original example, protection of colours is an elaboration of the durability of the painting, whereas in the switched condition the durability is the result of the protection of colors. One could argue that this change in meaning is an indication of thematic assignment. However, the nature of the change results in the opposite conclusion: it seems that the sense of the discourse relation does not arise from the discourse connective itself. Switching the arguments does not reverse the direction of the previous discourse relation, but results in a completely different meaning resulting from the contents and the ordering of the arguments themselves.

The argument structure of a syntactic predicate specifies the arity of the predicate, the syntactic properties of its arguments, and the semantic relation of the arguments to the predicate.

The arity of a discourse connective in most accounts, eg. in LDM and D-LTAG, is by definition two. Although the discourse adverbials take only one argument structurally in D-LTAG, they are still considered binary predicates.

There are some syntactic restrictions on the arguments of subordinating conjunctions. These conjunctions take arguments of certain finiteness and assign a case to the subordinate clauses or anaphoric items they take as second arguments. However, these restrictions come from their sentence-level syntactic properties, or in Grimes's terms, their status as lexical predicates. If we consider all the variations in (43) different manifestations of the same relation, we see that CAUSE relation does not restrict its arguments syntactically. The linear order, finiteness, and case of the arguments all differ across the variations, even within the subordinator variations. There are no restrictions on the first arguments of subordinators, and there seem to be no restrictions whatsoever on the arguments of coordinating conjunctions and discourse adverbials.

The lack of thematic assignment by itself does not necessarily mean that the discourse relation is not predicative. It merely shows that if the discourse connective is a predicate, it is of a different kind than sentence-level predicates. Grimes (1975) defines three kinds of semantic units: roles, lexical predicates, and rhetorical predicates. Roles, or cases, themselves are predicates that are selected and dominated by lexical predicates. Lexical predicates are what we traditionally think of as predicates, that assign roles. Finally, rhetorical predicates build rhetorical complexes by uniting the propositions built by the lexical predicates and roles; and larger complexes by recursively uniting rhetorical complexes. Thus Grimes differentiates the predicates that assign roles and predicates that express relations, but does not assign roles. Considering the fact that it is possible to represent operators as predicates, and that there are no corresponding operators for all discourse predicates, we consider representing discourse relations as predicates as preferable to representing some relations as predicates and some as operators, as it offers a unified approach. However, we restrict our use of discourse-level predicates to the non-case-assigning rhetorical predicates of Grimes.

CHAPTER 5

CONCLUSION

5.1 Summary and Conclusions

In this study we have presented our descriptive analysis of the discourse connectives and the structures they seem to anchor in the TDB 1.0 and STC Demo. Our extensive analysis of the relations in the corpora, along with comparison with the discussions of the discourse structure in various theories of discourse in English, has revealed some key properties of discourse relations, and has shed light onto the roles discourse connectives play with regards to discourse relations.

We observed that the discourse structure that is expressed by explicit connectives in written and spoken Turkish includes tree-conforming configurations such as *independent relations*, *full embedding* and *nested relations*, as well as tree violating configurations such as *shared argument*, *properly contained argument*, *properly contained relation*, *partially overlapping arguments*, and *pure crossing*.

We found out that *properly contained arguments* and *properly contained relations* are mostly due to the syntactic asymmetry between the arguments. We claim that these syntactic asymmetries do not apply at the semantic level. *Partially overlapping arguments* can be eliminated by reannotation. The few *pure crossing* configurations are accounted for by either surface crossing due to wrapping or by anaphoric discourse relations, and parentheticals.

The only tree violation at the semantic level that cannot be explained away by syntactic asymmetry and anaphora, and cannot be eliminated by reannotation are *shared arguments*. We argue that the final discourse model will include crossing, but should accommodate multiparenting. However, this is a limited sort of multiparenting, as the relations that share an argument are semantically independent, i.e., they are not composed over each other as for example control verbs and the verbs they control are composed over. Relations that share arguments are independently parsable and function application is sufficient for their processing.

Discourse relations (coherence relations, rhetorical relations) are a closed set. Although depending on the approach and the theory the number of these relations change, they are never treated as an open class. This means that when a new clause is introduced into the discourse, it can be related to the previous discourse only in a limited number of ways. We will call this the set of possible relations.

The discourse connectives that signal the discourse relations come from a variety of syntactic classes including subordinating and coordinating conjunctions and discourse adverbials

B. Webber & Joshi (1998); Zeyrek & Webber (2008). They can also be expressed by other means, as in AltLex in the PDTB, and phrasal expressions and other alternative lexicalizations in the TDB. Moreover, they can be complete absent from the text as in inserted implicit connectives in PDTB. In addition, they don't seem to impose any syntactic or semantic restrictions on their arguments.

Connective based approaches such as D-LTAG and DCCG treat discourse connectives as lexical predicates, whereas other theories mostly see them as clues that signal relations that exist independent of any lexical heads. We see DCCG as an improvement on CCG: it does not propose a new, independent discourse syntax, but fine tunes the lexical entries for discourse connectives in CCG. Instead of treating discourse adverbials the same as other sentential adverbs, for example, DCCG incorporates the anaphoric argument of the discourse adverbial to the derivation, giving a more complete account of the adverb at sentence level, too. It should be noted DCCG is, to the best of our knowledge, not concerned with implicit connectives. D-LTAG, on the other hand, emphasizes the similarities between the discourse syntax and sentence syntax, by proposing a sentence-like but independent syntax for discourse. LTAG and D-LTAG are not parts of the same syntax, but they are parallel syntaxes that share the same principles and work at different levels.

The strength of connective based approaches comes from the fact that discourse connectives make the discourse relations explicit. The audience can interpret the connection between a clause and the previous discourse in many different ways. It is likely to be cohesive with multiple previous clauses, or collection of clauses, which we call span as a blanket term. In addition, it can also be related to a single previous span in many different ways, although the ways it can be related is limited to the set of possible inferences. In the absence of a discourse connective, the audience selects at least one possible interpretation form the set, by means of other cohesive ties, world knowledge, as well as other discourse deictic aids such as definiteness (Von Heusinger, 2002) and tense (B. L. Webber, 1988). In the absence of explicit clues, the inferences may not be strong enough, and result in explicit questioning of the relation as we demonstrated in the example from the STC demo (11).

The presence of discourse connectives makes the intended relation explicit. Note that one relation can be expressed by a variety of connectives and non-connective expressions as in (43), and an instance of a connective can be interpreted as expressing multiple relations, as evidenced by multiple sense annotation in PDTB. In short, there is no one-to-one relation between a discourse connective and the sense it conveys.

Taking into consideration that (a) discourse connectives signal a closed set of relations, (b) they are optional when the inferences are strong enough, and (b) they do not have a one-to-one relationship with the relations they signal, our conclusion is that a discourse connective does not predicate the relation the way a verb builds the syntax of the clause. Instead, it explicitly selects among the predicative inferences that are present or possible between the new span and the previous discourse.

It should be noted that this is a theoretical discussion, which does not necessarily have practical implications for connective-based discourse banks such as PDTB and TDB. These resources provide valuable data that makes extensive qualitative research, including our investigations in chapters 3 and 4 possible, as well as providing enough real use data to profile discourse connectives for sentence-level syntax.

5.2 Limitations

This thesis is essentially built on a corpus-driven study and is mostly bound by the limitations of corpora in general, and PDTB/ TDB scheme and the data on the TDB 1.0 and the STC Demo in particular.

Corpora are resources of finite size, whereas the compositional nature of language results in infinite possibilities. As a result, there will always be the possibility of not being able to attest some linguistic patterns that is actually in the language. As the size and representativeness of the corpus increases, the probability of missing viable pattern will decrease. Nevertheless, as long as the study is conducted on a finite resource, it will never be a perfect representation of the infinite language. In our case, the 400,000-word TDB 1.0 is a sizable corpus, but we still had to construct examples (e.g (43)) in order to be able to convey some of our ideas. The STC is not released yet, and the 20,000-word STC Demo is limited in size. Because some rarer configurations occurs only once or twice, it is not statistically comparable to the TDB 1.0.

In addition to the possibility of the lack of total coverage, corpora may include data that is not in the language due to the performance and/or resource preparation errors, although in this study we did not encounter more than a handful of small errors thanks to the meticulous creation process of the TDB 1.0 which included several cycles of checks and proofs.

What has a larger impact on the study is that the TDB is an ongoing work. As mentioned several times before, the TDB 1.0 does not include implicit connectives. The annotation of AltLex relations were in progress as of writing this thesis. There are future plans for morphological analysis and disambiguation, which will make annotation of simplex subordinators and discourse particles possible.

Another limitation resulting from the corpora in question is a more fundamental one. The connective-based approach of the PDTB/TDB scheme limits the way the study can investigate the discourse structure. Specifically, the discourse connectives by definition require two and only two arguments. When there was the possibility of more than two arguments, we handled this possibility by choosing the shared argument or the fully embedded structures instead of a flat representation as discussed in 4.1.1. When there was the possibility of a single explicit argument, on the other hand, the annotation scheme and the tool did not allow them. These instances were left out as non-discursive uses of the token. However, we fear that we might have missed some discursive uses. The fact that there is no second argument present in the text does not necessarily mean that there is no second argument at all. If the the second argument is recovered from the world knowledge, or inferred from the previous discourse in general but cannot be pinpointed down to a specific span, discursive uses of connectives may have been dismissed as non-discursive. From personal experience we believe such cases are rare if present, but without further studies that allow extratextual arguments we cannot make a sound claim.

PDTB assumes a practical approach to language resource creation, which are more computationally oriented rather than cognitively oriented. For example, the inclusion of the NoRel relation makes sure that all sentences are connected, and results in a fully parsable discourse structure, although annotating relations that are not really there is neither necessary nor plausible from a cognitive standpoint.

To the best of our knowledge, there are no comparable corpora for Turkish annotated for

other discourse theories such as RST or DRT. As a result, we were not able to compare the structures resulting from different approaches to discourse representation.

Discourse as a field is underdefined. Approaches like D-LTAG and DCCG take a syntactic approach to discourse and put great weight in the linear order of the constituents that make up the discourse units, whereas the Coherence Theory, LDM, and DRT take a semantic approach. The Tripartite theory and the SDRT are hybrid approaches that take both the syntax and the semantics into account, although the former leans towards more syntactic approaches and latter to more semantic approaches. The RST and the PDTB take a functional approach, focusing on what the research program and NLP applications need and how the annotators can make faster and more accurate decisions. This various approaches to discourse is one was a limitation of this study because they are not directly comparable and the jargon of one approach does not transfer directly to the other. On the other hand, the availability of various approaches is in fact an advantage for the researcher, as once goal and the level of interest is set for the study, one can select the approach that works best for themselves.

Finally, the limitation with the greatest impact on this thesis is time and budget constraints. The STC Demo annotations and reannotations on both corpora are carried out by a single annotator and therefore do not have any inter-annotator or similar reliability metrics. In order to overcome this limitation, we include the full list of inter-relational configurations in both the TDB 1.0 and the STC Demo (see D). Interested researchers are welcome to replicate our analyses. Also due to time and budget considerations, reannotations only cover the attested tree-structure violations. Although we argue that the adjacent nature of simplex subordinators, discourse particles, and implicit connectives they cannot result in *pure crossing* configurations, it is possible that reannotations on the whole corpora may have caused more *shared arguments* and *properly contained arguments* and *relations*, and will have completely eliminated *independent* and *nested relations*.

5.3 Future Work

The most immediate work that should follow this study is to complete at least one more set of annotations for the STC Demo annotations and the reannotation work on both corpora. After the annotations are done, we would like to release the data together with the inter-annotator agreement statistics.

In order to reveal the true complexity of the discourse structure, we would like to remove the adjacency restriction from the implicit connectives. We expect this modification to reveal two distinct results. Firstly, non adjacent implicit connectives are the only relations that are not annotated on the TDB 1.0 and the STC Demo that may cause *pure crossing* configurations. Notice that explicit connectives do not have the adjacency requirement. We do not expect to see implicit connectives result in more complex structures than explicit connectives; however, we believe that the only way to have a sound claim on this matter is to remove the adjacency requirement for implicit connective annotations.

Secondly, the inter-annotator agreement statics of such annotation will provide a way to measure inference agreement. More specifically, the comparison of the inter-annotator agreements of explicit connectives and those of implicit connectives that do not have the adjacency requirement will reveal the true impact of having explicit discourse connectives on the perceived

structure of discourse.

As a complementary to this corpus-based study of inference agreement, multimodal psycholinguistic studies of inference and perceived discourse structure can be conducted by utilizing self paced reading and eye-tracking tasks.

Finally, we would like to explore the structure of discourse in a broader cognitive context. Steedman (2002) provides a framework for relating natural language grammar and planned action. He argues that both systems have applicative semantics, utilizing functional composition and type-raising. So far our investigations suggest that discourse has much simpler structure, as we observe that function application seems to be adequate for discourse processing. We have yet to need function composition at discourse level.

Bibliography

- Aktaş, B., Bozsahin, C., & Zeyrek, D. (2010). Discourse relation configurations in turkish and an annotation environment. In *Proceedings of the fourth linguistic annotation workshop* (pp. 202–206).
- Asher, N. (1993). *Reference to abstract objects in discourse* (Vol. 50). Springer.
- Baldrige, J., & Kruijff, G.-J. M. (2002). Coupling ccg and hybrid logic dependency semantics. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 319–326).
- Baldrige, J., & Lascarides, A. (2005). Annotating discourse structures for robust semantic interpretation. In *Proceedings of the 6th international workshop on computational semantics*.
- Calhoun, S., Carletta, J., Brenier, J. M., Mayo, N., Jurafsky, D., Steedman, M., & Beaver, D. (2010). The nxt-format switchboard corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language resources and evaluation*, 44(4), 387–419.
- Demirşahin, I. (2008). *Connective position, argument order and information structure of discourse connectives in written turkish texts* (Unpublished master's thesis). Middle East Technical University.
- Demirşahin, I. (2012). Discourse structure in simultaneous spoken turkish. In *Proceedings of acl 2012 student research workshop* (pp. 55–60).
- Demirşahin, I., Öztürel, A., Bozsahin, C., & Zeyrek, D. (2013). Applicative structures and immediate discourse in the turkish discourse bank. In *Proceedings of the fourth linguistic annotation workshop* (pp. 32–69).
- Demirşahin, I., Sevdik-Çallı, A., Balaban, H. Ö., Çakıcı, R., & Zeyrek, D. (2012). Turkish discourse bank: Ongoing developments. In *Proc. Irec 2012. the first turkic languages workshop*.
- Demirşahin, I., Yalçınkaya, İ., & Zeyrek, D. (2012). Pair annotation: Adaption of pair programming to corpus annotation. In *Proceedings of the sixth linguistic annotation workshop* (pp. 31–39).
- Demirşahin, I., & Zeyrek, D. (2014). Annotating discourse connectives in spoken turkish. *LAW VIII*, 105.
- Demirşahin, I., & Zeyrek, D. (in press). Pair annotation as a novel annotation procedure: The case of turkish discourse bank. In J. Pustejovsky & N. Ide (Eds.), *Handbook of linguistic annotation*. Springer Verlag.
- Egg, M., & Redeker, G. (2008). Underspecified discourse representation. In A. Benz & P. Kuhnlein (Eds.), *Constraints in discourse* (pp. 117–138). John Benjamins Publishing.

- Egg, M., & Redeker, G. (2010). How complex is discourse structure? In *In proceedings of the seventh international conference on language resources and evaluation (Irec)*.
- Forbes, K., Miltsakaki, E., Prasad, R., Sarkar, A., Joshi, A., & Webber, B. (2003). D-ltag system: Discourse parsing with a lexicalized tree-adjoining grammar. *Journal of Logic, Language and Information*, 12(3), 261–279.
- Forbes-Riley, K., Webber, B., & Joshi, A. (2006). Computing discourse semantics: The predicate-argument semantics of discourse connectives in d-ltag. *Journal of Semantics*, 23(1), 55–106.
- Grimes, J. E. (1975). *The thread of discourse* (Vol. 207). Walter de Gruyter.
- Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational linguistics*, 12(3), 175–204.
- Halliday, M. A., & Hasan, R. (1976). *Cohesion in english*. Longman.
- Hobbs, J. R. (1979). Coherence and coreference. *Cognitive science*, 3(1), 67–90.
- Hobbs, J. R. (1985). *On the coherence and structure of discourse* (Tech. Rep.). Report CSLI-85-37, Center for Study of Language and Information.
- Joshi, A. K. (1985). How much contextsensitivity is necessary for characterizing structural descriptions: Tree adjoining grammars. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing*. Cambridge University Press.
- Joshi, A. K. (1987). An introduction to tree adjoining grammars. *Mathematics of language*, 1, 87–115.
- Joshi, A. K. (2011). Some aspects of transition from sentence to discourse. In *Keynote address, informatics science festival, middle east technical university, ankara, june 9*.
- Joshi, A. K., & Schabes, Y. (1997). Tree-adjoining grammars. In *Handbook of formal languages* (pp. 69–123). Springer.
- Kamp, H. (1981). A theory of truth and semantic representation. *Formal semantics-the essential readings*, 189–222.
- Knott, A., Oberlander, J., O'Donnell, M., & Mellish, C. (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* (pp. 181–196). John Benjamins Publishing.
- Kornfilt, J. (2013). *Turkish*. Routledge.
- Kruijff, G.-J. (2001). *A categorial-modal logical architecture of informativity* (Unpublished doctoral dissertation). Citeseer.
- Lee, A., Prasad, R., Joshi, A., Dinesh, N., & Webber, B. (2006). Complexity of dependencies in discourse: Are dependencies in discourse more complex than in syntax. In *Proceedings of the 5th international workshop on treebanks and linguistic theories, prague, czech republic, december*.

- Lee, A., Prasad, R., Joshi, A., & Webber, B. (2008). Departures from tree structures in discourse: Shared arguments in the penn discourse treebank. In *Proceedings of the constraints in discourse iii workshop*.
- Longacre, R. E. (1976). *An anatomy of speech notions* (No. 3). Peter de Ridder Press Lisse.
- Mann, W. C., & Thompson, S. A. (1987). *Rhetorical structure theory: A theory of text organization*. (Tech. Rep.). DTIC Document.
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3), 243–281.
- McCarthy, J. (1963). *Situations, actions, and causal laws* (Tech. Rep.). DTIC Document.
- Nakatsu, C., & White, M. (2010). Generating with discourse combinatory categorial grammar. *Linguistic Issues in Language Technology*, 4(1), 1–62.
- Özge, U. (2003). *A tune-based account of turkish information structure* (Unpublished master's thesis). Middle East Technical University.
- Özge, U., & Bozsahin, C. (2010). Intonation in the grammar of turkish. *Lingua*, 120(1), 132–175.
- Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of pragmatics*, 12(5), 601–638.
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. K., & Webber, B. L. (2008). The penn discourse treebank 2.0. In *Lrec*.
- Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., & Webber, B. L. (2007). *The penn discourse treebank 2.0 annotation manual* (Tech. Rep.). IRCS Technical Reports Series.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A grammar of english*. Longman: London.
- Ruhi, Ş. (2009). The pragmatics of yani as a parenthetical marker in turkish: Evidence from the metu turkish corpus. *Working papers in corpus-based linguistics and language education*, 285–298.
- Say, B., Zeyrek, D., Oflazer, K., & Özge, U. (2002). Development of a corpus and a treebank for present-day written turkish. In *Proceedings of the eleventh international conference of turkish linguistics* (pp. 183–192).
- Shieber, S. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, 8, 333–343.
- Steedman, M. (2002). Plans, affordances, and combinatory grammar. *Linguistics and Philosophy*, 25(5-6), 723–753.
- Stent, A. (2000). Rhetorical structure in dialog. In *Proceedings of the first international conference on natural language generation-volume 14* (pp. 247–252).
- Tonelli, S., Riccardi, G., Prasad, R., & Joshi, A. K. (2010). Annotation of discourse relations for conversational spoken dialogs. In *Lrec*.

- Von Heusinger, K. (2002). Specificity and definiteness in sentence and discourse structure. *Journal of semantics*, 19(3), 245–274.
- Webber, B. (2004). D-Itag: Extending lexicalized tag to discourse. *Cognitive Science*, 28(5), 751–779.
- Webber, B. (2006). Accounting for discourse relations: Constituency and dependency. *Intelligent linguistic architectures*, 339–360.
- Webber, B., Egg, M., & Kordoni, V. (2012). Discourse structure and language technology. *Natural Language Engineering*, 18(4), 437–490.
- Webber, B., & Joshi, A. (1998). Anchoring a lexicalized tree-adjoining grammar for discourse. In *Coling/acl workshop on discourse relations and discourse markers* (pp. 86–92).
- Webber, B., Joshi, A., Miltsakaki, E., Prasad, R., Dinesh, N., Lee, A., & Forbes, K. (2006). A short introduction to the penn discourse tree bank. *COPENHAGEN STUDIES IN LANGUAGE*, 32, 9.
- Webber, B., Stone, M., Joshi, A., & Knott, A. (2003). Anaphora and discourse structure. *Computational Linguistics*, 29(4), 545–587.
- Webber, B. L. (1988). Tense as discourse anaphor. *Computational Linguistics*, 14(2), 61–73.
- Williams, L., Kessler, R. R., Cunningham, W., & Jeffries, R. (2000). Strengthening the case for pair programming. *IEEE software*, 17(4), 19–25.
- Wolf, F., & Gibson, E. (2004). Representing discourse coherence: a corpus-based analysis. In *Proceedings of the 20th international conference on computational linguistics* (p. 134).
- Wolf, F., & Gibson, E. (2005). Representing discourse coherence: A corpus-based study. *Computational Linguistics*, 31(2), 249–287.
- Zeyrek, D., Demirşahin, I., Sevdik-Çallı, A., Balaban, H. Ö., Yalçınkaya, İ., & Turan, Ü. D. (2010). The annotation scheme of the turkish discourse bank and an evaluation of inconsistent annotations. In *Proceedings of the fourth linguistic annotation workshop* (pp. 282–289).
- Zeyrek, D., Demirşahin, I., Sevdik-Çallı, A., & Çakıcı, R. (2013). Turkish discourse bank: Porting a discourse annotation style to a morphologically rich language. *Dialogue & Discourse*, 4(2), 174–184.
- Zeyrek, D., Turan, Ü. D., & Demirşahin, I. (2008). Structural and presuppositional connectives in turkish. In editor (Ed.), *Proceedings of the constraint in discourse iii, potsdam, germany* (pp. 131–137).
- Zeyrek, D., & Webber, B. L. (2008). A discourse resource for turkish: Annotating discourse connectives in the metu corpus. In *Proceedings of the the 6th workshop on asian language resources, the 3rd international joint conference on natural language processing (ijnlp)*, (pp. 65–72).

APPENDIX A

DESCRIPTIVES

Table A.1: The number of annotated connectives and their total number of occurrences in TDB 1.0.

	Search Token	Annotations	Total Occurences
1	<i>aksine</i>	13	21
2	<i>ama</i>	1024	1126
3	<i>amaçla</i>	11	16
4	<i>amacıyla</i>	64	77
5	<i>amacı ile</i>	1	2
6	<i>ancak</i>	419	525
7	<i>ardından</i>	71	207
8	<i>aslında</i>	81	127
9	<i>ayrıca</i>	108	125
10	<i>beraber</i>	6	39
11	<i>beri</i>	4	81
12	<i>birlikte</i>	33	363
13	<i>böylece</i>	85	97
14	<i>bu yana</i>	10	73
15	<i>çünkü</i>	300	305
16	<i>dahası</i>	10	13
17	<i>dolayı</i>	21	58
18	<i>dolayısı ile</i>	1	2
19	<i>dolayısıyla</i>	66	83
20	<i>ek olarak</i>	1	3
21	<i>fakat</i>	80	89
22	<i>fekat</i>	3	3
23	<i>gene de</i>	26	27
24	<i>gerek</i>	2	122
25	<i>gibi</i>	228	1503
26	<i>ha... ha</i>	2	4
27	<i>halbuki</i>	17	18
28	<i>halde</i>	61	70
29	<i>hem</i>	41	197
30	<i>hem... hem</i>	41	126
31	<i>için</i>	1102	2144

Continued on next page

Table A.1 – continued from previous page

	Search Token	Annotations	Total Occurrences
32	<i>içindir</i>	4	6
33	<i>iken</i>	22	22
34	<i>ister</i>	6	48
35	<i>kadar</i>	159	1033
36	<i>karşılık</i>	28	69
37	<i>karşın</i>	71	113
38	<i>mesela</i>	13	20
39	<i>ne... ne</i>	44	163
40	<i>ne ki</i>	14	16
41	<i>ne var ki</i>	32	34
42	<i>nedeni ile</i>	3	8
43	<i>nedeniyle</i>	42	220
44	<i>nedenle</i>	117	120
45	<i>nedenlerle</i>	4	13
46	<i>neticede</i>	1	1
47	<i>neticesinde</i>	1	2
48	<i>önce</i>	134	532
49	<i>örneğin</i>	64	83
50	<i>örnek olarak</i>	2	4
51	<i>ötürü</i>	11	20
52	<i>oysa</i>	136	137
53	<i>rağmen</i>	77	136
54	<i>sayede</i>	5	5
55	<i>sayesinde</i>	3	26
56	<i>sebeple</i>	1	2
57	<i>sözgelimi</i>	6	8
58	<i>söz gelimi</i>	1	2
59	<i>sonra</i>	713	1255
60	<i>sonuç olarak</i>	5	5
61	<i>sonuçta</i>	10	18
62	<i>sonucunda</i>	12	48
63	<i>taftan</i>	3	15
64	<i>tersine</i>	11	27
65	<i>ve</i>	2111	7486
66	<i>veya</i>	40	188
67	<i>veyahut</i>	4	6
68	<i>ya</i>	2	552
69	<i>ya... ya</i>	6	66
70	<i>ya da</i>	139	412
71	<i>yahut</i>	3	6
72	<i>yalnız</i>	12	123
73	<i>yandan</i>	70	102
74	<i>yine de</i>	65	67
75	<i>yoksa</i>	75	103
Continued on next page			

Table A.1 – continued from previous page

	Search Token	Annotations	Total Occurrences
76	<i>yüzden</i>	66	68
77	<i>yüziinden</i>	5	69
78	<i>zaman</i>	159	521
79	<i>zamanda</i>	39	84
	Total	8483	21710

APPENDIX B

A SAMPLE XML FILE FROM TDB

```
<?xml version "1.0" encoding="UTF-8"? > <Document >
  <Relation note="" type="EXPLICIT" >
    <Conn >
      <Span >
        <Text >aksine </Text >
        <BeginOffset >679 </BeginOffset >
        <EndOffset >685 </EndOffset >
      </Span >
    </Conn >
    <Mod >
      <Span >
        <Text >tam </Text >
        <BeginOffset >675 </BeginOffset >
        <EndOffset >678 </EndOffset >
      </Span >
    </Mod >
    <Arg1 >
      <Span >
        <Text >Adalet Bakanı Seyit Bey, maddeye ilişkin
          eleştirilere katıldığını belirtmiş </Text >
        <BeginOffset >563 </BeginOffset >
```

```
<EndOffset >638 </EndOffset >
</Span >
</Arg1 >
<Arg2 >
  <Span >
    <Text >Cebelibereket mebusu İhsan Bey ise </Text >
    <BeginOffset >640 </BeginOffset >
    <EndOffset >674 </EndOffset >
  </Span >
  <Span >
    <Text >“inkılâbın adaletinin” uygulanması istemiştir </Text >
    <BeginOffset >686 </BeginOffset >
    <EndOffset >731 </EndOffset >
  </Span >
</Arg2 >
</Relation >
</Document >
```

APPENDIX C

TOOLS

TDB Tools

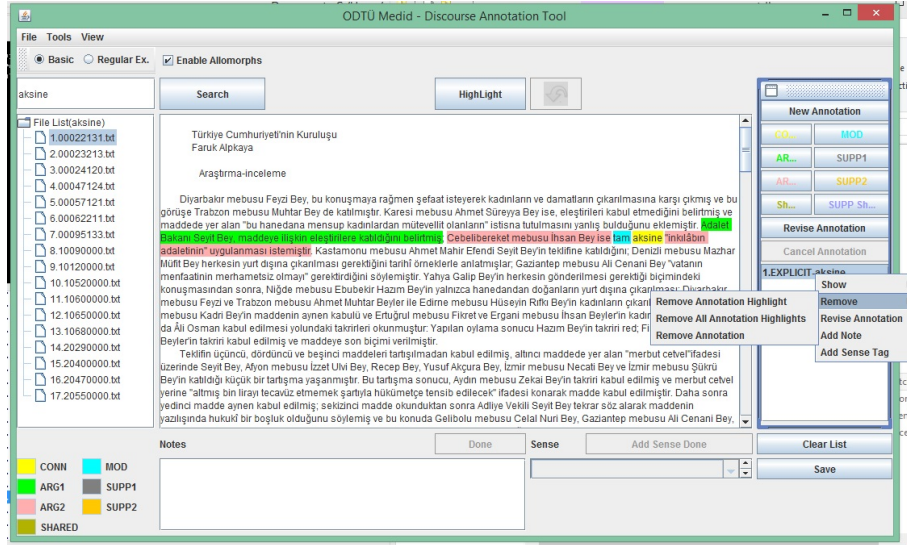


Figure C.1: Discourse Annotation Tool for Turkish

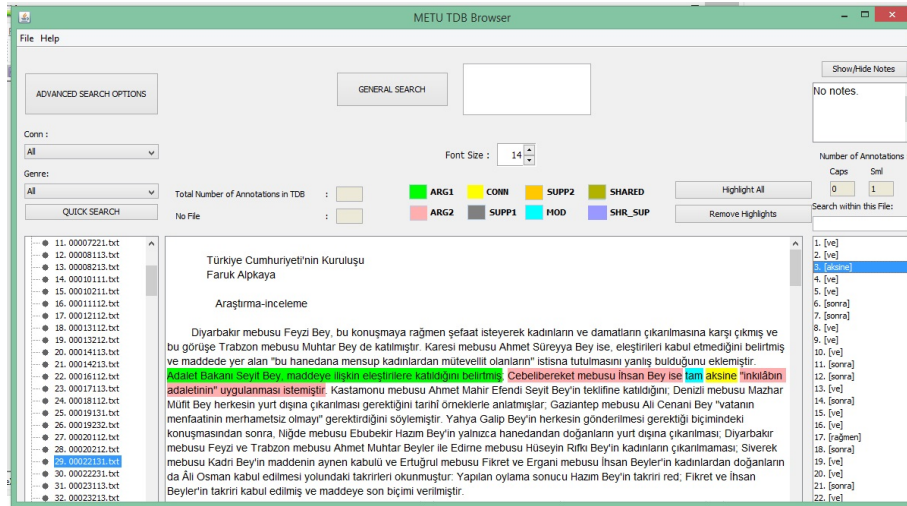


Figure C.2: Turkish Discourse Bank Browser

Tools used for STC Demo annotation

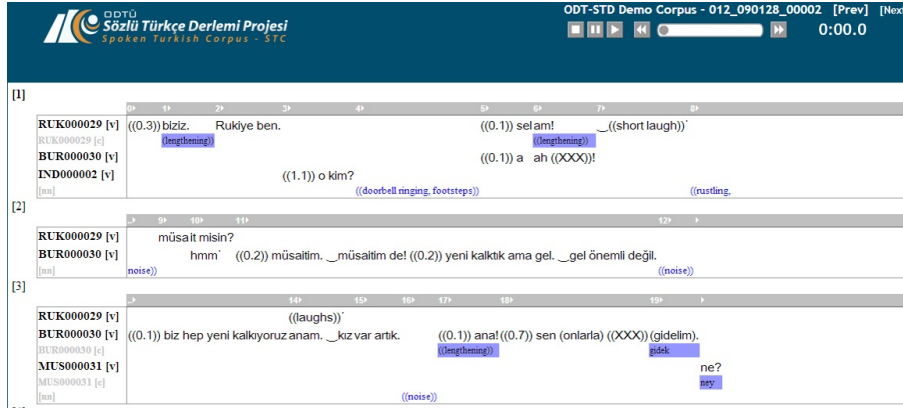


Figure C.3: Spoken Turkish Corpus Demo Exmeralda Interface

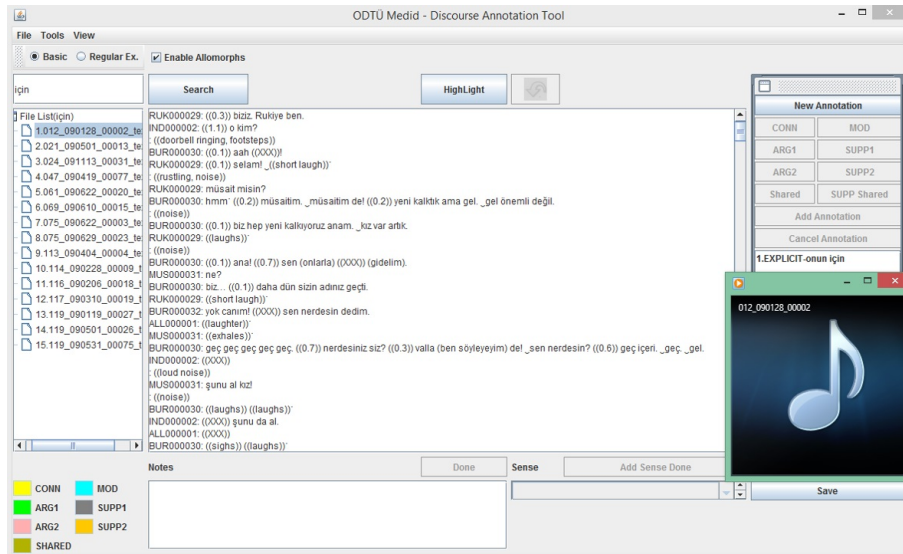


Figure C.4: Flat Spoken Turkish Corpus Transcriptions in Discourse Annotation for Turkish together with the audio on Windows Media Player

APPENDIX D

LIST OF ALL CONFIGURATIONS

Table D.1: List of all configurations, reasons for tree violations, and the results of reannotation in the TDB 1.0

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00001131	2	3	cor	cor	shared	interpret	embed
00001131	4	5	cor	cor	embed	-	embed
00001131	5	6	cor	cor	embed	-	embed
00001131	12	13	adv	adv	pc-arg	interpret	embed
00001131	18	19	cor	sub	embed	-	embed
00001131	27	28	sub	cor	pc-rel	syntactic	pc-rel
00001131	28	29	cor	sub	pc-rel	interpret	embed
00001131	32	33	cor	sub	embed	-	embed
00001131	40	41	adv	cor	embed	-	embed
00001131	42	43	cor	sub	pc-arg	interpret	embed
00001131	44	45	cor	cor	shared	semantic	shared
00001131	56	57	cor	sub	pc-rel	missing	embed
00001131	58	59	cor	cor	embed	-	embed
00001131	66	67	cor	sub	embed	-	embed
00001231	6	7	cor	adv	shared	semantic	shared
00001231	11	12	adv	cor	embed	-	embed
00001231	17	18	sub	cor	embed	-	embed
00001231	29	30	adv	cor	pc-arg	interpret	embed
00001231	31	32	adv	cor	pc-arg	interpret	shared
00001231	35	36	adv	cor	pc-rel	interpret	embed
00001231	35	37	adv	sub	pc-rel	interpret	embed
00001231	36	37	cor	sub	embed	-	embed
00001231	45	46	cor	sub	embed	-	embed
00002113	3	4	cor	adv	shared	multi	ident
00002113	5	6	cor	adv	shared	multi	ident
00002113	10	11	phr	sub	embed	-	embed
00002113	14	15	sub	adv	pc-rel	interpret	embed
00002113	23	24	cor	cor	shared	interpret	ident
00002113	27	28	cor	cor	shared	interpret	embed
00002213	12	13	sub	cor	embed	-	embed
00002213	23	24	adv	adv	nested	-	nested

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00003121	2	3	adv	cor	pc-arg	interpret	embed
00003121	3	4	cor	adv	shared	interpret	embed
00003121	7	8	cor	cor	pc-rel	syntactic	pc-rel
00003121	10	13	cor	adv	pc-rel	syntactic	pc-rel
00003121	11	12	cor	cor	shared	semantic	shared
00003121	11	13	cor	adv	pc-rel	syntactic	pc-rel
00003121	12	13	cor	adv	pc-rel	syntactic	pc-rel
00003121	14	15	cor	adv	embed	-	embed
00003121	15	16	adv	cor	pc-rel	syntactic	pc-rel
00003121	15	17	adv	cor	pc-rel	syntactic	pc-rel
00003121	16	17	cor	cor	embed	-	embed
00003121	21	22	sub	cor	pc-arg	missing	embed
00003121	25	26	cor	cor	shared	interpret	embed
00003121	25	27	cor	adv	pc-rel	interpret	embed
00003121	26	27	cor	adv	pc-rel	interpret	embed
00003121	27	28	adv	cor	shared	semantic	shared
00003121	42	43	cor	cor	pc-arg	interpret	embed
00003221	4	5	adv	adv	shared	semantic	shared
00003221	10	11	adv	cor	pc-rel	syntactic	pc-rel
00003221	15	16	adv	cor	shared	interpret	embed
00003221	19	20	cor	adv	pc-rel	syntactic	pc-rel
00003221	20	21	adv	cor	pc-rel	missing	embed
00003221	23	24	cor	cor	embed	-	embed
00003221	24	25	cor	cor	embed	-	embed
00003221	24	26	cor	sub	embed	interpret	embed
00003221	25	26	cor	sub	embed	-	embed
00003221	28	29	sub	sub	shared	missing	embed
00003221	28	30	sub	sub	shared	missing	embed
00003221	28	31	sub	sub	shared	missing	embed
00003221	28	32	sub	sub	shared	missing	embed
00003221	29	30	sub	sub	shared	missing	embed
00003221	29	31	sub	sub	shared	missing	embed
00003221	29	32	sub	sub	shared	missing	embed
00003221	30	31	sub	sub	shared	missing	embed
00003221	30	32	sub	sub	shared	missing	embed
00003221	31	32	sub	sub	shared	missing	embed
00003221	40	41	cor	adv	embed	-	embed
00003221	45	46	sub	cor	embed	-	embed
00003221	52	53	adv	cor	pc-arg	interpret	embed
00003221	55	56	cor	cor	shared	semantic	shared
00003221	56	57	cor	cor	pc-arg	interpret	embed
00005121	6	7	cor	phr	shared	multi	ident
00005121	11	12	cor	adv	pc-rel	syntactic	pc-rel
00005121	15	16	sub	adv	pc-rel	missing	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00005221	3	4	sub	phr	nested	-	nested
00005221	3	4	sub	phr	pc-rel	syntactic	pc-rel
00005221	8	9	cor	sub	embed	-	embed
00005221	17	18	cor	sub	embed	-	embed
00005221	17	19	cor	adv	shared	interpret	embed
00005221	17	20	cor	adv	pc-arg	interpret	embed
00005221	18	19	sub	adv	embed	-	embed
00005221	18	20	sub	adv	pc-rel	interpret	embed
00005221	19	20	adv	adv	pc-arg	interpret	embed
00005221	23	24	adv	cor	shared	interpret	embed
00005221	25	26	cor	adv	pc-arg	interpret	embed
00005221	30	31	cor	cor	pc-rel	syntactic	pc-rel
00005221	37	38	sub	cor	pc-rel	syntactic	pc-rel
00005221	37	39	sub	adv	pc-rel	syntactic	pc-rel
00005221	38	39	cor	adv	shared	multi	ident
00005221	42	43	sub	cor	embed	-	embed
00005221	49	50	adv	sub	embed	-	embed
00005221	59	60	cor	cor	embed	-	embed
00005221	63	64	cor	cor	nested	-	nested
00005221	64	65	cor	cor	embed	-	embed
00005221	67	68	adv	cor	shared	semantic	shared
00005221	70	71	cor	cor	embed	-	embed
00005221	72	73	cor	adv	shared	semantic	shared
00005221	74	75	cor	adv	shared	multi	ident
00006131	1	2	cor	adv	shared	interpret	ident
00006131	1	3	cor	adv	pc-arg	interpret	embed
00006131	1	4	cor	adv	pc-arg	interpret	embed
00006131	2	3	adv	adv	embed	-	embed
00006131	2	4	adv	adv	pc-arg	interpret	embed
00006131	3	4	adv	adv	shared	error	ident
00006131	13	14	cor	sub	embed	-	embed
00006131	18	19	sub	sub	shared	interpret	embed
00006131	33	34	adv	cor	pc-rel	interpret	embed
00006231	1	2	sub	adv	pc-rel	interpret	embed
00006231	3	4	adv	cor	embed	-	embed
00006231	3	5	adv	adv	nested	-	nested
00006231	4	5	cor	adv	nested	-	nested
00006231	11	12	adv	cor	embed	-	embed
00006231	15	16	cor	adv	pc-arg	interpret	ident
00006231	19	20	sub	cor	pc-rel	missing	embed
00006231	26	27	sub	adv	embed	-	embed
00006231	26	28	sub	adv	embed	-	embed
00006231	27	28	adv	adv	shared	multi	ident
00006231	32	33	sub	phr	pc-rel	missing	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00007121	5	6	cor	cor	embed	-	embed
00007121	7	8	cor	sub	embed	-	embed
00007121	11	12	cor	cor	shared	semantic	shared
00007121	12	13	cor	cor	pc-rel	syntactic	pc-rel
00007121	16	17	adv	cor	pc-rel	syntactic	pc-rel
00007121	27	28	cor	cor	pc-rel	syntactic	pc-rel
00007121	33	34	adv	cor	embed	-	embed
00007121	33	35	adv	adv	shared	semantic	shared
00007121	34	35	cor	adv	embed	-	embed
00007121	35	36	adv	sub	embed	-	embed
00007121	38	39	adv	adv	embed	leftout	embed
00007121	42	43	sub	sub	shared	error	ident
00007121	44	45	sub	cor	shared	interpret	embed
00007121	55	56	adv	adv	shared	interpret	embed
00007221	3	4	adv	cor	shared	interpret	embed
00007221	9	10	cor	sub	pc-rel	syntactic	pc-rel
00007221	16	17	sub	cor	pc-rel	syntactic	pc-rel
00007221	19	21	cor	adv	pc-rel	missing	embed
00007221	20	21	cor	adv	nested	-	nested
00007221	29	30	cor	adv	embed	-	embed
00007221	33	34	sub	cor	pc-rel	missing	embed
00007221	36	37	sub	adv	embed	-	embed
00007221	43	44	adv	adv	shared	interpret	embed
00007221	46	47	cor	adv	nested	-	nested
00007221	52	53	adv	adv	pc-arg	interpret	embed
00007221	55	56	cor	phr	embed	-	embed
00007221	61	62	cor	cor	shared	interpret	embed
00008113	2	3	cor	sub	embed	interpret	embed
00008113	2	4	cor	cor	embed	-	embed
00008113	2	5	cor	adv	embed	-	embed
00008113	2	6	cor	sub	embed	interpret	embed
00008113	3	4	sub	cor	embed	-	embed
00008113	3	5	sub	adv	embed	-	embed
00008113	4	5	cor	adv	shared	multi	ident
00008113	4	6	cor	sub	embed	-	embed
00008113	5	6	adv	sub	embed	-	embed
00008113	9	10	sub	sub	shared	interpret	embed
00008113	12	13	cor	sub	pc-rel	syntactic	pc-rel
00008113	14	15	cor	sub	pc-rel	syntactic	pc-rel
00008113	14	16	cor	cor	pc-rel	syntactic	pc-rel
00008113	15	16	sub	cor	embed	-	embed
00008113	18	19	cor	cor	pc-arg	interpret	embed
00008113	19	20	cor	cor	shared	interpret	embed
00008113	23	24	sub	cor	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00008113	28	29	sub	adv	embed	-	embed
00008113	34	35	sub	adv	embed	-	embed
00008113	34	36	sub	cor	pc-rel	interpret	embed
00008113	34	37	sub	adv	pc-rel	interpret	embed
00008113	35	36	adv	cor	embed	-	embed
00008113	35	37	adv	adv	embed	-	embed
00008113	35	38	adv	adv	shared	semantic	shared
00008113	36	37	cor	adv	shared	multi	ident
00008113	36	38	cor	adv	embed	-	embed
00008113	37	38	adv	adv	embed	-	embed
00008113	40	41	cor	cor	pc-rel	interpret	embed
00008113	49	50	cor	adv	shared	multi	ident
00008113	52	53	cor	sub	embed	-	embed
00008213	5	6	cor	adv	pc-rel	syntactic	pc-rel
00008213	8	9	sub	cor	shared	interpret	embed
00008213	23	24	cor	adv	pc-rel	interpret	embed
00008213	25	26	sub	adv	embed	-	embed
00008213	26	27	adv	adv	shared	interpret	embed
00008213	27	28	adv	cor	pc-rel	syntactic	pc-rel
00008213	33	36	cor	phr	pc-arg	missing	embed
00008213	34	35	cor	adv	pc-rel	syntactic	pc-rel
00008213	34	36	cor	phr	pc-rel	syntactic	pc-rel
00008213	35	36	adv	phr	pc-arg	missing	embed
00008213	37	38	sub	cor	pc-rel	interpret	embed
00008213	40	41	adv	cor	pc-arg	interpret	embed
00008213	42	43	adv	cor	shared	interpret	embed
00008213	44	45	cor	cor	shared	semantic	shared
00008213	50	51	cor	cor	pc-rel	missing	embed
00008213	51	52	cor	cor	embed	-	embed
00008213	54	55	cor	cor	embed	-	embed
00008213	55	56	cor	sub	embed	-	embed
00010111	6	7	cor	adv	shared	multi	ident
00010111	15	16	cor	adv	shared	missing	embed
00010111	24	25	adv	cor	embed	-	embed
00010111	31	32	cor	cor	pc-rel	missing	embed
00010111	31	33	cor	adv	pc-arg	interpret	indep
00010111	38	39	cor	cor	embed	-	embed
00010111	40	41	cor	cor	embed	-	embed
00010111	43	44	cor	adv	shared	interpret	embed
00010111	44	45	adv	cor	pc-rel	syntactic	pc-rel
00010111	47	48	cor	cor	shared	interpret	embed
00010111	48	49	cor	adv	pc-arg	interpret	embed
00010111	53	54	cor	phr	nested	-	nested
00010111	53	55	cor	cor	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00010111	54	55	phr	cor	cross	semantic	cross
00010111	58	59	sub	sub	pc-rel	interpret	embed
00010111	58	60	sub	phr	pc-arg	interpret	embed
00010111	59	60	sub	phr	pc-rel	missing	embed
00010211	2	3	adv	cor	shared	semantic	shared
00010211	3	4	cor	cor	embed	-	embed
00010211	6	7	phr	adv	shared	interpret	embed
00010211	7	8	adv	cor	pc-arg	semantic	shared
00010211	9	10	adv	adv	pc-arg	syntactic	pc-rel
00010211	9	11	adv	phr	pc-arg	syntactic	pc-arg
00010211	10	11	adv	phr	pc-arg	interpret	shared
00010211	11	12	phr	cor	pc-arg	interpret	shared
00010211	12	13	cor	adv	shared	semantic	shared
00010211	14	15	adv	cor	shared	interpret	embed
00010211	17	18	phr	cor	shared	interpret	embed
00010211	27	28	sub	adv	pc-rel	interpret	embed
00010211	29	30	adv	cor	shared	semantic	shared
00010211	34	35	cor	adv	pc-rel	syntactic	pc-rel
00010211	40	41	sub	cor	pc-rel	missing	embed
00010211	42	43	cor	cor	shared	error	ident
00010211	48	49	adv	cor	nested	-	nested
00010211	49	50	cor	phr	pc-arg	interpret	shared
00010211	50	51	phr	cor	shared	interpret	embed
00010211	50	52	phr	adv	pc-arg	missing	embed
00010211	51	52	cor	adv	pc-rel	missing	embed
00011112	1	2	sub	cor	embed	-	embed
00011112	2	3	cor	cor	pc-rel	interpret	embed
00011112	16	17	adv	adv	pc-rel	syntactic	pc-rel
00011112	16	18	adv	cor	shared	semantic	shared
00011112	17	18	adv	cor	pc-rel	syntactic	pc-rel
00011112	24	25	sub	cor	embed	-	embed
00011112	25	26	cor	cor	pc-rel	interpret	embed
00012112	3	4	adv	cor	shared	semantic	shared
00012112	8	9	adv	cor	pc-arg	interpret	embed
00012112	13	14	adv	cor	pc-rel	interpret	nested
00012112	17	18	cor	adv	embed	leftout	embed
00012112	19	20	sub	cor	shared	interpret	embed
00012112	21	22	adv	cor	pc-rel	syntactic	pc-rel
00012112	25	26	sub	cor	pc-rel	syntactic	pc-rel
00012112	25	27	sub	cor	nested	-	nested
00012112	26	27	cor	cor	nested	-	nested
00012112	27	28	cor	sub	pc-rel	interpret	indep
00012112	27	29	cor	adv	pc-arg	missing	embed
00012112	30	31	sub	cor	embed	leftout	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00012112	30	32	sub	cor	pc-rel	syntactic	pc-rel
00012112	31	32	cor	cor	pc-rel	syntactic	pc-rel
00012112	34	35	phr	adv	shared	interpret	embed
00012112	35	36	adv	phr	pc-arg	missing	embed
00012112	40	41	sub	cor	pc-rel	syntactic	pc-rel
00013112	3	4	cor	cor	pc-rel	syntactic	pc-rel
00013112	5	6	adv	sub	pc-rel	syntactic	pc-rel
00013112	13	17	cor	phr	pc-rel	missing	embed
00013112	14	15	cor	adv	embed	-	embed
00013112	14	17	cor	phr	pc-rel	missing	embed
00013112	15	17	adv	phr	pc-rel	missing	embed
00013112	16	17	cor	phr	pc-arg	interpret	ident
00013112	16	18	cor	adv	pc-arg	interpret	embed
00013112	17	18	phr	adv	embed	-	embed
00013112	21	22	cor	adv	shared	multi	ident
00013112	25	26	adv	cor	embed	-	embed
00013112	27	28	adv	cor	nested	-	nested
00013112	29	30	cor	adv	shared	interpret	embed
00013112	30	31	adv	adv	shared	interpret	embed
00013112	35	36	cor	sub	embed	-	embed
00013112	43	44	cor	adv	pc-rel	interpret	embed
00013112	48	49	phr	cor	embed	-	embed
00013112	60	61	cor	adv	shared	multi	ident
00013112	64	65	adv	cor	shared	semantic	shared
00013212	3	4	cor	adv	embed	-	embed
00013212	7	8	cor	sub	embed	-	embed
00013212	9	10	adv	cor	pc-rel	syntactic	pc-rel
00013212	13	14	cor	adv	shared	multi	ident
00013212	15	16	sub	cor	embed	-	embed
00013212	16	17	cor	phr	shared	semantic	shared
00013212	19	20	adv	cor	pc-rel	syntactic	pc-rel
00013212	22	23	adv	cor	pc-rel	syntactic	pc-rel
00013212	27	28	cor	phr	shared	multi	ident
00013212	31	32	cor	adv	shared	multi	ident
00013212	33	34	phr	phr	nested	-	nested
00014113	3	4	cor	sub	embed	-	embed
00014113	6	7	sub	adv	pc-rel	missing	embed
00014113	8	9	sub	cor	pc-rel	missing	embed
00014113	14	15	sub	phr	pc-rel	missing	embed
00014113	16	17	adv	sub	pc-rel	missing	embed
00014113	18	19	cor	cor	shared	interpret	embed
00014113	19	20	cor	phr	shared	semantic	shared
00014113	22	23	adv	adv	shared	semantic	shared
00014113	26	27	phr	cor	pc-rel	interpret	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00014213	2	3	cor	cor	embed	leftout	embed
00014213	2	4	cor	adv	embed	leftout	embed
00014213	3	4	cor	adv	shared	multi	ident
00014213	3	5	cor	cor	embed	-	embed
00014213	3	6	cor	phr	embed	interpret	embed
00014213	3	7	cor	cor	pc-rel	interpret	embed
00014213	4	5	adv	cor	shared	interpret	embed
00014213	5	6	cor	phr	pc-rel	interpret	embed
00014213	5	7	cor	cor	pc-rel	interpret	embed
00014213	6	7	phr	cor	embed	-	embed
00014213	16	17	cor	sub	pc-rel	missing	embed
00014213	22	23	cor	sub	embed	-	embed
00014213	22	24	cor	phr	pc-rel	syntactic	pc-rel
00014213	23	24	sub	phr	pc-rel	syntactic	pc-rel
00014213	25	26	sub	phr	embed	-	embed
00014213	27	28	cor	phr	pc-rel	missing	embed
00014213	28	29	phr	sub	pc-rel	missing	embed
00014213	28	30	phr	phr	shared	missing	embed
00014213	29	30	sub	phr	pc-rel	missing	embed
00014213	32	33	adv	cor	embed	leftout	embed
00014213	38	39	cor	cor	embed	-	embed
00014213	40	41	cor	adv	embed	-	embed
00014213	40	42	cor	cor	pc-rel	syntactic	pc-rel
00014213	40	43	cor	adv	pc-arg	interpret	embed
00014213	41	42	adv	cor	pc-rel	syntactic	pc-rel
00014213	41	43	adv	adv	shared	interpret	embed
00014213	42	43	cor	adv	pc-rel	syntactic	pc-rel
00014213	46	47	cor	cor	pc-rel	syntactic	pc-rel
00014213	46	48	cor	cor	pc-rel	syntactic	pc-rel
00014213	47	48	cor	cor	pc-rel	syntactic	pc-rel
00014213	50	51	cor	cor	shared	semantic	shared
00016112	7	8	cor	sub	embed	-	embed
00016112	9	10	adv	adv	embed	-	embed
00016112	10	11	adv	adv	shared	interpret	embed
00016112	11	12	adv	sub	embed	-	embed
00016112	15	16	cor	sub	embed	-	embed
00016112	22	23	cor	cor	pc-rel	syntactic	pc-rel
00016112	29	30	sub	cor	embed	-	embed
00016112	30	31	cor	cor	pc-rel	syntactic	pc-rel
00016112	32	33	adv	adv	pc-rel	syntactic	pc-rel
00016112	33	34	adv	cor	pc-rel	syntactic	pc-rel
00016112	35	36	cor	cor	embed	-	embed
00017113	1	2	sub	sub	embed	-	embed
00017113	3	4	sub	sub	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00017113	10	11	cor	phr	pc-rel	syntactic	pc-rel
00017113	16	17	cor	adv	embed	-	embed
00017113	17	18	adv	cor	embed	-	embed
00017113	36	37	cor	cor	pc-rel	syntactic	pc-rel
00017113	36	38	cor	phr	pc-rel	syntactic	pc-rel
00017113	37	38	cor	phr	pc-rel	syntactic	pc-rel
00018112	1	2	cor	cor	embed	-	embed
00018112	3	4	adv	cor	pc-rel	syntactic	pc-rel
00019131	7	8	cor	phr	shared	multi	ident
00019131	7	9	cor	sub	embed	-	embed
00019131	7	10	cor	phr	embed	interpret	embed
00019131	8	9	phr	sub	embed	-	embed
00019131	8	10	phr	phr	embed	interpret	embed
00019131	9	10	sub	phr	embed	-	embed
00019131	10	11	phr	adv	shared	interpret	embed
00019131	16	17	adv	adv	shared	semantic	shared
00019131	21	22	cor	phr	shared	multi	ident
00019131	27	28	sub	cor	embed	-	embed
00019131	35	36	cor	sub	embed	-	embed
00019131	37	38	cor	sub	pc-rel	interpret	embed
00019131	41	42	sub	adv	embed	-	embed
00019131	45	47	cor	adv	pc-arg	interpret	embed
00019131	46	47	sub	adv	pc-rel	syntactic	pc-rel
00019232	1	2	sub	adv	pc-rel	syntactic	pc-rel
00019232	4	5	sub	sub	pc-rel	missing	embed
00019232	21	22	phr	sub	embed	-	embed
00019232	28	29	phr	sub	pc-rel	missing	embed
00019232	31	32	cor	sub	pc-rel	interpret	embed
00019232	33	34	sub	cor	pc-rel	missing	embed
00019232	33	35	sub	cor	embed	interpret	embed
00019232	33	36	sub	sub	pc-rel	interpret	embed
00019232	34	35	cor	cor	embed	interpret	embed
00019232	34	36	cor	sub	pc-rel	interpret	embed
00019232	35	36	cor	sub	embed	-	embed
00019232	41	42	sub	adv	pc-arg	interpret	embed
00019232	42	43	adv	sub	pc-rel	interpret	embed
00019232	45	46	sub	adv	embed	-	embed
00020112	8	9	cor	cor	shared	multi	ident
00020112	8	10	cor	cor	pc-rel	missing	embed
00020112	9	10	cor	cor	pc-rel	missing	embed
00022131	5	6	cor	adv	pc-rel	missing	embed
00022131	7	8	sub	cor	shared	interpret	embed
00022131	11	12	sub	sub	pc-rel	syntactic	pc-rel
00022131	11	13	sub	cor	shared	interpret	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00022131	11	14	sub	adv	embed	interpret	embed
00022131	12	13	sub	cor	pc-rel	syntactic	pc-rel
00022131	12	14	sub	adv	pc-rel	syntactic	pc-rel
00022131	13	14	cor	adv	embed	interpret	embed
00022131	14	15	adv	cor	embed	-	embed
00022131	14	16	adv	cor	pc-rel	interpret	embed
00022131	14	17	adv	sub	embed	interpret	embed
00022131	15	16	cor	cor	embed	-	embed
00022131	15	17	cor	sub	embed	interpret	embed
00022131	16	17	cor	sub	pc-rel	syntactic	pc-rel
00022131	18	19	sub	cor	embed	-	embed
00022131	20	21	cor	adv	pc-rel	interpret	embed
00022131	21	22	adv	cor	pc-rel	interpret	shared
00022131	23	24	cor	sub	pc-rel	syntactic	pc-rel
00022131	23	25	cor	cor	pc-rel	syntactic	pc-rel
00022131	24	25	sub	cor	embed	-	embed
00022131	26	27	cor	cor	pc-rel	syntactic	pc-rel
00022131	28	29	cor	adv	pc-rel	interpret	embed
00022131	28	31	cor	adv	pc-rel	interpret	embed
00022131	29	30	adv	cor	embed	-	embed
00022131	29	31	adv	adv	embed	-	embed
00022131	30	31	cor	adv	embed	interpret	embed
00022131	32	33	sub	cor	embed	-	embed
00022131	38	39	sub	cor	shared	interpret	embed
00022131	43	44	sub	cor	pc-rel	syntactic	pc-rel
00022131	43	45	sub	cor	shared	interpret	embed
00022131	44	45	cor	cor	pc-rel	syntactic	pc-rel
00022131	47	48	cor	cor	pc-rel	missing	embed
00022131	47	49	cor	adv	pc-rel	missing	embed
00022131	48	49	cor	adv	shared	multi	ident
00022131	54	55	cor	sub	pc-rel	syntactic	pc-rel
00022131	56	57	cor	adv	pc-rel	missing	embed
00022131	60	61	cor	cor	pc-rel	syntactic	pc-rel
00022131	60	62	cor	cor	pc-rel	syntactic	pc-rel
00022131	60	63	cor	adv	pc-rel	missing	embed
00022131	61	62	cor	cor	pc-rel	interpret	embed
00022131	61	63	cor	adv	pc-rel	syntactic	pc-rel
00022131	62	63	cor	adv	pc-rel	syntactic	pc-rel
00022131	63	64	adv	cor	pc-rel	interpret	embed
00022131	65	66	cor	cor	pc-rel	syntactic	pc-rel
00022131	67	68	cor	cor	pc-rel	syntactic	pc-rel
00022131	70	71	sub	sub	embed	-	embed
00022131	70	72	sub	cor	pc-arg	interpret	embed
00022131	71	72	sub	cor	shared	interpret	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00022131	72	73	cor	sub	embed	-	embed
00022131	72	74	cor	adv	shared	interpret	embed
00022131	73	74	sub	adv	embed	-	embed
00022231	2	3	cor	sub	pc-arg	syntactic	pc-rel
00022231	2	4	cor	cor	pc-rel	syntactic	pc-rel
00022231	3	4	sub	cor	pc-rel	interpret	embed
00022231	5	6	cor	adv	pc-arg	syntactic	pc-rel
00022231	5	7	cor	cor	pc-rel	syntactic	pc-rel
00022231	6	7	adv	cor	pc-rel	syntactic	pc-rel
00022231	8	9	cor	cor	pc-rel	syntactic	pc-rel
00022231	10	11	cor	cor	pc-rel	syntactic	pc-rel
00022231	10	12	cor	cor	pc-rel	syntactic	pc-rel
00022231	11	12	cor	cor	pc-rel	syntactic	pc-rel
00022231	12	13	cor	cor	pc-rel	syntactic	pc-rel
00022231	14	15	cor	cor	pc-rel	syntactic	pc-rel
00022231	15	16	cor	cor	pc-rel	syntactic	pc-rel
00022231	19	20	sub	cor	pc-rel	interpret	embed
00022231	20	21	cor	cor	shared	syntactic	pc-arg
00022231	22	23	cor	cor	pc-rel	syntactic	pc-rel
00022231	24	25	cor	sub	pc-rel	interpret	embed
00022231	27	28	cor	cor	pc-rel	syntactic	pc-rel
00022231	28	29	cor	cor	pc-rel	syntactic	pc-rel
00022231	28	30	cor	cor	pc-rel	syntactic	pc-rel
00022231	29	30	cor	cor	pc-rel	syntactic	pc-rel
00022231	32	33	sub	cor	pc-rel	syntactic	pc-rel
00022231	34	35	cor	cor	pc-rel	syntactic	pc-rel
00022231	34	36	cor	cor	pc-rel	syntactic	pc-rel
00022231	41	42	cor	sub	pc-rel	syntactic	pc-rel
00022231	41	43	cor	cor	pc-rel	syntactic	pc-rel
00022231	41	44	cor	sub	pc-rel	syntactic	pc-rel
00022231	41	45	cor	sub	pc-rel	syntactic	pc-rel
00022231	43	44	cor	sub	embed	-	embed
00022231	47	48	cor	cor	shared	interpret	embed
00022231	48	49	cor	phr	shared	interpret	embed
00022231	49	50	phr	cor	pc-arg	interpret	embed
00022231	50	51	cor	cor	pc-rel	syntactic	pc-rel
00022231	55	56	sub	cor	pc-rel	syntactic	pc-rel
00022231	59	60	phr	cor	pc-rel	syntactic	pc-rel
00022231	61	62	cor	cor	pc-rel	syntactic	pc-rel
00022231	66	67	cor	cor	pc-arg	interpret	embed
00022231	66	68	cor	cor	pc-rel	syntactic	pc-rel
00022231	67	68	cor	cor	pc-rel	syntactic	pc-rel
00022231	68	69	cor	sub	pc-rel	syntactic	pc-rel
00022231	68	70	cor	cor	pc-rel	syntactic	pc-rel

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00022231	72	73	cor	phr	pc-rel	syntactic	pc-rel
00022231	73	74	phr	cor	embed	-	embed
00022231	75	76	sub	adv	shared	semantic	shared
00022231	75	77	sub	cor	pc-rel	syntactic	pc-rel
00022231	76	77	adv	cor	pc-rel	syntactic	pc-rel
00022231	79	80	cor	phr	shared	multi	ident
00022231	79	81	cor	cor	pc-rel	syntactic	pc-rel
00022231	80	81	phr	cor	pc-rel	syntactic	pc-rel
00022231	83	84	phr	cor	pc-rel	syntactic	pc-rel
00022231	85	87	cor	cor	pc-rel	syntactic	pc-rel
00022231	86	87	cor	cor	pc-rel	syntactic	pc-rel
00022231	88	89	sub	cor	pc-rel	syntactic	pc-rel
00022231	88	90	sub	cor	pc-rel	missing	embed
00022231	89	90	cor	cor	pc-rel	syntactic	pc-rel
00022231	91	92	cor	cor	pc-rel	syntactic	pc-rel
00022231	91	93	cor	cor	pc-rel	syntactic	pc-rel
00023113	1	2	cor	sub	pc-rel	interpret	embed
00023113	14	15	adv	sub	shared	multi	ident
00023113	17	18	cor	sub	pc-rel	syntactic	pc-rel
00023113	19	20	cor	sub	embed	-	embed
00023113	27	28	cor	cor	pc-rel	syntactic	pc-rel
00023113	27	29	cor	adv	pc-rel	syntactic	pc-rel
00023113	27	30	cor	adv	shared	semantic	shared
00023113	28	29	cor	adv	pc-arg	multi	ident
00023113	28	30	cor	adv	pc-rel	syntactic	pc-rel
00023113	29	30	adv	adv	pc-rel	syntactic	pc-rel
00023113	30	31	adv	sub	pc-rel	missing	embed
00023113	30	32	adv	adv	pc-rel	syntactic	pc-rel
00023113	31	32	sub	adv	pc-rel	missing	embed
00023113	40	41	cor	phr	embed	-	embed
00023113	41	42	phr	cor	pc-arg	missing	embed
00023113	42	43	cor	sub	pc-rel	syntactic	pc-rel
00023213	2	3	sub	sub	pc-rel	syntactic	pc-rel
00023213	2	4	sub	cor	pc-rel	missing	embed
00023213	2	5	sub	sub	pc-rel	missing	embed
00023213	3	4	sub	cor	pc-rel	syntactic	pc-rel
00023213	4	5	cor	sub	pc-rel	missing	embed
00023213	6	7	cor	adv	shared	interpret	embed
00023213	6	8	cor	adv	nested	-	nested
00023213	9	10	cor	sub	embed	-	embed
00023213	21	22	adv	cor	embed	-	embed
00023213	24	25	sub	cor	shared	interpret	embed
00023213	36	37	cor	cor	embed	-	embed
00024120	5	6	cor	cor	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00024120	9	10	adv	cor	shared	semantic	shared
00024120	14	15	cor	cor	pc-rel	syntactic	pc-rel
00024120	16	17	cor	cor	embed	interpret	embed
00024120	16	18	cor	sub	embed	-	embed
00024120	16	19	cor	cor	pc-rel	syntactic	pc-rel
00024120	17	18	cor	sub	embed	-	embed
00024120	18	19	sub	cor	pc-rel	syntactic	pc-rel
00024120	22	23	cor	sub	embed	-	embed
00024120	28	29	sub	sub	pc-rel	interpret	embed
00024120	28	30	sub	cor	pc-rel	syntactic	pc-rel
00024120	28	31	sub	sub	pc-rel	syntactic	pc-rel
00024120	29	30	sub	cor	embed	-	embed
00024120	30	31	cor	sub	embed	-	embed
00024120	32	33	adv	phr	pc-rel	syntactic	pc-rel
00024120	34	35	cor	phr	embed	-	embed
00024120	40	41	cor	cor	pc-rel	missing	embed
00024120	40	42	cor	cor	pc-rel	syntactic	pc-rel
00024120	40	43	cor	phr	shared	semantic	shared
00024120	41	42	cor	cor	pc-rel	missing	embed
00024120	41	43	cor	phr	pc-rel	missing	embed
00024120	42	43	cor	phr	pc-rel	syntactic	pc-rel
00024120	43	44	phr	adv	pc-rel	interpret	embed
00024120	46	47	cor	adv	shared	semantic	shared
00024220	1	2	sub	cor	shared	interpret	embed
00024220	7	8	sub	cor	pc-rel	syntactic	pc-rel
00024220	9	10	cor	adv	embed	-	embed
00024220	10	11	adv	cor	pc-rel	syntactic	pc-rel
00024220	10	12	adv	cor	shared	semantic	shared
00024220	11	12	cor	cor	pc-rel	syntactic	pc-rel
00024220	13	14	cor	sub	embed	-	embed
00024220	16	17	sub	cor	pc-rel	syntactic	pc-rel
00024220	18	19	cor	sub	pc-rel	missing	embed
00024220	21	22	sub	sub	pc-rel	syntactic	pc-rel
00024220	21	23	sub	adv	pc-rel	syntactic	pc-rel
00024220	21	24	sub	phr	pc-rel	syntactic	pc-rel
00024220	22	23	sub	adv	embed	-	embed
00024220	22	24	sub	phr	pc-rel	missing	embed
00024220	23	24	adv	phr	embed	-	embed
00024220	24	25	phr	cor	pc-rel	interpret	embed
00024220	31	32	sub	adv	embed	-	embed
00024220	31	33	sub	adv	pc-rel	missing	embed
00024220	32	33	adv	adv	pc-rel	missing	embed
00024220	37	38	cor	phr	shared	interpret	embed
00024220	46	47	cor	cor	pc-rel	syntactic	pc-rel

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00025120	5	6	cor	adv	shared	multi	ident
00025120	9	10	cor	cor	shared	interpret	embed
00025120	12	13	sub	cor	pc-arg	interpret	embed
00025120	12	14	sub	cor	pc-arg	interpret	embed
00025120	13	14	cor	cor	shared	multi	ident
00025120	13	15	cor	sub	embed	-	embed
00025120	14	15	cor	sub	embed	-	embed
00025120	18	20	cor	phr	nested	-	nested
00025120	19	20	cor	phr	nested	-	nested
00025120	29	30	sub	adv	pc-rel	missing	embed
00025220	9	10	cor	cor	embed	-	embed
00025220	11	12	cor	cor	nested	-	nested
00025220	17	18	cor	phr	embed	-	embed
00025220	22	23	cor	cor	pc-rel	syntactic	pc-rel
00025220	28	29	phr	cor	embed	-	embed
00026131	2	3	cor	phr	pc-rel	syntactic	pc-rel
00026131	3	4	phr	adv	pc-arg	interpret	embed
00026131	4	5	adv	cor	pc-rel	syntactic	pc-rel
00026131	8	9	cor	adv	pc-rel	syntactic	pc-rel
00026131	13	14	adv	cor	pc-rel	syntactic	pc-rel
00026131	15	16	adv	sub	embed	-	embed
00026131	18	19	cor	sub	embed	-	embed
00026131	22	23	cor	phr	shared	interpret	embed
00026131	24	25	cor	adv	shared	multi	ident
00026131	26	27	phr	cor	shared	interpret	embed
00026131	27	28	cor	adv	pc-rel	interpret	embed
00026131	30	31	cor	adv	pc-rel	syntactic	pc-rel
00026131	37	38	sub	cor	shared	interpret	embed
00026131	40	41	cor	cor	pc-rel	syntactic	pc-rel
00026131	42	43	sub	phr	embed	-	embed
00026131	43	44	phr	cor	pc-rel	syntactic	pc-rel
00026131	43	45	phr	sub	pc-rel	syntactic	pc-rel
00026131	44	45	cor	sub	pc-rel	interpret	embed
00026131	46	47	phr	sub	embed	-	embed
00026131	46	48	phr	cor	shared	interpret	embed
00026131	46	49	phr	phr	shared	interpret	embed
00026131	47	48	sub	cor	embed	-	embed
00026131	47	49	sub	phr	embed	-	embed
00026131	48	49	cor	phr	shared	multi	ident
00026131	55	56	cor	cor	shared	interpret	embed
00026131	57	58	cor	phr	shared	multi	ident
00026131	57	59	cor	adv	shared	interpret	embed
00026131	58	59	phr	adv	shared	interpret	embed
00026131	59	60	adv	cor	pc-rel	interpret	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00026131	66	67	cor	phr	shared	interpret	embed
00026131	67	68	phr	cor	embed	leftout	embed
00026131	67	69	phr	cor	embed	leftout	embed
00026131	68	69	cor	cor	embed	-	embed
00026131	72	73	cor	sub	embed	-	embed
00026231	2	3	cor	cor	pc-rel	syntactic	pc-rel
00026231	2	5	cor	phr	pc-rel	syntactic	pc-rel
00026231	3	4	cor	cor	shared	interpret	embed
00026231	3	5	cor	phr	shared	interpret	embed
00026231	4	5	cor	phr	pc-arg	multi	ident
00026231	7	8	cor	cor	pc-rel	syntactic	pc-rel
00026231	7	9	cor	sub	embed	-	embed
00026231	7	10	cor	phr	shared	interpret	embed
00026231	8	9	cor	sub	pc-rel	syntactic	pc-rel
00026231	8	10	cor	phr	pc-rel	syntactic	pc-rel
00026231	9	10	sub	phr	embed	-	embed
00026231	10	11	phr	sub	embed	-	embed
00026231	12	13	sub	cor	embed	leftout	embed
00026231	12	14	sub	phr	embed	-	embed
00026231	13	14	cor	phr	shared	multi	ident
00026231	15	16	cor	adv	shared	multi	ident
00026231	15	17	cor	sub	pc-rel	syntactic	pc-rel
00026231	15	18	cor	phr	shared	interpret	embed
00026231	16	17	adv	sub	pc-rel	syntactic	pc-rel
00026231	16	18	adv	phr	shared	interpret	embed
00026231	17	18	sub	phr	pc-rel	syntactic	pc-rel
00026231	18	20	phr	sub	shared	interpret	embed
00026231	19	20	sub	sub	embed	-	embed
00026231	21	22	adv	adv	shared	interpret	embed
00026231	24	25	cor	cor	embed	-	embed
00026231	24	26	cor	adv	pc-rel	interpret	embed
00026231	25	26	cor	adv	pc-rel	interpret	embed
00026231	26	27	adv	cor	pc-rel	syntactic	pc-rel
00026231	26	29	adv	adv	pc-arg	interpret	indep
00026231	27	29	cor	adv	pc-rel	interpret	indep
00026231	28	29	cor	adv	shared	multi	ident
00026231	30	31	cor	cor	pc-rel	syntactic	pc-rel
00026231	31	32	cor	cor	embed	-	embed
00026231	33	34	cor	adv	shared	interpret	embed
00026231	34	35	adv	sub	embed	-	embed
00026231	37	38	cor	cor	pc-rel	syntactic	pc-rel
00026231	38	39	cor	cor	pc-rel	syntactic	pc-rel
00026231	40	41	cor	sub	embed	-	embed
00026231	44	45	cor	adv	pc-rel	missing	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00026231	45	46	adv	cor	embed	-	embed
00026231	45	47	adv	adv	shared	interpret	embed
00026231	46	47	cor	adv	embed	-	embed
00026231	49	50	adv	sub	embed	-	embed
00026231	51	52	cor	cor	shared	interpret	embed
00026231	55	56	cor	phr	embed	-	embed
00026231	56	57	phr	cor	pc-rel	missing	embed
00026231	58	59	cor	adv	shared	multi	ident
00026231	60	61	phr	phr	shared	interpret	embed
00026231	65	66	cor	cor	shared	interpret	embed
00026231	66	67	cor	cor	pc-rel	syntactic	pc-rel
00026231	66	68	cor	phr	shared	interpret	embed
00026231	67	68	cor	phr	pc-rel	syntactic	pc-rel
00026231	68	69	phr	cor	embed	-	embed
00026231	71	72	cor	cor	pc-rel	syntactic	pc-rel
00026231	72	73	cor	cor	embed	-	embed
00026231	74	75	adv	cor	embed	-	embed
00026231	76	77	cor	cor	shared	interpret	embed
00026231	78	79	cor	cor	embed	-	embed
00026231	78	80	cor	cor	shared	interpret	embed
00026231	78	81	cor	adv	embed	-	embed
00026231	79	80	cor	cor	nested	-	nested
00026231	79	81	cor	adv	pc-rel	interpret	embed
00026231	80	81	cor	adv	pc-arg	multi	ident
00026231	80	82	cor	cor	pc-arg	interpret	embed
00026231	81	82	adv	cor	shared	interpret	embed
00026231	82	83	cor	sub	pc-rel	interpret	embed
00026231	86	87	cor	adv	shared	multi	ident
00026231	86	88	cor	phr	pc-rel	syntactic	pc-rel
00026231	87	88	adv	phr	pc-rel	syntactic	pc-rel
00026231	88	89	phr	cor	embed	-	embed
00027113	1	2	cor	adv	pc-rel	error	embed
00027113	2	3	adv	cor	pc-rel	interpret	embed
00027113	4	5	adv	cor	pc-arg	missing	embed
00027113	15	17	adv	cor	pc-rel	missing	embed
00027113	16	17	cor	cor	nested	-	nested
00027113	18	19	cor	adv	pc-rel	syntactic	pc-rel
00027113	24	25	cor	cor	pc-rel	syntactic	pc-rel
00027113	26	27	cor	adv	pc-rel	missing	embed
00027113	29	30	cor	sub	pc-rel	syntactic	pc-rel
00027113	29	31	cor	cor	pc-rel	syntactic	pc-rel
00027113	30	31	sub	cor	shared	interpret	embed
00027113	33	34	cor	sub	embed	-	embed
00027213	5	6	adv	cor	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00027213	10	11	cor	phr	pc-rel	interpret	embed
00027213	14	15	cor	sub	embed	-	embed
00027213	19	21	sub	sub	pc-rel	missing	embed
00027213	19	22	sub	cor	pc-rel	missing	embed
00027213	20	21	sub	sub	pc-rel	missing	embed
00027213	20	22	sub	cor	pc-rel	missing	embed
00027213	21	22	sub	cor	embed	-	embed
00027213	22	23	cor	cor	pc-rel	syntactic	pc-rel
00027213	28	29	sub	cor	embed	-	embed
00027213	29	30	cor	cor	pc-rel	syntactic	pc-rel
00027213	32	33	cor	sub	pc-rel	missing	embed
00027213	35	36	cor	sub	embed	-	embed
00027213	37	38	cor	phr	embed	-	embed
00027213	38	39	phr	cor	embed	-	embed
00028120	7	8	phr	sub	embed	-	embed
00028220	2	3	phr	adv	pc-rel	missing	embed
00028220	7	8	cor	cor	nested	-	nested
00028220	8	9	cor	cor	pc-rel	syntactic	pc-rel
00028220	13	14	cor	cor	shared	interpret	embed
00028220	15	16	cor	cor	embed	-	embed
00028220	20	21	cor	cor	pc-arg	interpret	embed
00028220	23	24	sub	cor	embed	-	embed
00030130	6	7	adv	cor	shared	multi	ident
00030130	12	13	adv	adv	shared	interpret	embed
00030130	14	15	cor	adv	pc-rel	syntactic	pc-rel
00030130	17	18	cor	cor	pc-rel	syntactic	pc-rel
00030130	20	21	cor	adv	pc-rel	interpret	embed
00030130	20	22	cor	adv	pc-rel	interpret	embed
00030130	20	23	cor	cor	embed	-	embed
00030130	21	22	adv	adv	pc-arg	interpret	embed
00030130	21	23	adv	cor	embed	-	embed
00030130	22	23	adv	cor	pc-rel	missing	embed
00030130	38	39	cor	adv	embed	-	embed
00030224	5	6	sub	sub	nested	-	nested
00030224	5	7	sub	sub	shared	missing	embed
00030224	14	15	cor	adv	pc-arg	interpret	embed
00030224	24	25	sub	adv	pc-rel	missing	embed
00030224	26	27	adv	adv	shared	interpret	embed
00030224	32	33	adv	cor	shared	interpret	embed
00030224	35	36	cor	adv	pc-rel	missing	embed
00030224	39	40	adv	cor	pc-arg	interpret	embed
00032161	1	2	sub	cor	embed	-	embed
00032161	4	5	cor	cor	embed	-	embed
00032161	19	20	cor	cor	pc-rel	missing	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00032161	23	24	cor	sub	pc-rel	missing	embed
00032161	26	27	cor	cor	embed	-	embed
00032161	28	29	cor	cor	shared	interpret	embed
00032161	35	36	cor	cor	embed	leftout	embed
00032161	35	37	cor	sub	pc-rel	syntactic	pc-rel
00032161	36	37	cor	sub	pc-rel	syntactic	pc-rel
00032261	13	14	sub	cor	pc-rel	missing	embed
00032261	17	20	cor	phr	pc-rel	syntactic	pc-rel
00032261	26	27	cor	sub	pc-rel	missing	embed
00032261	26	28	cor	adv	pc-rel	missing	embed
00032261	38	39	cor	cor	pc-rel	syntactic	pc-rel
00032261	40	41	cor	phr	shared	interpret	embed
00032261	52	53	cor	phr	shared	multi	ident
00032261	54	55	cor	sub	embed	-	embed
00033123	6	7	cor	adv	shared	multi	ident
00033123	8	9	cor	adv	embed	-	embed
00033123	13	14	adv	phr	nested	-	nested
00033123	18	19	cor	sub	pc-rel	interpret	embed
00033123	18	20	cor	adv	shared	multi	ident
00033123	19	20	sub	adv	pc-arg	interpret	embed
00033123	22	23	cor	phr	pc-rel	missing	embed
00033123	22	24	cor	adv	pc-arg	missing	embed
00033123	22	25	cor	cor	overlap	interpret	indep
00033123	23	24	phr	adv	shared	interpret	embed
00033123	23	25	phr	cor	pc-arg	interpret	indep
00033123	24	25	adv	cor	pc-arg	interpret	indep
00033123	30	31	adv	cor	nested	-	nested
00033123	36	37	adv	adv	shared	interpret	embed
00033223	8	9	sub	adv	pc-rel	missing	embed
00033223	9	10	adv	sub	embed	-	embed
00033223	11	12	cor	adv	pc-rel	missing	embed
00033223	12	13	adv	adv	pc-arg	interpret	embed
00033223	15	17	cor	cor	nested	-	nested
00033223	16	17	sub	cor	nested	-	nested
00033223	18	19	cor	adv	shared	interpret	indep
00033223	25	26	cor	cor	pc-rel	interpret	indep
00033223	28	29	sub	cor	embed	-	embed
00033223	29	30	cor	sub	pc-rel	syntactic	pc-rel
00033223	39	40	adv	adv	pc-rel	missing	embed
00035120	9	10	sub	adv	shared	multi	ident
00035220	9	10	adv	cor	pc-arg	interpret	embed
00035220	15	16	sub	cor	nested	-	nested
00035220	17	18	cor	adv	shared	multi	ident
00035220	26	27	adv	cor	pc-rel	interpret	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00035220	31	32	phr	sub	embed	-	embed
00045224	2	3	sub	phr	pc-rel	missing	embed
00045224	5	6	sub	phr	nested	-	nested
00045224	16	17	cor	cor	pc-rel	missing	embed
00046124	10	11	sub	adv	pc-arg	interpret	embed
00046124	14	15	adv	sub	embed	-	embed
00046224	3	4	sub	adv	pc-rel	syntactic	pc-rel
00047124	1	2	cor	cor	shared	interpret	embed
00047124	9	10	sub	sub	pc-rel	leftout	embed
00047124	16	17	cor	sub	embed	-	embed
00047124	20	21	cor	sub	pc-rel	syntactic	pc-rel
00047124	22	23	sub	sub	pc-rel	missing	embed
00047124	41	42	sub	cor	embed	-	embed
00047124	46	47	adv	cor	pc-rel	missing	embed
00047124	46	48	adv	adv	pc-rel	missing	embed
00047124	47	48	cor	adv	shared	multi	ident
00047124	47	49	cor	phr	shared	semantic	shared
00047124	48	49	adv	phr	shared	semantic	shared
00047124	51	52	sub	cor	shared	interpret	embed
00047124	56	57	cor	cor	embed	-	embed
00047124	57	58	cor	phr	pc-rel	missing	embed
00047224	6	7	cor	cor	shared	interpret	embed
00047224	6	8	cor	adv	shared	interpret	embed
00047224	7	8	cor	adv	shared	multi	ident
00047224	16	17	cor	cor	pc-rel	syntactic	pc-rel
00047224	16	18	cor	sub	pc-rel	syntactic	pc-rel
00047224	17	18	cor	sub	pc-rel	syntactic	pc-rel
00047224	19	20	cor	cor	embed	-	embed
00047224	21	22	sub	cor	pc-rel	leftout	embed
00047224	22	23	cor	sub	embed	-	embed
00047224	25	26	cor	cor	embed	-	embed
00047224	35	36	sub	cor	pc-arg	interpret	shared
00047224	35	38	sub	cor	pc-arg	interpret	embed
00047224	36	37	cor	sub	pc-rel	missing	embed
00047224	36	38	cor	cor	shared	interpret	embed
00047224	37	38	sub	cor	pc-rel	missing	embed
00047224	39	40	cor	sub	pc-rel	missing	embed
00047224	39	41	cor	adv	pc-rel	missing	embed
00047224	40	41	sub	adv	embed	-	embed
00047224	44	45	cor	adv	shared	leftout	embed
00047224	50	51	sub	sub	embed	-	embed
00047224	51	52	sub	sub	pc-rel	interpret	embed
00047224	59	60	sub	cor	embed	-	embed
00047224	61	62	phr	cor	pc-rel	syntactic	pc-rel

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00047224	64	65	sub	cor	nested	-	nested
00047224	66	67	cor	sub	embed	-	embed
00047224	68	69	cor	cor	pc-rel	syntactic	pc-rel
00047224	69	70	cor	cor	embed	-	embed
00047224	72	73	sub	adv	embed	-	embed
00047224	74	75	cor	sub	embed	-	embed
00048120	1	2	adv	cor	pc-arg	multi	ident
00048120	3	4	sub	cor	nested	-	nested
00048120	15	16	cor	phr	pc-rel	missing	embed
00048120	25	27	cor	phr	pc-arg	interpret	indep
00048120	26	27	cor	phr	shared	multi	ident
00048120	26	28	cor	cor	shared	semantic	shared
00048120	27	28	phr	cor	shared	semantic	shared
00048120	29	30	sub	cor	pc-rel	syntactic	pc-rel
00048120	34	35	cor	adv	shared	interpret	embed
00048220	2	4	sub	adv	pc-rel	interpret	embed
00048220	3	4	sub	adv	pc-rel	syntactic	pc-rel
00048220	6	7	adv	sub	embed	-	embed
00048220	6	8	adv	cor	shared	interpret	embed
00048220	6	10	adv	adv	shared	missing	embed
00048220	7	8	sub	cor	embed	-	embed
00048220	7	10	sub	adv	nested	-	nested
00048220	8	10	cor	adv	nested	-	nested
00048220	9	10	sub	adv	nested	-	nested
00048220	10	11	adv	sub	embed	-	embed
00048220	14	15	adv	sub	pc-rel	missing	embed
00048220	34	35	cor	cor	pc-rel	syntactic	pc-rel
00048220	34	36	cor	adv	pc-rel	missing	embed
00048220	35	36	cor	adv	pc-rel	syntactic	pc-rel
00048220	40	41	sub	cor	shared	semantic	shared
00048220	45	46	cor	adv	shared	multi	ident
00048220	45	47	cor	cor	embed	-	embed
00048220	46	47	adv	cor	embed	-	embed
00048220	48	49	adv	adv	pc-arg	interpret	embed
00048220	49	50	adv	cor	shared	interpret	embed
00048220	50	51	cor	adv	shared	interpret	embed
00048220	54	55	cor	sub	embed	-	embed
00048220	57	58	sub	cor	embed	-	embed
00048220	61	62	cor	cor	shared	interpret	embed
00048220	64	65	sub	cor	pc-rel	interpret	embed
00050120	2	3	cor	adv	shared	multi	ident
00050120	6	7	cor	phr	embed	-	embed
00050120	9	10	cor	cor	shared	interpret	embed
00050120	10	11	cor	phr	pc-arg	syntactic	pc-rel

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00050120	22	23	sub	cor	nested	-	nested
00050120	26	27	sub	adv	pc-rel	missing	embed
00050120	28	29	cor	adv	shared	interpret	embed
00050120	35	36	phr	cor	pc-arg	interpret	embed
00050120	43	44	sub	cor	embed	-	embed
00050220	11	12	cor	adv	shared	multi	ident
00050220	16	17	cor	sub	embed	-	embed
00050220	19	20	cor	cor	shared	interpret	embed
00050220	27	28	sub	sub	pc-rel	missing	embed
00050220	27	29	sub	sub	pc-rel	missing	embed
00050220	27	30	sub	phr	pc-rel	missing	embed
00050220	29	30	sub	phr	embed	-	embed
00050220	34	35	cor	cor	pc-rel	syntactic	pc-rel
00050220	47	48	adv	sub	pc-rel	missing	embed
00050220	47	49	adv	cor	shared	interpret	embed
00050220	47	50	adv	adv	shared	interpret	embed
00050220	48	49	sub	cor	pc-rel	missing	embed
00050220	48	50	sub	adv	pc-rel	missing	embed
00050220	49	50	cor	adv	shared	multi	ident
00051120	3	4	cor	cor	embed	-	embed
00051120	20	21	cor	adv	shared	multi	ident
00051120	24	25	sub	cor	pc-rel	missing	embed
00051120	34	35	sub	cor	pc-rel	missing	embed
00053123	1	2	sub	sub	embed	-	embed
00053123	5	6	sub	cor	pc-rel	syntactic	pc-rel
00053123	10	11	sub	adv	embed	leftout	embed
00053123	10	12	sub	adv	pc-arg	missing	embed
00053123	31	32	cor	adv	shared	multi	ident
00053123	35	36	sub	cor	embed	-	embed
00053123	40	41	adv	cor	pc-rel	syntactic	pc-rel
00053223	1	2	cor	adv	pc-arg	multi	ident
00053223	14	15	cor	sub	pc-rel	syntactic	pc-rel
00053223	16	17	cor	cor	shared	missing	embed
00053223	19	20	phr	phr	shared	error	ident
00053223	22	23	sub	adv	pc-rel	missing	embed
00053223	28	29	phr	phr	shared	missing	embed
00053223	28	30	phr	phr	shared	missing	embed
00053223	29	30	phr	phr	shared	missing	embed
00053223	32	34	sub	adv	pc-rel	missing	embed
00053223	33	34	cor	adv	nested	-	nested
00053223	36	37	sub	cor	pc-rel	leftout	embed
00053223	37	38	cor	sub	pc-rel	missing	embed
00053223	37	39	cor	phr	embed	-	embed
00053223	38	39	sub	phr	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00053223	40	43	phr	sub	pc-arg	semantic	shared
00053223	41	42	cor	phr	pc-arg	multi	ident
00053223	41	43	cor	sub	nested	-	nested
00053223	42	43	phr	sub	nested	-	nested
00054123	13	14	cor	sub	embed	-	embed
00054123	17	19	sub	adv	pc-rel	syntactic	pc-rel
00054123	18	19	cor	adv	pc-rel	syntactic	pc-rel
00054123	18	21	cor	adv	nested	-	nested
00054123	19	20	adv	cor	pc-arg	syntactic	pc-arg
00054123	19	21	adv	adv	nested	-	nested
00054123	20	21	cor	adv	nested	-	nested
00054123	21	22	adv	cor	shared	semantic	shared
00054123	24	25	cor	cor	pc-rel	syntactic	pc-rel
00054123	27	28	sub	cor	embed	interpret	embed
00054123	27	29	sub	sub	embed	-	embed
00054123	28	29	cor	sub	embed	-	embed
00054123	32	33	cor	sub	pc-rel	missing	embed
00054123	38	39	cor	sub	shared	interpret	embed
00054123	42	43	adv	adv	pc-rel	syntactic	pc-rel
00054123	43	44	adv	cor	embed	-	embed
00054123	43	45	adv	adv	embed	-	embed
00054123	44	45	cor	adv	shared	leftout	embed
00054123	50	52	cor	phr	pc-rel	interpret	embed
00054123	51	52	sub	phr	nested	-	nested
00054223	5	6	cor	adv	shared	multi	ident
00054223	5	7	cor	adv	shared	semantic	shared
00054223	6	7	adv	adv	shared	semantic	shared
00054223	8	9	cor	phr	pc-rel	syntactic	pc-rel
00054223	10	11	cor	phr	shared	multi	ident
00054223	15	16	cor	cor	nested	-	nested
00054223	29	30	sub	phr	pc-rel	missing	embed
00054223	38	39	adv	cor	pc-rel	missing	embed
00054223	38	40	adv	adv	pc-rel	missing	embed
00054223	38	41	adv	cor	pc-rel	missing	embed
00054223	39	40	cor	adv	pc-rel	interpret	embed
00054223	39	41	cor	cor	pc-rel	interpret	embed
00054223	40	41	adv	cor	embed	-	embed
00054223	46	47	adv	adv	embed	-	embed
00054223	50	51	adv	adv	nested	-	nested
00054223	53	54	cor	cor	pc-rel	missing	embed
00054223	57	58	sub	cor	pc-rel	interpret	embed
00054223	57	59	sub	adv	pc-rel	interpret	embed
00054223	58	59	cor	adv	embed	-	embed
00054223	61	62	cor	sub	pc-rel	missing	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00055121	1	2	cor	adv	shared	multi	ident
00055121	1	3	cor	adv	shared	multi	ident
00055121	2	3	adv	adv	shared	semantic	shared
00055121	7	8	sub	cor	pc-rel	missing	embed
00055121	27	28	cor	sub	embed	-	embed
00055121	27	29	cor	adv	pc-rel	missing	embed
00055121	28	29	sub	adv	pc-rel	missing	embed
00055121	29	30	adv	cor	embed	interpret	embed
00055121	29	31	adv	sub	embed	-	embed
00055121	30	31	cor	sub	embed	-	embed
00055221	3	4	sub	adv	embed	-	embed
00057121	7	8	cor	sub	embed	-	embed
00057121	10	11	adv	cor	pc-rel	missing	embed
00057121	12	13	cor	cor	embed	-	embed
00057121	12	14	cor	adv	embed	-	embed
00057121	13	14	cor	adv	shared	multi	ident
00057121	26	27	cor	adv	shared	multi	ident
00057121	26	28	cor	cor	pc-rel	syntactic	pc-rel
00057121	27	28	adv	cor	pc-rel	syntactic	pc-rel
00057121	30	31	cor	cor	embed	-	embed
00057121	30	32	cor	adv	pc-rel	syntactic	pc-rel
00057121	31	32	cor	adv	pc-rel	syntactic	pc-rel
00057121	32	33	adv	sub	embed	-	embed
00057121	39	41	sub	cor	embed	-	embed
00057121	40	41	sub	cor	nested	-	nested
00057121	44	45	cor	cor	pc-rel	missing	embed
00057121	46	47	cor	sub	embed	-	embed
00057221	3	4	cor	cor	embed	-	embed
00057221	5	6	adv	cor	embed	-	embed
00057221	8	9	cor	cor	embed	-	embed
00057221	14	15	sub	cor	embed	-	embed
00057221	17	18	cor	adv	shared	multi	ident
00057221	19	20	cor	cor	nested	-	nested
00057221	23	24	cor	adv	shared	multi	ident
00057221	25	26	adv	adv	shared	error	ident
00057221	27	28	cor	cor	shared	semantic	shared
00057221	36	37	adv	sub	pc-arg	interpret	embed
00057221	36	38	adv	cor	pc-arg	interpret	embed
00057221	36	39	adv	cor	pc-arg	interpret	embed
00057221	37	38	sub	cor	embed	-	embed
00057221	37	39	sub	cor	pc-rel	interpret	embed
00057221	38	39	cor	cor	embed	-	embed
00057221	46	47	cor	sub	pc-rel	missing	embed
00057221	48	49	cor	adv	shared	multi	ident

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00057221	48	50	cor	cor	shared	interpret	embed
00057221	49	50	adv	cor	shared	interpret	embed
00057221	52	53	cor	cor	shared	interpret	embed
00057221	54	55	cor	sub	embed	-	embed
00057221	54	56	cor	cor	embed	interpret	embed
00057221	55	56	sub	cor	embed	-	embed
00057221	57	58	sub	cor	embed	-	embed
00057221	57	59	sub	cor	pc-rel	syntactic	pc-rel
00057221	58	59	cor	cor	pc-rel	syntactic	pc-rel
00057221	60	61	cor	adv	shared	multi	ident
00057221	60	62	cor	cor	pc-arg	interpret	indep
00057221	60	63	cor	cor	shared	interpret	indep
00057221	60	64	cor	adv	shared	interpret	indep
00057221	61	62	adv	cor	pc-arg	interpret	indep
00057221	61	63	adv	cor	shared	interpret	indep
00057221	61	64	adv	adv	shared	interpret	indep
00057221	62	63	cor	cor	nested	-	nested
00057221	62	64	cor	adv	nested	-	nested
00057221	63	64	cor	adv	shared	multi	ident
00057221	67	68	sub	cor	pc-rel	syntactic	pc-rel
00057221	69	70	cor	cor	pc-rel	syntactic	pc-rel
00057221	75	76	cor	cor	embed	-	embed
00057221	80	81	cor	cor	nested	-	nested
00058111	1	2	sub	adv	pc-rel	syntactic	pc-rel
00058111	13	14	sub	phr	pc-rel	leftout	embed
00058111	24	25	cor	adv	shared	interpret	embed
00058111	25	26	adv	cor	pc-rel	syntactic	pc-rel
00058111	30	31	adv	cor	shared	interpret	embed
00058111	33	34	cor	adv	shared	interpret	embed
00058111	37	38	cor	adv	embed	-	embed
00058111	40	41	cor	adv	shared	semantic	shared
00058111	46	47	adv	adv	shared	missing	embed
00058111	49	50	phr	cor	shared	interpret	embed
00058111	50	51	cor	cor	shared	semantic	shared
00058211	6	7	adv	cor	shared	semantic	shared
00058211	22	23	sub	adv	embed	-	embed
00058211	38	39	sub	adv	embed	-	embed
00059131	3	4	cor	cor	shared	interpret	embed
00059131	3	5	cor	adv	shared	interpret	embed
00059131	4	5	cor	adv	shared	multi	ident
00059131	4	6	cor	sub	embed	-	embed
00059131	4	7	cor	adv	shared	interpret	embed
00059131	5	6	adv	sub	embed	-	embed
00059131	5	7	adv	adv	shared	interpret	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00059131	6	7	sub	adv	embed	-	embed
00059131	8	9	cor	cor	shared	interpret	embed
00059131	11	12	cor	sub	pc-rel	missing	embed
00059131	16	17	cor	sub	shared	interpret	embed
00059131	20	21	cor	cor	shared	interpret	embed
00059131	22	23	cor	cor	pc-arg	interpret	embed
00059131	26	27	cor	phr	shared	multi	ident
00059131	33	34	sub	cor	pc-rel	missing	embed
00059131	33	35	sub	cor	embed	-	embed
00059131	34	35	cor	cor	pc-rel	missing	embed
00059131	37	38	cor	cor	pc-rel	missing	embed
00059131	39	40	sub	adv	pc-rel	syntactic	pc-rel
00059131	39	43	sub	adv	pc-rel	syntactic	pc-rel
00059131	40	41	adv	cor	shared	interpret	embed
00059131	40	43	adv	adv	pc-rel	interpret	embed
00059131	41	42	cor	adv	shared	interpret	embed
00059131	41	43	cor	adv	pc-arg	interpret	embed
00059131	42	43	adv	adv	nested	-	nested
00059131	44	45	sub	cor	pc-rel	missing	embed
00059131	55	56	cor	sub	embed	-	embed
00059131	57	58	cor	adv	pc-rel	interpret	embed
00059131	57	61	cor	cor	overlap	interpret	indep
00059131	59	61	sub	cor	pc-rel	interpret	indep
00059131	60	61	cor	cor	pc-rel	interpret	shared
00059131	63	64	sub	cor	pc-rel	missing	embed
00059131	65	66	cor	adv	pc-arg	semantic	shared
00059131	66	67	adv	sub	pc-rel	leftout	embed
00059131	66	68	adv	cor	pc-arg	interpret	embed
00059131	67	68	sub	cor	pc-rel	interpret	embed
00059131	71	72	cor	cor	pc-rel	missing	embed
00059131	71	73	cor	cor	pc-rel	syntactic	pc-rel
00059131	72	73	cor	cor	pc-rel	syntactic	pc-rel
00059131	78	79	cor	sub	pc-rel	missing	embed
00059131	84	85	cor	sub	embed	-	embed
00059131	89	90	cor	sub	embed	-	embed
00059131	89	91	cor	cor	shared	semantic	shared
00059131	90	91	sub	cor	embed	-	embed
00059131	91	92	cor	sub	embed	-	embed
00059131	93	94	cor	phr	embed	-	embed
00059131	96	97	sub	cor	pc-rel	missing	embed
00059232	1	2	adv	sub	embed	-	embed
00059232	4	5	cor	cor	embed	-	embed
00059232	5	6	cor	adv	shared	semantic	shared
00059232	6	7	adv	cor	pc-rel	syntactic	pc-rel

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00059232	8	9	cor	phr	embed	-	embed
00059232	9	10	phr	cor	shared	interpret	embed
00059232	11	12	phr	cor	pc-arg	interpret	embed
00059232	15	16	cor	adv	embed	-	embed
00059232	16	17	adv	cor	embed	-	embed
00059232	18	19	cor	cor	embed	-	embed
00059232	20	21	cor	cor	shared	interpret	embed
00059232	20	22	cor	adv	pc-rel	interpret	embed
00059232	21	22	cor	adv	pc-rel	interpret	embed
00059232	22	23	adv	cor	pc-rel	syntactic	pc-rel
00059232	24	25	cor	cor	pc-rel	syntactic	pc-rel
00059232	26	27	cor	cor	pc-arg	leftout	shared
00059232	27	28	cor	cor	shared	interpret	embed
00059232	28	29	cor	sub	embed	-	embed
00059232	31	32	cor	cor	pc-rel	syntactic	pc-rel
00059232	35	36	cor	cor	pc-rel	syntactic	pc-rel
00059232	37	38	cor	cor	pc-rel	syntactic	pc-rel
00059232	37	39	cor	cor	pc-rel	interpret	embed
00059232	38	39	cor	cor	pc-rel	syntactic	pc-rel
00059232	39	40	cor	cor	pc-rel	missing	embed
00059232	39	41	cor	cor	embed	interpret	embed
00059232	40	41	cor	cor	embed	-	embed
00059232	43	44	cor	cor	pc-rel	syntactic	pc-rel
00059232	43	45	cor	sub	pc-rel	syntactic	pc-rel
00059232	43	46	cor	adv	pc-arg	interpret	embed
00059232	44	45	cor	sub	pc-rel	missing	embed
00059232	44	46	cor	adv	pc-rel	syntactic	pc-rel
00059232	45	46	sub	adv	pc-rel	syntactic	pc-rel
00059232	47	48	sub	cor	embed	-	embed
00059232	50	51	cor	cor	embed	-	embed
00059232	53	54	cor	cor	shared	interpret	embed
00059232	56	57	cor	cor	pc-rel	syntactic	pc-rel
00059232	58	59	cor	cor	shared	interpret	embed
00059232	59	60	cor	cor	shared	interpret	embed
00059232	60	61	cor	adv	shared	interpret	embed
00059232	61	62	adv	cor	shared	interpret	embed
00059232	62	63	cor	cor	shared	interpret	embed
00059232	62	65	cor	adv	pc-rel	missing	embed
00059232	63	64	cor	cor	shared	interpret	embed
00059232	63	65	cor	adv	pc-rel	missing	embed
00059232	64	65	cor	adv	pc-rel	missing	embed
00059232	65	66	adv	cor	pc-rel	syntactic	pc-rel
00059232	65	67	adv	cor	pc-arg	interpret	embed
00059232	65	68	adv	adv	pc-arg	interpret	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00059232	65	69	adv	adv	pc-rel	missing	embed
00059232	65	70	adv	cor	pc-rel	missing	embed
00059232	66	67	cor	cor	pc-rel	syntactic	pc-rel
00059232	66	68	cor	adv	pc-rel	syntactic	pc-rel
00059232	67	68	cor	adv	shared	multi	ident
00059232	67	69	cor	adv	pc-rel	missing	embed
00059232	67	70	cor	cor	pc-rel	missing	embed
00059232	68	69	adv	adv	pc-rel	missing	embed
00059232	68	70	adv	cor	pc-rel	missing	embed
00059232	69	70	adv	cor	shared	interpret	embed
00059232	74	75	sub	cor	embed	-	embed
00059232	75	76	cor	cor	pc-rel	syntactic	pc-rel
00059232	81	82	cor	cor	embed	-	embed
00059232	82	83	cor	adv	pc-rel	interpret	embed
00059232	84	85	cor	sub	embed	interpret	embed
00059232	84	86	cor	sub	embed	interpret	embed
00059232	84	87	cor	cor	pc-rel	syntactic	pc-rel
00059232	84	88	cor	cor	pc-arg	interpret	embed
00059232	85	86	sub	sub	shared	missing	embed
00059232	85	87	sub	cor	pc-rel	syntactic	pc-rel
00059232	85	88	sub	cor	overlap	syntactic	pc-rel
00059232	86	87	sub	cor	pc-rel	syntactic	pc-rel
00059232	86	88	sub	cor	overlap	interpret	embed
00059232	89	90	cor	cor	shared	interpret	embed
00059232	92	93	cor	cor	shared	interpret	embed
00059232	95	96	cor	cor	pc-rel	syntactic	pc-rel
00059232	100	101	cor	cor	pc-arg	interpret	embed
00059232	100	102	cor	cor	overlap	interpret	embed
00059232	100	103	cor	cor	pc-arg	interpret	embed
00059232	101	102	cor	cor	pc-rel	interpret	embed
00059232	101	103	cor	cor	pc-rel	interpret	embed
00059232	102	103	cor	cor	embed	-	embed
00059232	103	104	cor	cor	shared	interpret	embed
00059232	106	107	cor	adv	shared	multi	ident
00059232	106	108	cor	cor	pc-arg	interpret	embed
00059232	107	108	adv	cor	pc-arg	interpret	embed
00060111	1	2	sub	cor	nested	-	nested
00060111	3	4	cor	phr	pc-arg	syntactic	pc-arg
00060111	10	11	cor	phr	shared	multi	ident
00060111	17	18	cor	cor	pc-rel	syntactic	pc-rel
00060111	21	22	cor	cor	pc-rel	syntactic	pc-rel
00060111	21	23	cor	sub	pc-rel	syntactic	pc-rel
00060111	22	23	cor	sub	embed	-	embed
00060111	25	26	cor	cor	pc-rel	missing	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00060111	26	27	cor	sub	pc-rel	missing	embed
00060111	26	28	cor	adv	pc-rel	missing	embed
00060111	27	28	sub	adv	embed	-	embed
00060111	27	29	sub	cor	pc-rel	missing	embed
00060111	27	30	sub	adv	pc-rel	missing	embed
00060111	28	29	adv	cor	pc-rel	missing	embed
00060111	28	30	adv	adv	pc-rel	missing	embed
00060111	29	30	cor	adv	shared	multi	ident
00060111	31	32	cor	cor	pc-rel	missing	embed
00060111	35	36	adv	sub	pc-rel	missing	embed
00060111	42	43	cor	sub	embed	-	embed
00060111	42	44	cor	cor	pc-rel	syntactic	pc-rel
00060111	42	45	cor	cor	shared	interpret	embed
00060111	43	44	sub	cor	pc-rel	syntactic	pc-rel
00060111	43	45	sub	cor	embed	-	embed
00060111	44	45	cor	cor	pc-rel	syntactic	pc-rel
00060111	46	49	sub	phr	pc-rel	missing	embed
00060111	47	49	cor	phr	pc-rel	missing	embed
00060111	48	49	sub	phr	pc-rel	missing	embed
00060111	60	61	cor	adv	nested	-	nested
00060111	61	62	adv	sub	embed	interpret	embed
00060111	61	63	adv	cor	embed	-	embed
00060111	62	63	sub	cor	embed	-	embed
00060111	66	67	sub	adv	nested	-	nested
00060211	4	5	cor	adv	embed	-	embed
00060211	8	9	sub	cor	pc-rel	missing	embed
00060211	16	17	sub	sub	pc-arg	interpret	embed
00060211	21	22	adv	cor	pc-rel	syntactic	pc-rel
00060211	25	26	cor	cor	shared	interpret	embed
00062111	3	4	adv	adv	shared	interpret	embed
00062111	12	13	sub	cor	shared	interpret	embed
00062111	16	17	sub	sub	pc-rel	missing	embed
00062111	23	24	phr	cor	shared	interpret	nested
00062111	31	32	cor	adv	pc-rel	interpret	embed
00062111	34	35	cor	cor	embed	-	embed
00062111	38	39	cor	adv	nested	-	nested
00062111	39	40	adv	cor	embed	-	embed
00062211	12	13	cor	cor	embed	-	embed
00062211	22	23	adv	cor	shared	interpret	embed
00062211	25	26	cor	cor	pc-rel	syntactic	pc-rel
00062211	35	36	sub	cor	pc-rel	missing	embed
00062211	37	38	sub	sub	pc-arg	interpret	embed
00062211	39	40	cor	cor	embed	-	embed
00062211	40	41	cor	phr	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00062211	43	44	cor	adv	shared	multi	ident
00062211	53	54	sub	cor	embed	-	embed
00063160	3	4	cor	sub	embed	-	embed
00063160	9	10	sub	cor	shared	interpret	embed
00063160	12	13	cor	sub	pc-rel	syntactic	pc-rel
00063160	18	19	cor	cor	embed	-	embed
00063160	18	20	cor	cor	pc-arg	interpret	embed
00063160	19	20	cor	cor	embed	-	embed
00063260	1	2	adv	cor	embed	-	embed
00063260	13	14	cor	cor	pc-rel	missing	embed
00063260	15	16	sub	cor	pc-rel	missing	embed
00064111	3	6	sub	adv	pc-rel	interpret	embed
00064111	4	5	cor	cor	pc-rel	interpret	embed
00064111	4	6	cor	adv	nested	-	nested
00064111	5	6	cor	adv	nested	-	nested
00064111	7	8	adv	sub	pc-rel	syntactic	pc-rel
00064111	10	11	cor	cor	shared	interpret	embed
00064111	10	12	cor	cor	pc-rel	missing	embed
00064111	11	12	cor	cor	pc-rel	missing	embed
00064111	12	13	cor	sub	embed	-	embed
00064111	15	16	phr	adv	shared	interpret	embed
00064111	16	17	adv	sub	embed	-	embed
00064111	20	21	adv	cor	embed	-	embed
00064111	26	27	adv	adv	shared	semantic	shared
00064111	32	33	cor	cor	shared	interpret	embed
00064111	40	41	cor	cor	embed	-	embed
00064111	49	50	cor	sub	embed	-	embed
00064211	1	2	cor	adv	embed	-	embed
00064211	3	4	cor	adv	shared	multi	ident
00064211	6	7	adv	sub	embed	-	embed
00064211	10	11	cor	cor	shared	interpret	embed
00064211	12	13	cor	sub	embed	-	embed
00064211	14	15	adv	phr	pc-arg	interpret	embed
00064211	14	16	adv	cor	pc-arg	interpret	embed
00064211	14	17	adv	adv	pc-arg	interpret	shared
00064211	15	16	phr	cor	embed	-	embed
00064211	17	18	adv	sub	pc-rel	missing	embed
00064211	17	19	adv	adv	shared	interpret	embed
00064211	18	19	sub	adv	pc-rel	missing	embed
00064211	19	20	adv	adv	shared	interpret	embed
00064211	24	25	sub	cor	pc-rel	syntactic	pc-rel
00064211	29	30	adv	sub	embed	-	embed
00064211	34	35	sub	adv	embed	-	embed
00064211	36	37	sub	sub	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00064211	36	38	sub	cor	embed	-	embed
00064211	37	38	sub	cor	embed	interpret	embed
00064211	38	39	cor	cor	pc-rel	missing	embed
00064211	42	43	adv	adv	pc-rel	syntactic	pc-rel
00064211	45	46	sub	adv	nested	-	nested
00064211	46	47	adv	adv	pc-arg	interpret	embed
00064211	53	54	sub	cor	pc-rel	missing	embed
00064211	56	57	cor	adv	embed	-	embed
00064211	57	58	adv	adv	shared	interpret	embed
00064211	57	59	adv	adv	shared	interpret	embed
00064211	58	59	adv	adv	shared	interpret	embed
00065111	4	5	sub	sub	embed	-	embed
00065111	11	12	sub	cor	embed	-	embed
00065111	19	20	sub	cor	pc-rel	leftout	embed
00065111	20	21	cor	sub	embed	-	embed
00065111	25	26	sub	sub	embed	-	embed
00065111	34	35	cor	cor	shared	interpret	embed
00065111	38	39	sub	cor	embed	-	embed
00065111	40	41	adv	sub	embed	-	embed
00068131	8	9	cor	sub	pc-rel	missing	embed
00068131	27	28	cor	adv	shared	interpret	embed
00068131	29	30	cor	sub	pc-rel	syntactic	pc-rel
00068131	36	37	sub	phr	pc-rel	missing	embed
00068231	2	3	cor	adv	nested	-	nested
00068231	7	8	sub	adv	embed	-	embed
00068231	10	11	cor	cor	embed	-	embed
00068231	17	18	cor	sub	pc-rel	syntactic	pc-rel
00068231	20	23	adv	cor	nested	-	nested
00068231	21	22	cor	adv	pc-rel	syntactic	pc-rel
00068231	21	23	cor	cor	nested	-	nested
00068231	22	23	adv	cor	nested	-	nested
00068231	27	28	cor	sub	pc-rel	missing	embed
00068231	29	30	sub	cor	embed	-	embed
00068231	30	31	cor	adv	shared	interpret	embed
00068231	31	32	adv	cor	pc-rel	syntactic	pc-rel
00068231	31	33	adv	adv	shared	semantic	shared
00068231	32	33	cor	adv	pc-rel	syntactic	pc-rel
00068231	36	37	adv	sub	embed	-	embed
00068231	42	43	sub	adv	embed	-	embed
00075133	2	3	cor	adv	shared	interpret	embed
00075133	6	7	adv	cor	shared	semantic	shared
00075133	16	17	sub	adv	embed	-	embed
00075133	18	19	phr	sub	pc-rel	missing	embed
00075133	18	20	phr	sub	pc-rel	missing	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00075133	19	20	sub	sub	shared	interpret	embed
00075133	21	22	cor	phr	shared	multi	ident
00075133	23	24	cor	adv	shared	multi	ident
00075133	38	39	sub	adv	pc-arg	interpret	embed
00075133	45	46	cor	sub	pc-rel	syntactic	pc-rel
00075133	48	49	cor	sub	pc-rel	syntactic	pc-rel
00075233	10	11	adv	adv	embed	-	embed
00075233	11	12	adv	sub	embed	-	embed
00075233	22	23	cor	adv	shared	multi	ident
00075233	26	28	adv	adv	pc-rel	syntactic	pc-rel
00075233	27	28	cor	adv	pc-rel	syntactic	pc-rel
00075233	36	37	cor	adv	pc-arg	interpret	shared
00075233	39	40	sub	cor	pc-rel	missing	embed
00075233	43	44	sub	cor	embed	-	embed
00075233	44	45	cor	cor	shared	interpret	embed
00075233	46	47	cor	cor	shared	interpret	embed
00075233	47	48	cor	cor	shared	semantic	shared
00075233	50	51	sub	cor	pc-rel	syntactic	pc-rel
00075233	50	52	sub	cor	pc-rel	syntactic	pc-rel
00075233	51	52	cor	cor	pc-rel	syntactic	pc-rel
00077111	12	13	sub	adv	pc-rel	syntactic	pc-rel
00077111	23	24	cor	cor	pc-arg	interpret	shared
00077111	27	28	sub	cor	pc-rel	syntactic	pc-rel
00077211	1	2	cor	sub	pc-rel	syntactic	pc-rel
00077211	5	6	cor	adv	shared	interpret	embed
00077211	11	12	sub	cor	pc-rel	syntactic	pc-rel
00077211	19	20	sub	sub	pc-rel	missing	embed
00077211	20	21	sub	sub	pc-rel	syntactic	pc-rel
00077211	22	23	adv	cor	embed	-	embed
00077211	38	39	sub	sub	shared	missing	embed
00077211	38	40	sub	cor	embed	-	embed
00077211	39	40	sub	cor	embed	-	embed
00077211	42	43	sub	adv	embed	-	embed
00077211	46	47	cor	sub	pc-rel	syntactic	pc-rel
00077211	48	49	cor	sub	pc-rel	interpret	embed
00095133	1	2	cor	cor	embed	-	embed
00095133	7	10	sub	cor	pc-rel	interpret	embed
00095133	8	9	cor	cor	pc-rel	syntactic	pc-rel
00095133	8	10	cor	cor	pc-rel	syntactic	pc-rel
00095133	9	10	cor	cor	embed	-	embed
00095133	16	17	sub	cor	pc-rel	missing	embed
00095133	21	22	sub	adv	pc-rel	missing	embed
00095133	22	23	adv	cor	pc-rel	missing	embed
00095133	22	24	adv	sub	pc-rel	missing	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
00095133	22	25	adv	cor	embed	-	embed
00095133	23	24	cor	sub	embed	-	embed
00095133	23	25	cor	cor	embed	interpret	embed
00095133	24	25	sub	cor	embed	-	embed
00095133	26	27	sub	sub	pc-arg	missing	embed
00095133	30	31	sub	cor	embed	-	embed
00095133	34	35	sub	cor	pc-rel	interpret	embed
00095133	38	39	cor	phr	embed	-	embed
00095133	40	41	sub	sub	embed	-	embed
00095133	46	47	sub	cor	embed	-	embed
00095133	48	51	cor	adv	shared	interpret	embed
00095133	49	51	sub	adv	nested	-	nested
00095133	50	51	cor	adv	nested	-	nested
00095133	54	57	adv	adv	shared	interpret	embed
00095133	55	56	cor	cor	shared	interpret	embed
00095133	55	57	cor	adv	nested	-	nested
00095133	56	57	cor	adv	nested	-	nested
00095133	58	59	sub	cor	pc-rel	syntactic	pc-rel
00095133	64	65	cor	adv	pc-rel	missing	embed
00095133	65	66	adv	cor	embed	-	embed
00095133	69	70	cor	cor	pc-rel	syntactic	pc-rel
00095133	74	75	cor	sub	embed	-	embed
00199170	16	17	sub	adv	embed	-	embed
00199170	16	18	sub	cor	pc-arg	interpret	embed
00199170	17	18	adv	cor	overlap	interpret	embed
00199170	30	31	sub	adv	nested	-	nested
10010000	2	3	phr	cor	embed	-	embed
10010000	8	11	adv	adv	nested	-	nested
10010000	9	11	adv	adv	nested	-	nested
10010000	10	11	sub	adv	nested	-	nested
10010000	15	16	phr	adv	shared	interpret	embed
10010000	16	17	adv	cor	pc-rel	syntactic	pc-rel
10010000	25	26	cor	sub	embed	-	embed
10010000	30	31	cor	cor	nested	-	nested
10010000	31	32	cor	sub	embed	-	embed
10010000	34	35	cor	cor	pc-rel	missing	embed
10010000	37	38	adv	adv	shared	interpret	embed
10020000	12	13	sub	phr	embed	-	embed
10020000	12	14	sub	phr	embed	-	embed
10020000	13	14	phr	phr	shared	semantic	shared
10020000	17	18	cor	adv	pc-rel	syntactic	pc-rel
10020000	20	21	cor	phr	shared	multi	ident
10020000	24	25	cor	adv	embed	-	embed
10020000	25	26	adv	cor	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10020000	30	31	sub	sub	pc-rel	syntactic	pc-rel
10030000	4	5	sub	sub	shared	interpret	embed
10030000	13	14	cor	sub	embed	-	embed
10030000	24	25	sub	cor	pc-rel	syntactic	pc-rel
10030000	38	39	sub	phr	embed	-	embed
10040000	2	3	adv	cor	pc-arg	interpret	embed
10040000	8	9	cor	cor	pc-rel	syntactic	pc-rel
10040000	9	10	cor	sub	embed	-	embed
10040000	12	13	cor	cor	pc-rel	syntactic	pc-rel
10040000	16	17	cor	adv	shared	multi	ident
10040000	18	19	cor	cor	pc-rel	leftout	embed
10040000	19	20	cor	cor	pc-arg	interpret	embed
10050000	5	6	cor	sub	embed	-	embed
10050000	8	9	sub	cor	embed	-	embed
10050000	9	10	cor	sub	pc-rel	syntactic	pc-rel
10050000	17	19	cor	cor	pc-rel	missing	embed
10050000	18	19	cor	cor	pc-rel	missing	embed
10050000	23	24	cor	sub	embed	-	embed
10050000	29	30	cor	sub	pc-rel	leftout	embed
10050000	37	39	cor	adv	nested	-	nested
10050000	38	39	cor	adv	nested	-	nested
10060000	7	8	adv	sub	shared	interpret	embed
10060000	25	26	phr	cor	pc-rel	syntactic	pc-rel
10070000	5	7	sub	adv	pc-rel	syntactic	pc-rel
10070000	6	7	sub	adv	pc-rel	missing	embed
10070000	11	12	cor	phr	shared	syntactic	pc-rel
10070000	14	15	cor	adv	shared	multi	ident
10070000	18	19	sub	phr	embed	-	embed
10070000	21	22	cor	sub	embed	-	embed
10070000	23	24	cor	sub	embed	-	embed
10080000	1	2	sub	cor	embed	-	embed
10080000	3	4	sub	cor	embed	-	embed
10080000	5	6	sub	sub	pc-arg	syntactic	pc-rel
10080000	10	11	adv	sub	pc-rel	missing	embed
10080000	10	12	adv	cor	shared	interpret	embed
10080000	11	12	sub	cor	pc-rel	interpret	embed
10080000	13	14	cor	adv	shared	multi	ident
10080000	22	23	sub	sub	pc-rel	syntactic	pc-rel
10080000	23	24	sub	cor	pc-rel	syntactic	pc-rel
10080000	23	25	sub	cor	pc-rel	syntactic	pc-rel
10080000	23	26	sub	adv	pc-rel	syntactic	pc-rel
10080000	25	26	cor	adv	shared	multi	ident
10080000	29	30	cor	adv	embed	-	embed
10080000	34	35	adv	adv	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10080000	38	39	adv	sub	embed	-	embed
10080000	40	41	sub	cor	pc-rel	error	embed
10080000	40	42	sub	adv	embed	-	embed
10080000	41	42	cor	adv	pc-rel	interpret	embed
10090000	3	4	cor	cor	shared	interpret	embed
10090000	5	6	cor	sub	pc-rel	missing	embed
10090000	5	7	cor	sub	pc-rel	missing	embed
10090000	5	8	cor	cor	pc-rel	missing	embed
10090000	5	9	cor	sub	pc-rel	missing	embed
10090000	5	10	cor	cor	pc-rel	missing	embed
10090000	5	11	cor	cor	pc-arg	interpret	embed
10090000	5	13	cor	adv	overlap	interpret	embed
10090000	7	8	sub	cor	embed	-	embed
10090000	8	9	cor	sub	embed	-	embed
10090000	8	10	cor	cor	pc-rel	missing	embed
10090000	8	11	cor	cor	shared	semantic	shared
10090000	8	13	cor	adv	pc-arg	interpret	embed
10090000	9	10	sub	cor	pc-rel	missing	embed
10090000	9	11	sub	cor	embed	-	embed
10090000	9	13	sub	adv	pc-rel	interpret	embed
10090000	10	11	cor	cor	pc-rel	missing	embed
10090000	10	13	cor	adv	pc-rel	missing	embed
10090000	11	12	cor	cor	embed	-	embed
10090000	11	13	cor	adv	embed	-	embed
10090000	12	13	cor	adv	pc-rel	interpret	embed
10090000	14	15	adv	phr	pc-rel	missing	embed
10090000	21	22	adv	cor	pc-rel	syntactic	pc-rel
10090000	21	23	adv	cor	pc-arg	interpret	embed
10090000	22	23	cor	cor	pc-rel	syntactic	pc-rel
10090000	23	24	cor	cor	shared	semantic	shared
10090000	33	34	sub	sub	pc-rel	syntactic	pc-rel
10100000	1	2	sub	cor	nested	-	nested
10100000	1	6	sub	adv	nested	-	nested
10100000	2	6	cor	adv	nested	-	nested
10100000	3	6	cor	adv	nested	-	nested
10100000	4	5	cor	adv	shared	multi	ident
10100000	4	6	cor	adv	nested	-	nested
10100000	5	6	adv	adv	nested	-	nested
10100000	10	11	cor	adv	nested	-	nested
10100000	15	16	cor	cor	pc-rel	syntactic	pc-rel
10100000	26	27	adv	sub	embed	-	embed
10100000	30	31	sub	cor	embed	-	embed
10100000	37	38	cor	cor	embed	-	embed
10100000	42	43	sub	cor	pc-rel	syntactic	pc-rel

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10100000	45	46	cor	sub	pc-rel	syntactic	pc-rel
10100000	47	48	sub	cor	pc-rel	leftout	embed
10100000	48	49	cor	sub	embed	-	embed
10110000	9	10	sub	cor	pc-rel	syntactic	pc-rel
10110000	11	12	cor	adv	nested	-	nested
10110000	17	18	cor	phr	embed	-	embed
10110000	19	20	cor	cor	pc-rel	syntactic	pc-rel
10120000	3	4	sub	sub	pc-arg	interpret	embed
10120000	5	6	sub	cor	embed	-	embed
10120000	11	12	cor	sub	pc-rel	syntactic	pc-rel
10120000	13	14	cor	cor	pc-rel	syntactic	pc-rel
10120000	15	16	cor	phr	shared	multi	ident
10120000	19	20	cor	sub	embed	-	embed
10120000	23	24	cor	phr	nested	-	nested
10130000	8	9	adv	cor	pc-rel	syntactic	pc-rel
10130000	9	10	cor	cor	embed	-	embed
10130000	14	15	cor	cor	shared	semantic	shared
10130000	17	18	phr	adv	pc-rel	syntactic	pc-rel
10130000	18	19	adv	adv	shared	interpret	embed
10130000	20	21	sub	cor	pc-rel	syntactic	pc-rel
10130000	26	27	cor	adv	pc-rel	missing	embed
10130000	27	28	adv	phr	shared	interpret	embed
10130000	36	37	cor	sub	pc-rel	leftout	embed
10130000	36	38	cor	sub	pc-rel	missing	embed
10130000	43	44	cor	phr	shared	interpret	embed
10130000	44	45	phr	phr	shared	interpret	embed
10130000	45	46	phr	phr	embed	-	embed
10130000	48	49	cor	sub	embed	-	embed
10140000	6	7	cor	cor	embed	-	embed
10140000	9	10	cor	adv	shared	multi	ident
10140000	11	12	cor	cor	pc-rel	syntactic	pc-rel
10140000	17	18	cor	sub	pc-rel	syntactic	pc-rel
10140000	20	21	cor	cor	pc-rel	syntactic	pc-rel
10140000	26	27	sub	cor	pc-rel	syntactic	pc-rel
10140000	33	34	sub	cor	pc-rel	syntactic	pc-rel
10150000	11	12	sub	sub	shared	missing	embed
10150000	11	13	sub	sub	shared	missing	embed
10150000	12	13	sub	sub	shared	missing	embed
10150000	18	19	sub	sub	pc-arg	syntactic	pc-rel
10150000	21	22	phr	cor	embed	-	embed
10150000	26	27	sub	cor	embed	-	embed
10150000	29	30	sub	cor	pc-arg	interpret	embed
10160000	9	10	adv	cor	embed	-	embed
10160000	12	13	cor	cor	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10160000	14	15	phr	cor	pc-rel	syntactic	pc-rel
10160000	28	29	sub	adv	pc-rel	leftout	embed
10170000	4	5	cor	sub	embed	-	embed
10170000	6	7	cor	sub	embed	-	embed
10170000	6	8	cor	cor	embed	-	embed
10170000	7	8	sub	cor	pc-arg	interpret	embed
10170000	9	10	sub	adv	nested	-	nested
10170000	12	13	phr	phr	shared	interpret	embed
10170000	20	21	cor	cor	pc-rel	syntactic	pc-rel
10170000	25	26	cor	cor	shared	multi	ident
10170000	28	29	cor	cor	pc-rel	syntactic	pc-rel
10170000	28	30	cor	cor	shared	interpret	embed
10170000	29	30	cor	cor	pc-rel	syntactic	pc-rel
10170000	31	32	adv	cor	pc-rel	syntactic	pc-rel
10170000	33	34	adv	adv	embed	-	embed
10170000	33	35	adv	cor	shared	interpret	embed
10170000	34	35	adv	cor	nested	-	nested
10170000	37	38	cor	cor	pc-rel	syntactic	pc-rel
10170000	42	43	adv	sub	pc-arg	leftout	embed
10180000	9	10	sub	adv	embed	-	embed
10180000	10	11	adv	sub	embed	-	embed
10180000	21	22	cor	adv	pc-rel	syntactic	pc-rel
10180000	23	24	sub	adv	pc-rel	missing	embed
10190000	4	5	cor	adv	shared	multi	ident
10190000	8	9	cor	cor	pc-rel	syntactic	pc-rel
10190000	10	11	cor	adv	embed	-	embed
10190000	15	16	adv	adv	pc-rel	syntactic	pc-rel
10190000	19	20	phr	phr	shared	missing	embed
10190000	23	24	sub	cor	embed	-	embed
10190000	24	25	cor	sub	embed	-	embed
10200000	4	5	cor	sub	pc-rel	syntactic	pc-rel
10200000	4	6	cor	cor	pc-rel	syntactic	pc-rel
10200000	5	6	sub	cor	pc-arg	interpret	embed
10200000	12	13	cor	cor	pc-rel	syntactic	pc-rel
10200000	15	16	sub	cor	pc-rel	missing	embed
10200000	15	17	sub	phr	embed	-	embed
10200000	16	17	cor	phr	shared	multi	ident
10200000	16	18	cor	cor	shared	interpret	embed
10200000	17	18	phr	cor	shared	interpret	embed
10210000	2	3	cor	cor	pc-rel	interpret	embed
10210000	4	5	cor	adv	pc-rel	missing	embed
10210000	4	6	cor	adv	pc-arg	missing	embed
10210000	5	6	adv	adv	pc-arg	interpret	embed
10210000	8	9	adv	adv	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10210000	9	10	adv	cor	pc-rel	leftout	embed
10210000	9	11	adv	phr	embed	-	embed
10210000	10	11	cor	phr	embed	-	embed
10210000	10	26	cor	phr	pc-rel	interpret	embed
10210000	11	26	phr	phr	embed	-	embed
10210000	12	13	sub	sub	embed	-	embed
10210000	14	15	cor	phr	embed	-	embed
10210000	14	26	cor	phr	pc-rel	interpret	embed
10210000	15	16	phr	adv	shared	semantic	shared
10210000	15	26	phr	phr	pc-rel	interpret	embed
10210000	16	26	adv	phr	pc-rel	interpret	embed
10210000	17	18	cor	adv	shared	multi	ident
10210000	17	19	cor	adv	embed	-	embed
10210000	17	26	cor	phr	pc-rel	interpret	embed
10210000	18	19	adv	adv	embed	-	embed
10210000	18	26	adv	phr	pc-rel	interpret	embed
10210000	19	26	adv	phr	pc-rel	interpret	embed
10210000	20	26	sub	phr	pc-rel	interpret	embed
10210000	21	22	adv	cor	embed	-	embed
10210000	23	24	cor	cor	pc-rel	interpret	embed
10210000	23	25	cor	adv	pc-rel	interpret	embed
10210000	24	25	cor	adv	shared	interpret	embed
10210000	35	36	cor	cor	pc-rel	interpret	embed
10210000	38	39	sub	adv	pc-rel	leftout	embed
10210000	42	43	sub	cor	pc-rel	missing	embed
10220000	2	3	cor	adv	embed	-	embed
10220000	3	4	adv	cor	shared	interpret	embed
10220000	10	11	adv	cor	pc-arg	interpret	embed
10220000	15	16	adv	sub	embed	-	embed
10220000	17	18	phr	cor	shared	interpret	embed
10220000	23	24	sub	adv	embed	-	embed
10220000	31	32	cor	cor	pc-rel	syntactic	pc-rel
10220000	37	38	adv	cor	pc-rel	missing	embed
10220000	40	41	sub	sub	pc-rel	syntactic	pc-rel
10220000	42	43	phr	cor	pc-rel	syntactic	pc-rel
10230000	12	13	sub	cor	pc-rel	syntactic	pc-rel
10230000	17	18	cor	adv	embed	interpret	embed
10230000	22	23	cor	sub	embed	-	embed
10230000	29	30	cor	sub	pc-rel	interpret	embed
10240000	5	7	sub	adv	embed	-	embed
10240000	6	7	cor	adv	nested	-	nested
10240000	19	20	sub	cor	embed	-	embed
10250000	4	5	sub	cor	pc-rel	syntactic	pc-rel
10250000	6	7	adv	adv	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10250000	8	9	sub	adv	embed	-	embed
10250000	9	10	adv	sub	pc-rel	missing	embed
10250000	16	17	sub	cor	pc-arg	missing	embed
10250000	16	18	sub	sub	shared	missing	embed
10250000	17	18	cor	sub	pc-arg	missing	embed
10250000	22	23	adv	cor	pc-rel	syntactic	pc-rel
10250000	29	30	sub	cor	embed	-	embed
10250000	29	31	sub	cor	pc-rel	interpret	embed
10250000	30	31	cor	cor	embed	-	embed
10260000	2	3	sub	phr	pc-rel	syntactic	pc-rel
10260000	6	7	cor	sub	embed	-	embed
10260000	6	8	cor	cor	pc-rel	syntactic	pc-rel
10260000	6	13	cor	cor	nested	-	nested
10260000	7	8	sub	cor	embed	-	embed
10260000	7	13	sub	cor	nested	-	nested
10260000	8	13	cor	cor	nested	-	nested
10260000	9	13	adv	cor	nested	-	nested
10260000	10	11	cor	cor	embed	-	embed
10260000	10	13	cor	cor	nested	-	nested
10260000	11	13	cor	cor	nested	-	nested
10260000	12	13	cor	cor	nested	-	nested
10260000	14	15	cor	cor	pc-rel	syntactic	pc-rel
10260000	14	16	cor	cor	embed	-	embed
10260000	15	16	cor	cor	pc-rel	syntactic	pc-rel
10260000	19	20	cor	adv	shared	multi	ident
10260000	21	22	sub	sub	embed	-	embed
10260000	21	23	sub	adv	shared	interpret	embed
10260000	22	23	sub	adv	embed	-	embed
10260000	23	24	adv	cor	pc-rel	missing	embed
10260000	30	31	cor	cor	embed	-	embed
10270000	3	4	cor	sub	embed	-	embed
10270000	7	8	sub	sub	shared	missing	embed
10270000	7	9	sub	cor	pc-rel	syntactic	pc-rel
10270000	8	9	sub	cor	pc-rel	syntactic	pc-rel
10270000	14	15	cor	cor	pc-rel	syntactic	pc-rel
10270000	17	18	sub	cor	pc-arg	interpret	embed
10270000	19	20	cor	sub	pc-rel	semantic	shared
10270000	22	23	cor	sub	pc-rel	syntactic	pc-rel
10280000	4	5	cor	adv	embed	-	embed
10280000	15	16	sub	cor	embed	-	embed
10280000	21	22	cor	cor	shared	missing	embed
10280000	23	24	adv	sub	pc-rel	syntactic	pc-rel
10280000	28	30	cor	cor	shared	semantic	shared
10280000	29	30	cor	cor	nested	-	nested

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10290000	4	5	sub	phr	pc-rel	syntactic	pc-rel
10300000	2	3	cor	cor	pc-rel	missing	embed
10300000	6	7	cor	cor	pc-rel	syntactic	pc-rel
10300000	7	8	cor	adv	pc-rel	syntactic	pc-rel
10300000	9	10	adv	adv	nested	-	nested
10300000	12	13	adv	cor	pc-rel	syntactic	pc-rel
10300000	19	20	adv	cor	pc-rel	syntactic	pc-rel
10310000	13	14	cor	adv	pc-arg	interpret	embed
10310000	14	15	adv	cor	pc-arg	interpret	embed
10310000	23	24	cor	cor	embed	-	embed
10310000	27	28	cor	cor	pc-rel	syntactic	pc-rel
10310000	27	29	cor	sub	pc-rel	syntactic	pc-rel
10310000	28	29	cor	sub	pc-rel	interpret	embed
10320000	4	5	adv	sub	embed	-	embed
10320000	6	7	sub	cor	shared	interpret	embed
10320000	8	9	cor	cor	nested	-	nested
10320000	9	10	cor	cor	pc-rel	missing	embed
10320000	9	11	cor	sub	pc-rel	missing	embed
10320000	10	11	cor	sub	embed	-	embed
10320000	13	14	adv	cor	pc-rel	missing	embed
10320000	16	17	cor	adv	shared	multi	ident
10320000	23	24	cor	cor	pc-rel	syntactic	pc-rel
10320000	34	35	cor	cor	pc-rel	missing	embed
10320000	34	36	cor	adv	pc-rel	missing	embed
10320000	37	38	cor	adv	pc-rel	syntactic	pc-arg
10320000	37	39	cor	cor	pc-rel	missing	embed
10320000	38	39	adv	cor	embed	-	embed
10330000	1	2	cor	sub	overlap	syntactic	pc-rel
10330000	14	15	sub	phr	embed	-	embed
10340000	2	3	sub	sub	pc-arg	interpret	pc-rel
10340000	2	4	sub	sub	pc-rel	syntactic	pc-rel
10340000	3	4	sub	sub	pc-rel	syntactic	pc-rel
10340000	7	8	adv	cor	shared	multi	ident
10340000	8	9	cor	adv	shared	multi	ident
10340000	12	13	cor	cor	embed	-	embed
10340000	15	16	sub	phr	embed	-	embed
10340000	17	18	adv	sub	pc-rel	leftout	embed
10340000	17	19	adv	cor	embed	-	embed
10340000	18	19	sub	cor	embed	-	embed
10340000	23	25	cor	adv	pc-rel	syntactic	pc-rel
10340000	24	25	cor	adv	nested	-	nested
10340000	26	27	cor	cor	embed	-	embed
10340000	28	29	cor	cor	embed	-	embed
10340000	30	31	sub	cor	pc-rel	syntactic	pc-rel

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10340000	33	34	sub	cor	pc-arg	interpret	embed
10340000	36	37	adv	cor	embed	-	embed
10340000	39	40	phr	cor	pc-rel	interpret	embed
10340000	39	41	phr	adv	pc-arg	interpret	embed
10340000	40	41	cor	adv	pc-rel	interpret	embed
10350000	7	8	adv	adv	shared	interpret	embed
10350000	8	9	adv	cor	shared	interpret	embed
10350000	9	10	cor	cor	pc-rel	syntactic	pc-rel
10350000	11	12	cor	cor	shared	error	ident
10350000	14	15	cor	phr	pc-rel	interpret	embed
10350000	15	16	phr	sub	pc-rel	interpret	embed
10350000	15	17	phr	cor	pc-rel	syntactic	pc-rel
10350000	15	18	phr	sub	pc-rel	syntactic	pc-rel
10350000	16	17	sub	cor	pc-rel	syntactic	pc-rel
10350000	16	18	sub	sub	embed	-	embed
10350000	17	18	cor	sub	pc-rel	syntactic	pc-rel
10350000	19	21	sub	cor	embed	-	embed
10350000	20	21	cor	cor	nested	-	nested
10350000	32	33	sub	adv	embed	-	embed
10350000	33	34	adv	cor	pc-rel	syntactic	pc-rel
10350000	36	37	adv	cor	embed	-	embed
10350000	38	39	cor	sub	embed	-	embed
10350000	43	44	sub	cor	pc-rel	missing	embed
10360000	7	8	sub	cor	pc-arg	interpret	embed
10360000	12	13	adv	adv	shared	interpret	embed
10360000	13	14	adv	cor	shared	semantic	shared
10360000	14	15	cor	adv	pc-arg	semantic	shared
10360000	15	16	adv	cor	pc-rel	syntactic	pc-rel
10360000	15	17	adv	adv	shared	interpret	embed
10360000	16	17	cor	adv	pc-rel	syntactic	pc-rel
10360000	19	20	sub	adv	pc-rel	syntactic	pc-rel
10360000	22	23	cor	cor	shared	interpret	embed
10370000	7	8	cor	phr	shared	multi	ident
10370000	22	23	cor	sub	embed	-	embed
10370000	32	33	cor	sub	pc-rel	interpret	embed
10380000	1	2	cor	sub	shared	interpret	embed
10380000	5	6	sub	cor	embed	-	embed
10380000	11	12	sub	cor	embed	-	embed
10380000	11	13	sub	adv	pc-rel	missing	embed
10380000	12	13	cor	adv	embed	-	embed
10380000	13	14	adv	cor	embed	-	embed
10380000	15	16	cor	adv	pc-arg	syntactic	pc-arg
10380000	17	18	cor	cor	embed	-	embed
10380000	17	19	cor	sub	embed	interpret	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10380000	18	19	cor	sub	embed	-	embed
10380000	24	25	cor	sub	pc-rel	syntactic	pc-rel
10380000	26	27	sub	phr	pc-rel	missing	embed
10380000	26	27	sub	phr	nested	-	nested
10380000	33	34	cor	adv	pc-rel	missing	embed
10390000	3	4	adv	cor	pc-rel	missing	embed
10390000	3	5	adv	cor	embed	-	embed
10390000	6	7	cor	sub	pc-arg	interpret	embed
10390000	6	8	cor	adv	pc-arg	interpret	embed
10390000	7	8	sub	adv	embed	leftout	embed
10390000	8	9	adv	cor	embed	-	embed
10390000	10	11	cor	sub	embed	-	embed
10390000	10	12	cor	cor	pc-rel	leftout	embed
10390000	11	12	sub	cor	embed	-	embed
10390000	12	13	cor	sub	embed	-	embed
10390000	14	15	cor	cor	pc-rel	syntactic	pc-rel
10390000	15	16	cor	cor	pc-rel	missing	embed
10390000	17	18	cor	sub	embed	-	embed
10390000	20	21	sub	adv	pc-rel	syntactic	pc-rel
10390000	28	29	adv	adv	shared	interpret	embed
10390000	34	35	cor	sub	shared	multi	ident
10390000	39	40	sub	cor	embed	-	embed
10390000	41	42	sub	cor	embed	-	embed
10400000	6	7	sub	cor	embed	-	embed
10400000	15	16	cor	cor	shared	missing	embed
10400000	15	17	cor	cor	shared	missing	embed
10400000	16	17	cor	cor	shared	missing	embed
10400000	18	19	cor	cor	embed	-	embed
10400000	31	32	sub	sub	shared	missing	embed
10400000	33	34	cor	adv	pc-rel	syntactic	pc-rel
10400000	36	37	cor	cor	pc-rel	missing	embed
10400000	37	38	cor	cor	embed	-	embed
10510000	4	5	cor	cor	embed	-	embed
10510000	8	9	sub	adv	pc-rel	interpret	embed
10510000	11	12	phr	cor	embed	-	embed
10510000	19	20	cor	cor	shared	semantic	shared
10510000	20	21	cor	adv	shared	semantic	shared
10510000	21	22	adv	sub	embed	-	embed
10510000	21	23	adv	cor	pc-arg	interpret	embed
10510000	21	24	adv	cor	overlap	interpret	embed
10510000	22	23	sub	cor	pc-arg	interpret	embed
10510000	22	24	sub	cor	overlap	interpret	embed
10510000	23	24	cor	cor	pc-arg	interpret	embed
10510000	29	30	sub	cor	pc-arg	interpret	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10510000	38	42	sub	cor	pc-rel	syntactic	pc-rel
10510000	39	42	sub	cor	nested	-	nested
10510000	40	41	cor	cor	pc-arg	leftout	embed
10510000	40	42	cor	cor	nested	-	nested
10510000	41	42	cor	cor	nested	-	nested
10520000	4	5	cor	cor	pc-rel	missing	embed
10520000	11	12	cor	cor	embed	-	embed
10520000	16	17	cor	adv	shared	multi	ident
10520000	18	19	sub	adv	embed	-	embed
10520000	22	23	adv	adv	pc-rel	missing	embed
10520000	29	30	sub	cor	pc-rel	syntactic	pc-rel
10520000	34	35	adv	sub	pc-rel	syntactic	pc-rel
10520000	34	37	adv	cor	shared	interpret	embed
10520000	35	37	sub	cor	pc-rel	syntactic	pc-rel
10520000	36	37	cor	cor	nested	-	nested
10520000	36	37	cor	cor	pc-rel	syntactic	pc-rel
10530000	14	15	cor	sub	embed	-	embed
10530000	14	16	cor	sub	embed	-	embed
10530000	15	16	sub	sub	shared	missing	embed
10530000	16	17	sub	cor	shared	missing	embed
10530000	19	20	cor	sub	embed	-	embed
10530000	22	23	sub	cor	pc-rel	interpret	embed
10530000	23	24	cor	cor	shared	interpret	embed
10530000	26	27	cor	sub	embed	-	embed
10530000	34	35	sub	sub	pc-rel	syntactic	pc-rel
10530000	36	37	cor	cor	pc-rel	syntactic	pc-rel
10550000	1	4	cor	adv	nested	-	nested
10550000	2	4	cor	adv	nested	-	nested
10550000	3	4	sub	adv	nested	-	nested
10550000	4	5	adv	cor	pc-rel	syntactic	pc-rel
10550000	8	9	sub	sub	pc-rel	syntactic	pc-rel
10550000	14	15	cor	cor	pc-rel	syntactic	pc-rel
10550000	16	17	sub	cor	pc-rel	syntactic	pc-rel
10550000	28	29	sub	phr	embed	-	embed
10560000	2	3	cor	adv	pc-rel	syntactic	pc-rel
10560000	7	8	cor	cor	pc-rel	syntactic	pc-rel
10560000	8	9	cor	sub	embed	-	embed
10560000	30	31	sub	sub	pc-rel	interpret	embed
10560000	33	34	cor	adv	pc-rel	syntactic	pc-rel
10570000	3	4	cor	cor	pc-rel	missing	embed
10570000	6	7	cor	cor	pc-rel	syntactic	pc-rel
10570000	8	11	cor	cor	pc-rel	syntactic	pc-rel
10570000	9	10	adv	sub	pc-rel	syntactic	pc-rel
10570000	9	11	adv	cor	nested	-	nested

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10570000	10	11	sub	cor	nested	-	nested
10570000	14	15	sub	cor	embed	-	embed
10570000	20	21	sub	cor	embed	-	embed
10570000	20	22	sub	adv	embed	-	embed
10570000	21	22	cor	adv	shared	multi	ident
10570000	26	27	cor	phr	shared	multi	ident
10570000	37	38	sub	cor	pc-rel	missing	embed
10570000	37	39	sub	phr	pc-rel	missing	embed
10570000	38	39	cor	phr	pc-rel	missing	embed
10570000	39	40	phr	cor	embed	interpret	embed
10570000	39	41	phr	sub	pc-arg	error	embed
10580000	9	10	cor	cor	pc-rel	syntactic	pc-rel
10580000	18	19	cor	sub	embed	-	embed
10590000	1	3	cor	cor	pc-rel	syntactic	pc-rel
10590000	2	3	adv	cor	pc-rel	syntactic	pc-rel
10590000	10	11	sub	cor	embed	-	embed
10590000	10	12	sub	cor	embed	-	embed
10590000	11	12	cor	cor	shared	interpret	embed
10590000	15	16	adv	cor	pc-rel	syntactic	pc-rel
10590000	17	19	cor	phr	pc-rel	syntactic	pc-rel
10590000	18	19	cor	phr	pc-rel	syntactic	pc-rel
10590000	19	20	phr	cor	pc-rel	syntactic	pc-rel
10590000	19	21	phr	sub	embed	-	embed
10590000	20	21	cor	sub	pc-rel	interpret	embed
10590000	25	26	sub	cor	embed	-	embed
10590000	30	31	cor	sub	embed	-	embed
10590000	30	32	cor	adv	embed	interpret	embed
10590000	31	32	sub	adv	embed	-	embed
10590000	35	36	cor	cor	pc-rel	missing	embed
10590000	35	37	cor	cor	embed	-	embed
10590000	42	43	sub	cor	pc-arg	interpret	embed
10590000	42	44	sub	cor	pc-rel	syntactic	pc-rel
10590000	42	45	sub	cor	pc-rel	interpret	embed
10590000	43	44	cor	cor	pc-rel	syntactic	pc-rel
10590000	43	45	cor	cor	pc-rel	syntactic	pc-rel
10590000	44	45	cor	cor	pc-rel	missing	embed
10590000	46	47	phr	cor	embed	-	embed
10590000	46	48	phr	cor	shared	interpret	embed
10590000	47	48	cor	cor	embed	-	embed
10590000	50	51	cor	sub	shared	interpret	embed
10590000	53	54	cor	cor	pc-rel	syntactic	pc-rel
10590000	56	57	sub	cor	pc-rel	syntactic	pc-rel
10590000	58	59	cor	sub	embed	-	embed
10590000	58	60	cor	cor	embed	interpret	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10590000	59	60	sub	cor	embed	-	embed
10590000	61	62	cor	cor	shared	interpret	embed
10600000	6	7	sub	cor	pc-rel	leftout	embed
10600000	17	18	adv	phr	shared	multi	ident
10600000	27	28	cor	sub	pc-rel	interpret	embed
10600000	30	31	cor	cor	pc-rel	syntactic	pc-rel
10610000	6	7	adv	adv	shared	interpret	embed
10610000	7	8	adv	sub	pc-rel	syntactic	pc-rel
10610000	13	14	cor	sub	embed	-	embed
10610000	13	15	cor	cor	embed	-	embed
10610000	13	16	cor	adv	embed	-	embed
10610000	14	15	sub	cor	pc-rel	interpret	embed
10610000	14	16	sub	adv	pc-rel	interpret	embed
10610000	15	16	cor	adv	shared	multi	ident
10610000	15	17	cor	cor	pc-rel	syntactic	pc-rel
10610000	16	17	adv	cor	pc-rel	syntactic	pc-rel
10610000	18	32	cor	adv	nested	-	nested
10610000	19	20	cor	sub	embed	-	embed
10610000	19	32	cor	adv	nested	-	nested
10610000	20	32	sub	adv	nested	-	nested
10610000	21	22	cor	sub	embed	-	embed
10610000	21	23	cor	cor	embed	-	embed
10610000	21	24	cor	adv	embed	-	embed
10610000	21	32	cor	adv	nested	-	nested
10610000	22	23	sub	cor	pc-rel	interpret	embed
10610000	22	24	sub	adv	pc-rel	interpret	embed
10610000	22	32	sub	adv	nested	-	nested
10610000	23	24	cor	adv	shared	multi	ident
10610000	23	25	cor	cor	pc-rel	syntactic	pc-rel
10610000	23	32	cor	adv	nested	-	nested
10610000	24	25	adv	cor	pc-rel	syntactic	pc-rel
10610000	24	32	adv	adv	nested	-	nested
10610000	25	32	cor	adv	nested	-	nested
10610000	26	32	sub	adv	nested	-	nested
10610000	27	32	cor	adv	nested	-	nested
10610000	28	32	cor	adv	nested	-	nested
10610000	29	30	cor	cor	embed	-	embed
10610000	29	31	cor	cor	pc-arg	interpret	embed
10610000	29	32	cor	adv	nested	-	nested
10610000	30	31	cor	cor	shared	interpret	embed
10610000	30	32	cor	adv	nested	-	nested
10610000	31	32	cor	adv	nested	-	nested
10620000	2	3	adv	cor	pc-rel	syntactic	pc-rel
10620000	2	4	adv	sub	pc-rel	syntactic	pc-rel

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10620000	3	4	cor	sub	pc-rel	syntactic	pc-rel
10620000	7	8	cor	sub	embed	-	embed
10620000	12	13	sub	cor	pc-rel	missing	embed
10620000	22	24	sub	adv	pc-rel	syntactic	pc-rel
10620000	22	26	sub	adv	pc-rel	syntactic	pc-rel
10620000	24	25	adv	cor	embed	-	embed
10620000	24	26	adv	adv	embed	-	embed
10620000	25	26	cor	adv	embed	interpret	embed
10620000	27	28	cor	phr	pc-rel	missing	embed
10620000	28	29	phr	cor	embed	-	embed
10620000	34	35	cor	sub	embed	-	embed
10630000	4	5	cor	sub	embed	-	embed
10630000	6	7	adv	phr	embed	-	embed
10630000	8	9	sub	sub	shared	missing	embed
10630000	8	10	sub	cor	shared	missing	embed
10630000	8	12	sub	phr	pc-rel	missing	embed
10630000	9	10	sub	cor	shared	missing	embed
10630000	9	11	sub	sub	shared	missing	embed
10630000	9	12	sub	phr	pc-rel	missing	embed
10630000	10	11	cor	sub	shared	missing	embed
10630000	10	12	cor	phr	pc-rel	missing	embed
10630000	11	12	sub	phr	pc-rel	missing	embed
10630000	12	13	phr	adv	embed	-	embed
10630000	15	16	cor	cor	pc-rel	missing	embed
10630000	15	17	cor	sub	pc-rel	interpret	embed
10630000	16	17	cor	sub	pc-rel	missing	embed
10630000	17	18	sub	cor	pc-rel	interpret	embed
10630000	17	19	sub	sub	pc-rel	interpret	embed
10630000	17	20	sub	cor	pc-rel	interpret	embed
10630000	18	19	cor	sub	embed	-	embed
10630000	19	20	sub	cor	embed	-	embed
10630000	24	25	cor	phr	shared	interpret	embed
10630000	28	29	cor	sub	pc-rel	missing	embed
10630000	28	30	cor	cor	pc-arg	missing	embed
10630000	29	30	sub	cor	embed	-	embed
10630000	34	35	adv	adv	shared	interpret	embed
10630000	35	36	adv	cor	pc-rel	missing	embed
10630000	50	51	adv	cor	embed	-	embed
10630000	54	56	sub	adv	nested	-	nested
10630000	55	56	sub	adv	nested	-	nested
10630000	56	57	adv	cor	shared	interpret	embed
10630000	57	58	cor	cor	embed	-	embed
10640000	12	13	cor	adv	embed	-	embed
10640000	21	22	adv	adv	shared	missing	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10640000	21	25	adv	adv	shared	missing	embed
10640000	22	23	adv	cor	pc-rel	syntactic	pc-rel
10640000	22	25	adv	adv	shared	missing	embed
10640000	23	25	cor	adv	nested	-	nested
10640000	24	25	adv	adv	nested	-	nested
10640000	27	28	sub	cor	pc-rel	interpret	embed
10640000	27	29	sub	cor	pc-rel	syntactic	pc-rel
10640000	27	30	sub	cor	pc-rel	syntactic	pc-rel
10640000	28	29	cor	cor	embed	-	embed
10640000	28	30	cor	cor	pc-rel	syntactic	pc-rel
10640000	29	30	cor	cor	embed	-	embed
10640000	31	32	cor	adv	pc-arg	interpret	embed
10640000	32	33	adv	adv	shared	interpret	embed
10640000	35	36	cor	phr	pc-rel	missing	embed
10640000	36	37	phr	adv	pc-arg	interpret	shared
10640000	38	39	cor	cor	pc-arg	interpret	embed
10640000	41	42	cor	cor	pc-arg	interpret	embed
10650000	2	3	adv	adv	shared	interpret	embed
10650000	3	4	adv	cor	pc-rel	missing	embed
10650000	5	6	cor	cor	pc-arg	interpret	embed
10650000	6	9	cor	phr	pc-arg	missing	shared
10650000	7	8	cor	adv	pc-rel	syntactic	pc-rel
10650000	7	9	cor	phr	pc-rel	syntactic	pc-rel
10650000	8	9	adv	phr	pc-rel	missing	embed
10650000	9	10	phr	sub	pc-rel	missing	embed
10650000	9	11	phr	cor	pc-rel	syntactic	pc-rel
10650000	10	11	sub	cor	pc-rel	syntactic	pc-rel
10650000	12	13	sub	adv	embed	-	embed
10650000	15	16	cor	adv	shared	interpret	embed
10650000	16	17	adv	sub	pc-rel	syntactic	pc-rel
10650000	16	18	adv	sub	pc-rel	missing	embed
10650000	19	20	adv	phr	shared	interpret	embed
10650000	22	23	cor	adv	shared	multi	ident
10650000	24	25	sub	cor	embed	-	embed
10650000	25	26	cor	phr	embed	-	embed
10650000	25	27	cor	adv	pc-arg	interpret	embed
10650000	26	27	phr	adv	pc-arg	interpret	embed
10650000	27	28	adv	adv	shared	interpret	embed
10650000	28	29	adv	sub	pc-rel	missing	embed
10650000	33	34	sub	cor	pc-rel	interpret	embed
10650000	33	35	sub	sub	embed	-	embed
10650000	34	35	cor	sub	embed	-	embed
10650000	40	41	adv	cor	shared	interpret	embed
10650000	42	43	cor	adv	shared	multi	ident

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10650000	45	48	phr	phr	shared	interpret	embed
10650000	46	47	adv	sub	pc-rel	missing	embed
10650000	46	48	adv	phr	nested	-	nested
10650000	47	48	sub	phr	nested	-	nested
10650000	48	49	phr	sub	pc-rel	missing	embed
10650000	48	50	phr	sub	pc-rel	missing	embed
10660000	4	5	cor	cor	pc-rel	syntactic	pc-rel
10660000	6	7	sub	cor	embed	-	embed
10660000	6	8	sub	cor	embed	-	embed
10660000	7	8	cor	cor	pc-rel	interpret	embed
10660000	12	13	cor	adv	nested	-	nested
10660000	13	14	adv	cor	pc-rel	interpret	embed
10660000	15	16	sub	adv	embed	-	embed
10660000	17	18	cor	cor	pc-rel	syntactic	pc-rel
10660000	20	21	adv	cor	embed	-	embed
10660000	23	24	cor	cor	pc-rel	syntactic	pc-rel
10660000	24	25	cor	cor	pc-rel	interpret	embed
10660000	29	30	adv	cor	pc-arg	interpret	embed
10670000	6	7	cor	adv	pc-arg	interpret	embed
10670000	9	10	cor	sub	embed	-	embed
10670000	14	15	sub	sub	shared	error	ident
10670000	14	16	sub	cor	embed	-	embed
10670000	15	16	sub	cor	embed	-	embed
10670000	25	26	cor	sub	pc-rel	interpret	embed
10680000	1	2	phr	sub	pc-rel	syntactic	pc-rel
10680000	4	5	sub	cor	embed	-	embed
10680000	17	18	cor	phr	shared	multi	ident
10680000	20	21	sub	cor	embed	-	embed
10680000	21	22	cor	sub	embed	-	embed
10680000	27	28	cor	cor	pc-rel	syntactic	pc-rel
10680000	28	29	cor	cor	pc-rel	syntactic	pc-rel
10680000	30	31	cor	phr	embed	-	embed
10680000	31	32	phr	cor	embed	-	embed
10680000	34	35	cor	sub	embed	interpret	embed
10680000	34	36	cor	sub	embed	-	embed
10680000	34	37	cor	adv	embed	-	embed
10680000	35	36	sub	sub	embed	-	embed
10680000	35	37	sub	adv	embed	interpret	embed
10680000	36	37	sub	adv	embed	interpret	embed
10680000	37	38	adv	sub	pc-rel	interpret	embed
10680000	40	41	phr	cor	shared	interpret	embed
10680000	41	42	cor	cor	pc-rel	syntactic	pc-rel
10680000	41	43	cor	adv	pc-arg	interpret	embed
10680000	42	43	cor	adv	nested	-	nested

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
10680000	42	43	cor	adv	pc-rel	syntactic	pc-rel
10680000	42	44	cor	phr	pc-rel	syntactic	pc-rel
10690000	12	13	cor	cor	embed	-	embed
10690000	15	16	phr	adv	shared	interpret	embed
10690000	15	19	phr	adv	pc-rel	interpret	embed
10690000	16	17	adv	cor	pc-rel	leftout	embed
10690000	16	18	adv	cor	embed	-	embed
10690000	16	19	adv	adv	embed	-	embed
10690000	17	18	cor	cor	embed	-	embed
10690000	17	19	cor	adv	pc-rel	leftout	embed
10690000	18	19	cor	adv	pc-rel	interpret	embed
10690000	24	25	cor	cor	pc-rel	syntactic	pc-rel
10690000	26	27	sub	cor	pc-rel	syntactic	pc-rel
10690000	29	30	adv	cor	embed	-	embed
10690000	31	32	phr	cor	pc-rel	syntactic	pc-rel
10690000	31	34	phr	adv	pc-arg	missing	embed
10690000	32	34	cor	adv	pc-rel	syntactic	pc-rel
10690000	33	34	cor	adv	pc-rel	syntactic	pc-rel
10700000	11	12	sub	cor	pc-rel	syntactic	pc-rel
10700000	11	13	sub	cor	pc-rel	syntactic	pc-rel
10700000	12	13	cor	cor	pc-rel	syntactic	pc-rel
10700000	15	16	sub	sub	pc-rel	syntactic	pc-rel
10700000	16	17	sub	sub	embed	-	embed
10700000	18	19	cor	cor	pc-rel	syntactic	pc-rel
10700000	20	21	sub	cor	pc-rel	syntactic	pc-rel
10700000	24	25	cor	cor	pc-rel	interpret	embed
10700000	24	26	cor	adv	pc-rel	interpret	embed
10700000	25	26	cor	adv	embed	-	embed
10700000	26	27	adv	cor	embed	-	embed
10700000	29	30	cor	adv	nested	-	nested
10700000	30	31	adv	cor	pc-rel	leftout	embed
10700000	30	32	adv	sub	embed	-	embed
10700000	30	33	adv	phr	shared	interpret	embed
10700000	31	32	cor	sub	embed	-	embed
10700000	31	33	cor	phr	pc-rel	leftout	embed
10700000	32	33	sub	phr	embed	-	embed
10700000	39	40	cor	adv	embed	-	embed
10700000	40	41	adv	cor	pc-arg	interpret	embed
10700000	43	44	cor	phr	shared	multi	ident
10700000	47	48	phr	adv	embed	-	embed
10700000	49	50	cor	phr	pc-rel	syntactic	pc-rel
10700000	52	53	cor	cor	pc-rel	syntactic	pc-rel
10700000	54	55	sub	phr	pc-rel	syntactic	pc-rel
20190000	12	13	sub	cor	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
20190000	13	14	cor	cor	shared	interpret	embed
20190000	18	19	cor	cor	pc-rel	syntactic	pc-rel
20190000	22	23	cor	adv	embed	-	embed
20190000	35	36	cor	sub	pc-rel	syntactic	pc-rel
20200000	10	11	cor	cor	shared	semantic	shared
20200000	12	13	phr	cor	pc-rel	interpret	embed
20200000	21	22	cor	sub	embed	-	embed
20210000	2	3	cor	adv	pc-rel	missing	embed
20210000	4	5	cor	cor	pc-rel	syntactic	pc-rel
20210000	6	7	cor	phr	pc-rel	syntactic	pc-rel
20210000	10	11	phr	sub	embed	-	embed
20210000	19	20	cor	sub	pc-rel	syntactic	pc-rel
20210000	19	21	cor	cor	pc-rel	syntactic	pc-rel
20210000	20	21	sub	cor	embed	-	embed
20210000	24	25	cor	cor	embed	-	embed
20220000	9	10	cor	adv	embed	-	embed
20220000	11	12	sub	cor	pc-rel	missing	embed
20220000	12	13	cor	sub	pc-rel	missing	embed
20220000	15	16	cor	sub	pc-rel	missing	embed
20220000	17	18	adv	sub	pc-rel	missing	embed
20220000	19	20	cor	adv	shared	interpret	embed
20220000	27	28	adv	cor	embed	-	embed
20220000	29	30	adv	sub	embed	-	embed
20230000	4	5	cor	sub	pc-rel	syntactic	pc-rel
20230000	6	8	cor	cor	pc-rel	missing	embed
20230000	7	8	sub	cor	pc-rel	missing	embed
20230000	11	12	sub	phr	pc-rel	missing	embed
20230000	14	15	cor	adv	shared	multi	ident
20230000	14	16	cor	cor	pc-rel	syntactic	pc-rel
20230000	15	16	adv	cor	pc-rel	syntactic	pc-rel
20230000	20	21	sub	cor	embed	-	embed
20230000	23	24	cor	adv	shared	multi	ident
20230000	25	26	cor	adv	shared	multi	ident
20230000	29	30	cor	adv	embed	-	embed
20240000	1	2	cor	sub	embed	-	embed
20240000	7	8	cor	adv	shared	interpret	embed
20240000	8	9	adv	cor	shared	interpret	embed
20240000	9	10	cor	sub	embed	-	embed
20240000	15	16	sub	adv	pc-rel	missing	embed
20240000	16	17	adv	sub	pc-rel	missing	embed
20240000	18	19	cor	phr	shared	multi	ident
20240000	22	23	adv	cor	embed	-	embed
20240000	28	29	sub	cor	pc-rel	interpret	embed
20240000	28	30	sub	cor	pc-rel	interpret	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
20240000	29	30	cor	cor	shared	error	ident
20240000	32	33	adv	cor	nested	-	nested
20240000	39	41	sub	adv	embed	-	embed
20240000	40	41	cor	adv	nested	-	nested
20240000	41	42	adv	cor	pc-rel	syntactic	pc-rel
20240000	41	43	adv	cor	pc-arg	missing	embed
20240000	42	43	cor	cor	pc-rel	syntactic	pc-rel
20240000	43	44	cor	sub	pc-rel	interpret	embed
20240000	46	47	sub	cor	embed	-	embed
20240000	47	48	cor	cor	shared	interpret	embed
20240000	50	51	cor	adv	shared	interpret	embed
20240000	55	56	cor	sub	embed	-	embed
20240000	55	57	cor	adv	nested	-	nested
20240000	56	57	sub	adv	nested	-	nested
20240000	57	58	adv	cor	embed	-	embed
20240000	57	59	adv	sub	pc-rel	missing	embed
20240000	58	59	cor	sub	embed	-	embed
20250000	3	4	adv	cor	embed	-	embed
20250000	6	7	cor	adv	pc-rel	missing	embed
20250000	8	9	sub	adv	pc-rel	missing	embed
20250000	11	12	adv	cor	shared	interpret	embed
20250000	22	23	sub	cor	embed	-	embed
20250000	24	25	sub	phr	pc-rel	missing	embed
20250000	28	29	cor	sub	embed	-	embed
20250000	31	32	sub	adv	embed	-	embed
20250000	36	37	sub	sub	pc-rel	syntactic	pc-rel
20250000	42	43	sub	cor	pc-rel	syntactic	pc-rel
20250000	46	47	cor	sub	embed	-	embed
20260000	5	6	sub	cor	pc-rel	syntactic	pc-rel
20260000	5	7	sub	sub	pc-rel	syntactic	pc-rel
20260000	10	11	adv	cor	embed	-	embed
20260000	13	14	cor	cor	pc-rel	missing	embed
20260000	19	20	sub	cor	pc-rel	syntactic	pc-rel
20260000	19	21	sub	sub	pc-rel	syntactic	pc-rel
20260000	24	25	adv	cor	embed	-	embed
20260000	27	28	cor	cor	pc-rel	missing	embed
20270000	2	3	cor	sub	embed	-	embed
20270000	17	18	cor	phr	shared	interpret	embed
20270000	19	20	sub	cor	pc-arg	interpret	embed
20270000	19	21	sub	cor	embed	-	embed
20270000	19	22	sub	adv	embed	interpret	embed
20270000	20	21	cor	cor	embed	-	embed
20270000	20	22	cor	adv	pc-rel	interpret	embed
20270000	21	22	cor	adv	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
20280000	1	2	phr	cor	pc-arg	interpret	embed
20280000	4	5	adv	cor	shared	semantic	shared
20280000	10	11	cor	cor	embed	-	embed
20280000	13	14	cor	cor	pc-arg	interpret	embed
20280000	17	18	adv	sub	pc-rel	syntactic	pc-rel
20280000	21	22	sub	cor	pc-rel	syntactic	pc-rel
20280000	31	32	sub	cor	embed	-	embed
20290000	5	6	adv	phr	embed	-	embed
20290000	8	9	adv	cor	embed	-	embed
20290000	26	27	sub	cor	pc-arg	interpret	embed
20290000	34	35	cor	phr	shared	multi	ident
20290000	39	40	cor	phr	shared	multi	ident
20290000	43	44	cor	cor	shared	interpret	embed
20300000	2	3	cor	sub	embed	-	embed
20300000	7	8	cor	cor	pc-rel	syntactic	pc-rel
20300000	14	15	sub	cor	embed	-	embed
20300000	18	19	cor	sub	embed	-	embed
20300000	18	20	cor	adv	pc-rel	missing	embed
20300000	19	20	sub	adv	pc-rel	missing	embed
20300000	26	27	cor	cor	embed	-	embed
20300000	27	28	cor	sub	pc-rel	leftout	embed
20300000	27	29	cor	cor	pc-rel	interpret	embed
20300000	27	30	cor	sub	pc-rel	leftout	embed
20300000	27	31	cor	cor	pc-rel	interpret	embed
20300000	28	29	sub	cor	pc-arg	interpret	embed
20300000	28	30	sub	sub	shared	leftout	embed
20300000	28	31	sub	cor	shared	interpret	embed
20300000	29	30	cor	sub	shared	interpret	embed
20300000	30	31	sub	cor	shared	interpret	embed
20300000	32	33	cor	adv	shared	interpret	embed
20300000	34	35	cor	cor	pc-rel	missing	embed
20300000	34	36	cor	adv	pc-rel	missing	embed
20300000	35	36	cor	adv	shared	multi	ident
20300000	40	41	adv	sub	embed	-	embed
20310000	2	3	cor	adv	shared	multi	ident
20310000	2	4	cor	cor	embed	-	embed
20310000	3	4	adv	cor	embed	-	embed
20310000	4	5	cor	cor	embed	-	embed
20310000	13	14	cor	phr	shared	multi	ident
20310000	16	17	cor	cor	pc-rel	missing	embed
20310000	20	21	cor	cor	embed	-	embed
20310000	22	23	cor	cor	embed	-	embed
20310000	24	25	cor	cor	embed	-	embed
20310000	24	26	cor	phr	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
20310000	25	26	cor	phr	embed	interpret	embed
20310000	30	31	sub	adv	pc-rel	missing	embed
20310000	32	33	sub	cor	pc-rel	syntactic	pc-rel
20310000	42	43	sub	cor	embed	-	embed
20310000	49	50	adv	sub	pc-rel	syntactic	pc-rel
20320000	11	12	adv	cor	pc-rel	syntactic	pc-rel
20330000	4	5	sub	cor	embed	-	embed
20330000	16	17	sub	phr	embed	-	embed
20330000	24	25	sub	cor	embed	-	embed
20330000	25	26	cor	cor	embed	interpret	embed
20330000	25	27	cor	sub	embed	-	embed
20330000	26	27	cor	sub	embed	-	embed
20330000	33	34	cor	sub	embed	-	embed
20340000	7	8	sub	phr	embed	-	embed
20340000	30	31	sub	cor	pc-rel	syntactic	pc-rel
20350000	2	3	phr	sub	embed	-	embed
20350000	4	5	phr	sub	embed	-	embed
20350000	8	9	adv	sub	embed	-	embed
20350000	11	12	sub	cor	pc-rel	syntactic	pc-rel
20350000	17	18	cor	sub	pc-rel	syntactic	pc-rel
20350000	20	21	cor	phr	shared	multi	ident
20350000	22	23	sub	phr	embed	-	embed
20350000	24	25	cor	adv	embed	-	embed
20350000	25	26	adv	sub	embed	-	embed
20350000	28	29	sub	cor	embed	-	embed
20350000	31	32	sub	cor	pc-rel	interpret	embed
20350000	31	33	sub	sub	pc-rel	interpret	embed
20350000	31	34	sub	adv	pc-rel	interpret	embed
20350000	32	33	cor	sub	pc-arg	interpret	embed
20350000	32	34	cor	adv	pc-rel	interpret	embed
20350000	33	34	sub	adv	pc-rel	interpret	embed
20350000	35	36	adv	sub	embed	-	embed
20350000	35	37	adv	adv	shared	interpret	embed
20350000	36	37	sub	adv	embed	-	embed
20350000	37	38	adv	cor	pc-rel	syntactic	pc-rel
20350000	37	39	adv	cor	embed	-	embed
20350000	38	39	cor	cor	pc-rel	syntactic	pc-rel
20360000	1	2	sub	adv	embed	-	embed
20360000	19	20	cor	sub	embed	-	embed
20360000	26	27	cor	adv	shared	interpret	embed
20360000	27	28	adv	cor	embed	-	embed
20360000	34	35	cor	sub	pc-rel	missing	embed
20360000	43	44	sub	sub	pc-rel	syntactic	pc-rel
20360000	43	45	sub	cor	pc-rel	syntactic	pc-rel

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
20360000	43	47	sub	adv	pc-rel	syntactic	pc-rel
20360000	44	45	sub	cor	pc-rel	leftout	embed
20360000	44	47	sub	adv	pc-rel	leftout	embed
20360000	45	46	cor	cor	pc-rel	syntactic	pc-rel
20360000	45	47	cor	adv	embed	-	embed
20360000	46	47	cor	adv	pc-rel	syntactic	pc-rel
20360000	52	53	sub	cor	pc-arg	interpret	embed
20370000	2	3	cor	adv	pc-rel	syntactic	pc-rel
20370000	12	13	cor	phr	shared	multi	ident
20370000	12	14	cor	sub	embed	-	embed
20370000	13	14	phr	sub	pc-arg	leftout	embed
20370000	15	16	sub	cor	embed	-	embed
20370000	18	19	sub	sub	shared	missing	embed
20370000	23	24	sub	adv	pc-rel	missing	embed
20370000	24	25	adv	cor	pc-arg	interpret	embed
20370000	26	27	sub	phr	embed	-	embed
20370000	27	28	phr	cor	pc-arg	interpret	embed
20370000	28	29	cor	sub	pc-rel	syntactic	pc-rel
20370000	38	39	cor	sub	pc-rel	syntactic	pc-rel
20380000	10	11	cor	adv	pc-rel	syntactic	pc-rel
20380000	12	13	sub	cor	embed	-	embed
20380000	15	16	sub	adv	pc-rel	syntactic	pc-rel
20380000	17	19	sub	adv	pc-rel	syntactic	pc-rel
20380000	18	19	cor	adv	nested	-	nested
20380000	19	20	adv	sub	pc-rel	syntactic	pc-rel
20380000	21	22	cor	adv	pc-arg	syntactic	pc-arg
20380000	22	23	adv	sub	pc-rel	missing	embed
20390000	6	7	cor	sub	pc-rel	syntactic	pc-rel
20390000	10	11	sub	cor	embed	-	embed
20390000	16	17	adv	cor	pc-arg	multi	ident
20390000	19	20	sub	adv	embed	-	embed
20390000	21	22	sub	cor	pc-rel	missing	embed
20390000	33	34	cor	adv	embed	-	embed
20400000	2	3	adv	cor	embed	-	embed
20400000	2	4	adv	adv	shared	interpret	embed
20400000	3	4	cor	adv	embed	-	embed
20400000	8	9	cor	sub	embed	-	embed
20400000	8	10	cor	adv	pc-rel	syntactic	pc-rel
20400000	9	10	sub	adv	pc-rel	syntactic	pc-rel
20400000	17	18	sub	cor	pc-arg	interpret	embed
20400000	24	25	adv	sub	embed	-	embed
20400000	35	36	sub	cor	pc-rel	interpret	embed
20400000	35	37	sub	adv	pc-rel	syntactic	pc-rel
20400000	36	37	cor	adv	pc-rel	syntactic	pc-rel

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
20400000	40	41	cor	phr	embed	-	embed
20410000	8	9	cor	cor	pc-rel	syntactic	pc-rel
20410000	9	10	cor	sub	embed	-	embed
20410000	15	16	sub	cor	embed	-	embed
20420000	8	9	cor	cor	pc-rel	missing	embed
20420000	10	11	cor	adv	shared	interpret	embed
20420000	12	13	sub	cor	embed	-	embed
20420000	13	14	cor	sub	embed	-	embed
20420000	18	19	cor	adv	embed	-	embed
20420000	21	22	cor	adv	pc-rel	missing	embed
20420000	38	39	adv	adv	shared	multi	ident
20420000	38	40	adv	cor	shared	interpret	embed
20420000	43	44	sub	phr	embed	-	embed
20420000	45	46	cor	sub	embed	-	embed
20430000	8	9	cor	cor	embed	-	embed
20430000	11	12	cor	cor	shared	semantic	shared
20430000	25	26	cor	adv	shared	interpret	embed
20440000	4	5	cor	adv	pc-rel	syntactic	pc-rel
20440000	5	6	adv	adv	embed	-	embed
20440000	17	18	cor	phr	pc-rel	syntactic	pc-rel
20440000	18	19	phr	phr	shared	interpret	embed
20440000	26	27	cor	sub	pc-rel	interpret	embed
20440000	26	28	cor	adv	pc-arg	interpret	embed
20440000	27	28	sub	adv	embed	-	embed
20440000	31	32	cor	sub	pc-rel	syntactic	pc-rel
20440000	47	48	cor	cor	pc-rel	syntactic	pc-rel
20440000	51	52	adv	adv	pc-rel	syntactic	pc-rel
20450000	2	3	cor	adv	embed	-	embed
20450000	3	4	adv	cor	shared	interpret	embed
20450000	3	5	adv	adv	shared	interpret	embed
20450000	4	5	cor	adv	shared	multi	ident
20450000	4	6	cor	cor	embed	-	embed
20450000	5	6	adv	cor	pc-arg	interpret	embed
20450000	7	8	cor	cor	pc-rel	missing	embed
20450000	7	9	cor	adv	pc-rel	missing	embed
20450000	8	9	cor	adv	shared	multi	ident
20450000	11	12	phr	cor	embed	-	embed
20460000	18	19	cor	sub	embed	-	embed
20460000	26	27	sub	cor	embed	-	embed
20460000	29	30	cor	sub	embed	-	embed
20460000	29	32	cor	adv	pc-rel	syntactic	pc-rel
20460000	30	32	sub	adv	pc-rel	syntactic	pc-rel
20460000	31	32	phr	adv	nested	-	nested
20460000	34	35	cor	sub	pc-rel	syntactic	pc-rel

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
20460000	38	39	cor	adv	pc-rel	missing	embed
20470000	1	2	sub	adv	pc-rel	missing	embed
20470000	21	22	phr	sub	pc-rel	syntactic	embed
20470000	21	23	phr	cor	shared	interpret	embed
20470000	22	23	sub	cor	pc-rel	syntactic	pc-rel
20470000	25	26	sub	sub	shared	missing	embed
20480000	1	2	cor	phr	pc-rel	syntactic	pc-rel
20480000	1	3	cor	cor	pc-rel	syntactic	pc-rel
20480000	2	3	phr	cor	shared	interpret	embed
20480000	8	9	adv	sub	embed	-	embed
20480000	13	14	sub	cor	shared	interpret	embed
20480000	15	16	cor	sub	embed	-	embed
20480000	18	19	phr	adv	shared	interpret	embed
20480000	20	21	adv	cor	pc-rel	missing	embed
20480000	29	30	phr	sub	embed	-	embed
20480000	37	38	cor	cor	embed	-	embed
20480000	39	40	cor	phr	embed	-	embed
20490000	5	6	adv	adv	pc-rel	missing	embed
20490000	11	12	sub	cor	embed	-	embed
20490000	11	13	sub	adv	embed	-	embed
20490000	12	13	cor	adv	shared	multi	ident
20490000	16	17	sub	cor	embed	-	embed
20490000	26	27	sub	cor	embed	-	embed
20490000	27	28	cor	sub	pc-rel	syntactic	pc-rel
20490000	37	38	adv	cor	shared	interpret	embed
20500000	6	7	sub	cor	embed	-	embed
20500000	13	14	adv	cor	pc-rel	syntactic	pc-rel
20500000	19	20	cor	sub	embed	-	embed
20500000	25	26	sub	cor	pc-rel	syntactic	pc-rel
20500000	25	27	sub	sub	pc-rel	syntactic	pc-rel
20500000	26	27	cor	sub	pc-rel	syntactic	pc-rel
20500000	30	31	cor	sub	pc-arg	interpret	embed
20500000	40	41	sub	cor	embed	-	embed
20510000	2	3	cor	adv	shared	interpret	embed
20510000	5	6	cor	cor	embed	-	embed
20510000	13	14	cor	cor	embed	-	embed
20510000	15	16	cor	adv	shared	interpret	embed
20510000	15	17	cor	adv	pc-rel	interpret	embed
20510000	16	17	adv	adv	pc-arg	interpret	embed
20510000	23	24	cor	adv	nested	-	nested
20510000	29	30	adv	adv	shared	error	ident
20510000	31	34	sub	sub	shared	interpret	embed
20510000	32	33	cor	sub	cross	-	embed
20510000	32	34	cor	sub	shared	interpret	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
20510000	33	34	sub	sub	embed	-	embed
20510000	43	44	cor	adv	pc-arg	interpret	embed
20520000	2	3	sub	cor	embed	-	embed
20520000	3	4	cor	phr	pc-arg	interpret	embed
20520000	4	5	phr	cor	pc-rel	syntactic	pc-rel
20520000	4	6	phr	phr	shared	semantic	shared
20520000	5	6	cor	phr	pc-rel	syntactic	pc-rel
20520000	19	20	sub	cor	embed	-	embed
20520000	23	24	sub	phr	embed	-	embed
20530000	6	7	adv	sub	pc-rel	interpret	embed
20530000	6	8	adv	adv	pc-arg	missing	embed
20530000	7	8	sub	adv	pc-rel	missing	embed
20530000	8	9	adv	adv	pc-arg	interpret	embed
20530000	20	21	cor	cor	pc-rel	interpret	embed
20530000	27	28	adv	cor	embed	-	embed
20530000	28	29	cor	cor	embed	-	embed
20530000	36	37	sub	cor	embed	-	embed
20530000	38	39	sub	cor	pc-rel	syntactic	pc-rel
20530000	39	40	cor	cor	shared	interpret	embed
20530000	41	42	sub	cor	embed	-	embed
20530000	46	47	sub	cor	embed	-	embed
20530000	47	48	cor	sub	embed	-	embed
20530000	50	51	cor	cor	nested	-	nested
20530000	52	53	cor	cor	embed	-	embed
20530000	52	54	cor	adv	embed	-	embed
20530000	53	54	cor	adv	shared	multi	ident
20540000	4	5	sub	adv	pc-rel	missing	embed
20540000	5	6	adv	sub	embed	-	embed
20540000	11	12	cor	sub	embed	-	embed
20540000	21	22	cor	phr	shared	multi	ident
20540000	29	30	sub	adv	embed	-	embed
20540000	41	42	sub	cor	embed	-	embed
20550000	9	10	sub	cor	embed	-	embed
20550000	15	16	sub	cor	embed	-	embed
20550000	20	21	adv	cor	shared	interpret	embed
20550000	23	25	adv	cor	pc-rel	multi	ident
20550000	23	26	adv	adv	shared	interpret	embed
20550000	23	27	adv	cor	pc-rel	interpret	embed
20550000	24	25	adv	cor	pc-rel	syntactic	pc-rel
20550000	24	27	adv	cor	embed	-	embed
20550000	25	26	cor	adv	pc-rel	syntactic	pc-rel
20550000	25	27	cor	cor	pc-rel	syntactic	pc-rel
20550000	26	27	adv	cor	pc-arg	interpret	embed
20550000	27	28	cor	cor	shared	interpret	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
20550000	31	32	sub	cor	embed	-	embed
20550000	37	38	cor	adv	embed	-	embed
20550000	38	39	adv	cor	pc-rel	syntactic	pc-rel
20560000	1	2	sub	cor	pc-rel	syntactic	pc-rel
20560000	1	3	sub	sub	overlap	syntactic	pc-rel
20560000	1	4	sub	phr	pc-rel	syntactic	pc-rel
20560000	2	3	cor	sub	embed	-	embed
20560000	2	4	cor	phr	pc-rel	syntactic	pc-rel
20560000	3	4	sub	phr	pc-rel	syntactic	pc-rel
20560000	4	5	phr	cor	shared	interpret	embed
20560000	7	8	phr	cor	pc-rel	syntactic	pc-rel
20560000	9	10	cor	phr	shared	multi	ident
20560000	9	11	cor	sub	pc-rel	interpret	embed
20560000	10	11	phr	sub	pc-rel	interpret	embed
20560000	15	16	cor	sub	embed	-	embed
20560000	24	25	cor	sub	embed	-	embed
20560000	29	30	cor	adv	shared	interpret	embed
20560000	31	32	cor	phr	embed	-	embed
20560000	35	36	sub	cor	pc-rel	syntactic	pc-rel
20560000	37	38	cor	sub	pc-rel	syntactic	pc-rel
20560000	42	43	cor	cor	pc-rel	syntactic	pc-rel
20560000	43	44	cor	adv	shared	interpret	embed
20560000	45	46	adv	adv	shared	interpret	embed
20560000	46	47	adv	adv	pc-rel	missing	embed
20560000	49	50	cor	phr	pc-rel	missing	embed
20570000	19	20	sub	sub	pc-arg	interpret	embed
20570000	19	22	sub	adv	embed	-	embed
20570000	20	22	sub	adv	embed	-	embed
20570000	21	22	adv	adv	nested	-	nested
20570000	22	23	adv	sub	pc-rel	syntactic	pc-rel
20570000	24	25	sub	sub	embed	-	embed
20570000	34	35	cor	cor	embed	-	embed
20570000	37	38	phr	cor	embed	-	embed
20580000	1	2	cor	cor	pc-rel	syntactic	pc-rel
20580000	3	4	sub	adv	pc-rel	syntactic	pc-rel
20580000	11	12	cor	cor	nested	-	nested
20580000	26	27	cor	cor	embed	-	embed
20590000	2	3	cor	adv	pc-rel	interpret	embed
20590000	8	9	cor	cor	pc-rel	syntactic	pc-rel
20590000	14	15	cor	sub	embed	-	embed
20590000	18	19	cor	cor	shared	interpret	embed
20600000	3	4	cor	cor	embed	-	embed
20600000	10	11	sub	phr	embed	-	embed
20600000	12	13	sub	phr	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
20600000	16	17	cor	cor	pc-arg	interpret	embed
20600000	16	18	cor	adv	shared	syntactic	pc-rel
20600000	16	19	cor	cor	pc-arg	interpret	embed
20600000	16	20	cor	adv	pc-rel	interpret	embed
20600000	17	18	cor	adv	pc-arg	syntactic	pc-rel
20600000	17	19	cor	cor	embed	-	embed
20600000	17	20	cor	adv	pc-rel	interpret	embed
20600000	18	19	adv	cor	pc-rel	interpret	embed
20600000	18	20	adv	adv	pc-rel	syntactic	pc-rel
20600000	19	20	cor	adv	pc-arg	multi	ident
20600000	19	21	cor	cor	embed	-	embed
20600000	20	21	adv	cor	pc-rel	leftout	embed
20600000	23	24	cor	adv	shared	multi	ident
20600000	33	34	cor	adv	embed	-	embed
20600000	36	37	cor	phr	shared	interpret	embed
20600000	38	39	sub	phr	embed	-	embed
20600000	40	41	cor	sub	shared	interpret	embed
20600000	48	49	sub	cor	pc-arg	interpret	embed
20600000	51	52	sub	cor	embed	-	embed
20610000	3	4	cor	cor	embed	-	embed
20610000	5	6	adv	sub	embed	-	embed
20610000	11	12	phr	sub	embed	-	embed
20610000	20	21	cor	phr	embed	-	embed
20610000	28	29	sub	cor	embed	-	embed
20610000	29	30	cor	phr	shared	interpret	embed
20620000	1	2	sub	adv	embed	-	embed
20620000	2	3	adv	sub	pc-arg	interpret	embed
20620000	9	10	cor	cor	embed	-	embed
20620000	13	15	sub	adv	pc-rel	missing	embed
20620000	14	15	sub	adv	nested	-	nested
20630000	2	3	cor	sub	embed	-	embed
20630000	7	8	sub	cor	pc-rel	syntactic	pc-rel
20630000	11	12	sub	cor	embed	-	embed
20630000	14	15	sub	cor	pc-rel	syntactic	pc-rel
20630000	14	16	sub	phr	pc-rel	syntactic	pc-rel
20630000	15	16	cor	phr	shared	multi	ident
20630000	17	18	sub	sub	pc-rel	syntactic	pc-rel
20630000	23	24	adv	sub	embed	-	embed
20630000	26	29	sub	adv	pc-rel	syntactic	pc-rel
20630000	27	29	sub	adv	pc-rel	syntactic	pc-rel
20630000	28	29	cor	adv	pc-rel	syntactic	pc-rel
20630000	29	30	adv	sub	pc-rel	syntactic	pc-rel
20630000	31	32	sub	cor	pc-rel	syntactic	pc-rel
20630000	34	35	sub	cor	embed	-	embed

Continued on next page

Table D.1 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
20630000	35	36	cor	sub	pc-rel	missing	embed
20630000	37	38	cor	cor	pc-rel	syntactic	pc-rel
20630000	44	45	sub	sub	overlap	interpret	embed
20630000	49	50	cor	sub	embed	-	embed
20630000	49	51	cor	adv	embed	-	embed
20630000	50	51	sub	adv	embed	interpret	embed
20640000	13	14	cor	cor	pc-rel	syntactic	pc-rel
20640000	17	18	cor	sub	pc-rel	syntactic	pc-rel

Table D.2: List of all configurations, reasons for tree violations, and the results of reannotation in the STC Demo

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
012_090128_00002	8	9	cor	cor	pc-arg	interpret	embed
021_090501_00013	3	4	adv	cor	pc-arg	interpret	shared
021_090501_00013	4	5	cor	sub	embed	-	embed
021_090501_00013	4	6	cor	adv	pc-arg	missing	embed
021_090501_00013	5	6	sub	adv	pc-rel	missing	embed
024_091113_00031	1	2	phr	phr	shared	interpret	embed
024_091113_00031	3	4	phr	cor	shared	interpret	embed
024_091113_00031	10	11	sub	phr	pc-rel	leftout	embed
024_091113_00031	16	17	adv	cor	cross	-	cross
024_091113_00031	19	20	cor	cor	shared	semantic	shared
024_091113_00031	24	25	cor	cor	shared	semantic	shared
024_091113_00031	26	27	cor	cor	pc-rel	interpret	embed
052_090819_00016	5	9	cor	adv	nested	-	nested
052_090819_00016	5	11	cor	adv	nested	-	nested
052_090819_00016	6	9	adv	adv	nested	-	nested
052_090819_00016	6	11	adv	adv	nested	-	nested
052_090819_00016	7	9	adv	adv	nested	-	nested
052_090819_00016	7	11	adv	adv	nested	-	nested
052_090819_00016	8	9	cor	adv	nested	-	nested
052_090819_00016	8	11	cor	adv	nested	-	nested
052_090819_00016	9	11	adv	adv	pc-arg	missing	embed
052_090819_00016	10	11	sub	adv	nested	-	nested
052_090819_00016	13	16	cor	adv	nested	-	nested
052_090819_00016	14	16	cor	adv	nested	-	nested
052_090819_00016	15	16	sub	adv	nested	-	nested
061_090622_00020	7	8	sub	phr	embed	-	embed
061_090622_00020	10	11	cor	adv	nested	-	nested
061_090622_00020	13	14	cor	phr	embed	-	embed
061_090622_00020	15	17	cor	adv	embed	-	embed
061_090622_00020	16	17	cor	adv	pc-arg	interpret	embed
061_090622_00020	18	19	adv	cor	pc-arg	interpret	shared
061_090622_00020	20	21	sub	phr	pc-rel	interpret	embed
061_090622_00020	20	22	sub	cor	pc-rel	interpret	embed
061_090622_00020	21	22	phr	cor	shared	interpret	embed
061_090622_00020	24	25	cor	adv	embed	-	embed
061_090622_00020	29	30	cor	cor	pc-rel	missing	embed
061_090622_00020	35	36	adv	adv	shared	missing	embed
061_090622_00020	39	40	cor	cor	pc-arg	missing	embed
061_090622_00020	43	44	cor	cor	shared	missing	embed
061_090622_00020	47	49	cor	cor	nested	-	nested
061_090622_00020	48	49	adv	cor	nested	-	nested
061_090622_00020	50	51	adv	cor	pc-arg	interpret	embed
061_090622_00020	52	53	cor	cor	shared	semantic	shared
Continued on next page							

Table D.2 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
061_090622_00020	55	56	adv	adv	pc-rel	missing	embed
061_090622_00020	55	57	adv	adv	pc-rel	missing	embed
061_090622_00020	55	58	adv	adv	pc-rel	missing	embed
061_090622_00020	55	59	adv	adv	pc-arg	missing	embed
061_090622_00020	55	60	adv	adv	pc-arg	missing	embed
061_090622_00020	55	61	adv	adv	pc-arg	missing	embed
061_090622_00020	56	57	adv	adv	shared	missing	embed
061_090622_00020	57	58	adv	adv	pc-arg	interpret	embed
061_090622_00020	57	59	adv	adv	pc-arg	missing	embed
061_090622_00020	57	60	adv	adv	pc-arg	missing	embed
061_090622_00020	57	61	adv	adv	pc-arg	missing	embed
061_090622_00020	58	59	adv	adv	embed	-	embed
061_090622_00020	58	60	adv	adv	embed	-	embed
061_090622_00020	58	61	adv	adv	embed	-	embed
061_090622_00020	59	60	adv	adv	shared	missing	embed
061_090622_00020	59	61	adv	adv	shared	missing	embed
061_090622_00020	60	61	adv	adv	shared	missing	embed
061_090622_00020	62	63	cor	adv	shared	interpret	embed
061_090622_00020	62	64	cor	adv	shared	missing	embed
061_090622_00020	63	64	adv	adv	shared	missing	embed
061_090622_00020	65	66	cor	adv	pc-arg	multi	ident
061_090622_00020	67	70	adv	adv	shared	missing	embed
061_090622_00020	68	69	adv	adv	shared	missing	embed
061_090622_00020	68	70	adv	adv	nested	-	nested
061_090622_00020	69	70	adv	adv	nested	-	nested
061_090622_00020	72	73	cor	cor	pc-rel	syntactic	pc-rel
061_090622_00020	72	85	cor	adv	pc-rel	syntactic	pc-rel
061_090622_00020	73	85	cor	adv	shared	interpret	embed
061_090622_00020	74	75	sub	adv	embed	-	embed
061_090622_00020	74	85	sub	adv	nested	-	nested
061_090622_00020	75	76	adv	cor	embed	-	embed
061_090622_00020	75	77	adv	sub	pc-rel	interpret	embed
061_090622_00020	75	85	adv	adv	nested	-	nested
061_090622_00020	76	77	cor	sub	embed	-	embed
061_090622_00020	76	85	cor	adv	nested	-	nested
061_090622_00020	77	85	sub	adv	nested	-	nested
061_090622_00020	78	80	cor	adv	pc-arg	interpret	embed
061_090622_00020	78	85	cor	adv	nested	-	nested
061_090622_00020	79	80	cor	adv	pc-arg	interpret	embed
061_090622_00020	79	81	cor	adv	nested	-	nested
061_090622_00020	79	85	cor	adv	nested	-	nested
061_090622_00020	80	81	adv	adv	shared	missing	embed
061_090622_00020	80	85	adv	adv	nested	-	nested
061_090622_00020	81	85	adv	adv	nested	-	nested

Continued on next page

Table D.2 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
061_090622_00020	82	85	cor	adv	nested	-	nested
061_090622_00020	83	85	cor	adv	nested	-	nested
061_090622_00020	84	85	cor	adv	nested	-	nested
061_090622_00020	85	86	adv	adv	pc-arg	interpret	embed
061_090622_00020	86	87	adv	adv	pc-arg	interpret	embed
061_090622_00020	87	88	adv	cor	pc-rel	interpret	embed
061_090622_00020	90	91	adv	adv	shared	missing	embed
061_090622_00020	91	92	adv	adv	shared	missing	embed
061_090622_00020	95	96	adv	adv	pc-arg	interpret	shared
061_090622_00020	96	97	adv	sub	embed	-	embed
061_090622_00020	102	103	adv	adv	shared	interpret	embed
061_090622_00020	106	107	phr	cor	embed	-	embed
061_090622_00020	106	108	phr	cor	embed	-	embed
061_090622_00020	107	108	cor	cor	shared	interpret	embed
061_090622_00020	109	110	cor	phr	nested	-	nested
061_090622_00020	110	111	phr	sub	embed	-	embed
061_090622_00020	110	112	phr	adv	shared	leftout	embed
061_090622_00020	111	112	sub	adv	embed	-	embed
061_090622_00020	112	113	adv	adv	shared	interpret	embed
069_090610_00015	1	2	phr	phr	shared	semantic	shared
072_090913_00006	6	7	cor	phr	pc-arg	interpret	embed
075_090622_00003	2	3	cor	phr	shared	interpret	embed
075_090622_00003	5	6	cor	phr	shared	multi	ident
075_090629_00023	1	2	adv	phr	shared	interpret	embed
112_090217_00001	2	3	cor	cor	embed	-	embed
112_090217_00001	7	8	cor	adv	shared	multi	ident
113_090404_00004	2	3	adv	adv	embed	-	embed
113_090404_00004	8	9	cor	cor	embed	-	embed
116_090206_00018	1	2	sub	cor	embed	-	embed
117_090310_00019	20	21	sub	phr	pc-rel	syntactic	pc-rel
117_090310_00019	21	23	phr	adv	pc-arg	interpret	embed
117_090310_00019	22	23	cor	adv	pc-arg	multi	ident
119_090119_00027	3	4	cor	cor	pc-arg	interpret	embed
119_090119_00027	7	8	cor	cor	pc-rel	syntactic	pc-rel
119_090119_00027	7	9	cor	cor	shared	interpret	embed
119_090119_00027	8	9	cor	cor	pc-rel	syntactic	pc-rel
119_090119_00027	14	15	cor	cor	embed	-	embed
119_090119_00027	18	19	cor	cor	embed	-	embed
119_090119_00027	18	20	cor	cor	pc-rel	interpret	embed
119_090119_00027	18	21	cor	cor	pc-arg	interpret	embed
119_090119_00027	19	20	cor	cor	embed	-	embed
119_090119_00027	19	21	cor	cor	pc-arg	interpret	embed
119_090119_00027	20	21	cor	cor	shared	interpret	embed
119_090119_00027	24	25	cor	cor	overlap	missing	embed

Continued on next page

Table D.2 – continued from previous page

File No	Rel1	Rel2	Type1	Type2	Initial	Reason	Final
119_090119_00027	24	26	cor	phr	pc-arg	interpret	embed
119_090119_00027	25	26	cor	phr	pc-arg	multi	ident
119_090119_00027	34	35	cor	cor	pc-rel	syntactic	pc-rel
119_090119_00027	36	37	cor	sub	pc-rel	interpret	embed
119_090119_00027	36	38	cor	adv	pc-rel	syntactic	pc-rel
119_090119_00027	37	38	sub	adv	pc-rel	missing	embed
119_090119_00027	38	39	adv	cor	shared	interpret	embed
119_090119_00027	41	43	cor	adv	pc-rel	missing	embed
119_090119_00027	42	43	cor	adv	pc-rel	missing	embed
119_090123_00029	1	2	sub	adv	pc-arg	interpret	embed
119_090123_00029	2	3	adv	cor	shared	interpret	embed
119_090123_00029	8	9	sub	cor	pc-rel	syntactic	pc-rel
119_090123_00029	9	10	cor	sub	pc-rel	missing	embed
119_090123_00029	12	13	cor	adv	embed	-	embed
119_090123_00029	14	15	cor	adv	overlap	interpret	embed
119_090501_00026	2	3	cor	adv	shared	interpret	embed
119_090501_00026	7	9	sub	cor	embed	-	embed
119_090501_00026	8	9	cor	cor	nested	-	nested
119_090501_00026	10	11	adv	cor	nested	-	nested
119_090501_00026	13	14	cor	sub	pc-rel	interpret	embed
119_090501_00026	16	17	sub	cor	embed	-	embed
119_090501_00026	19	20	sub	adv	embed	-	embed
119_090501_00026	22	23	cor	cor	pc-rel	syntactic	pc-rel
119_090501_00026	25	26	cor	cor	pc-rel	syntactic	pc-rel
119_090501_00026	27	28	sub	phr	embed	-	embed
119_090501_00026	31	32	sub	sub	shared	missing	embed
119_090501_00026	31	33	sub	sub	shared	missing	embed
119_090501_00026	32	33	sub	sub	shared	missing	embed
119_090531_00075	1	2	cor	sub	embed	-	embed
119_090531_00075	10	12	cor	adv	pc-rel	syntactic	pc-rel
119_090531_00075	19	20	cor	adv	pc-rel	syntactic	pc-rel
119_090531_00075	20	21	adv	cor	pc-rel	syntactic	pc-rel
119_090531_00075	27	28	cor	cor	pc-rel	syntactic	pc-rel

Legend:

Syntactic types of discourse connectives:

cor: Coordinating conjunction

sub: Subordinating conjunction

adv: Discourse adverbial

phr: Phrasal expression

Initial and final configurations:

indep: Independent relations

embed: Fully embedded relations

nested: Nested relations

shared: Shared argument

pc-arg: Properly contained argument

pc-rel: Properly contained relation

partial: Partially overlapping arguments

cross: Pure crossing

Reasons for tree violations:

missing: Relations yet unannotated

multi: Multiple connectives

leftout: Material left out due to guidelines

error: Annotation error

interpret: Reinterpretable relations

syntactic: Syntactic asymmetry

semantic: Semantic tree violation

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Demirşahin Işın

Nationality: Turkish (TC)

Date and Place of Birth: September 27th, Bursa, TURKEY

Marital Status: Single

Phone: +90 312 210 38 09

EDUCATION

Degree	Institution	Year of Graduation
M.S.	Middle East Technical University Cognitive Science	2008
B.S.	Middle East Technical University Computer Education and Instructional Technologies	2004
High School	Eskişehir Fatih Fen Lisesi	1999

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
Nov 2014 - Present	Google via ManAsset	Analytical Linguistic Project Manager
May 2014 - Nov 2014	TextLink	Researcher Web Administrator
Jan 2009- Dec 2013	Middle East Technical University	Research Assistant Researcher for projects
Apr 2011-Dec 2013	Turkish Discourse Bank METU BAP	BAP-07-04-2011-005 BAP-07-04-2012-001 BAP-07-04-2013-003
Oct 2007 - Feb 2011	Turkish Discourse Bank METU BIDEB	Researcher for TUBITAK project 107E156
May 2005 - Oct 2005	Bilemek Information-Education Co. Ltd.	Instructional Technologist

PUBLICATIONS

Journals

Zeyrek, D., Demirşahin, I., Sevdik-Çallı, A. B., Çakıcı, R. (2013). Turkish Discourse Bank: Porting a discourse annotation style to a morphologically rich language. *Dialog & Discourse* 4 (2) pp. 174-184.

Refereed Conferences

Demirşahin, I., Zeyrek, D. (2014). Annotating Discourse Connectives in Spoken Turkish. In *Proceedings of the COLING2014. LAW VIII. The 8th Linguistic Annotation Workshop*.

Demirşahin, I., Oztürel, A., Bozşahin, C., Zeyrek, D. (2013). Applicative Structures and Immediate Discourse in the Turkish Discourse Bank. In *Proceedings of the ACL 2013. LAW VII&ID. The 7th Linguistic Annotation Workshop & Interoperability with Discourse*.

Demirşahin, I. (2012). Discourse Structure in Simultaneous Spoken Turkish. In *Proceedings of the ACL2012 Student Research Workshop*.

Demirşahin, I., Yalçınkaya, İ., Zeyrek, D. (2012). Pair Annotation: Adaption of Pair Programming to Corpus Annotation. In *Proceedings of the ACL 2012. LAW VI. The Sixth Linguistic Annotation Workshop*.

Demirşahin, I., Sevdik-Çallı, A., Balaban, H. O., Çakıcı, R., Zeyrek, D. (2012). Turkish Discourse Bank: Ongoing Developments. In *Proceedings of LREC 2012. The First Turkic Languages Workshop*.

Zeyrek, D., Demirşahin, I., Sevdik-Çallı, A., Balaban, H. O., Yalçınkaya, İ., Turan, U. D. (2010). The Annotation Scheme of the Turkish Discourse Bank and Evaluation of Inconsistent Annotations. In *Proceedings of the ACL 2010. LAW IV. The Fourth Linguistic Annotation Workshop*.

Demirşahin, I. (2010) Information Structural Properties of Turkish Discourse Connectives. In *Proceedings of the ICTL2010 15th International Conference on Turkish Linguistics*.

Zeyrek, D., Demirşahin, I., Sevdik Çallı, A. B., Ogel Balaban, H. (2010). Bu, şu, o and Their Referent types in Turkish Discourse Bank. In *Proceedings of the ICTL2010 15th International Conference on Turkish Linguistics*.

Bozşahin, C., Zeyrek, D., Demirsahin, I. (2010) Soylem ve Yapı [Structure and Discourse]. 24. Ulusal Dilbilim Kurultayı. [In *Proceedings of the 24th Annual Meeting of Linguistics*.]

Zeyrek, D., Turan, U. D., Bozşahin, C., Çakıcı, R., Sevdik-Çallı, A., Demirşahin, Aktaş, B., Yalçınkaya, İ., Ogel, H. (2009). Annotating Subordinators in the Turkish Discourse Bank. In *Proceedings of the ACL-IJCNLP, LAW III, The Third Linguistic Annotation Workshop*.

Zeyrek, D., Demirşahin, I., Sevdik-Çallı, A.B. (2008). ODTU Metin Düzeyinde İşaretlenmiş Derlem Projesi Tanıtımı [Introduction to Turkish Discourse Bank Project]. In *Proceedings of the Mersin Symposium*.

Zeyrek, D., Turan, U. D., Demirşahin, (2008). Structural and presuppositional connectives in Turkish. In Proceedings of the CID III, Constraints in Discourse 3.

Book Chapters

Zeyrek, D., Demirşahin, I., Bozşahin, C. (Forthcoming) Turkish Discourse Bank: Connectives and Their Configurations. In Kemal Oflazer and Murat Saraçlar (eds.) Studies in Turkish Language Processing. Springer Verlag.

Demirşahin, I., Zeyrek, D. (Forthcoming) Turkish Discourse Bank. In Nancy and James Pustejovsky (eds.) Handbook of Linguistic Annotation. Springer.

Zeyrek, D., Demirşahin, I., Turan, U. D., Çakıcı, R. (2012) A corpus-based analysis of Fakat, Yoksa, Ayrıca. In Anton Benz, Peter Kuehnllein, Manfred Stede (eds). Constraints in Discourse III Amsterdam, The Netherlands: John Benjamins.

Masters Thesis

Demirşahin, I. (2008). Connective Position, Argument Order and Structure of Discourse Connectives in Written Turkish Texts. MSc Thesis, ODTÜ, Ankara.

TECHNICAL SKILLS

Research Tools: Turkish Discourse Treebank, METU Turkish Corpus, METU-Sabancı Turkish Treebank, METU Spoken Turkish Corpus, TextSTAT, SPSS

Programming Literacy: C++, Python, XML, HTML, PHP, SQL

Operating Systems: OS X, Linux (Ubuntu, Goobuntu), Windows

Other: Eclipse, Graphviz, SVN, Google Docs, Office Programs, LaTeX, Adobe Photoshop

LANGUAGES

Turkish (Native)
Crimean Tatar (Native)
English (Fluent)
Japanese (Intermediate)
Karaim (Intermediate)
French (Intermediate)
German (Beginner)
Chinese (Beginner)

HONORS AND AWARDS

ACL Student Research Fellow (2012)
LOT Winter School International Graduate Fellow (2011)
TÜBİTAK Domestic Graduate Fellow (2005-2007)

ACADEMIC MEMBERSHIPS

Association for Computational Linguistics (2012)
Laboratory for Computational Studies of Language (Since 2007)
Ankara Linguistic Circle (Since 2006)

EXTRACURRICULAR MEMBERSHIPS

METU Office of Sports - Yoga (2012 - 2014)
METU Office of Sports - Free-style Combat (2011 - 2014)
METU Conficius Institute - Tai Chi (2011-2012)
METU Office of Sports - Pilates (2008 - 2014)
METU Office of Sports - Sports Nutrition Certificate Program (2008)
METU Science Fiction and Fantasy Society Head of the Board of Directives (2003)
METU Science Fiction and Fantasy Society Member (1999 - 2008)

OTHER INTERESTS

Science Fiction and Fantasy Literature
Role Playing Games
Board Games
Computer and Mobile Games
Nutrition and Fitness
Environment and Sustainability (WWF supporter since 2009)
Human Rights (Amnesty International supporter since 2014)