Intercoder agreement in discourse analysis

Seminar in methodology and statistics
13th May 2009

Ildikó Berzlánovich, Myrthe Faber
University of Groningen
Center for Language and Cognition Groningen

Outline

PART 1: Intercoder agreement for the study of discourse structure (Ildikó)

PART 2: Intercoder agreement in conversation analysis (Myrthe)

Outline – Discourse structure

- Own research
- Annotation problems
- Intercoder agreement measures (percentage agreement, Cohen's kappa)
- Practice so far
- Decisions for own research

General aim:

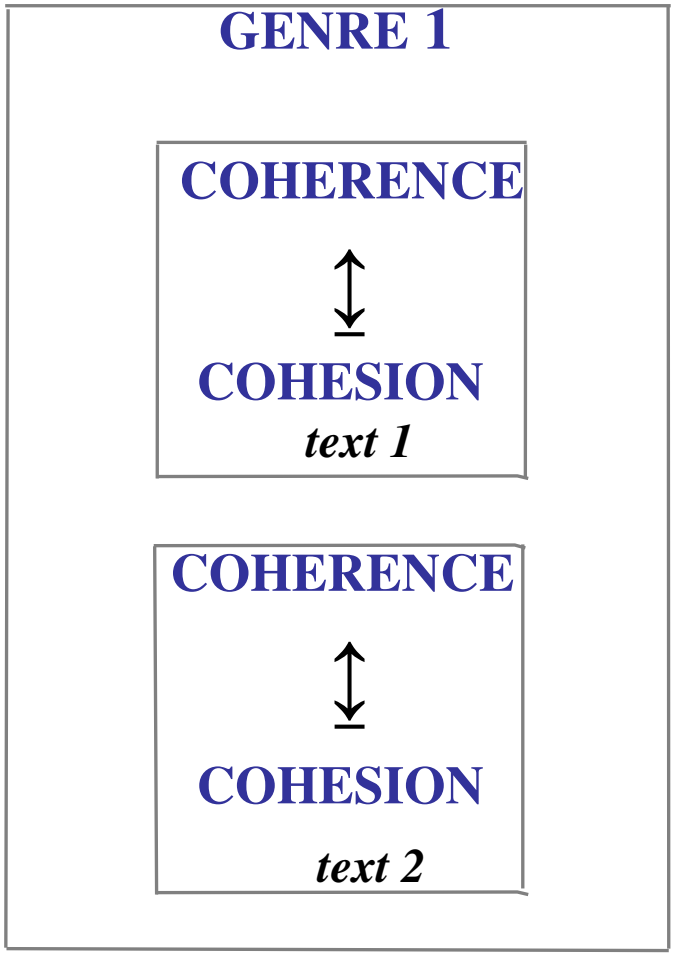
interaction between coherence and lexical cohesion across genres

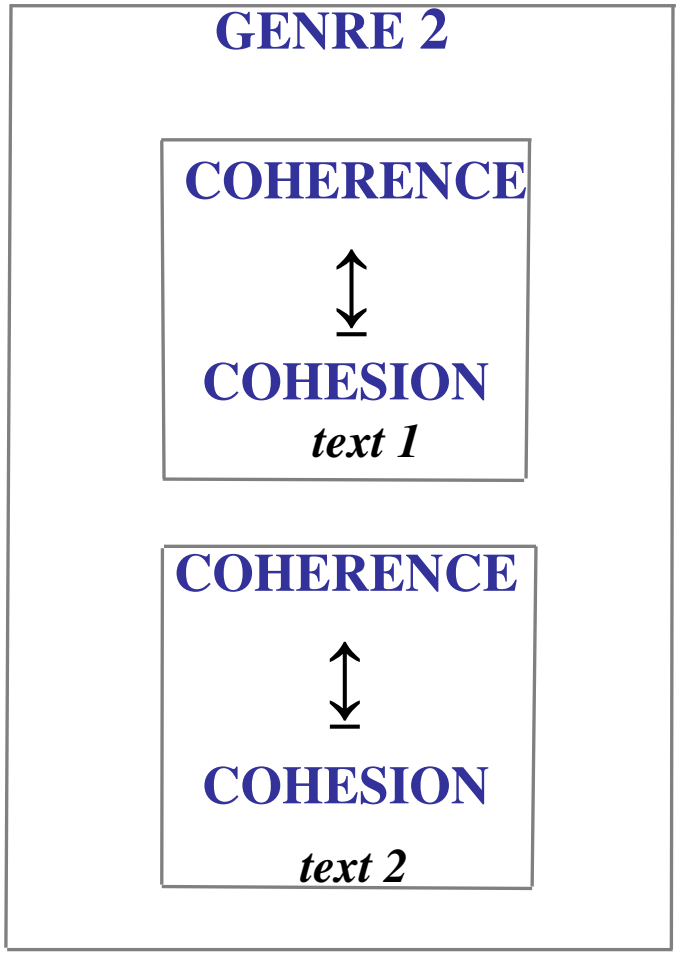
Alignment hypothesis:

Lexical cohesion is more closely aligned with coherence in thematically organized texts than in intentionally organized texts.

Specifically:

- close alignment in expository texts
- less or no alignment in persuasive texts


GENRE 1
COHERENCE
COHESION
text 1
COHERENCE
COHESION
text 2


GENRE 2
COHERENCE
COHESION
text 1
COHERENCE
COHESION
text 2

Genre

- class of communicative events with common communicative purposes shared in a discourse community (Swales 1990)
- genre-specific move structure

Coherence

- underlying relations between discourse units in text

Cohesion

- semantic relations between surface elements in text

Texts

- expository texts: encyclopedia entries (EE01, EE02)
- persuasive texts: fundraising letters (FL01, FL02)

Encyclopedia entries

1. name the object
2. define the object
3. describe in general (e.g., size, age, category)
4. describe details (e.g., surface, past/future development, discovery)

Fundraising letters - seven moves (Upton 2002)

1. get attention
2. introduce the cause and/or establish credentials of organization
3. solicit response
4. offer incentive
5. reference insert
6. express gratitude
7. conclude with pleasantries

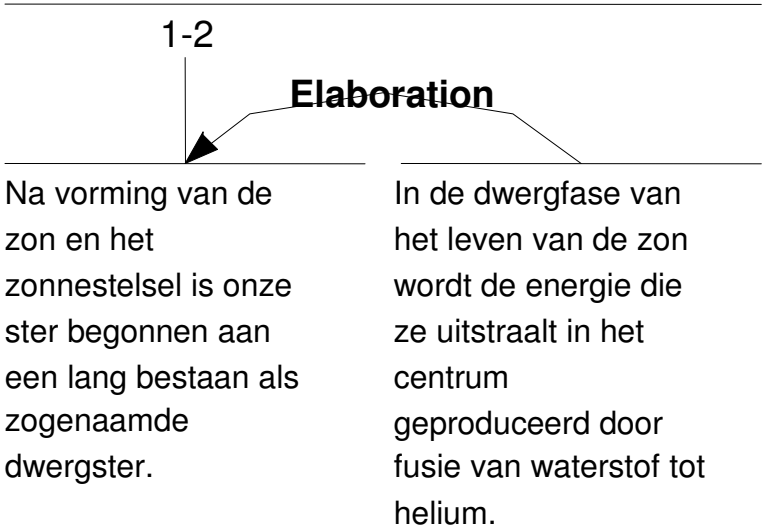
Fragment from EE01 (*De Zon*)

Na vorming van de zon en het zonnestelsel is onze ster begonnen aan een lang bestaan als zogenaamde dwergster. In de dwergfase van het leven van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium. De zon is nu ongeveer vijf miljard jaar oud en heeft nog brandstof voor nog eens vijf miljard jaar.

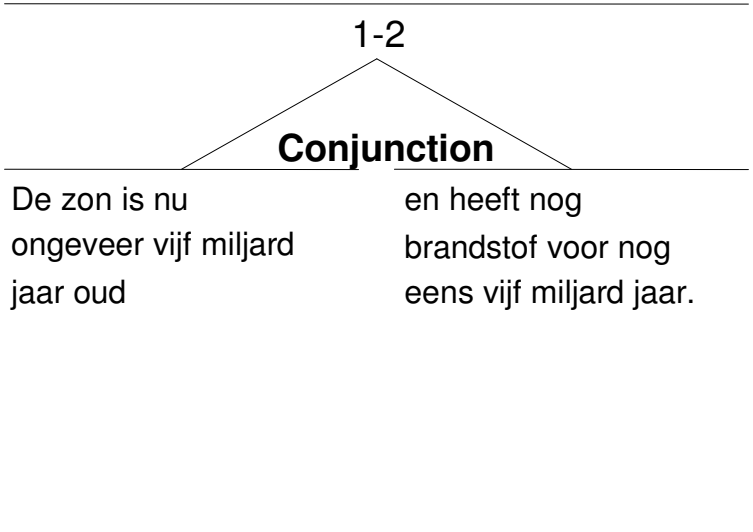
After the forming of the sun and the solar system, our star began its long existence as a so-called dwarf star. In the dwarf phase of its life, the energy that the sun gives off is generated in its core through the fusion of hydrogen into helium. The sun is about five billion years old now and it still has enough fuel for another five billion years.

Rhetorical Structure Theory (Mann & Thompson 1988)

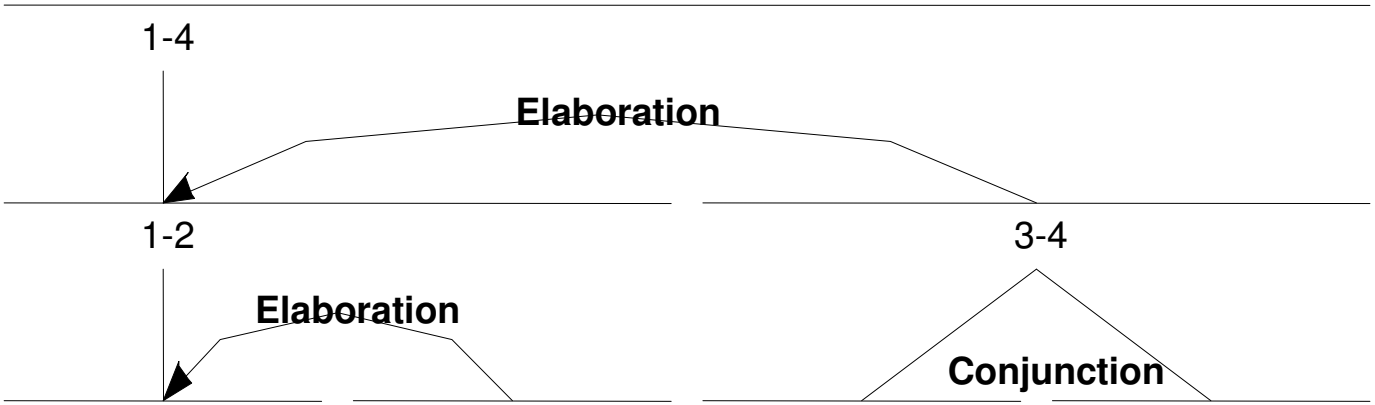
- functional relations between propositions
- reconstruction of writer's purposes
- subject matter vs. presentational relations
- mononuclear relations



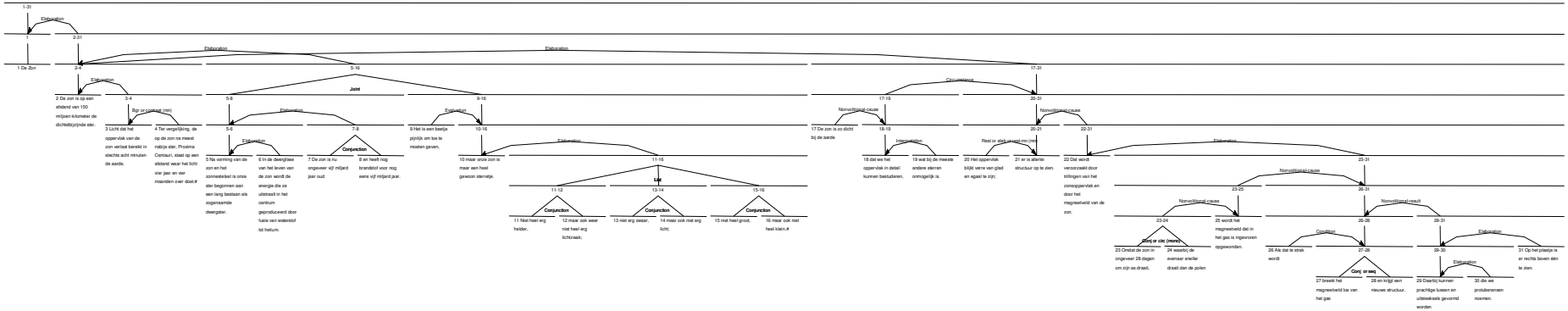
multinuclear relations



hierarchy


1-4
Elaboration
1-2
3-4
Elaboration
Conjunction
Na vorming van de zon en het zonnestelsel is onze ster begonnen aan een lang bestaan als zogenaamde dwergster.
In de dwergfase van het leven van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.
De zon is nu ongeveer vijf miljard jaar oud
en heeft nog brandstof voor nog eens vijf miljard jaar.

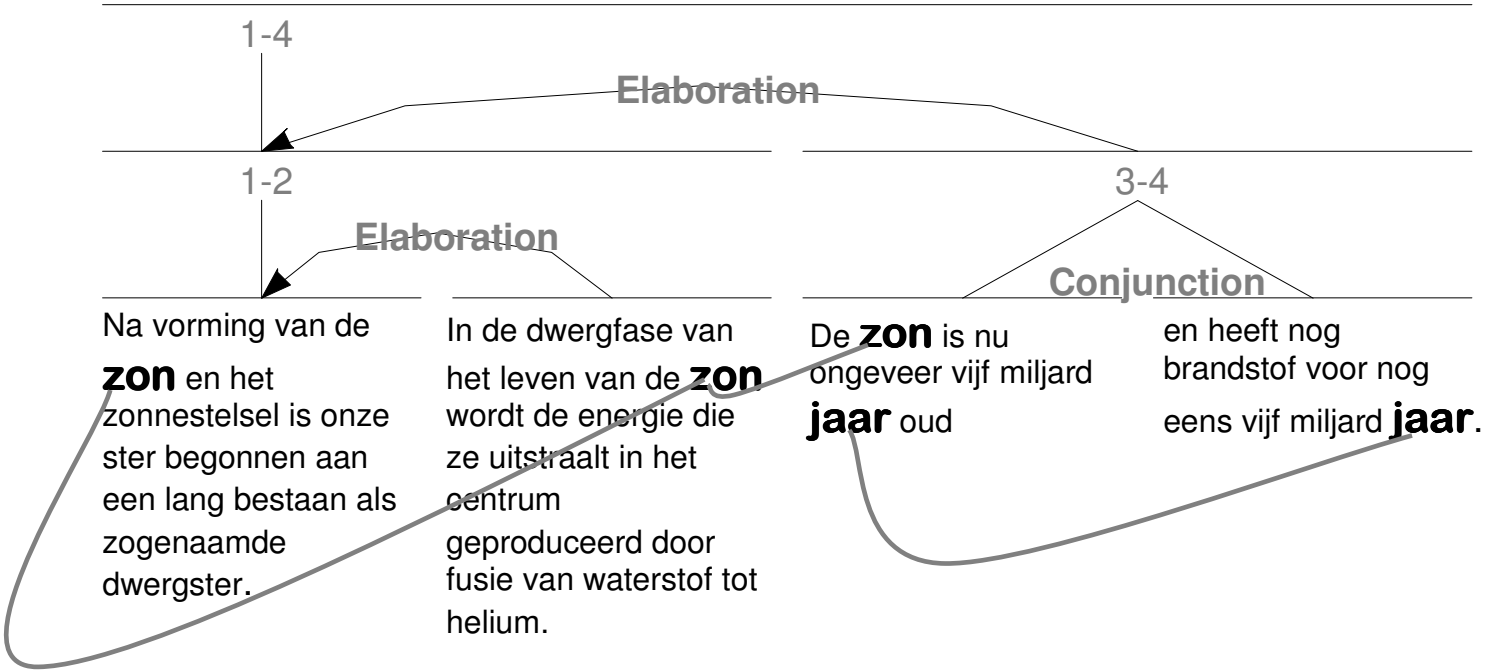
hierarchy


1-31
Elaboration
1 De Zon
2-31
Elaboration
2-4
2 De zon is op een afstand van 150 miljoen kilometer de dichtstbijzijnde ster.
Elaboration
2-4
3 Licht dat het oppervlak van de zon verlaat bereikt in slechts acht minuten de aarde.
Rpr or contrast (rm)
4 Ter vergelijking, de op de zon na meest nabije ster, Proxima Centauri, staat op een afstand waar het licht vier jaar en vier maanden over doet.#
Elaboration
5-16
Joint
5-8
5 Na vorming van de zon en het zonnestelsel is onze ster begonnen aan een lang bestaan als zogenaamde dwergster.
Elaboration
5-6
6 In de dwergfase van het leven van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.
Conjunction
7-8
7 De zon is nu ongeveer vijf miljard jaar oud
8 en heeft nog brandstof voor nog eens vijf miljard jaar.
Evaluation
9-16
9 Het is een beetje pijnlijk om toe te moeten geven,
Elaboration
10-16
10 maar onze zon is maar een heel gewoon sterretje.
Elaboration
13-16
List
11-12
Conjunction
11 Niet heel erg heldere,
12 maar ook weer niet heel erg lichtzwak;
13-14
Conjunction
13 niet erg zwaar,
14 maar ook niet erg licht;
15-16
Conjunction
15 niet heel groot,
16 maar ook niet heel klein.#
Circumstance
17-31
17-19
Nonvolitional cause
17 De zon is zo dicht bij de aarde
18 dat we het oppervlak in detail kunnen bestuderen,
Interpretation
19 wat bij de meeste andere sterren onmogelijk is.
20-31
Nonvolitional cause
20-21
Rest or elab or expl rm (rm)
20 Het oppervlak blijkt vanw van glad en egaal te zijn;
21 er is alleriei structuur op te zien.
22-31
22 Dat wordt veroorzaakt door trillingen van het zonsoppervlak en door het magneetveld van de zon.
Elaboration
23-31
Nonvolitional cause
23-25
Nonvolitional cause
23-24
Conj or circ (more)
23 Omdat de zon in ongeveer 28 dagen om zijn as draait,
24 waarbij de evenaar sneller draait dan de polen
25 wordt het magneetveld dat in het gas is ingevroren opgewonden.
26-31
Nonvolitional result
26-28
Condition
26 Als dat te sterk wordt
27-28
Conj or seq
27 breekt het magneetveld los van het gas
28 en krijgt een nieuwe structuur.
29-31
Elaboration
29-30
Elaboration
29 Daarbij kunnen prachtige lussen en uitbekwaste gevormd worden
30 die we protuberansen noemen.
31 Op het plaatje is er rechts boven één te zien.

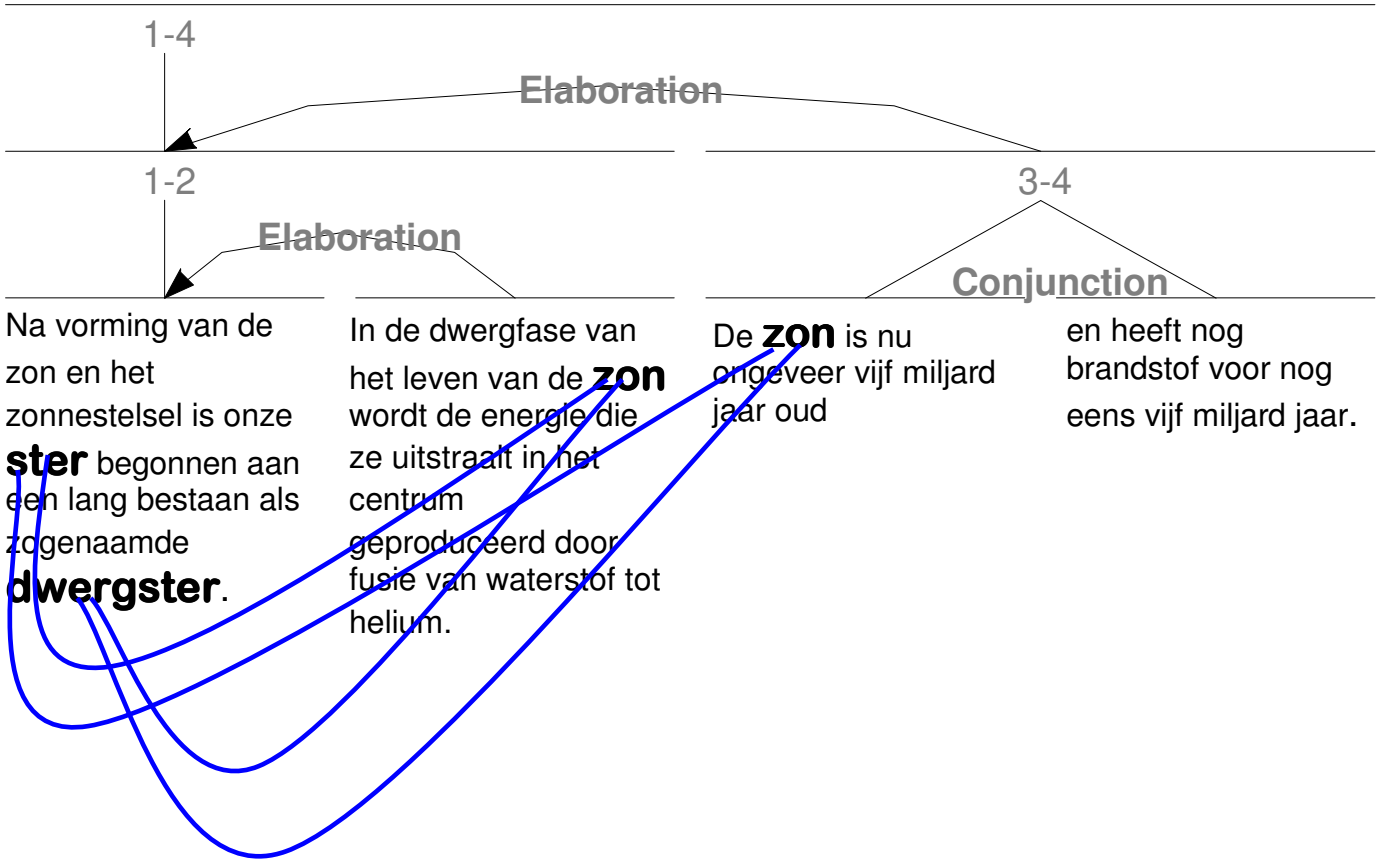
- lexical cohesion
- network of relations
- lexical cohesive relations

  - repetition
  - systematic semantic relations
    - hyponymy, hyperonymy, co-hyponymy
    - meronymy, holonymy, co-meronymy
    - synonymy
    - antonymy
  - collocation

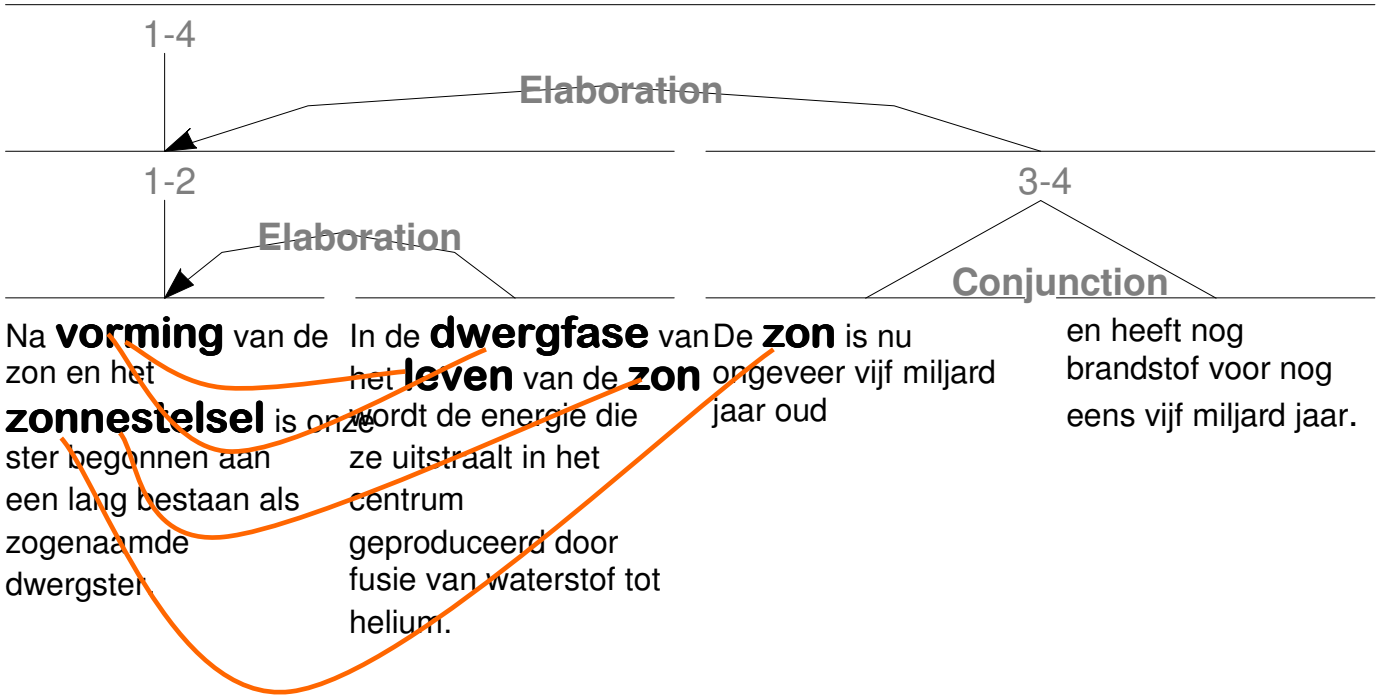
Repetition


1-4
Elaboration
1-2
Elaboration
3-4
Conjunction
Na vorming van de zon en het zonnestelsel is onze ster begonnen aan een lang bestaan als zogenaamde dwergster.
In de dwergfase van het leven van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.
De zon is nu ongeveer vijf miljard jaar oud
en heeft nog brandstof voor nog eens vijf miljard jaar.

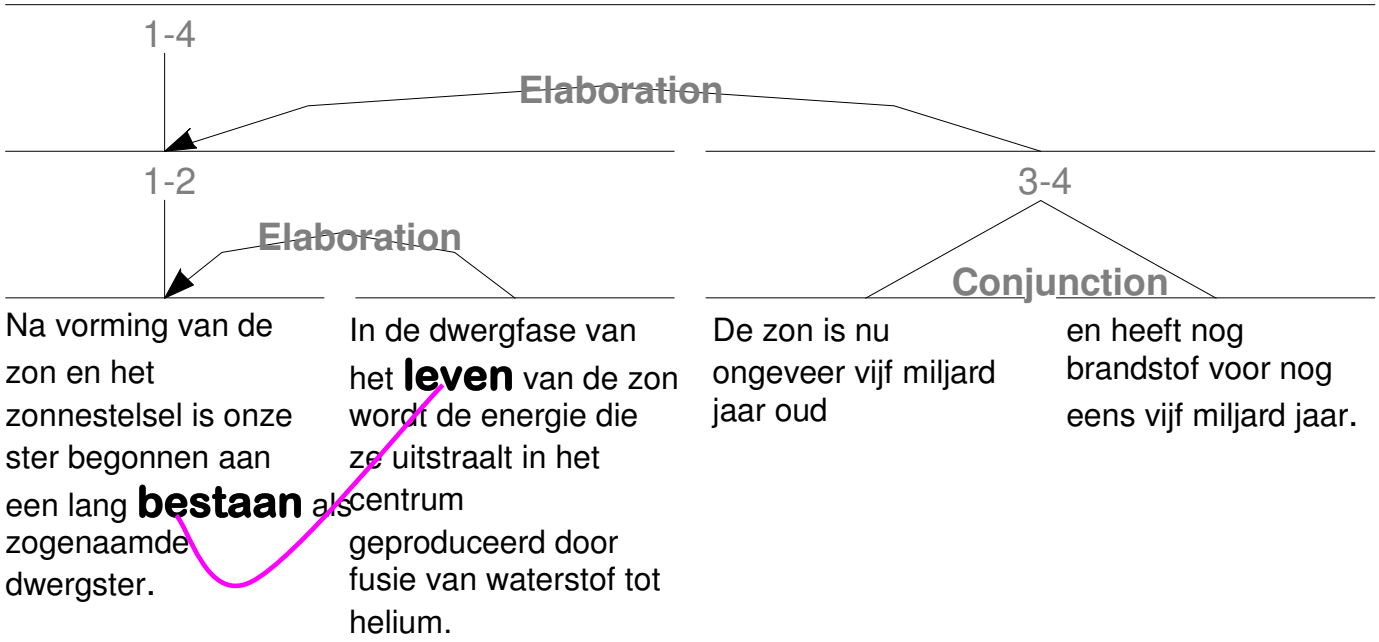
Hyponymy


1-4
Elaboration
1-2
Elaboration
3-4
Conjunction
Na vorming van de zon en het zonnestelsel is onze **ster** begonnen aan een lang bestaan als zogenaamde **dwergster**.
In de dwergfase van het leven van de **zon** wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.
De **zon** is nu ongeveer vijf miljard jaar oud
en heeft nog brandstof voor nog eens vijf miljard jaar.

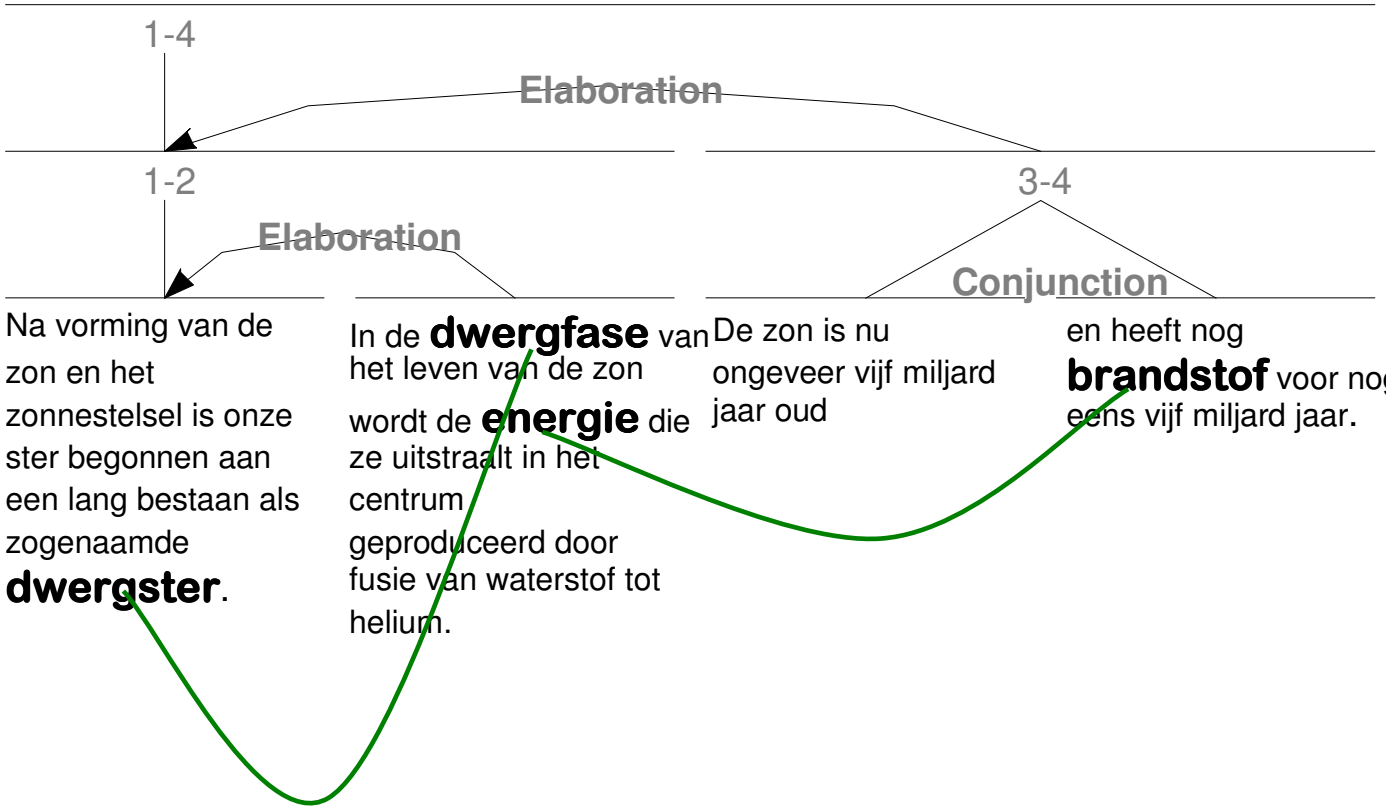
Meronymy

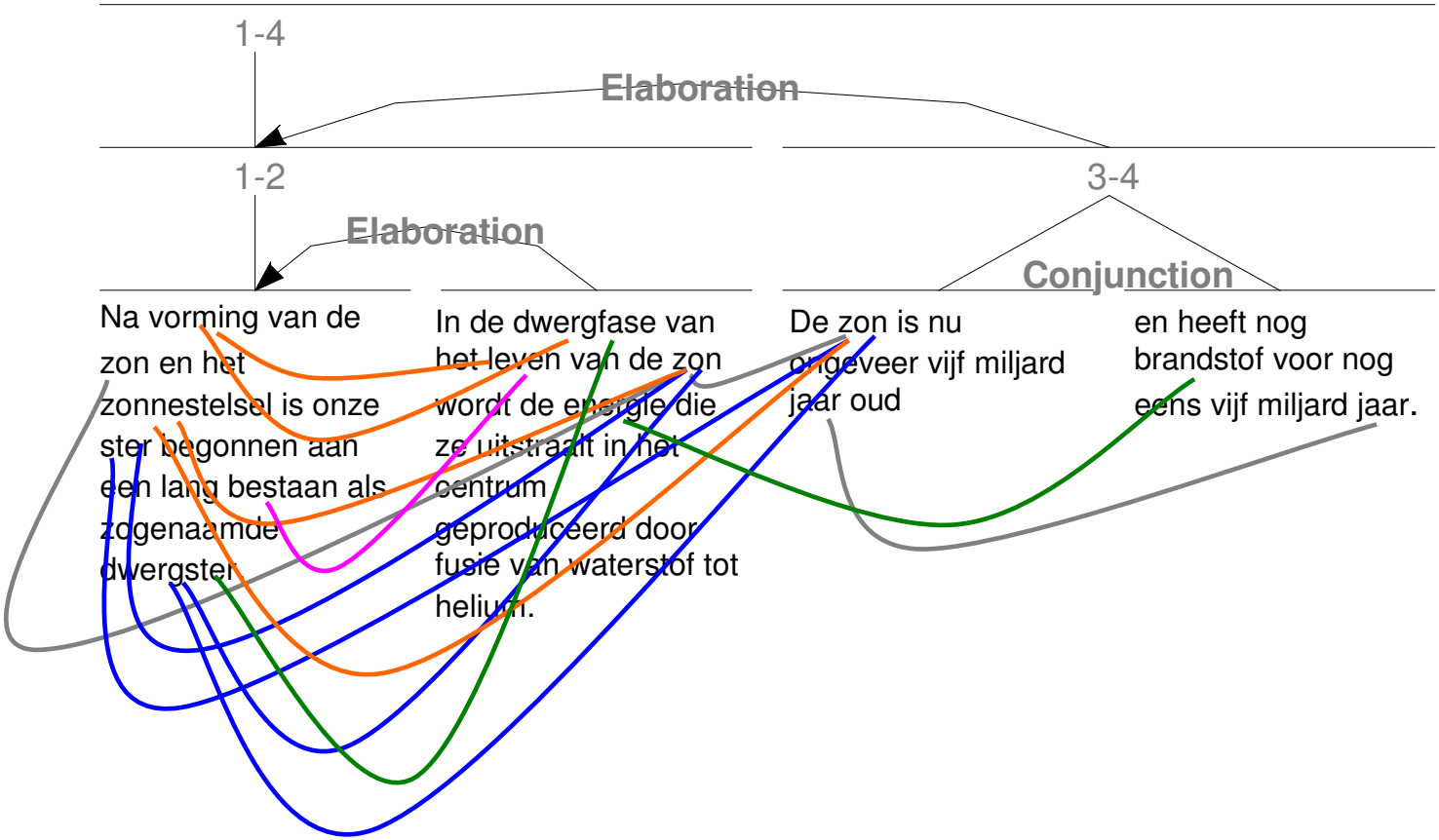

1-4
Elaboration
1-2
Elaboration
3-4
Conjunction
Na **vorming** van de zon en het **zonnestelsel** is onze ster begonnen aan een lang bestaan als zogenaamde dwergster.
In de **dwergfase** van het **leven** van de **zon** wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.
De **zon** is nu ongeveer vijf miljard jaar oud
en heeft nog brandstof voor nog eens vijf miljard jaar.

Synonymy


1-4
Elaboration
1-2
Elaboration
3-4
Conjunction
Na vorming van de zon en het zonnestelsel is onze ster begonnen aan een lang **bestaan** als zogenaamde dwergster.
In de dwergfase van het **leven** van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.
De zon is nu ongeveer vijf miljard jaar oud
en heeft nog brandstof voor nog eens vijf miljard jaar.

Collocation

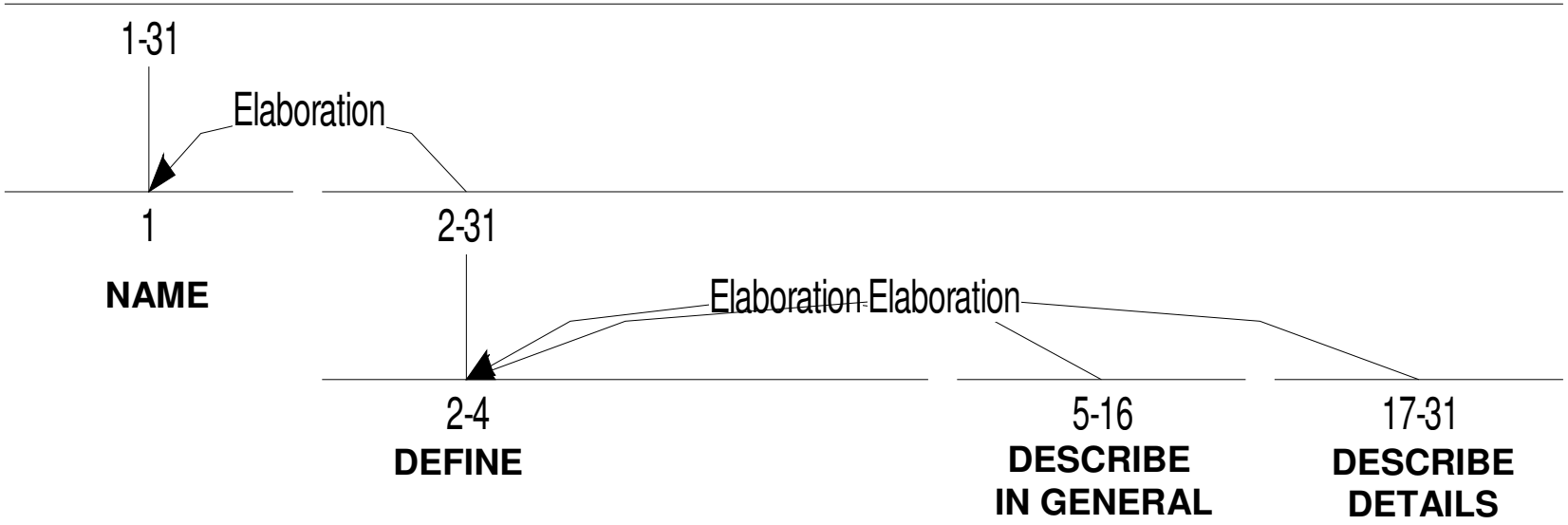



1-4
Elaboration
1-2
Elaboration
3-4
Conjunction
Na vorming van de zon en het zonnestelsel is onze ster begonnen aan een lang bestaan als zogenaamde dwergster.
In de dwergfase van het leven van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.
De zon is nu ongeveer vijf miljard jaar oud
en heeft nog brandstof voor nog eens vijf miljard jaar.

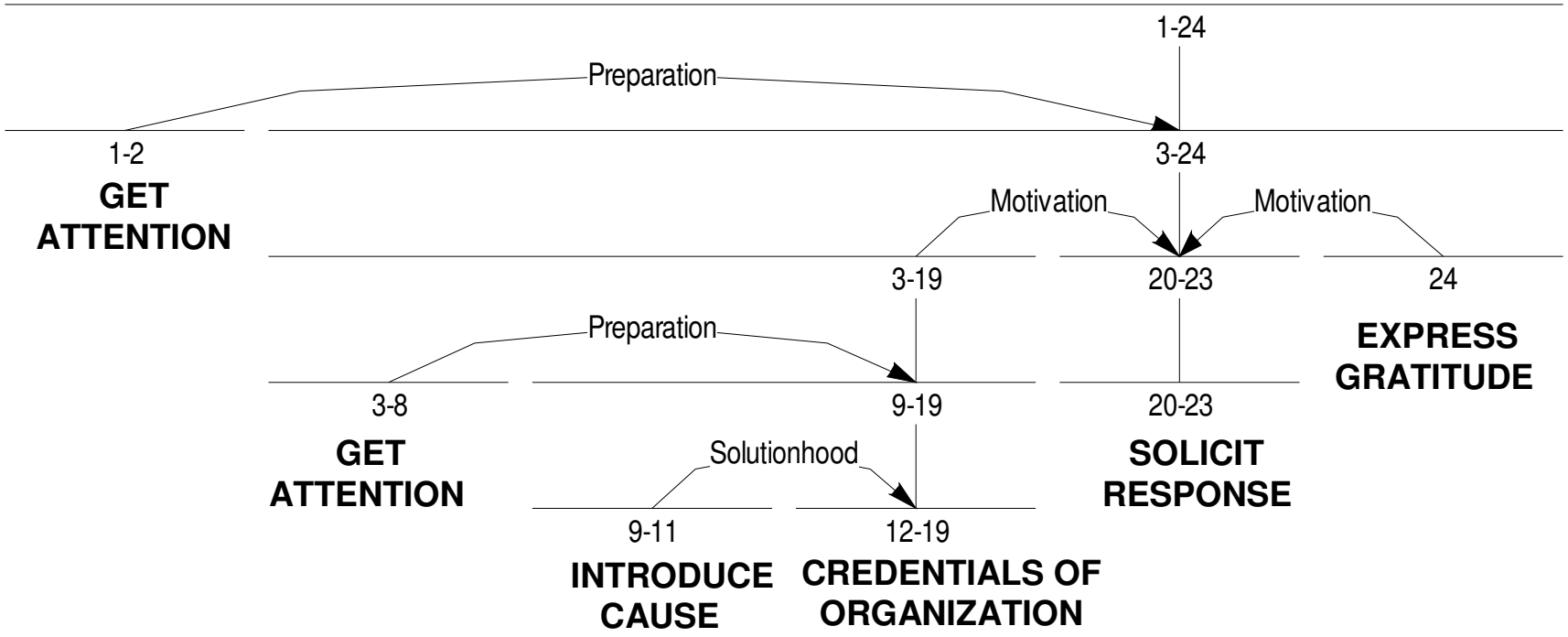
Centrality of discourse units

	EDUs	***Moves***
in coherence structure	smallest units in RST tree	top level of RST tree
in lexical cohesion	lexical cohesive links per EDU	average lexical cohesion density

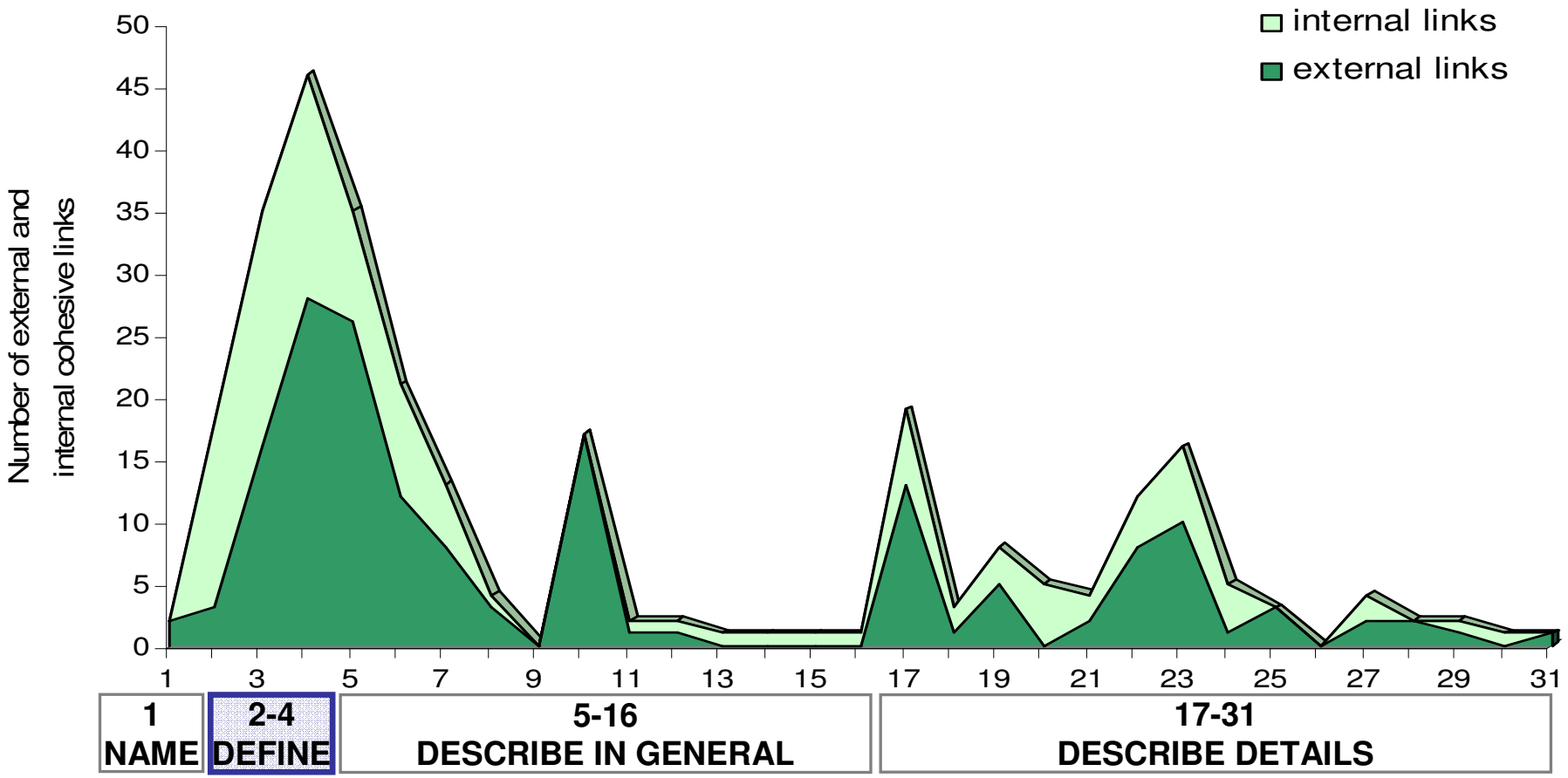
EE01


1-31
Elaboration
1
NAME
2-31
Elaboration
Elaboration
2-4
DEFINE
5-16
DESCRIBE
IN GENERAL
17-31
DESCRIBE
DETAILS

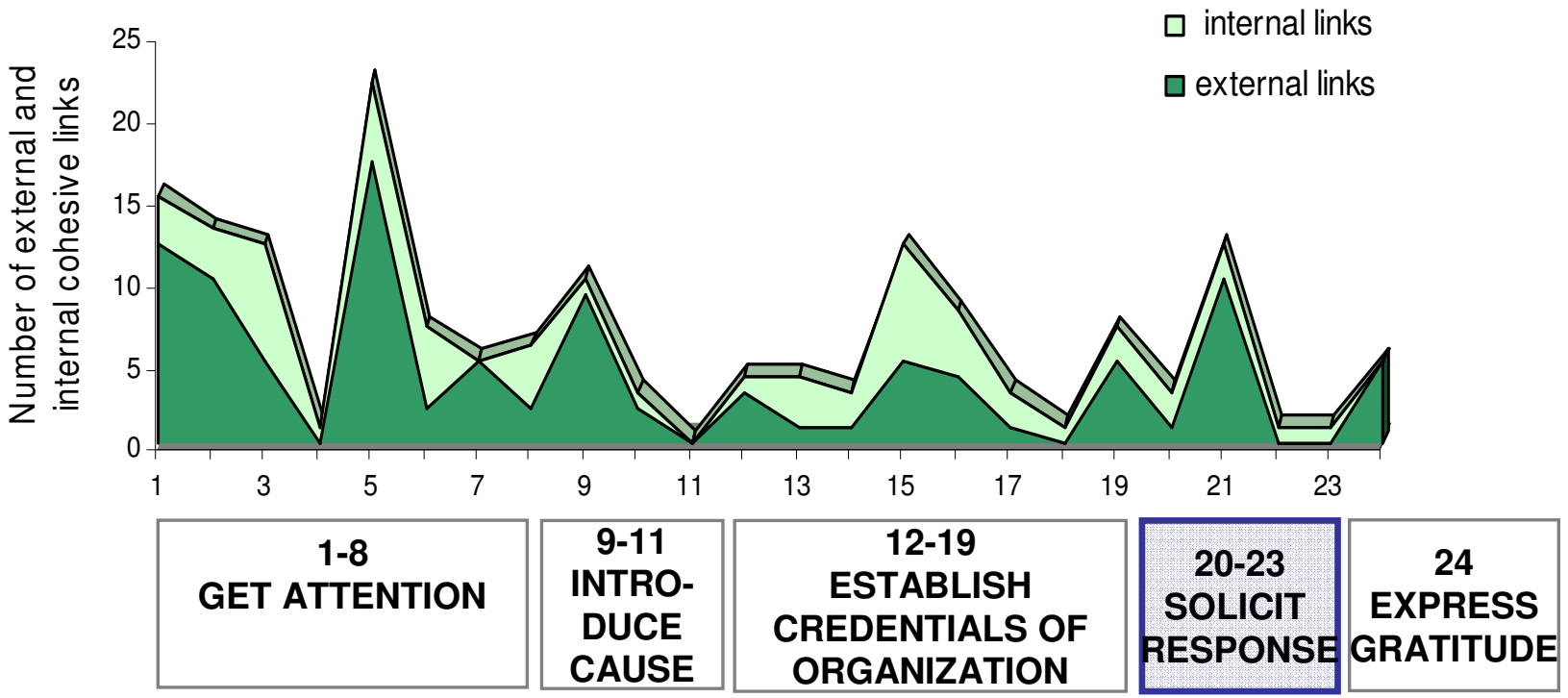
FL01


1-24
Preparation
1-2
3-24
GET ATTENTION
Motivation
Motivation
3-19
20-23
24
Preparation
EXPRESS GRATITUDE
3-8
9-19
20-23
GET ATTENTION
Solutionhood
SOLICIT RESPONSE
9-11
12-19
INTRODUCE CAUSE
CREDENTIALS OF ORGANIZATION

EE01


internal links
external links
Number of external and internal cohesive links
0
5
10
15
20
25
30
35
40
45
50
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
1 NAME
2-4 DEFINE
5-16 DESCRIBE IN GENERAL
17-31 DESCRIBE DETAILS

FL01


internal links
external links
Number of external and internal cohesive links
25
20
15
10
5
0
1
3
5
7
9
11
13
15
17
19
21
23
1-8 GET ATTENTION
9-11 INTRO-DUCE CAUSE
12-19 ESTABLISH CREDENTIALS OF ORGANIZATION
20-23 SOLICIT RESPONSE
24 EXPRESS GRATITUDE

- EE: *Define* is the central move both in coherence and in lexical cohesion
- FL: *Solicit response* is the central move in coherence, but no central move was found for lexical cohesion

→ alignment much closer for EE than for FL

- Move analysis
- Segmentation
- Cohesion analysis
- Coherence analysis

Structural segments + moves

- [Met vriendelijke groet] [en alvast heel hartelijk dank,]

Overlapping moves

- *Dankzij uw donaties* aan de Nierstichting kunnen wij ons inzetten voor een betere kwaliteit van leven voor nierpatienten.

Two functions in one segment

- De *bijgevoegde* kaarten zijn een *bedankje* voor het lezen van mijn brief over de malaria-epidemie in Afrika.

EDU (elementary discourse unit) = clause

[*Na vorming van de zon en het zonnestelsel is onze ster begonnen aan een lang bestaan als zogenaamde dwergster.*] [*In de dwergfase van het leven van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.*] [*De zon is nu ongeveer vijf miljard jaar oud*] [*en heeft nog brandstof voor nog eens vijf miljard jaar.*]

EDU = clause

[*Na vorming van de zon en het zonnestelsel is onze ster begonnen aan een lang bestaan als zogenaamde dwergster.*] [*In de dwergfase van het leven van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.*] [*De zon is nu ongeveer vijf miljard jaar oud*] [*en heeft nog brandstof voor nog eens vijf miljard jaar.*]

EDU = clause

[*Na vorming van de zon en het zonnestelsel is onze ster begonnen aan een lang bestaan als zogenaamde dwergster.*] [*In de dwergfase van het leven van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.*] [*De zon is nu ongeveer vijf miljard jaar oud*] [*en heeft nog brandstof voor nog eens vijf miljard jaar.*]

EDU = clause

[*Na vorming van de zon en het zonnestelsel is onze ster begonnen aan een lang bestaan als zogenaamde dwergster.*] [*In de dwergfase van het leven van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.*] [*De zon is nu ongeveer vijf miljard jaar oud*] [*en heeft nog brandstof voor nog eens vijf miljard jaar.*]

[*Na vorming van de zon en het zonnestelsel*] [*is onze ster begonnen aan een lang bestaan als zogenaamde dwergster.*] [*In de dwergfase van het leven van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd*] [*door fusie van waterstof tot helium.*] [*De zon is nu ongeveer vijf miljard jaar oud*] [*en heeft nog brandstof voor nog eens vijf miljard jaar.*]

Comparisons

- [De zon is zo dicht bij de aarde] [dat we het oppervlak in detail kunnen bestuderen.]
- [Een neutronenster is ongeveer anderhalf keer zo zwaar als de zon,]
- [Echter, Saturnus produceert meer licht] [dan hij van de zon ontvangt.]
- [De atmosferische druk is op het oppervlak zo'n 90 keer groter dan op Aarde.]

Embedded EDUs

- 12 [In kraters nabij de polen van Mercurius, […13…] bestaat misschien zelfs ijs. /

  13 [waar nooit zonlicht komt,]

Parentheticals

- 14 De binnenste maan […15…] beweegt iets sneller dan de buitenste / 15 [(van 2002 tot 2005 is dat Epimetheus)]

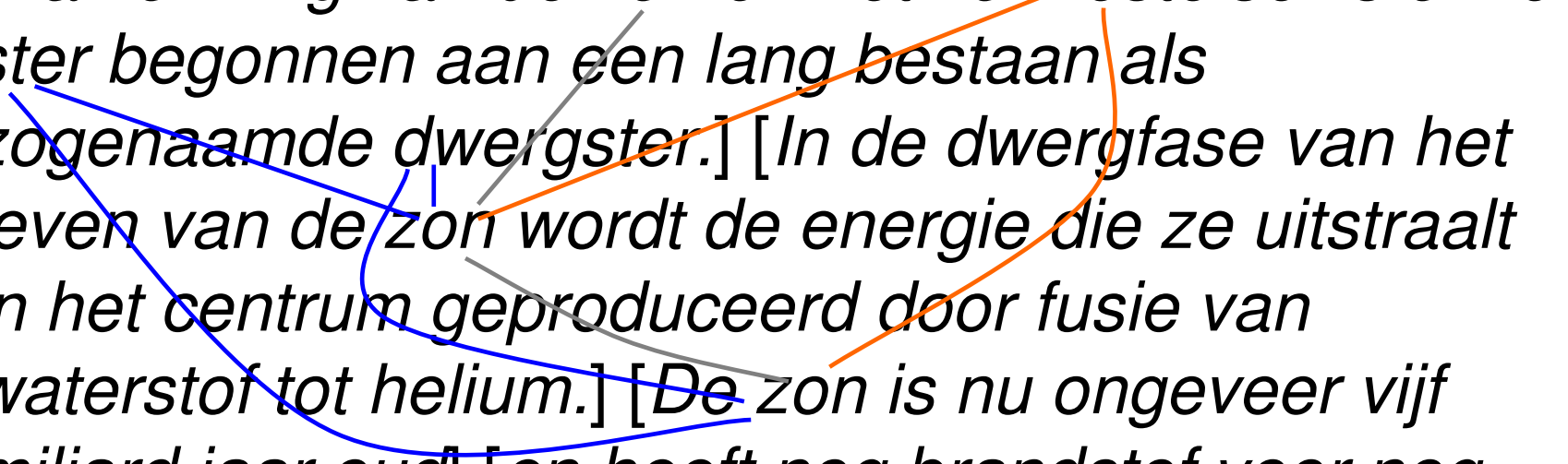
[*Na vorming van de zon en het zonnestelsel is onze ster begonnen aan een lang bestaan als zogenaamde dwergster.*] [*In de dwergfase van het leven van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.*] [*De zon is nu ongeveer vijf miljard jaar oud*] [*en heeft nog brandstof voor nog eens vijf miljard jaar.*]

⟶ link only with the closest preceding item

[*Na vorming van de zon en het zonnestelsel is onze ster begonnen aan een lang bestaan als zogenaamde dwergster.*] [*In de dwergfase van het leven van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.*] [*De zon is nu ongeveer vijf miljard jaar oud*] [*en heeft nog brandstof voor nog eens vijf miljard jaar.*]

⟶ all links identified (same item to more preceding items)

[*Na vorming van de zon en het zonnestelsel is onze ster begonnen aan een lang bestaan als zogenaamde dwergster.*] [*In de dwergfase van het leven van de zon wordt de energie die ze uitstraalt in het centrum geproduceerd door fusie van waterstof tot helium.*] [*De zon is nu ongeveer vijf miljard jaar oud*] [*en heeft nog brandstof voor nog eens vijf miljard jaar.*]



⟶ all links identified (same preceding item to more succeeding items)

Multiple relations

- *gasplaneet – Aarde:* co-meronymy OR co-hyponymy

Context

- *Mercurius – as* / *polen* / *krater*

Word forms

- *Aards – Aarde*: repetition

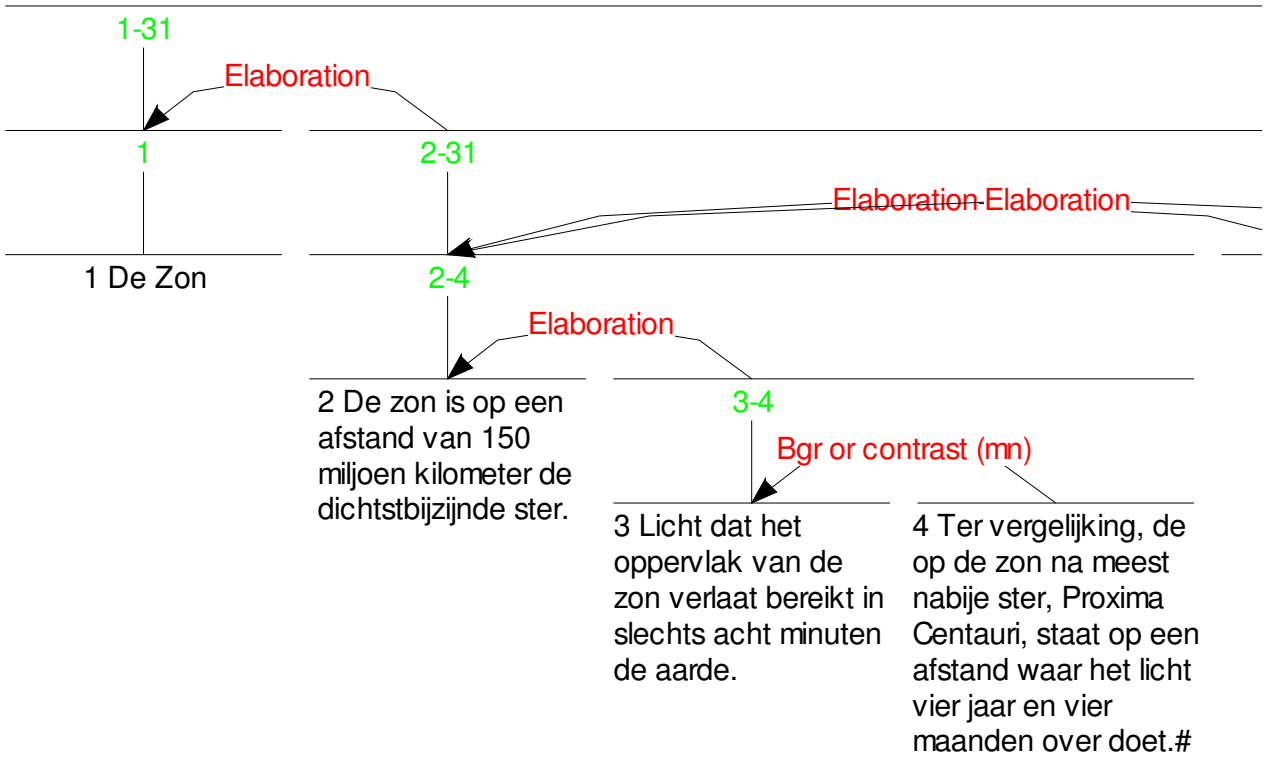
Abbreviations

- *€ – euro*: repetition; *€ – bedrag*: collocation; *H_2 – waterstof*: synonymy

Multi-word units

- *dwergster – dwergfase*: collocation; *Proxima Centauri*;
- [De meest spectaculaire structuur op het oppervlak van Jupiter is een *grote rode vlek* iets ten zuiden van de evenaar.] [Deze structuur wordt de *Grote Rode Vlek* genoemd.]

- RST allows parallel analyses
- hierarchical structure (decisions of the annotator at a given step affects the decisions made at subsequent steps)



Measures

- for annotation purposes
- reliability of data
- validity of coding scheme
- reproducibility of coding

$\rightarrow$ detailed reference manual

- no tradition to report agreement
- about 10% of the corpus
- 2 coders / more than 2 coders
- naive vs. trained coders

Measures – precentage agreement

= observed agreement

		Coder A		
		Stat	IReq	Total
Coder B	Stat	20	20	40
	IReq	10	50	60
	Total	30	70	100

$A_o = (20 + 50) / 100 = 0.7$

- not correct for chance agreement (no comparability, biased)
- not correct for distribution of items among categories

Measures – precentage agreement

Time Period	Ratings by Psychologist 1	Ratings by Psychologist 2
1	0	0
2	0	0
3	0	0
4	0	0
5	1	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	1
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0

$A_o = 18/20 = 90\%$

$A_e = 1/20 \text{ X } 1/20 + 19/20 \text{ X } 19/20 =$

$.0025 + .9025 =$

$.9050$

Measures – Cohen's kappa

$$\kappa = \frac{A_o - A_e}{1 - A_e}$$

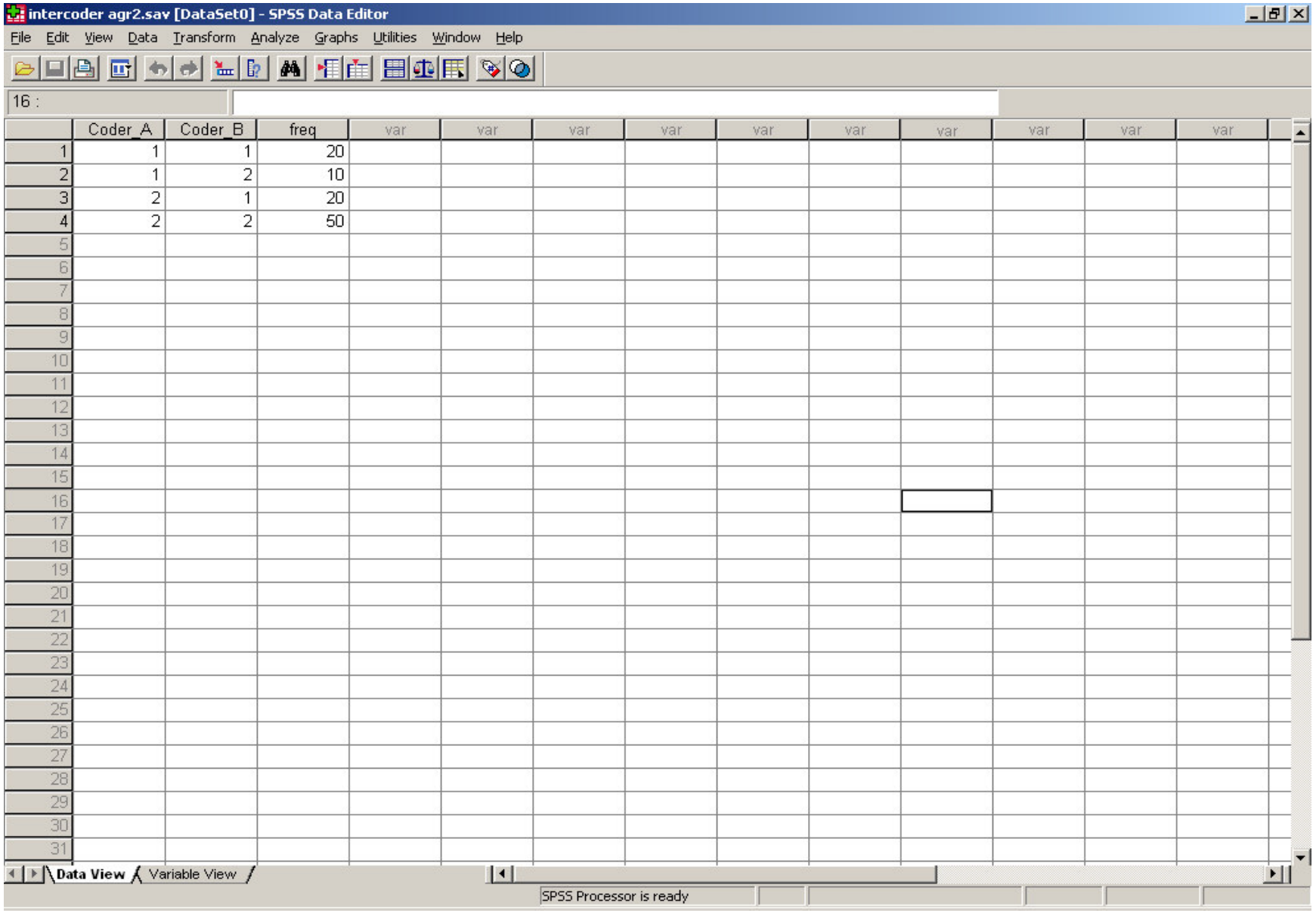
Ao = (20+50) / 100 = 0.70
Ae = (30/100) X (40/100) + (70/100) X (60/100) = 0.54
κ = (0.70 – 0.54) / (1 – 0.54) = 0.348

$$P(k|c_i) = \hat{P}(k|c_i) = \frac{\mathbf{n}_{c_i k}}{\mathbf{i}}$$

		Coder A		
		Stat	IReq	Total
Coder B	Stat	20	20	40
	IReq	10	50	60
	Total	30	70	100

- category judgments
- chance-corrected
- individual coder distributions
- -1 < K < 1

Measures – Cohen's kappa



Measures – Cohen's kappa

Coder_A * Coder_B Crosstabulation

Count

		Coder_B		
		statement	info-req	Total
Coder_A	statement	20	10	30
	info-req	20	50	70
Total		40	60	100

Symmetric Measures

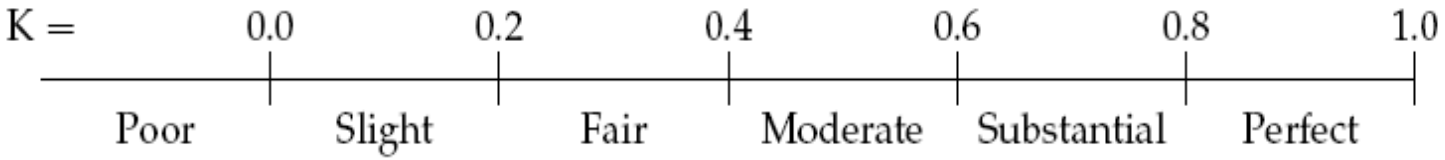
		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Measure of Agreement	Kappa	,348	,095	3,563	,000
N of Valid Cases		100			

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis.

Interpretation

Krippendorff (1980): .67 < K < .80 tentative; .80 <K definite conclusions

Rietveld & van Hout (1993):



Craggs & Wood (2005): no general threshold

Artstein & Poesio (2008): .80 (.70)

Spooren & Degand (2009): .70 (for coherence relations)

Measures – Cohen's kappa

Category prevalence

Coder 1 \ Coder 2	Accept	Ack	
Accept	90	5	95
Ack	5	0	5
	95	5	100

$P(A) = 0.90, P(E) = 0.905$

$\kappa_{Co} = -0.048$

Coder 1 \ Coder 2	Accept	Ack	
Accept	45	5	50
Ack	5	45	50
	50	50	100

$P(A) = 0.90, P(E) = 0.5$

$\kappa_{Co} = 0.80$

Bias

Coder 1 \ Coder 2	Accept	Ack	
Accept	40	15	55
Ack	20	25	45
	60	40	100

$P(A) = 0.65, P(E) = 0.52$

$\kappa_{Co} = 0.27$

Coder 1 \ Coder 2	Accept	Ack	
Accept	40	35	75
Ack	0	25	25
	40	60	100

$P(A) = 0.65, P(E) = 0.45$

$\kappa_{Co} = 0.418$

Other measures

- same distribution of coders → Scott's π
- more than two coders → multi-π, multi-K
- types of disagreements differentiated → weighted coefficients: weighted K, Krippendorff's α

- annotation of discourse structure

- complexity of annotation tasks
  - segmentation
  - move analysis
  - lexical cohesion analysis
  - coherence analysis

Practice so far – segmentation

- proposition / sentence / clause / turn
- CL: automatic segmentation
- early studies: percentage agreement; later: K
- agreement on ”bulk” of segments, but disagreement on exact boundaries
- broad vs. finer segments

Fewer boundaries, higher expected agreement.

Case 1: Broad segments
$A_o = 0.96, A_e = 0.89, K = 0.65$

		Coder A		
		Boundary	No Boundary	Total
Coder B	Boundary	2	1	3
	No Boundary	1	46	47
	Total	3	47	50

Case 2: Fine discourse units
$A_o = 0.88, A_e = 0.53, K = 0.75$

		Coder A		
		Boundary	No Boundary	Total
Coder B	Boundary	16	3	19
	No Boundary	3	28	31
	Total	19	31	50

Marcu et al. (1999)

- EDU boundaries for RST
- K calculated in two ways:

  K_w: boundary can be after any word

  K_u: boundary where at least one coder identified a boundary

  $K_w > K_u$

Practice so far – move analysis

- few studies: percentage agreement ("interrater reliability was calculated at 84%"; "the two raters had an agreement rate of 92% in identifying and categorizing the moves")

- **de Groot (2008)**

  10% of corpus, 2 trained coders, Cohen's kappa

Table 6.4: Evaluation of intercoder agreement for combinations of moves+strategies in the written texts and combinations of moves+strategies in the photographs.

Section/ text type	*Qualification*	*Mean Kappa*	*SD*	*Min. – max.*	*Agreement %*
Management statements/TEXT	Moderate	.58	0.04	.50 – .66	60.63
Management statements/PHOTO	Almost perfect	.92	0.17	.33 – 1.00	93.75
Profiles/ TEXT	Substantial	.73	0.14	.55 – .84	77.21
Profiles/PHOTO	Slight	.13	–	.13 – .13	14.29
Operational reviews/TEXT	Substantial	.67	0.09	.55 – .77	68.27
Operational reviews/PHOTO	Fair	.25	0.45	-.25 – .61	61.48

Practice so far – lexical cohesion

- agreement not reported for lexical cohesion
- percentage agreement for word pairs
- word-sense tagging: rely on dictionaries, hierarchical tagsets (e.g., WordNet)
- ? corpus annotated for lexical cohesion

Spooren & Degand (2009)

- category labels for coherence relations bw two fragments
- two expert coders
- change in coding manual, fewer variables, more fragments
  → K improved (.60)

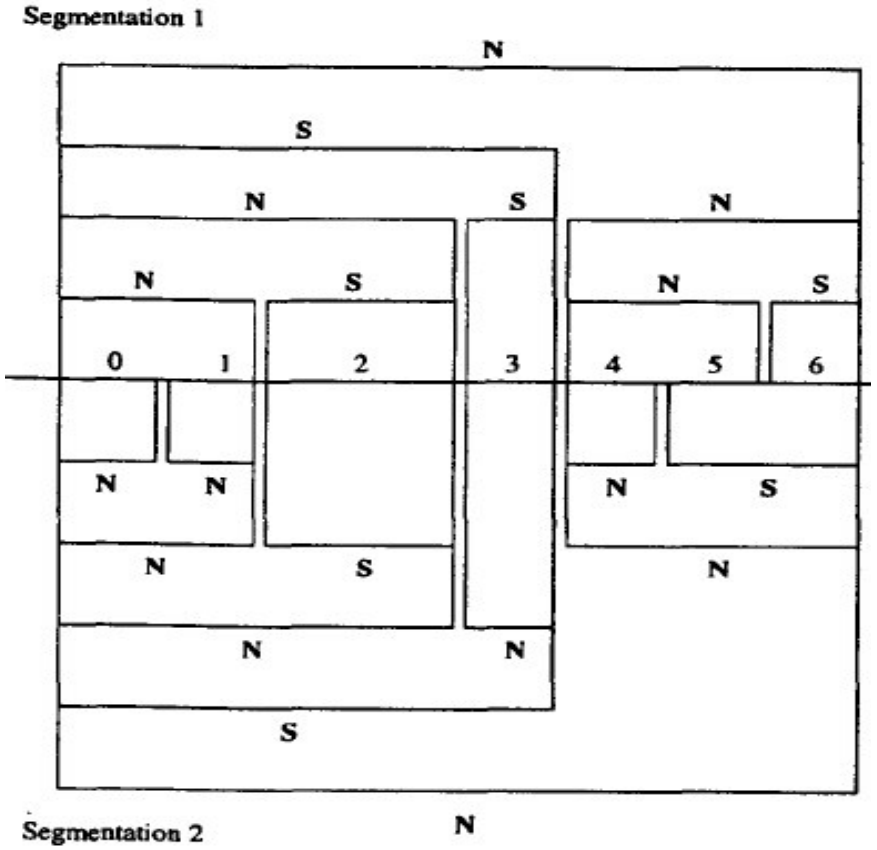
RST: more complex (N vs. S; relation labels; hierarchical tree structure)

Marcu et al. (1999), Carlson et al. (2003)

- for RST
- 2-3 expert coders
- long training phase
- method: mapping hierarchical structures into sets of units that are labeled with category judgments
- label for each EDU which is identified by at least one coder

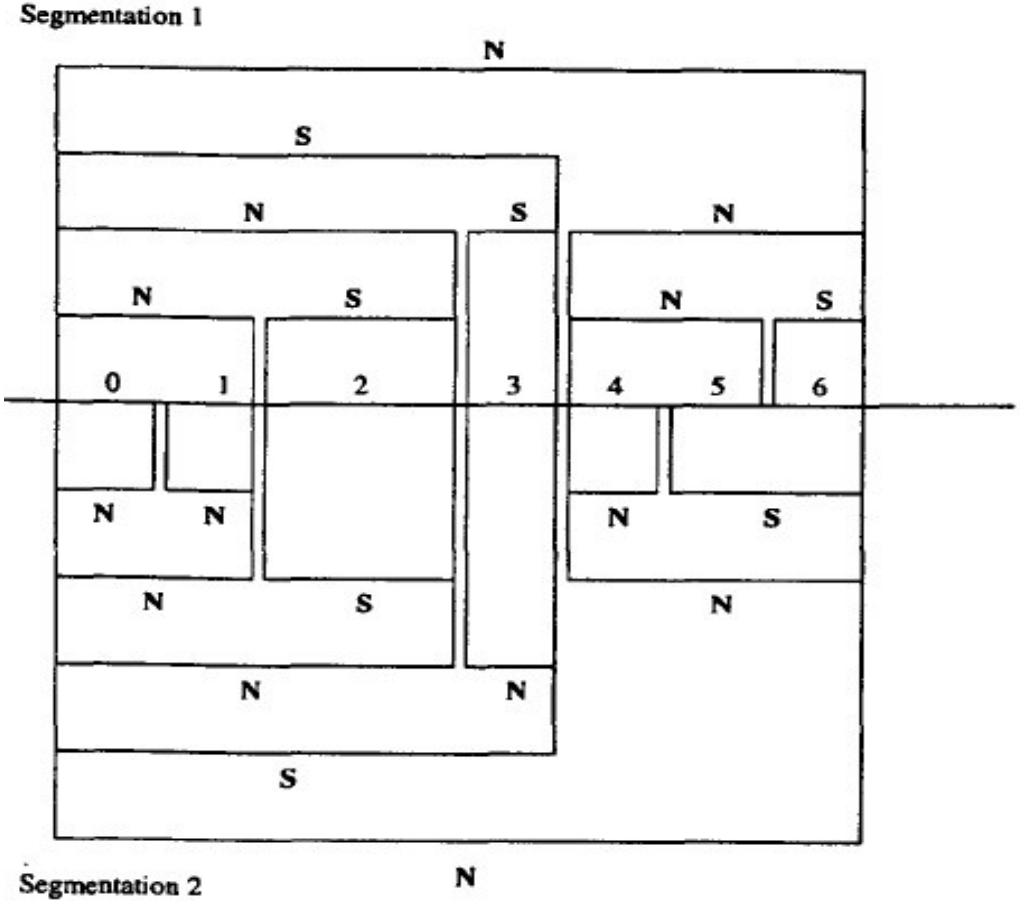
coder 1: [0,1] [2,2] [3,3] [4,5] [6,6]

coder 2: [0,0] [1,1] [2,2] [3,3] [4,4] [5,6]



Practice so far – coherence

- mapping for active and non-acitive EDUs → NONE label



Segment	Segmentation 1	Segmentation 2
[0,0]	none	N
[0,1]	N	N
[0,2]	N	N
[0,3]	S	S
[0,4]	none	none
[0,5]	none	none
[0,6]	N	N
[1,1]	none	N
[1,2]	none	none
[1,3]	none	none
[1,4]	none	none
...		
[4,4]	none	N
[4,5]	N	none
[4,6]	N	N
[5,5]	none	none
[5,6]	none	S
[6,6]	S	none

- K_n for nuclearity
- K_r for coherence relations
- K_{rr} for a reduced set of coherence relations

Practice so far – coherence

Problems:

- violation of independence assumption bw categorical judgments
- NONE agreements make K artificially high
- agreements of different importance
- not for diagnosing disagreements

Q: how to calculate agreement on hierarchical annotation?

Decisions – general

- corpus: 150 texts (25 texts per genre)
  to annotate for lexical cohesion, RST
- two coders
- expert coders
- detailed coding manual
- coding manual refined as corpus grows
- 20% of the corpus (5 texts per genre)
- training phase: orientation (principles, tools); independent codings (compare $\rightarrow$ revise manual); final phase (reduce differences)
- presegmented texts for RST analysis and cohesion analysis
- agreement measured per genre as corpus grows
- be explicit (process, scores)

Decisions – segmentation

- two coders: GR + IB
- Cohen's kappa: calculate (1) for EDU boundary, (2) for word boundary (following Marcu et al. 1999)

  YES / NO categories for each possible location

  example: segmentation of nine encyclopedia entries (EEs)

	EDU level			Word level		
	A_o	A_e	K	A_o	A_e	K
EE17	0.95	0.82	0.72	0.99	0.82	0.94
EE18	0.86	0.74	0.46	0.98	0.83	0.88
EE19	0.84	0.76	0.33	0.98	0.80	0.90
EE20	0.92	0.79	0.62	0.99	0.84	0.94
EE21	0.93	0.92	0.14	0.99	0.84	0.94
EE22	0.89	0.88	0.08	0.99	0.84	0.94
EE23	0.96	0.88	0.67	0.99	0.83	0.93
EE24	0.86	0.78	0.36	0.99	0.91	0.89
EE25	0.91	0.73	0.67	0.99	0.82	0.94

- human errors
- new rules added to coding manual

  OR

- percentage agreement (following Carletta et al. 1997)

Decisions – move analysis

- two coders: IB + MR
- training the co-coder
- six different move structures for six different genres → time-consuming
- Cohen's kappa (following de Groot 2008)

  ? two-step procedure:

  1: identifying move boundaries – percentage agreement OR K

  2: labeling moves identified by both coders – K

Decisions – cohesion

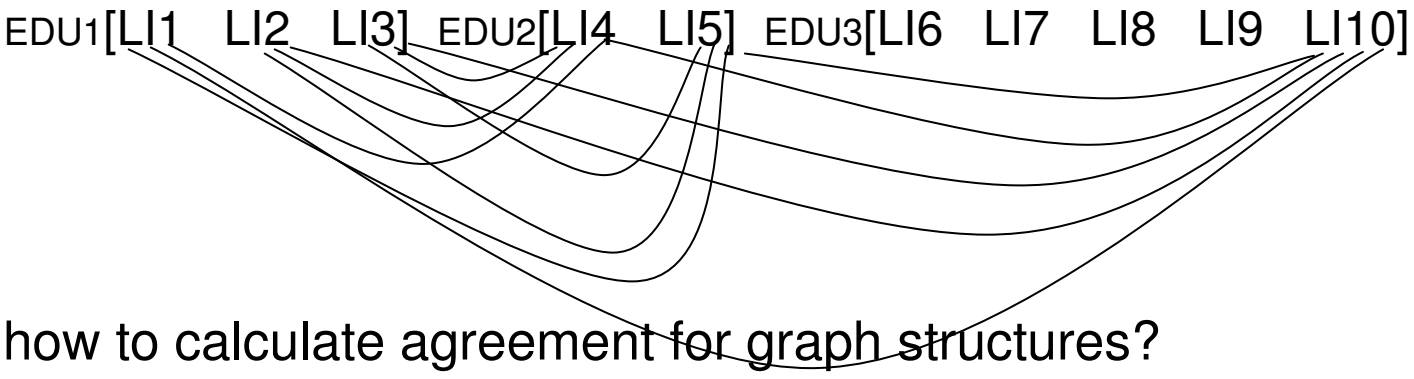
- two coders: IB + MR
- training the co-coder
- texts presegmented, potential lexical items preselected
- MMAX2 tool, CORNETTO
- ? Cohen's kappa

  mutually exclusive categories

  ? equally distinct categories (? weighted coefficient)

  lexical cohesive relations across EDU boundaries

  multiple relations allowed (= multi-tagged lexical items)



LI = lexical item

Q: how to calculate agreement for graph structures?

Decisions – coherence

- two coders: IB + GR
- texts from early training phase not included in the corpus
- ? following Marcu et al. (1999)

Reading

Artstein, R. & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics 34*(4), 555-596.

References (intercoder agreement)

Carletta, J. et al. (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics 23*(1): 13-31.

Carlson, L., Marcu, D. & Okurowski, M.E. (2003). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In J. Kuppevelt & R.W. Smith (Eds.). *Current and new directions in discourse and dialogue*. (pp. 85-112). Dordrecht: Kluwer.

Craggs, Richard & Wood, Mary McGee (2005). Evaluating discourse and dialogue coding schemes. *Computational Linguistics 31*(3): 289-295.

de Groot, E. (2008). *English annual reports in Europe.* Utrecht: LOT.

Krippendorff, K. (1980). *Content analysis: an introduction to its methodology. Chapter 12.* Beverly Hills, CA: Sage.

Marcu, D, Amorrortu, E. & Romera, M. (1999). Experiments in constructing a corpus of discourse trees. In *Proceedings of the ACL Workshop on Standards and Tools for Discourse Tagging*. (pp. 48-57). College Park, MD.

Rietveld, T. & van Hout, R. (1993). *Statistical techniques for the study of language and language behaviour*. Berlin: de Gruyter.

Spooren, W. & Degand, L. (2009). Coding coherence relations: reliability and validity. Unpublished manuscript.

References (IB's research)

Mann, W.C. & Thompson, S.A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text 8*(3), 243-281.

Swales, John (1990). *Genre analysis. English in academic and research settings.* Cambridge: Cambridge University Press.

Upton, Thomas A. (2002). Understanding direct mail letters as a genre. *International Journal of Corpus Linguistics 7*(1), 65-85.