



Deliverable D5.1

Text Summarisation Tools

Grant Agreement number: 250467

Project acronym: ATLAS

Project type: Pilot B

Deliverable D5.1 Text Summarisation Tools 31.08.2012

Project coordinator name, title and organisation:

Anelia Belogay, CEO, Diman Karagiozov, CTO,

Tetracom Interactive Solutions

Tel: +35924950444

Fax: +35924950443

E-mail: anelia@tetracom.com, diman@tetracom.com

Project website address: www.atlasproject.eu

Authors:

Polivios Raxis, Ioannidis Dimosthenis (Atlantis), Svetla Koeva, Angel Genov, Ivelina Stoyanova (IBL-DCL), Maciej Ogrodniczuk, Adam Przepiórkowski, Mateusz Kopeć (ICS PAS), Anelia Belogay, Emil Stoyanov, Diman Karagiozov (Tetracom), Cristina Vertan (UHH), Dan Cristea, Eugen Ignat, Daniel Anechitei, Sabina Deliu (UAIC)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	X



C	Confidential, only for members of the consortium and the Commission Services	
---	--	--

Revision	Date	Author, organisation	Description
1	March 2012	D.Cristea, E.Ignat, D.Anechitei, S.Deliu	WP5 Summarisation Guide
2	1 July 2012	D.Anechitei	Multilingual Discourse Processing – master thesis, “Alexandru Ioan Cuza” University of Iași
3	11 August 2012	D.Cristea, D.Anechitei, E.Ignat	Preliminary version of D5.1-v0 distributed to partners
4	28 August 2012	D.Cristea, D.Anechitei, E.Ignat	Enhanced version D5.1-v4
5	30 August 2012	D.Cristea, D.Anechitei, E.Ignat	D5.1-v5
6	31 August 2012	D.Cristea, D.Anechitei, E.Ignat	D5.1-v8
7	2 September 2012	D.Cristea, D.Anechitei, E.Ignat	D5.1-v9
8	3 September 2012	D.Cristea, D.Anechitei, E.Ignat	First complete outline of the document: D5.1-v10
9	7 September 2012	M. Ogradniczuk, A. Przepiórkowski, Mateusz Kopeć (ICS PAS), D. Cristea, E. Ignat, D.Anechitei, S.Deliu (UAIC), D. Karagiozov, E. Stoyanov, A. Belogay (Tetracom), S. Koeva, A. Genov, I. Stoyanova (IBL-DCL), D. Ioannidis, P. Raxis (Atlantis), C. Vertan (UHH)	Filled in the tables with data for all languages – pre-final version
10	8 September 2012	S. Koeva	Modification of the Bulgarian tables
11	10 September 2012	A. Belogay, D. Karagiozov	UIMA integration of summarisation chains - observations
12	13 September 2012	E. Ignat and D.Cristea	Final version

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.



TABLE	OF	CONTENTS
TABLE OF CONTENTS		4
List of figures		6
List of TABLES		8
1 Introduction		10
1.1 <i>State of the art</i>		10
1.2 <i>The ATLAS approach</i>		11
1.3 <i>The ATLAS summarisation architecture</i>		11
2 survey OF THE LANGUAGE PROCESSING CHAINS acting as prerequisites for summarisation		13
2.1 <i>Common pre-processing tools</i>		13
2.2 <i>Language specific prerequisite tools</i>		13
2.2.1 <i>Bulgarian</i>		13
2.2.2 <i>English</i>		13
2.2.3 <i>German</i>		13
2.2.4 <i>Greek</i>		14
2.2.5 <i>Polish</i>		14
2.2.6 <i>Romanian</i>		14
3 THE PROPER SUMMARISATION TOOLS		15
3.1 <i>Anaphora resolution</i>		15
3.1.1 <i>The model and the engine</i>		15
3.1.2 <i>Localisation of RARE</i>		17
3.2 <i>Segmentation of sentences at clause boundaries</i>		23
3.2.1 <i>State of the art</i>		23
3.2.2 <i>Methodology</i>		23
3.2.3 <i>The CS Training module</i>		25
3.2.4 <i>The Clause Segmenter module</i>		32
3.2.5 <i>The CS Evaluation Module</i>		36
3.2.6 <i>The CS Calibration system</i>		38
3.3 <i>Discourse parsing</i>		42
3.3.1 <i>Introduction</i>		42
3.3.2 <i>Methodology</i>		45
3.3.3 <i>Incremental parsing of the discourse</i>		45
3.3.4 <i>Incremental parsing at the sentence level</i>		50
3.3.5 <i>Combinatorics at the sentence level</i>		52
3.3.6 <i>Heuristics: Centering</i>		59
3.3.7 <i>Heuristics: Veins Theory</i>		60
3.3.8 <i>Heuristics: favour lower adjunction levels</i>		63
3.3.9 <i>Heuristics: opening the minimum referentiality domain</i>		63
3.3.10 <i>Heuristics: maximum referentiality score</i>		64
3.3.11 <i>Heuristic: if a substitution node is open, consume it!</i>		65
3.3.12 <i>Heuristic: limit the length of the summary</i>		65
3.3.13 <i>Heuristics: concluding remarks</i>		65
3.3.14 <i>Parameters for the DP module</i>		65
3.4 <i>The Summariser for short texts</i>		69
3.4.1 <i>Types of summaries</i>		70



3.4.2 The Extract Summary module.....	70
3.4.3 The Evaluate Summary system.....	73
3.4.4 The DP calibration system.....	75
3.5 <i>The Summarization tools integrated in u-Compare</i>	78
3.5.1 u-Compare.....	78
3.5.2 Integrating the summarization chain in u-Compare.....	78
4 Auxiliary software	80
4.1 <i>Graphical interface for viewing and editing discourse markers</i>	80
4.2 <i>Graphical interface for viewing and evaluating the discourse trees</i>	85
5 building corpora	87
5.1 <i>Building the clause segmentation and markers corpora</i>	87
5.2 <i>Building the corpora of summaries</i>	94
6 Corpora and language models	96
6.1 <i>Bulgarian</i>	99
6.2 <i>English</i>	102
6.3 <i>German</i>	104
6.4 <i>Greek</i>	106
6.5 <i>Polish</i>	111
6.6 <i>Romanian</i>	113
7 Conclusions	116
7.1 <i>Critical analysis of the ATLAS Summarisation system</i>	116
7.2 <i>Further improvements and developments</i>	118
References	120
Annex I: Example of rules to be used by RARE for English	123
ANNEX II: summaries of short texts	126



LIST OF FIGURES

Figure 1: The block architecture of the summarisation process in ATLAS.....	10
Figure 2: Details of the summarisation prerequisite chain (WP4).....	11
Figure 3: Details of the summarisation chain (WP5).....	11
Figure 4: Two-layer representation of a co-referential anaphoric relation.....	14
Figure 5: Three layers representation of co-referencing REs (Cristea and Dima, 2001).....	15
Figure 6: The Training module.....	24
Figure 7: The Clause Segmenter Module	31
Figure 8 – The Evaluation Module.....	35
Figure 9: The CS Calibration System.....	39
Figure 10: Adjunction operation	45
Figure 11 – Substitution operation.....	46
Figure 12 – Internal right frontier	46
Figure 13: Auxiliary trees	47
Figure 14: Workflow of the incremental parsing at paragraph level.....	48
Figure 15: Initial discourse unit raises an expectation.....	49
Figure 16: The second unit fulfils the open expectation while raising another one	50
Figure 17: The third unit still leaves open the opened expectation	50
Figure 18: Finally, the forth unit signals the opened expectation.....	50
Figure 19: Combinatoric workflow for generating markers argument lists.....	53
Figure 20: Valid discourse tree structure for the considered example.....	54
Figure 21: Automatically generated tree structures.....	55
Figure 22: Total number of trees with two clauses	57
Figure 23: Total number of trees with three clauses and only one (and the same) projection type.....	57
Figure 24: Rooted binary trees with 1, 2, 3 and 4 leaf nodes (Davis, 2010)	58
Figure 25: Veins expressions for all cases.....	62
Figure 26: Computed vein expressions for the exemplification	63
Figure 27: Approximating a gold summary out of a number of human - produced summaries	75
Figure 28: Discourse Parser Calibration System	77
Figure 29: Markers' Editor.....	81
Figure 30: File menu of the Markers Editor.....	82



Figure 31: Add menu.....	82
Figure 32: Add menu - marker type.....	82
Figure 33: Add menu - marker type, connect attribute	83
Figure 34: Error message - marker field.....	83
Figure 35: Error message - relation field.....	83
Figure 36: Error message - adding existing marker	84
Figure 37: Edit Menu.....	84
Figure 38: Warning dialog	85
Figure 39: Search marker results.....	85
Figure 40: Error finding marker	85
Figure 41: Delete Menu	86
Figure 42: Draw Trees Graphical Interface	87
Figure 43: Example of a text annotated with PALinkA	94
Figure 44: RARE error caused by a matching of 2 pronouns.....	119
Figure 45: Segmenter error	120



LIST	OF	TABLES
Table 1: Example of manually annotated sentence at clauses and markers.....		27
Table 2: Example of positive marker context		28
Table 3: Example of negative marker context.....		28
Table 4: Example of clause boundary without marker		29
Table 5: Example of markers type		29
Table 6: Overview of the segmentation algorithm.....		34
Table 7: Example of a configuration file		39
Table 8: A general overview of the calibration system		40
Table 9: A general overview of the incremental parsing at the discourse level		48
Table 10: A general overview of the incremental parsing at the paragraph level.....		50
Table 11: A general overview of the incremental parsing at sentence level.....		52
Table 12: Markers with their patterns and computed argument lists		53
Table 13: Markers with their patterns and computed argument lists missing the marker “and”		55
Table 14: Total number of possible trees relative to the number of clauses.....		57
Table 15: A general overview of the combinatory system at sentence level.....		59
Table 16: Types of transitions in Centering theory		59
Table 17: RARE localisation files.....		97
Table 18: The AR annotated corpus		97
Table 19: The CS and DP corpus		98
Table 20: The summaries corpus		99
Table 21-BG: RARE localisation files for Bulgarian		100
Table 22-BG: The CS and DP corpus for Bulgarian		100
Table 23-BG: The Bulgarian summaries corpus. File names are given by code and code – file name correspondence is presented in the table below.		101
Table 24-BG: Code – file name correspondence for Table 22-BG and Table 23-BG.....		101
Table 25-EN: RARE localisation files for English		102
Table 26-EN: The CS and DP corpus for English.....		102
Table 27-EN: The CS and DP corpus for English (continued)		103
Table 28-EN: The English summaries corpus		103
Table 29-DE: RARE localisation files.....		104
Table 30-DE: CS and DP corpus for German. File names are given by code and code – file name correspondence is presented in the table below.		105



Table 31-DE: The German summaries corpus	105
Table 32-DE: Code – file name correspondence for Table 30-DE and Table 31-DE	106
Table 33-EL: RARE localisation files.....	107
Table 34-EL: the CS and DP corpus for Greek	109
Table 35-EL: the summaries corpus for Greek.....	110
Table 36-PL: RARE localisation files	111
Table 37-PL: The CS and DP corpus for Polish.....	111
Table 38-PL: The summaries corpus for Polish	112
Table 39-RO: RARE localisation files for Romanian	113
Table 40-RO: The CS and DP corpus for Romanian.....	113
Table 41-RO: The CS and DP corpus for Romanian (continued).....	114
Table 42-RO: The CS and DP corpus for Romanian (continued).....	114
Table 43-RO: The summaries corpus for Romanian	115



1 INTRODUCTION

D5.1 is the main deliverable of the WP5 “Summarisation” in the ATLAS project. It describes the text summarization tools for Bulgarian, German, Greek, English, Polish, Romanian. The document includes the following sections:

- Section 1: a short state-of-the-art, the summarisation approach in ATLAS, and the general architecture of the summary processing chain;
- Section 2: summary (detailing formats) of the LPCs (language processing chains) for the languages in ATLAS (Bulgarian, English, German, Greek, Polish and Romanian), which are considered as prerequisites for the proper summarisation modules;
- Section 3: detailed description of the summarisation modules (anaphora resolver, clause segmenter, discourse parser and summariser);
- Section 3.5: description of two auxiliary pieces of software built to support the development of the summarisation system;
- Section 5: partner’s manual for building corpora and preparing the summarisation chain corpus;
- Section 6: detailed description of the corpora used for the summarisation chain for each languages and the evaluation data;
- Section 0: concluding remarks and further developments.

1.1 State of the art

Discourse structures have a central role in several computational tasks, such as question-answering, dialogue generation, summarisation, information extraction, etc. The HILDA discourse parser (Hernault et al., 2010) is the first fully implemented text-level discourse parser with state-of-the-art performance. The system was trained on a variety of lexical and syntactic features extracted from the input text. While some of their features were borrowed from (Şoricut and Marcu, 2003), they also proposed the novel idea of discourse tree building by using two classifiers in cascade – a binary structure classifier to determine whether two adjacent text units should be merged to form a new sub-tree, and a multi-class classifier to



determine which discourse relation label should be assigned to the new sub-tree (Feng and Hirst, 2012).

1.2 The ATLAS approach

Our approach to summarization follows the one presented in (Cristea et al, 2005). The summary of a short text is obtained as a sequence of discourse clauses extracted from the original text, after obtaining the discourse structure of the text and exploiting cohesion and coherence properties. Our discourse structures resemble the RST (Rhetorical Structure Theory) trees (Mann and Thomson, 1987). The difference is that our trees are binary and lack relation names. Apart from obtaining general summaries, the output of the parsing process can be used to obtain excerpt-type summaries focused on individual characters mentioned in the text. A combined, pipe-line/parallel/incremental, type of processing is employed. To master the combinatorial explosion when discourse trees are being built, a beam-search processing is employed.

In the following section the architecture of the summarization system is discussed. The evaluation is realised by comparing the summaries outputted by the system against those collected from human subjects.

1.3 The ATLAS summarisation architecture

The general summarisation architecture in ATLAS includes the main parts displayed in Figure 1.

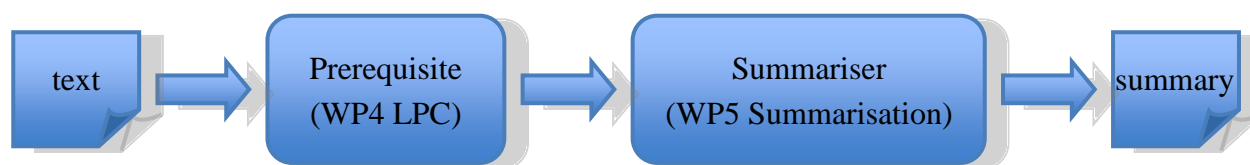


Figure 1: The block architecture of the summarisation process in ATLAS

As such, the summarisation process on a text starts after the language processing chain has been completed.

Going inside into the main modules of Figure 1: The block architecture of the summarisation process in ATLAS Figure 1, their internal architecture is shown in Figure 2 and Figure 3.

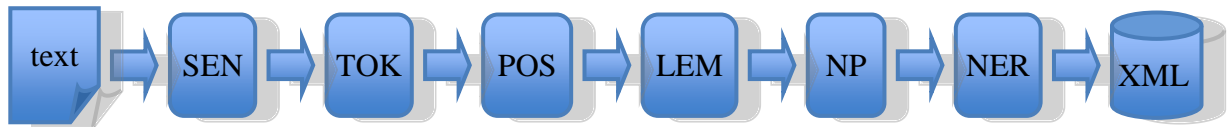


Figure 2: Details of the summarisation prerequisite chain (WP4).

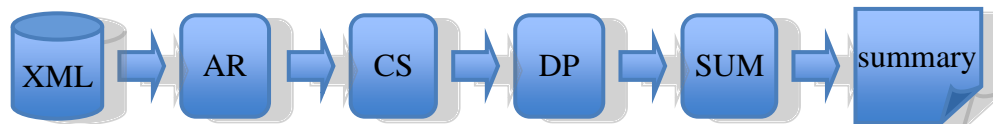


Figure 3: Details of the summarisation chain (WP5)

The abbreviations in Figures 2 and 3 have the following meaning:

SEN = sentence splitter

TOK = tokeniser

POS = part of speech tagger

LEM = lemmatiser

NP = noun phrase chunker

NER = name entity recogniser

AR = anaphora resolver

CS = clause splitter

DP = discourse parser

SUM = summariser

The intermediate format between the prerequisite modules and the proper summarisation chain is seen here in XML format, although it can be represented as UIMA CAS objects in the integrated version. All modules could be implemented two ways: either language independent, in which case they are accessing language specific models, or directly language specific. We will report for each individual language of the project the type these modules belong to.



2 SURVEY OF THE LANGUAGE PROCESSING CHAINS ACTING AS PREREQUISITES FOR SUMMARISATION

2.1 Common pre-processing tools

The common pre-processing tools for the ATLAS summarisation task (not shown in Figure 2) could include: the MIME type detector, a text extractor and a language recogniser (See ATLAS D4.1).

2.2 Language specific prerequisite tools

The language processing chains, acting as prerequisites of the summarisation task, for each language in the project have been thoroughly described in the deliverable D4.1 ATLAS Language Processing Chains. They are listed below.

2.2.1 Bulgarian

The following language processing tools make part of the Bulgarian summarisation prerequisite chain: BG sentence splitter, BG tokenizer, BG Part of Speech tagger, BG lemmatizer, BG Noun phrase extractor, BG Named entity recognizer. The tools exchange data among the chain using the so called vertical format.

2.2.2 English

The following language processing tools make part of the English summarisation prerequisite chain: EN sentence splitter, EN tokenizer, EN POS Tagger, EN lemmatizer, EN noun phrase extractor, EN Named entities recognizer.

2.2.3 German

The following language processing tools, part of the Weighted Constraint Dependency Parser (WCDP) and open source (GATE) modules, make part of the German summarisation prerequisite chain: DE tokenizer modified to recognize also sentence boundaries, an initial POS tagger whose output is corrected by the WCDG, lemmatizer, Proper Names extractor, NP-



chunker and NE recogniser.

2.2.4 Greek

The following language processing tools make part of the Greek summarisation prerequisite chain: EL tokeniser and sentence splitter, EL part-of-speech tagger, EL lemmatizer, EL noun phrase extractor, EL Named Entity Extractor.

2.2.5 Polish

The following language processing tools make part of the Polish summarisation prerequisite chain: PL sentence splitter, tokenizer, lemmatizer, morphological analyser, NP extractor, NE recognizer.

2.2.6 Romanian

The following language processing tools make part of the Polish summarisation prerequisite chain: a pre-processing module performing sentence splitting, tokenization, lemmatization and POS tagging, RO NP extractor, and RO NE recognizer.



3 THE PROPER SUMMARISATION TOOLS

3.1 Anaphora resolution

3.1.1 The model and the engine

Anaphora is the phenomenon of reiteration of an entity (called “antecedent”) by a reference (called “anaphor”) that points back to that entity. For practical reasons, we will call “referential expressions” (RE) both participants in an anaphoric relation. The pointing back (reference) is called *anaphor* and the entity to which it refers is called *antecedent*. As such, *anaphora resolution* (AR) will be understood as the process of identifying the antecedent of an anaphor. Actually, during the reading of a text, it is very likely that an anaphor becomes, in its turn, an antecedent for another co-referential anaphor that follows it.

We will call *discourse entity* (DE) a semantic representation of a referential expression (see Figure 4).

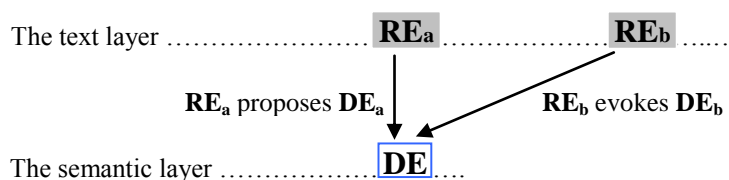


Figure 4: Two-layer representation of a co-referential anaphoric relation

Figure 4 displays the most common situation of two referential expressions, which are co-referential. The antecedent and the anaphor are located on the text layer (RE_a , respectively RE_b). On a deeper semantic layer, their common representation is marked as DE_a . As the reading moves on, a semantic representation is first born when RE_a is encountered. Then, at a later moment, when RE_b is read, it evokes the DE already built by RE_a . Any subsequent co-referring referential expression will in its turn evoke the same DE. One way of representing discourse entities in NLP systems is as feature structures, which are lists of attribute-value pairs. The exact configuration of these attributes, as well as their types (range of accepted values) should be evidenced by the anaphora resolution model.

**Ex. 1**

Mary went to the store. She bought some fruits.

In this example the pronoun *She* is the anaphor and *Mary* is the antecedent. An antecedent is usually a noun phrase and is positioned before the anaphor. Linguists call the relation between a pronoun and the expression it refers, being they in the same sentence or in different sentences, when the pronoun is positioned before the referred expression as cataphora (see Ex. 2).

Ex. 2

10 minutes after he arrived home, John went to sleep.

Anaphora resolution is defined in (Orăsan et. al, 2008) as the process of resolving an anaphoric expression to the expression it refers to. In (Cristea and Dima, 2001) the anaphora resolution process implies three layers (Figure 5):

- a *text layer*, containing referential expressions (RE) as they appear in the discourse;
- an *intermediate, projection, layer*, containing any specific information that can be extracted from the corresponding referential expression, usually represented in feature structures (FS);
- a *semantic layer*, containing descriptions of the discourse entities (DE). Here the information contributed by chains of referential expressions are accumulated.

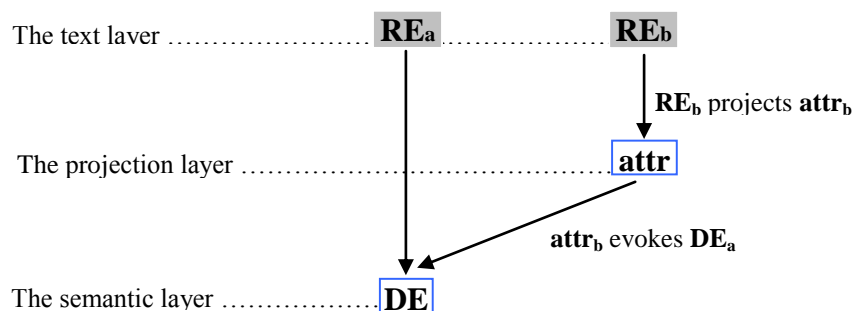


Figure 5: Three layers representation of co-referencing REs (Cristea and Dima, 2001)

Human beings read texts linearly and this makes that the process of interpretation of anaphora be also a linear one (or incremental). Most of the anaphors are resolved at the time of



reading them, although, sometimes, when an ambiguity occurs, the decision is postponed in the hope that the following text will provide more data to help the disambiguation process.

In (Cristea and Dima, 2001), the way the projecting and evoking mechanisms function is defined as a model described four components:

- a *set of primary attributes*, sensed on the projection layer and reported on the projection and semantic layer;
- a *set of knowledge sources*, manifesting as virtual processors and implemented in the actions of the prerequisite processors, fetching values for the primary attributes of the representations of the text layer;
- a *set of heuristics* or rules intended to co-operate in order to decide if an RE introduces a new DE or refers to an already existing DE;
- a *set of rules that configure the domain of referential accessibility* – this component has to establish the set of Des against which the resolution should be decided and their order according to prominence and discourse structure criteria.

This model is implemented as the RARE (*Robust Anaphora Resolution Engine*) module in ATLAS. RARE is a tool able to compute anaphora resolution co-reference chains, proprietary of UAIC-FII. It uses a mixed approach (applies a collection of symbolic rules whose weights are optimised with genetic algorithms). The core of the system is language independent, but in order to localise it to one language or another it requires specific resources (see section 3.1.2 of this document).

Anaphora resolution is one of the key steps of the discourse parser, by resolving anaphoric pronouns, automatically generated summaries may be more cohesive and, thus, more coherent. Calculating scores for references and transitions would be impossible without the proper identification of the co-referential chains. Moreover, in our summarization task resolving anaphors is also essential in order to avoid the problem of dangling pronouns – pronouns that do not have an antecedent in the summary.

3.1.2 Localisation of RARE

All partners that decided to use RARE in their approach for AR, had to adapt the general RARE machinery to work for their languages, a process called localisation. In order to make it more easily parameterisable for different languages, recently RARE has undergone major transformations. The most important changes referred to the manner in which the rules



incorporating matching conditions between anaphor and antecedent are defined. These rules are responsible to decide whether a referential expression introduces a new discourse entity, not mentioned before, or it refers (evokes) one already mentioned, and in this case – which one. In RARE, the text is processed left-to-right and a decision is taken each time a new referential expressions is met. The motor leaves behind chains of co-referential expressions. Each such chain is characterized by a feature structure that sums up all features of the Res in the chain. This is what we call a discourse entity (DE) – see also above. To give an example, suppose John Smith, a professor of computer science of 70 years of age, has been referred in a short text, as: *John Smith, Mr. Smith, he, John, him, he, a professor of computer science, the old man, the professor, John Smith, and he* again. At the end of the text, RARE ideally leaves behind a DE which includes the following feature structure: [ID = DE009; SEM = person, professor of computer science, old man; GEN = male; NAME = John Smith] as well as links to all the corresponding Res on the surface text.

The process runs as follows. The text is unfolded left to right. When a new RE, say RE_x , is met, its set of morphological, syntactic and semantic features is tested against the most recently referred Des which have been left behind by the motor. If, among them, there is one, say DE_y , for which the matching score of the pair (RE_x, DE_y) individualises significantly well, then the actual RE_x is added to the already existing chain of referential expressions attached to DE_y , or else a new DE is generated (proposed – in Figure 1), the RE_x 's features are copied onto it and RE_x becomes the first referential expression of a new chain.

There are three types of rules used for matching the pair (RE_x, DE_y) :

- **certifying rules**, which, if evaluated to TRUE on a pair (RE, DE), certify without ambiguity the DE as a referent for the RE. For instance, identical proper name usually denote the same person. In the example above, due to such a rule, the last RE *John Smith* is decided to refer the DE identified by DE009 above;
- **demolishing rules**, which rule out a possible DE as referent candidate of a RE. RARE includes a cabled demolishing rule invalidating the attempt to establish a co-referential link between nested referential expressions. In the example above, this rule invalidates a co-referential link between *computer science* and *a professor of computer science*;
- **promoting/demoting rules**, which increase/decrease a resolution score associated with a pair (RE, DE). A match of the condition expressed in such a rule adds to the overall resolution score associated with the pair (RE, DE) a positive value. If the current RE has



no certifying rule with any recorded DE, its candidate DE will be the one having the best overall score among those with which the RE experiences no triggered demolishing rule.

In order to make RARE work for a language L, the language engineer has to build two types of resources: a set of external files to be used in a pre-RARE phase, and a set of files accessed by RARE. The partners have been instructed to do the following steps:

- Copy from the ATLAS server the two executables: `pre-RARE.jar` and `RARE.jar`, as well as the `runRARE.bat` file and a short readme showing how to run RARE from the command line (for those not having Windows), and place them on your computer in an empty directory – say X.
- Make a sub-directory of X called `Resources`.
- Build a file called `copulativeVerbs.txt`, containing the list of copulative verbs¹ in L and place it in the `Resources` directory. This is a text file containing one lemma of a copulative verb on each line.
- Build a file called `determiners.xml`, containing the list of determiners (articles) in L and place it in the `Resources` directory. This is a text XML file including 3 types of articles: demonstrative, definite and indefinite. The tags should be: `<demDET>`, `<defDET>`, `<undefDET>`. Under each such tag you should place one such determiner, for example: `<defDET>the</defDET>`.
- Build a file called `familyNames.txt`, containing a list of family names in L and place it in the `Resources` directory.
- Build a file called `femaleNames.txt`, containing a list of female names in L and place it in the `Resources` directory.
- Build a file called `maleNames.txt`, containing a list of male names in L and place it in the `Resources` directory.
- Build an XML file called `rules.xml` describing specific co-reference constraints in L, and place it in the `Resources` directory. This file should contain a set of rule XML elements. The order of these elements in your file is not important.

¹ A *copulative verb* (often called a **linking verb**) is a verb that links a subject to a predicative noun. Examples: *John is a doctor. Marry will become a teacher. Her mother was very educated.*



The attributes of a `rule` element are:

- `ID` – a unique string of characters, having only a mnemonic role;
- `type` – should be “certifying”, “demolishing” or “promoting”. “certifying” type rules describe RE-DE conditions that will impose coreferential links; “demolishing” type rules describe RE-DE conditions for impossible co-referential links; “certifying” and “demolishing” type rules should not include “score” attributes. The “promoting” type rules describe conditions that contribute to establishing an RE-DE pair as coreferential: the value of the “score” attribute will be added to other scores contributed by other rules on the same pair;
- `score` – a positive number. In case the constraints defined by the rule prove to be matched between an RE and a DE, this score modifies the overall matching value of the pair (RE, DE) by addition. As such, for the current RE, that DE will be considered as the best candidate as an antecedent which has the biggest score among all.

The matching conditions of a `rule` element are defined by sub-elements of the following two types, which could be in any number and in any order:

- `constraint` – defines exact matching conditions on features of either the RE or the DE. If a rule element includes more constraint sub-elements, then all the constraints declared should be satisfied on the same (RE, DE) pair in order for the matching score associated with this pair to be modified by the score. A constraint element should have the following attributes compulsorily filled-in:
 - `on` – with the possible values: “RE” or “DE”. If `on` = “RE” the constraint applies to the referential expression, and if `on` = “DE” – it refers to the discourse entity;
 - `attr` – indicating the name of a (morphological, syntactic or semantic) feature (attribute) the constraint applies to. Example: `attr` = “POS”;
 - `val` – indicating the value the `attr` attribute should have in order for the constraint to be evaluated to TRUE. If the `val` attribute contains more than one value, the condition is satisfied if one of them is the value of the feature `attr` (therefore the values are OR-ed).
- `constraintNEG` – defines a negated condition: if the conditions described by the pair `attr-val` on the RE or DE, as set by the `on` attribute, are evaluated to FALSE, then the



`constraintNEG` is evaluated to `TRUE`.

- `relation` – states a relation that should be verified between features belonging to the RE and the DE. Its compulsory attributes are:
 - `pred` – sets the name of the predicate that should be evaluated to `TRUE` on the attributes (of RE and DE) defined in the `attrList`. For instance: `"equal"`;
 - `attrList` – indicates a list of features (same, both on the RE as well as on the DE) that the `pred` predicate applies to. All the attributes in the list should observe the restriction (therefore the relation acts as an AND). For instance: `"NUM GEN"`;

The sequence of `constrain`, `constrainNEG` and/or `relation` declarations should involve an RE as well as a DE. Only this way a score could be associated to such a pair.

Annex 1 contains a commented example for English.

Other needed resources are:

- `stopwords.txt` – a list of stop words;
- `tagset.xml` – the list of tags;
- `window.xml` – giving the length of a window where the antecedent should be looked for.

In order to run RARE, place your input file in the X directory and click on the `runRARE.bat` file.



RARE technical record

- Required language resources: for the localisation of RARE a set of files are needed, as will be described in section 3.1.2.
- Programming language: Java, platform independent.
- OS compatibility: Windows and Linux.
- Access conditions: the core RARE code is free for non-commercial use.
- Copyright: UAIC (core RARE), each partner for their localisation.
- Input:
 - for the stand alone package (at the exit of the LPC): XML text containing TOK (with attributes: ID, POS, LEMMA, NUMBER, optionally GENDER, PERSONTYPE; language specific variations are accepted), NP (with attributes ID, HEADID), S (with the attribute ID), P (with the attribute ID). Coding is UTF-8;
 - for the UIMA package: jCas containing types: RichToken (with attributes BEGIN, END, POS, BASE), Sentence (with attributes BEGIN, END), NpChunkWithHead (with attributes BEGIN, END, HeadToken, Parrent, Children).
- Resources: copulativeVerbs.txt, rules.xml, stopwords.txt, tagset.xml, window.xml
- Output format (only contributed by RARE):
 - as XML file: an UTF-8 coding containing the co-referent chains found in the input file: DE (with attributes: ID, ReList); NP (with attributes REF: ID of a DE element);
 - as UIMA data type for Markable annotation: uaic.uimatypes.CoreferenceChain (with the attribute DE, containing itself a list of types NpChunkWithHead).
- Status: stable, localised versions under constant improvement.



3.2 *Segmentation of sentences at clause boundaries*

3.2.1 State of the art

A clause is a grammatical unit of a sentence that includes, minimally, a predicate and an explicit or implied subject, and expresses a proposition (Nguyen, 2009). The identification of clause boundaries is important for a number of natural language processing (NLP) applications, such as machine translation, text-to-speech systems, parallel text alignment, building the discourse structure and automatic summarization. Numerous techniques are used to recognize clause boundaries for different languages, where some are rule based (Leffa, 1988), where the idea is to reduce clauses to a noun, an adjective or an adverb, and others are hybrid methods, like in (Parven *et al.*, 2011) and (Orăsan, 2000), where the results of a machine learning algorithm, trained on an annotated corpus, are processed by a shallow rule-based module in order to improve the accuracy of the method. Pușcașu (2004) presents a modified version of Orăsan's method. She transfers the technique that Orăsan describes for English to the Romanian sentence with good results. In (Șoricuț and Marcu, 2003) the discourse segmentation task is formulated as a binary classification problem of deciding whether or not to insert a segment boundary after each word in the sentence. In (Subba and Di Eugenio, 2007) a discourse segmenter based on artificial neural networks is presented and represents a first step to create a discourse parser that uses the discourse segmenter.

In our approach to discourse parsing and summarisation we will consider clauses as the elementary discourse units (*EDU*) of the discourse structure and the building blocks of the summaries.

3.2.2 Methodology

Our approach to discourse segmentation starts from the assumption that a clause is headed by a main verb or a verbal compound. Verbs and verb compounds are considered pivots and clause boundaries are looked for in-between them. Verb compounds are sequences of more than one verb in which one is the main verb and the others are auxiliaries, infinitives, conjunctives that complement the main verb and the semantics of the main verb in context obliges to take the whole construction together. An example of a compound verb is “like to



swim” (Ex. 3) where “like” and “swim” are two verbs between which we don’t want to place a clause boundary because the predicate “like” takes as a compulsory argument an object, here “to swim”. Identically, “managed to annotate” (Ex. 4) should be considered as being a compound verb.

Ex. 3

<When I go to the river,> <I like to swim with friends.>

Ex. 4

<As I expected,> <because he is very skilful,> <John managed to annotate the whole file.>

The exact place of a clause boundary between verbal phrases is, in many cases, indicated by discourse markers (key words or expressions) like in Ex. 5 and Ex. 6. A discourse marker, or, simply, a marker, is a word or a group of words having the function to signal a clause boundary and/or to signal a rhetorical relation between two text spans. When they are missing, like between the 1st and the 2nd clauses in Ex. 7, boundaries can be indicated by punctuation marks and can be put in evidence automatically by statistical methods. Such methods are trained on explicit annotations given in manually built files.

Ex. 5

<Verbs and verb compounds are considered pivots> <and clause boundaries are looked for in-between them.>

Ex. 6

<Markers are good>^[1] <because they can give information on the discourse structure.>^[2]

Ex. 7

<Although the snow was falling uninterruptedly,>^[1] <the Azuga slope was in pretty good condition.>^[2]

Ex. 8

<On the one hand, the system should be easy to use by anyone.>^[1] | <On the other, it should be robust.>^[2]

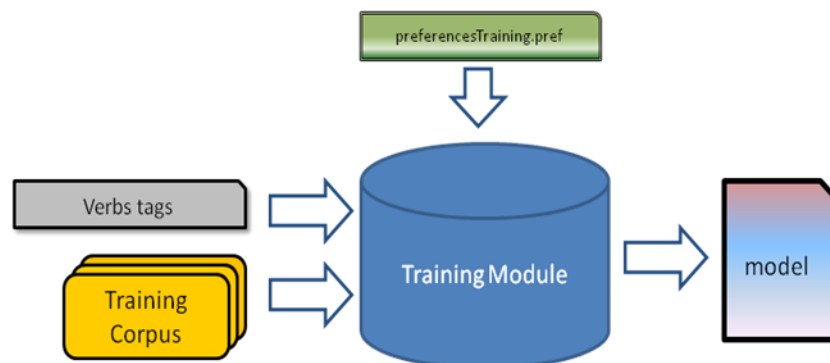
**Ex. 9**

<When Marvin was ten years old,>^[1] <his father took him through the long, echoing corridors>^[2] <that led up through Administration and Power,>^[3] <until at last they came to the uppermost levels of all>^[4] <and were among the swiftly growing vegetation of the Farmlands.>^[5]

In order to recognize the markers, a training is organized by using a window of n POS tags to the left of the marker and m POS tags to the right. For the cases in which we don't have any marker at the boundary between clauses a symmetrical window of l POS tags is used. The values of the three parameters m , n and l are set in calibration experiments.

3.2.3 The CS Training module

Based on the manual annotated files, the *Training module* of the CS, presented in Figure 6, extracts a model which will be used by the *Segmenter module*. The model incorporates patterns of use of markers used to decide the segmentation boundaries.

**Figure 6: The Training module**

Six files make up the model for a language, and for exemplification we will use the next example along with the annotation at clause and marker level in XML format presented in Table 1.

Ex. 10

<In times past there lived a king and queen,>^[1] <who said to each other every day of their lives,>^[2] <“Would that we had a child”!>^[3]



```
<S ID="S0">
  <CLAUSE CONTINUE="" ID="CLAUSE1">
    <W ID="W5" LEMMA="in" POS="IN">In</W>
    <W ID="W6" LEMMA="time" POS="NNS">times</W>
    <W ID="W7" LEMMA="past" POS="IN">past</W>
    <W ID="W8" LEMMA="there" POS="EX">there</W>
    <W ID="W9" LEMMA="live" POS="VBD">lived</W>
    <W ID="W10" LEMMA="a" POS="DT">a</W>
    <W ID="W11" LEMMA="king" POS="NN">king</W>
    <W ID="W12" LEMMA="and" POS="CC">and</W>
    <W ID="W13" LEMMA="queen" POS="NN">queen</W>
    <W ID="W14" LEMMA="," POS=",">,</W>
  </CLAUSE>
  <CLAUSE CONTINUE="" ID="CLAUSE2">
    <MARKER CONNECT="" ID="MARKER1" NUC="N_S" TYPE="">
      <W ID="W15" LEMMA="who" POS="WP">who</W>
    </MARKER>
    <W ID="W16" LEMMA="say" POS="VBD">said</W>
    <W ID="W17" LEMMA="to" POS="TO">to</W>
    <W ID="W18" LEMMA="each" POS="DT">each</W>
    <W ID="W19" LEMMA="other" POS="JJ">other</W>
    <W ID="W20" LEMMA="every" POS="DT">every</W>
    <W ID="W21" LEMMA="day" POS="NN">day</W>
    <W ID="W22" LEMMA="of" POS="IN">of</W>
    <W ID="W23" LEMMA="their" POS="PRP$">their</W>
    <W ID="W24" LEMMA="life" POS="NNS">lives</W>
    <W ID="W25" LEMMA="," POS=",">,</W>
  </CLAUSE>
  <CLAUSE CONTINUE="" ID="CLAUSE3">
    <W ID="W26" LEMMA="&quot;" POS="``">"</W>
    <W ID="W27" LEMMA="will" POS="MD">Would</W>
    <W ID="W28" LEMMA="that" POS="IN">that</W>
    <W ID="W29" LEMMA="we" POS="PRP">we</W>
```



```

<W ID="W30" LEMMA="have" POS="VBD">had</W>
<W ID="W31" LEMMA="a" POS="DT">a</W>
<W ID="W32" LEMMA="child" POS="NN">child</W>
<W ID="W33" LEMMA="!" POS=".">!</W>
<W ID="W34" LEMMA="&quot;;" POS="'">"</W>

</CLAUSE>

</S>

```

Table 1: Example of manually annotated sentence at clauses and markers

- positive examples (**Error! Reference source not found.**, file name: `positive_markers.xml`) – contexts of occurrence of words or sequences of words which should be considered markers; the contexts are expressed as configurable length sequences of POS tags notated as XML files; in this example were used the configuration with three POS tags to the left and one POS tag to the right of the marker (*and queen* , *who said*).

```

<LENGTH LEFT="3" RIGHT="1"/>
<MARKER ID="M0" LENGTH="1" TYPE="FRONT">
  <TEXT>
    <W ID="W0" LEMMA="who" />
  </TEXT>
  <PATTERN ID="P0" FREQUENCY ="1">
    <LEFT>
      <W ID="W1" POS="CC"/>
      <W ID="W2" POS="NN"/>
      <W ID="W3" POS="," />
    </LEFT>
    <RIGHT>
      <W ID="W4" POS="VBD"/>
    </RIGHT>
  <DISTANCE D1="5"/>

```



```
</PATTERN>
```

Table 2: Example of positive marker context

- negative examples (Table 3, file name: `negative_markers.xml`) – identical as above, but the contexts show examples where the words/sequences should not be considered markers; the list of negative examples is computed after all positive examples were identified within the corpus because we are looking only for contexts of words/sequences of words that were already annotated as markers (*lived a king and queen*);

```
<LENGTH LEFT="3" RIGHT="1"/>
  <MARKER ID="M0" LENGTH="1">
    <TEXT>
      <W ID="W0" LEMMA="and" />
    </TEXT>
  <PATTERN ID="P0" FREQUENCY = "1" >
    <LEFT>
      <W ID="W1" POS="VBD"/>
      <W ID="W2" POS="DT"/>
      <W ID="W3" POS="NN"/>
    </LEFT>
    <RIGHT>
      <W ID="W4" POS="NN"/>
    </RIGHT>
    <DISTANCE D1="3"/>
  </PATTERN>
```

Table 3: Example of negative marker context

- clause boundaries (Table 4, file named: `training_mv.xml`) – contexts, also of configurable length, of boundaries without explicit markers noted as XML files: please refer to Ex. 10: *of their lives , “Would that;*



```

<BOUNDARY ID="28" FREQUENCY="1">
  <LEFT>
    <W ID="112" POS="NN"/>
    <W ID="113" POS=",""/>
  </LEFT>
  <RIGHT>
    <W ID="114" POS="PRP"/>
    <W ID="115" POS="MD"/>
  </RIGHT>
</BOUNDARY>

```

Table 4: Example of clause boundary without marker

- markers types (Table 5, file named: markers_type.xml) – statistics on frequencies of occurrence in the corpus of the markers and their arguments noted as XML files; in the example of Table 5 we can notice that the “and” marker occurs in the corpus 205 times with the N_N pattern and the “which” marker occurs 35 times with the N_S pattern, and in only one case with the N_N pattern.

```

<MARKER ID="M0" LENGTH="1" TYPE="int" CONNECT="">
  <TEXT>
    <W ID="W0" LEMMA="and" />
  </TEXT>
  <T NUC="N_N" CASES="205" />
</MARKER>
<MARKER ID="M1" LENGTH="1" TYPE="int" CONNECT="">
  <TEXT>
    <W ID="W1" LEMMA="which" />
  </TEXT>
  <T NUC="N_N" CASES="1" />
  <T NUC="N_S" CASES="35" />
</MARKER>

```

Table 5: Example of markers type

- compound verbs model (file named verb_model.bin.gz) – features used to build this model are distances between the two verbs, their POS tags and the punctuation



between them. The model is built using the *MaxEnt*² library;

- markers model (file named `marker_model.bin.gz`) – features used to build this model are: lemma and the context of the marker expressed as configurable length sequences of POS tags and distance from the first verb identified within the sentence.

These 6 files are filled in automatically by the *Training module* from a manually annotated training corpus. The corpus consists of short stories manually annotated at clauses and markers, as described in Section 5.1. In (Hilbert et. al, 2006) the list of discourse markers which are indicating possible rhetorical relations is manually developed. Supplementary to this corpus, also needed are the following files considered resources:

- a list of the POS tags characterizing verbal predicates, useful for the compound verbs classification (file named `trainingVerbs.txt`) – compulsory;
- a list of POS of particles that usually glue together components of compound verbs (for instance: *to* in English, in constructions like *likes to swim*, or *să* in Romanian (*îi place să înoate*) (file named `trainingConj.txt`) – optional;
- a list of POS tags representing punctuation (file named `trainingPunct.txt`).

The parameters of the preferences file, called by default `preferencesTraining.pref`, are:

- `inputDirectory`: path to a folder containing the manually annotated corpus (described in Section 5.1);
- `trainingVerbs`: path towards the training verbs file (see above);
- `trainingConj`: path towards the gluing particles file (see above);
- `trainingPunct`: path towards the punctuation POS file (see above).
- `outputDirectory`: path towards a folder where the model will be written;
- `markerLeftLength`: length of a window to the left of the marker defining a learning context;
- `markerRightLength`: length of a window to the right of the marker defining a learning context;
- `verbLength`: length of a simitrical window defining the context of a no-marker signal boundary.

² The maximum Entropy Framework: <http://maxent.sourceforge.net/about.html>



The CS-training module technical record

- Required language resources: for the localisation of the CS-training module a set of files are needed, as will be described in section 5.1.
- Programming language: Java, platform independent.
- OS compatibility: Windows and Linux.
- Access conditions: the CS-training module is free for non-commercial use.
- Copyright: UAIC.
- Input (the module runs only as a stand-alone package):
 - A set of TRAINING GOLD files: XML manually annotated texts containing TOK (with attributes: ID, POS, LEMMA), S (with attribute ID), CLAUSE (with ID, and optionally CONTINUE), MARKER (with ID, NUC, TYPE, CONNECT). Coding is UTF-8;
- Preferences (one file, either created by the Calibration system, as described in Section 3.2.6, or filled in manually, see above):
 - `preferencesTraining.pref`.
- Resources (3 txt files, see above):
 - `trainingVerbs.txt`;
 - `trainingConj.txt`;
 - `trainingPunct.txt`.
- Output: 6 files (configuring the segmentation language model, as described above):
 - `positive_markers.xml`;
 - `negative_markers.xml`;
 - `training_mv.xml`;
 - `markers_type.xml`;
 - `verb_model.bin.gz`;
 - `marker_model.bin.gz`.
- Status: stable, finalised.



3.2.4 The Clause Segmenter module

The *Segmenter module* (

Figure 7) segments the input by applying a machine learning algorithm, to classify pairs of verbs as being or not compound verbs and, after that, applying rules and heuristics based on pattern matching or machine learning algorithms to identify the clause boundary.

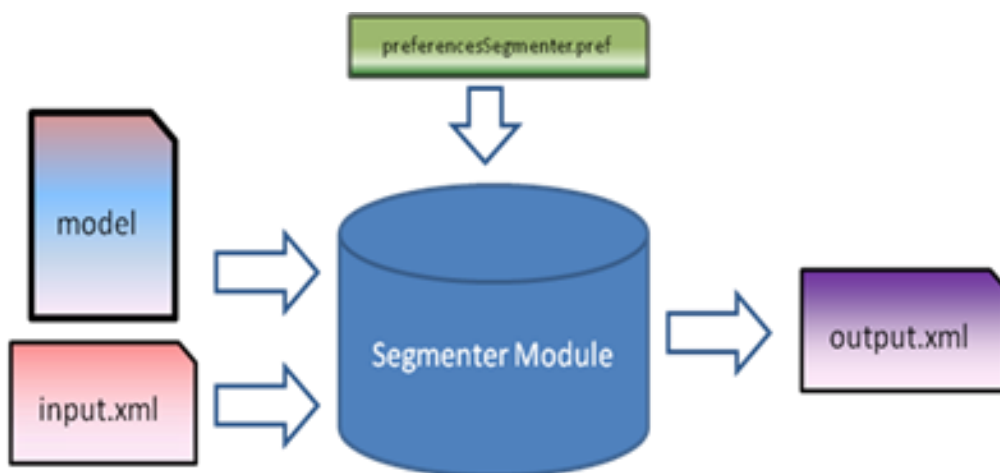


Figure 7: The Clause Segmenter Module

If a marker is selected as a clause boundary the exact place of the segmentation (before or after the marker position) is decided, depending on how it was annotated.

When the *Training module* generates the markers model, for each marker an attribute named `TYPE`, which may take only two values: `FRONT` and `BACK`, representing the position of the marker relative to the boundaries founded in the corpus, will be set. If a marker is annotated in both ways, the value of the `TYPE` attribute will be decided by the cases with greater frequency. For example, if in the manually annotated corpus we have n cases in which the clause boundary was in front of the "and" marker and m cases in which the clause boundary was after the marker, and $n \geq m$, then the value of the `TYPE` attribute will be `FRONT` and the segmentation would look like in Ex. 11:

**Ex. 11**

<Verbs and verb compounds are considered pivots> <and clause boundaries are looked for in-between them.>

Contrary if $n < m$ the value of the `TYPE` attribute will be `BACK` and the segmentation it would be like in Ex. 12:

Ex. 12

<Verbs and verb compounds are considered pivots and > < clause boundaries are looked for in-between them.>

In cases when between two not compound verbs we have several markers it will be selected the one who has the best fit with regard to the *matchingFactor*. Because some markers like “*and*” may occur more times in the corpus than others, thus having more chances to have a bigger score, the *matchingFactor* takes into account also the negative examples of occurrences of the marker context (if this option is enabled), decreasing the score in case of a bigger matching for a negative example. In this way we can avoid some cases like in the next example where, using the negative examples, the score for the “*who*” marker becomes bigger than the score for the “*and*” marker, and the segmentation is made correctly.

Ex. 13

<In times past there lived a king ~~and~~ queen,><who said to each other every day of their lives ...>

A threshold, represented by the *minimumFrequency* parameter, decides where to place a boundary in cases when both example of markers as “*that*” and boundaries without markers as “ ” are found in text between two verbs which were classified as not being compound, like in the next example:

Ex. 14

<...who said to each other every day of their lives,^[2]> <“Would ~~that~~ we had a child”!^[3]>



When we deal with cases when we have no marker the system will choose the case with the highest frequency from the corpus regarding the examples of clause boundaries without markers. In cases where neither markers, nor boundaries without explicit markers are found between a pair of verbs, the text will not be segmented. The algorithm is as follows (Table 6):

Algorithm 1 – Segmentation algorithm

```
1: load model
2: initialize parameters
3: boundaryList = {}
4: for each sentence in input
5:     for each pair of verbs v1,v2
6:         classify(v1,v2)
7:         if v1,v2 are not compound verbs then
8:             newBoundaries=searchBoundaryBetween(v1,v2)
9:         end if
10:    end for each pair of verbs
11:    boundaryList.add(newBoundaries)
12: end for each sentence
13: Output = Segment the input according to boundaryList
```

Table 6: Overview of the segmentation algorithm

The parameters of the preferences file for the CS module, called by default `preferencesSegmenter.pref`, are:

- `inputDirectory`: path to a folder containing the model (described in Section 3.2.3);
- `trainingVerbs`: path towards the training verbs file (see Section 3.2.3);
- `trainingConj`: path towards the gluing particles file (see Section 3.2.3);
- `trainingPunct`: path towards the punctuation POS file (see Section 3.2.3);
- `inputFile`: path towards the file wanted to be segmented;
- `outputFile`: path towards a file where the output file will be located;
- `includeNegativeMarker`: `false/true` taking in consideration or not the negative



examples;

- `includeMarkerEntropy`: true if a machine learning algorithm will be used; false if heuristics will be used;
- `matchingFactor`: defining #of tokens to be matched in the left-to-right total contextual window in order for a boundary to be taken (x10) (filled in by the calibration module);
- `minFrequency`: threshold above which a marker is to be considered;
- `tag`: defines the name of the TOK attribute where the training feature is found (usually POS or MSD).

The CS module technical record

- Programming language: Java, platform independent.
- OS compatibility: Windows and Linux.
- Access conditions: the CS module is free for non-commercial use.
- Copyright: UAIC for the core module; Each partner for their localised versions.
- Input:
 - for the stand-alone package (at the exit of RARE): XML text containing TOK (with attributes: ID, POS, LEMMA), S (with the attribute ID), P (with the attribute ID). Also, the input can tolerate: NP (with attributes ID, HEADID, REF: ID of a DE element), DE (with attributes: ID, ReList). Coding is UTF-8;
 - for the UIMA package: jCas containing types: RichToken (with attributes BEGIN, END, POS, BASE), Sentence (with attributes BEGIN, END). Also, the input can tolerate: NpChunkWithHead (with attributes BEGIN, END, HeadToken, Parrent, Children), CoreferenceChain (with attribute DE, containing itself a list of types NpChunkWithHead).
- Preferences (one file, either created by the Calibration system, as described in Section 3.2.6, or filled in manually, see above):
 - `preferencesSegmenter.pref`.
- Resources (see Section 3.2.3):
 - `trainingVerbs.txt`;
 - `trainingConj.txt`;
 - `trainingPunct.txt`.



- `positive_markers.xml`;
 - `negative_markers.xml`;
 - `training_mv.xml`;
 - `verb_model.bin.gz`;
 - `marker_model.bin.gz`.
- Output (only contributed by the CS module):
- an XML file in UTF-8 containing the new elements: `CLAUSE` markings (with the attribute `ID`);
 - UIMA data type for `CLAUSE` annotation, namely `uaic.uimatypes.Clause` (with attributes `BEGIN` and `END`).
- Status: stable. Localised versions could be improved if more gold data is provided.

3.2.5 The CS Evaluation Module

The *Evaluation module* (Figure 8) needs a test file (the output from the Segmenter module) and a gold file (annotated at least at clause level).

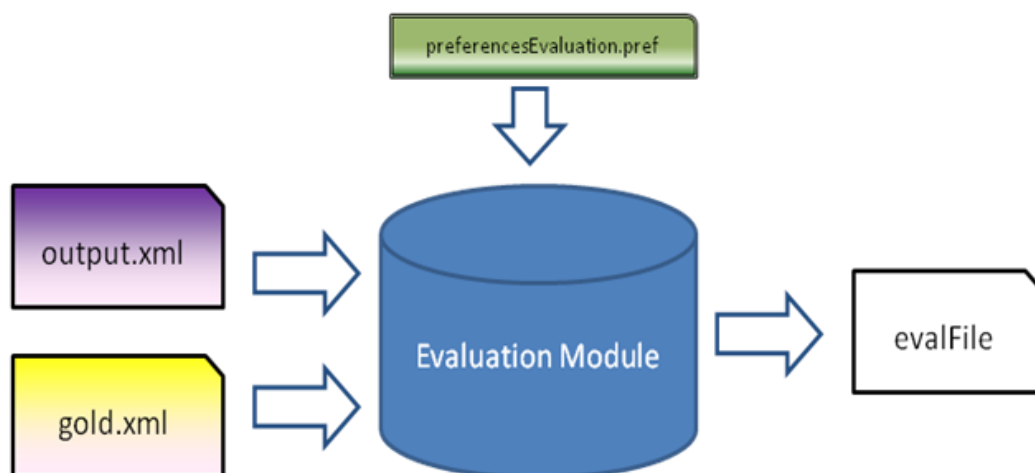


Figure 8 – The Evaluation Module

This module evaluates test files against gold files. Two metrics for evaluation have been considered:

The first metric calculates Precision, Recall and F-Measure, as follows:



$$\textit{Precision} = \frac{\textit{correct clause boundaries in the test file}}{\textit{total \# of clause boundaries in the test file}} ;$$

$$\textit{Recall} = \frac{\textit{correct clause boundaries in the test file}}{\textit{total \# of clause boundaries in the gold file}} ;$$

$$\textit{F - measure} = \frac{2 * \textit{Precision} * \textit{Recall}}{\textit{Precision} + \textit{Recall}} ;$$

The second metric is less restrictive and can be called *Accuracy*:

$$A = \frac{S}{N} ; \text{ where } S = \sum_{i=1}^N s_i ;$$

$$s_i = \min\left(w_i, \frac{1}{w_i}\right) ; \text{ where } w_i = \frac{L_T^{w_i}}{L_G^{w_i}} ;$$

$L_T^{w_i}$ – represents the length of the clause the word w_i belongs to, in the test file;

$L_G^{w_i}$ – represents the length of the clause the word w_i belongs to, in the gold file;

s_i – represents a score attached to the word w_i , the same for all words belonging to the same clause;

N – represents the total number of words in the test file;

S – represents the sum of scores of all words;

A – represents the accuracy, a sub unitary number (this number will be closer to 1 when the test and the gold files tend to include similar boundaries).

The Preferences file has the following fields:

- goldFile: path the GOLD file;
- testFile: path the TEST file;
- evalOutputFile: path the output file;
- evaluationMetric: F (the first metric described above, computing Precision, Recall and F-measure), or A (the second metric of above, computing Accuracy).

The Output file in the case of the first metric has the following fields:

- total gold: <nnn> – indicates the number of segments in the GOLD file;



- `total test: <nnn>` – indicates the number of segments in the TEST file;
- `correct segments: <nnn>` – indicates the number of correct segments in the TEST file;
- a log containing two sections:
 - `misses` – IDs of tokens to the right of the missed boundary;
 - `over` – IDs of tokens to the right of the wrongly indicated boundary.

The Output file in the case of the second metric has the following fields:

- S (as above);
- N (as above);
- A (as above);
- a log containing a list of groups with the following content, for each word of a wrongly segmented clause, separated by “----“:
 - ID of word;
 - Lt (as above);
 - Lg (as above);
 - s (as above).

The CS Evaluation module technical record

- Programming language: Java, platform independent.
- OS compatibility: Windows and Linux.
- Access conditions: free for non-commercial use.
- Copyright: UAIC.
- Input (the CS Evaluation module runs only on XML formatted texts):
 - The GOLD file: an XML manually annotated text containing TOK (with attributes: ID, POS, LEMMA), S (with the attribute ID), CLAUSE (with the attributes ID, CONTINUE);
 - The TEST file: same text as in the GOLD file, XML annotated as an output of the CS module.
- Preferences (one file, filled in manually, see above):
 - `preferencesEvaluation.pref`.



- Output:
 - a text file in UTF-8 containing the fields described above.
- Status: stable, finalised.

3.2.6 The CS Calibration system

At this point, with all three modules that support configurable parameters, normally we could try different configurations of parameters and notice the configuration that achieves the best results. The disadvantage of this method is the time required to perform manual runs, along with the analysis of the results. Considering that several languages are involved and each with specific configuration parameters, this process is cumbersome. The purpose of this calibration system is to ease our work and to identify the combination of parameters that will achieve the best results.

This module needs a configuration file, an input file, a gold file, a training corpus and the Segmenter modules described above (training, segmenter and evaluation).

In the configuration file we must fill in the minimum and maximum values of 5 parameters: *leftMarkerLength*, *rightMarkerLength*, *verbLength*, *Frequency*. Also this file will contain the measure we want to maximize: *Precision*, *Recall*, *Fmeasure* – as conforming to the first defined measure in Section 3.2.5, or *Accuracy* – conforming to the second measure, as defined in Section 3.2.5), a *tag* parameter (defining the name of the attribute giving the lexical category of the tokens in the training phase – POS or MSD), and all the methods that the *Segmenter module* supports (pattern matching, maximum entropy). An



example of the configuration file is given in Table 7.

```

minimLeftMarkerLength = "1"
maximLeftMarkerLength = "3"
minimRightMarkerLength = "1"
maximRightMarkerLength = "3"
minimVerbLength = "1"
maximVerbLength = "2"
minimFrequency = "5"
maximFrequency = "10"
measure = "Fmeasure"
outputFile = "outBG.txt"
tag = "POS"

```

Table 7: Example of a configuration file

The calibration algorithm is as follows (Table 8):

Algorithm 2 – Calibration System

```

1: initialize minimum and maximum values for parameters
2: optimalValues = {}
3: maximumResult = 0
4: for each segmentation method
5:   for each combination of parameters
6:     Run the training, segmenter and evaluation modules
7:     Analyze results
8:     if (result > maximumResult) then
9:       optimalValues = this combination of parameters
10:      maximumResult = result
12:   end if
13: end for each combination of parameters
15: end for each segmentation method
16: Show maximumResult and optimalValues

```

Table 8: A general overview of the calibration system

At the end, the calibration system identifies the method achieving the best results and the corresponding configuration of values of the parameters. Figure 9 gives an overview of the *Calibration System*.

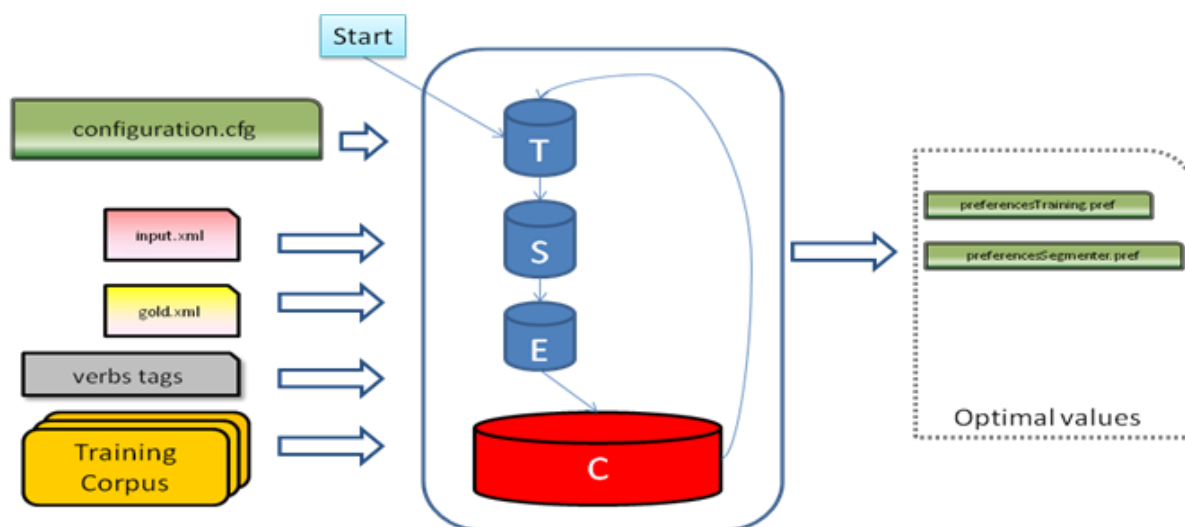


Figure 9: The CS Calibration System

The CS Calibration module technical record

- Programming language: Java, platform independent.
- OS compatibility: Windows.
- Access conditions: free for non-commercial use.
- Copyright: UAIC.
- Input (the CS Calibration module runs only on XML formatted texts):
 - A set of TRAINING GOLD files: XML manually annotated texts containing TOK (with attributes: ID, POS, LEMMA), S (with attribute ID), CLAUSE (with ID, and optionally CONTINUE), MARKER (with ID, NUC, TYPE, CONNECT). Coding is UTF-8;
 - The EVALUATION GOLD file (should not be one of the TRAINING GOLD files): an XML manually annotated text containing TOK (with attributes: ID, POS, LEMMA), S (with the attribute ID), CLAUSE (with the attributes ID, CONTINUE);
 - The INPUT file: an XML text containing: TOK (with attributes: ID, POS, LEMMA), S (with the attribute ID), P (with the attribute ID). Also, the input can tolerate: NP (with attributes ID, HEADID, REF: ID of a DE element), DE (with attributes: ID, ReList). Coding is UTF-8;
- Resources (3 txt files, see above):



- trainingVerbs.txt;
- trainingConj.txt;
- trainingPunct.txt.
- Output (as defined in Section 3.2.3 and Section 3.2.4):
 - preferencesTraining.pref
 - preferencesSegmenter.pref
- Status: stable, finalised.



3.3 *Discourse parsing*

3.3.1 Introduction

Discourse parsing is the process of building a hierarchical model of a discourse from its basic elements (sentences or clauses), as one would build a parse of a sentence from its words (Bangalore and Stent, 2009).

Rhetorical Structure Theory (Mann and Thompson, 1988) is one of the most popular discourse theories. In RST the discourse segments (spans) are plain text, whereas the aggregation of spans presupposes knowledge about the meaning of these spans. This theory uses relationships between segments of text to explicate the coherence of a narrative. In RST a text segment assumes one of two roles in a relationship: the nucleus (N) or satellite (S). Nuclei express what is more essential to the understanding of the narrative than the satellites. The size of a text segment is arbitrary but each should have independent functional integrity (Silva and Henderson, 2005). Relations hold between non-overlapping text segments and are of two kinds: *hypotactic*, and *paratactic*. Hypotactic relations connect one nucleus and one satellite. Paratactic relations hold between text segments of equal importance; that is multiple nuclei.

Noun phrase, also noted *NP* = nouns with all their modifiers (excepting relative clauses), pronouns, proper nouns. The NP notation is contributed by your NP-chunker. Noun phrases could include other noun phrases. The notation for NP should contain a head attribute, that reveals the central element. Examples (NPs in brackets, heads – underlined): <John Smith>, <him>, <<her> hat>, <<the University>'s building>, <two cats>, <a wonderful blonde in <a blue car>>...

Sentence = a grammatical unit of a language. Following the strict rules of grammaticality, a sentence begins with a capital letter and is ended by a punctuation mark, such as a dot (.), a question (?), an exclamation (!) mark, but in colloquial writing (such as chats, emails, etc.), these rules are often disobeyed. A sentence includes at least one clause, and can be as simple as an interjection or as complex as a long construction in which several clauses (see next) are in relation of coordination and/or subordination. It may span from less than a line to several pages.

Clause = a continuous or interrupted text span, which, at a syntactic level, is a constituent of a sentence. At a semantic level a clause includes a statement/event, therefore it should include a verbal construction and its semantic roles.



Elementary discourse unit (EDU) = a span of text (mainly contiguous, but sometimes also interrupted) which is the basic constituent of a discourse structure (also called a rhetorical tree in Rhetorical Structure Theory – see Mann and Thomson). Syntactically, an EDU is a simple sentence containing just one predication or only a clause of a sentence. Actually you can recognize an EDU as the span of text grouped around a main verb (or a verbal construct) and including its semantic roles. In the examples that follow EDUs are marked in sharp brackets.

Nucleus = a span of text (EDU) which, if deleted from its context, makes the text incoherent.

Satellite = a span of text (EDU) which, if deleted from its context, does not harm the overall understandability, even if some details are now missing.

To clarify the difference between nuclei and satellites, please see the following examples. In all the examples that follow, the clauses are enclosed in angular brackets and numbered in square brackets, and vertical bars indicate sentence borders.

Ex. 15

<Although it was raining,>^[1] <when leaving his house,>^[2] <John left home his umbrella.>^[3]

Try to delete any combination of clauses [1], [2], [3]: as you see, the following shorted versions are ok:

<when leaving his house,>^[2] <John left home his umbrella.>^[3]

<Although it was raining,>^[1] <John left home his umbrella.>^[3]

<John left home his umbrella.>^[3]

but the following versions are not ok:

<Although it was raining,>^[1] <when leaving his house,>^[2]

<when leaving his house,>^[2]

<Although it was raining,>^[1]

which shows that the clause [3] should be considered a nuclear EDU, while clauses [1] and [2] should be considered satellites.

Ex. 16

<A nucleus is a span of text which,>^[1] <if deleted from its context,>^[2] <makes the text incoherent.>^[3]



If anyone of these clauses would be deleted, the text loses its sense, therefore all should be considered nuclei.³

Ex. 17

<A satellite is a span of text which,>^[1] <if deleted from its context,>^[2] <does not harm the overall comprehension,>^[3] <even if some details are now missing.>^[4]

In this sentence the only clause that can be deleted is [4], so this should be taken as a satellite.⁴

A narration usually contains a sequence of events, which, mainly, are equally important. It is however possible that a narration includes also some details of some of the steps. In this case the details should be taken as satellites.

Ex. 18

<Early in the morning, John left home.>^[1] | <He intended to go to the lab,>^[2] <curious to find out the results of a night-long run.>^[3] | <When his car arrived at the intersection>^[4] <he find out that the main street was blocked by a demonstration.>^[5] | <He had to take only secondary streets,>^[6] <which made him finally arrive in the lab very late.>^[7]

In Ex. 4 the main line of the story develops as [1]+[2]+[5]+[6], which are nuclei, while [3] is a satellite to [2], [4] – a satellite to [5], and [7] – a satellite to [6].

A *discourse marker* or, simply, a *marker* should be understood as a word or a group of words having the function to signal a clause boundary and/or to signal a rhetorical relation between two text spans. In the following 4 examples, markers are underlined:

Ex. 19

<Markers are good>^[1] <because they can give information on the discourse structure.>^[2]

Ex. 20

<Although the snow was falling uninterruptedly,>^[1] <the Azuga slope was in pretty good condition.>^[2]

³ To be precise, we should notice that in this example, *which* actually belongs to the 3rd EDU, this one being interrupted by the second clause.

⁴ Same notice as above.

**Ex. 21**

<On the one hand, the system should be easy to use by anyone.>^[1] | <On the other, it should be robust.>^[2]

Ex. 22

<When Marvin was ten years old,>^[1] <his father took him through the long, echoing corridors>^[2] <that led up through Administration and Power,>^[3] <until at last they came to the uppermost levels of all>^[4] <and were among the swiftly growing vegetation of the Farmlands.>^[5]

3.3.2 Methodology

The *Discourse Parser* uses a symbolic approach and produces discourse trees, which include nuclearity, but lacking rhetorical relation names: intermediate nodes in the discourse tree have no name and terminal nodes are elementary discourse units, mainly clauses. It adopts an incremental policy in developing the trees, on three levels. First, an elementary discourse tree (EDT) of each sentence is produced, by using a combinatorics algorithm, described in (Cristea et al., 2003) or using incremental parsing. Both methods will be detailed in the following sections. Secondly, the elementary discourse tree of the paragraph is produced by attaching the EDT of the last sentence to the already developed tree on the right frontier (Cristea and Webber, 1997). The same approach is used at discourse level by attaching the paragraph tree of each paragraph to the already developed tree.

At each level, the parser goes on with a forest of developing trees in parallel, ranking them by a global score based on heuristics that are suggested by both Veins Theory (Cristea et al., 1998) and Centering Theory (Grosz et al., 1995). The trees used at the next step are only the best placed trees. The aim of this filtering step is to reduce the exponential explosion of the obtained trees.

3.3.3 Incremental parsing of the discourse

This section gives a description of the basic incremental parsing approach. The input to the parser is the text augmented with information about:

- SEN – sentence boundaries
- TOK + POS + LEMMA – tokens with specification of POS and lemma



- NP – noun phrases (or referential expressions)
- DE – coreference chains
- CLAUSE – clause boundaries (or discourse units)

as resulted from the prerequisites modules, RARE and the clause splitter (see Figure 2 and Figure 3).

The principle of sequentiality is reproduced bellow from (Marcu, 2000):

A left to right reading of the terminal frontier of the tree associated with a discourse must correspond to the span of text it analyses in the same left-to-right order.

The approach we present borrows the two operations used in (L)TAG (*lexicalized tree-adjointing grammar*) (Joshi and Schabes, 1997): *adjunction* and *substitution*.

The adjunction operation takes an initial tree or a developing tree (D-tree) and creates a new developing tree by combining the D-tree with an auxiliary tree (A-tree). The auxiliary tree has a special node called “foot node”, denoted by the * sign and placed on the terminal frontier of the auxiliary node. The sub-tree cropped by the adjunction operation is removed temporarily from the D-tree and is attached to the A-tree by replacing the foot node. Figure 10 depicts this idea.

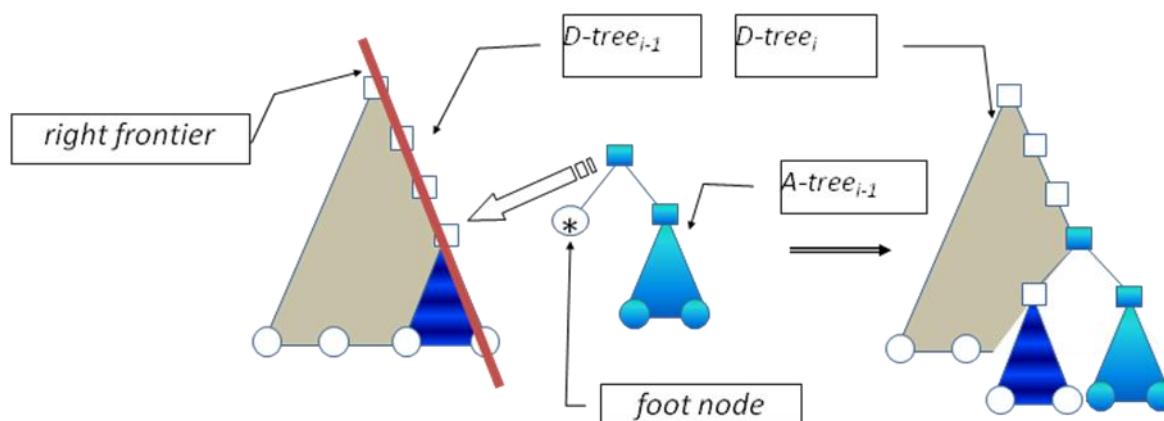


Figure 10: Adjunction operation

An auxiliary tree whose foot node is placed as a left child of its parent node is called a *left footed auxiliary tree* and in our approach only this kind of A-tree has been used, because only by using this type of A-trees the adjunction on the right frontier results in proper built D-trees at each step (Mitocariu et al, 2012). The right frontier represents the path from the root of the



tree to its rightmost leaf node.

The substitution operation replaces a node on the terminal frontier, called “*substitution node*”, and denoted by the \Downarrow sign, with an auxiliary tree (Figure 11).

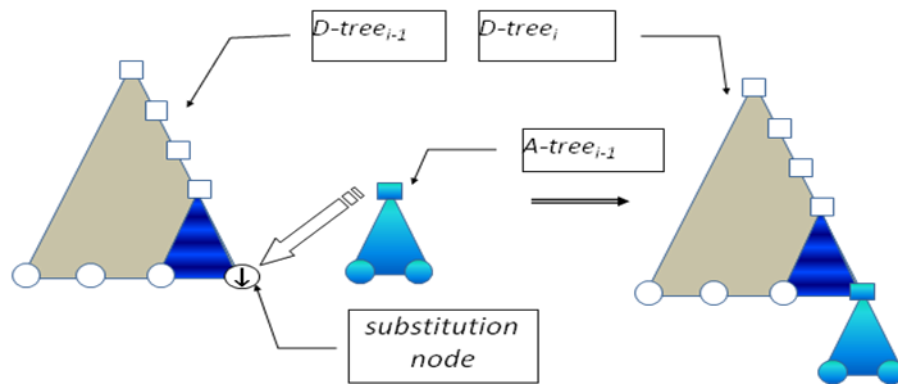


Figure 11 – Substitution operation

In a discourse structure with no substitution nodes, the adjunction operation is limited only to the *right frontier*. Since, during the incremental parsing, adjunction operations could occur when there are open also one or more substitution nodes, Cristea and Webber (1998) defines also the *internal right frontier*. The internal right frontier is the right frontier of the sibling node of the most inner substitution node, as illustrated in Figure 12.

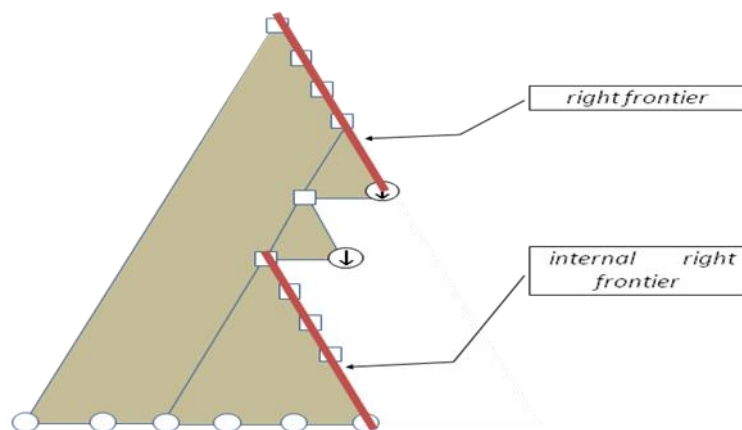


Figure 12 – Internal right frontier

We start from the assumption that a discourse consists of several paragraphs, but at least one is mandatory. Each paragraph has one or more sentences, which in turn has one or more clauses. All auxiliary trees describing the structure of a paragraph that can be built are presented in Figure 13. Their root nodes can have any of the types: N_N , N_S and S_N . We will see that at the paragraph level several types of auxiliary trees describing a sentence will be introduced. The general incremental parsing algorithm at the discourse level is described in



Table 9 where the *discourseWorkflow* is composed only of adjunction operations on the right frontier:

Algorithm 3 – Incremental Parsing at the discourse level

```

1: initialize PDTs = {0} - the set of partial developing trees
2: initialize PDTg = {0} - generated trees
3: for each paragraph in discourse
4:   create listOfParagraphTrees using paragraph level parsing;
5:   for each D-tree in PDTs
6:     for each A-tree in listOfParagraphTrees
7:       signal the discourseWorkflow to generate the next generation of D-
       trees
8:       add generated trees to PDTg
9:     end for each A-tree
10:  end for each D-tree
11:  evaluate PDTg using heuristics and keep the best ones
12:  PDTs = PDTg
13:end for each sentence
14:evaluate PDTs using heuristics and keep the best one

```

Table 9: A general overview of the incremental parsing at the discourse level

In order to create the paragraph trees we use a different approach in building the auxiliary trees by using information provided by internal markers. As was said before, external markers are markers that relate a sentence to another span of text outside itself. According to the type of the marker, four types of auxiliary trees can be built, as shown in Figure 13.

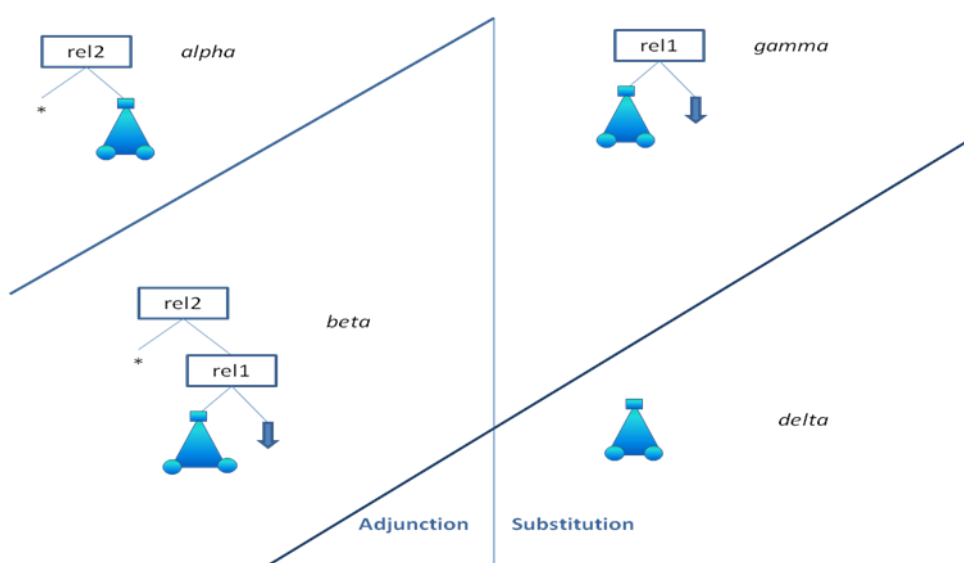


Figure 13: Auxiliary trees



The uses of different types of auxiliary trees is determined by two factors:

- the type of operation by which they can be glued on the developing structure: *alpha* and *beta* are used only for adjunction operations, while *gamma* and *delta* for substitution operations;
- the auxiliary tree introduces or not an expectation: *beta* and *gamma* are auxiliary trees that raise an expectation and *alpha* and *delta* are auxiliary trees which do not raise an expectation.

The workflow of the incremental parsing at paragraph level is illustrated in Figure 14.

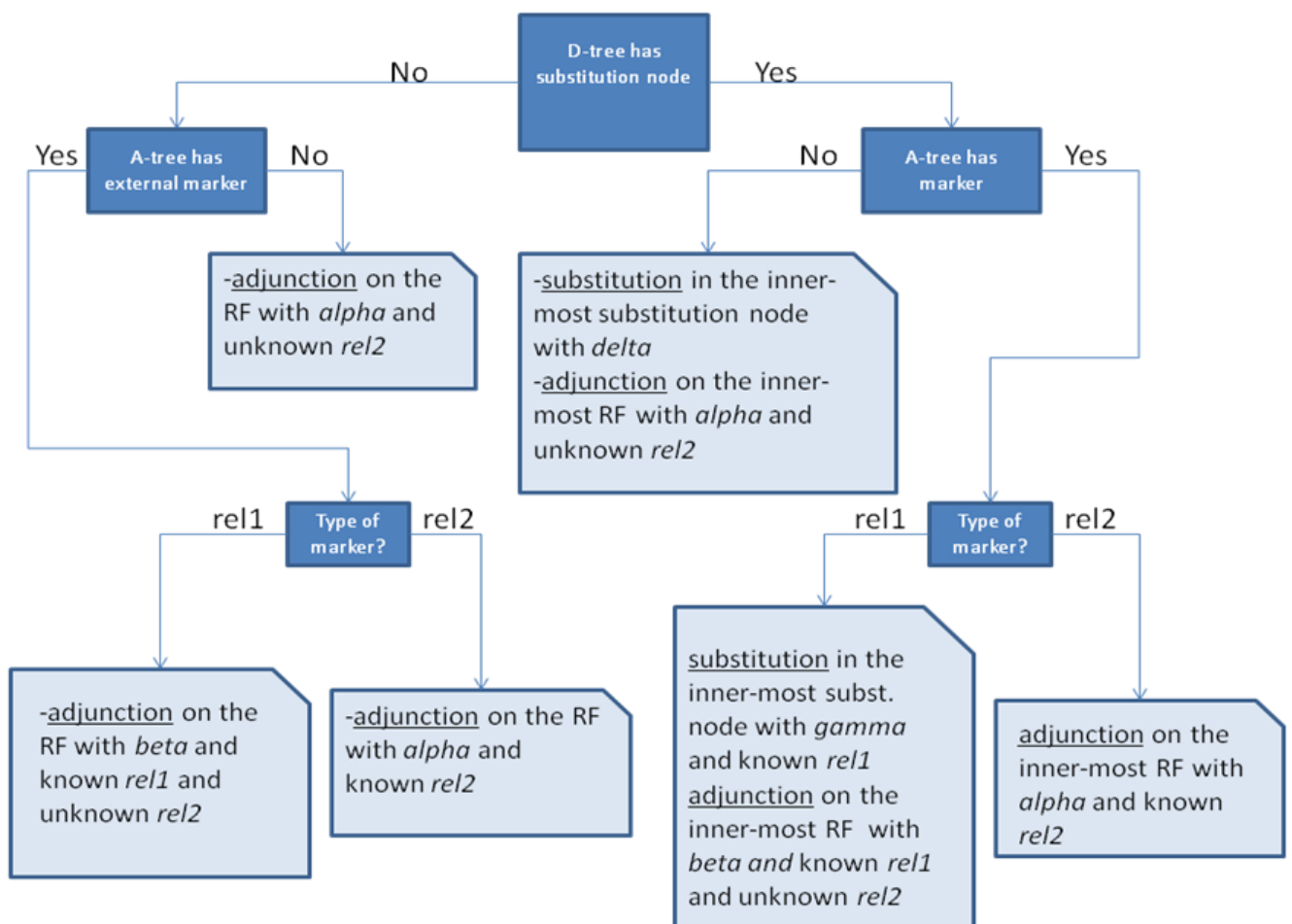


Figure 14: Workflow of the incremental parsing at paragraph level

Based on the *paragraphWorkflow*, and using the same idea as parsing at the discourse level, but with different types of auxiliary trees, the incremental parsing at the paragraph level is described in Table 10.

Algorithm 4 – Incremental Parsing at paragraph level



```

1: initialize PDTs = {0} - partial developing trees
2: initialize PDTg = {0} - generated trees
3: for each sentence in paragraph
4: create listOfSentenceTrees using incremental parsing or combinatorics;
5: for each D-tree in PDTs
6:   for each A-tree in listOfSentenceTrees
7:     signal the paragraphWorkflow to generate next generation of D-
      trees
8:     add generated trees to PDTg
9:   end for each A-tree
10: end for each D-tree
11: evaluate PDTG using heuristics and keep the best ones
12: PDTs = PDTg
13: end for each sentence
14: evaluate PDTs using heuristics and keep the best one

```

Table 10: A general overview of the incremental parsing at the paragraph level

3.3.4 Incremental parsing at the sentence level

At the sentence level, the incremental parsing seems to be quite the same as at the paragraph level, the only difference being that the internal markers give information of how the auxiliary trees must be computed. To understand better how parsing at sentence level works, let us consider Ex. 23:

Ex. 23

<John is determined to pass the NLP exam so,> <because he had missed many courses> <and was only vaguely implicated in the working sessions,> <he will have a hard time until summer.>

1. [*John is determined to pass the NLP exam so,*]

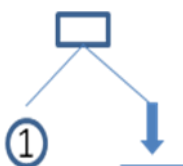


Figure 15: Initial discourse unit raises an expectation

2. [*because he has missed many courses*]

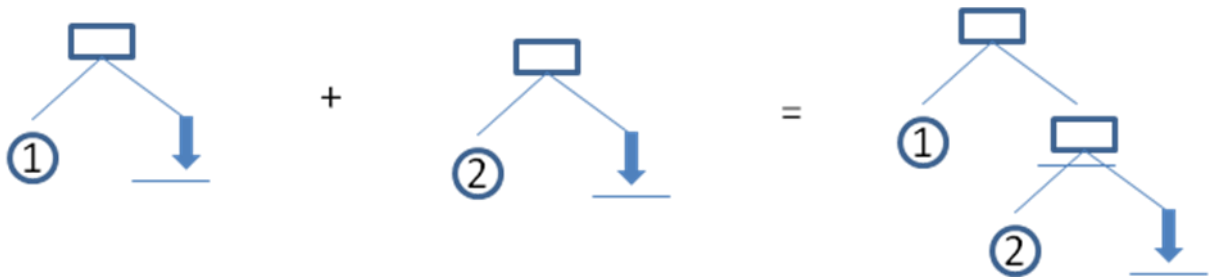


Figure 16: The second unit fulfils the open expectation while raising another one

3. [*and was only vaguely implicated in the working sessions,*]

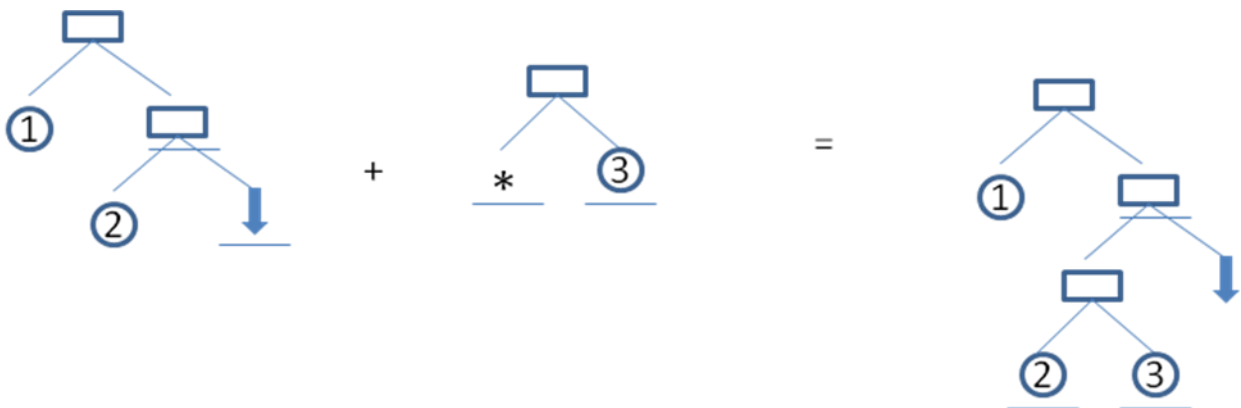


Figure 17: The third unit still leaves open the opened expectation

4. [*he will have a hard time until summer.*]



Figure 18: Finally, the forth unit signals the opened expectation

Using the same idea as parsing at paragraph level, but with different types of auxiliary trees, given by the internal markers, the incremental parsing at sentence level is described in Table 11.



Algorithm 5 – Incremental Parsing at sentence level

```

1: initialize PDTs = {0} - partial developing trees
2: initialize PDTg = {0} - generated trees
3: for each clause in sentence
4:   create listOfAuxiliaryTrees;
5:   for each D-tree in PDTs
6:     for each A-tree in listOfAuxiliaryTrees
7:       signal the sentenceWorkflow to generate next generation of D-trees
8:       add generated trees to PDTg
9:     end for each A-tree
10:  end for each D-tree
11:   evaluate PDTg using heuristics and keep the best ones
12:   PDTs = PDTg
13: end for each sentence
14: evaluate PDTs using heuristics and keep the best one

```

Table 11: A general overview of the incremental parsing at sentence level

3.3.5 Combinatorics at the sentence level

In this section it will be described a method of building the discourse tree structure covering mainly a sentence. This method is described in (Cristea et. al, 2003) and exploits all shortcomings (backtracking) of the incremental parsing but it cannot be applied on larger sentences for reasons that will be discussed next. Let us consider Ex. 24, where the notation indicates the segmentation in *EDUs* (clauses) and the markers are underlined.

Ex. 24

<John is determined to pass the NLP exam so,^[1]>< because he had missed many courses^[2] >
 <and was only vaguely implicated in the working sessions,^[3]>< he will have a hard time until
 summer.^[4]>

In a first phase of building the discourse tree, a module handles the recognition of the markers. Thus they are searched at the beginning and at the end of the *EDUs* ignoring punctuation based on the `markers_type.xml` file which is automatically generated by the *Training Module* of the *Segmenter System*. After the markers are identified, based on the marker patterns, the lists of possible combinations for each marker with each pattern are computed (Table 12). As we can see, a marker can have different patterns depending on the position it



occupies within the clause in the gold corpus. For example the marker “*because*” has two different argument lists for each of its patterns ($_NS$ and N_S). The underlined sequences of *EDUs* are considered as being nuclei and the others – as being satellites.

So		because		And	
<u>[1]</u> – [2,3,4]	N_S	<u>[2]</u> – [3,4]	_NS	<u>[1,2]</u> – <u>[3,4]</u>	N_N
<u>[1]</u> – [2,3]		<u>[2]</u> – [3]		<u>[2]</u> – [3,4]	
<u>[1]</u> – [2]		<u>[2,3]</u> – [4]		<u>[1,2]</u> – [3]	
		<u>[1]</u> – [2,3,4]	N_S	<u>[2]</u> – [3]	
		<u>[1]</u> – [2,3]			
		<u>[1]</u> – [2]			

Table 12: Markers with their patterns and computed argument lists

In Figure 19 is presented the workflow of computing lists of patterns’ arguments depending on the position of the marker within the sentence and its pattern. The pattern notation used in the workflow is used to simplify the figure and to suggest the marker’s relative position. For example, instead of a $_NN$ pattern, a marker can have also the patterns $_NS$ and $_SN$. The same happens in other cases where was used only one general notation N_N (for N_S and S_N) and $NN_$ (for $NS_$ and $SN_$).

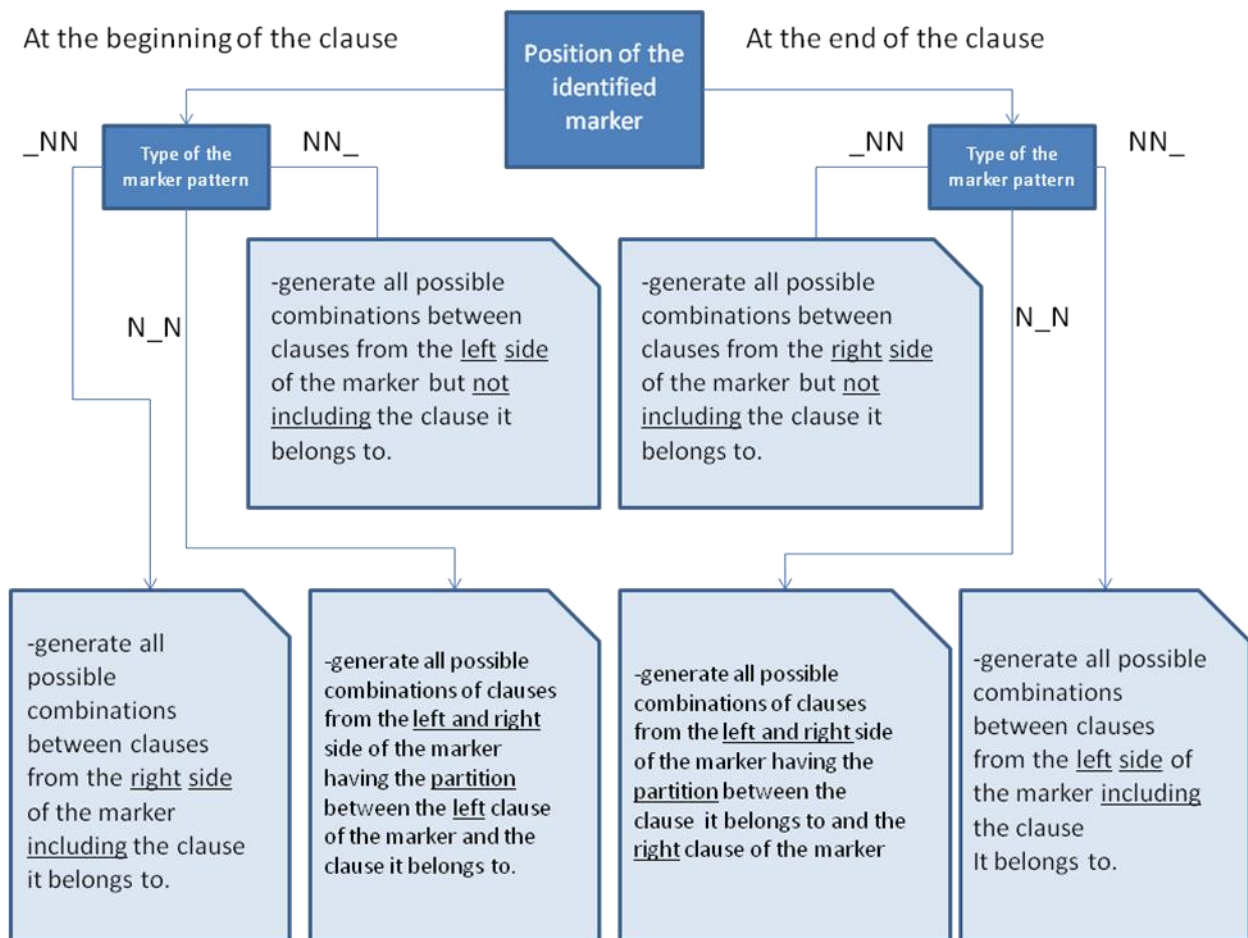


Figure 19: Combinatoric workflow for generating markers argument lists

After the lists are computed, for each combination of the lists these must be checked which of these are consistent (meaning that a tree structure can be generated).

A first idea is to generate all the combinations of tree structures for the sentence and, after that, for each example, to verify if each combination makes up a valid tree structure or not. Such a method is not used in practice because of the huge number of trees possibilities. To solve this problem a different approach, is proposed in (Cristea et. al, 2003) which will generate tree structures for valid combinations of lists. In order to validate a combination, a set of rules are defined:

- It is impossible to have two distinct markers which cover the same sequence of *EDUs*. The combination [1] – [2,3,4] for *so* and [1,2] – [3,4] for *and* do not obey this rule.
- If a sequence of *EDUs* appears in the left (right) side of two different combinations, then the right (left) side of one combination is included in the other's combination



list. This rule states that it is impossible to have nesting arguments. The combination [1] – [2,3] for *so* and [1,2] – [3] for *and* do not obey this rule because even if for the units 1 and 3 the rule is fulfilled (because [1] ⊆ [1,2] and [3] ⊆ [2,3] but for the unit 2, which appears in [2,3] and [1,2], neither [1,2] ⊆ [2,3], nor [2,3] ⊆ [1,2]).

After the combinations are validated with these rules a list of valid combinations is obtained. For the above example, among all possible combinations, only one combination is valid: [1] – [2,3,4] for *so*, [2,3] – [4] for *because* and [2] – [3] for *and*. The generated tree structure of this combination is presented in Figure 20.

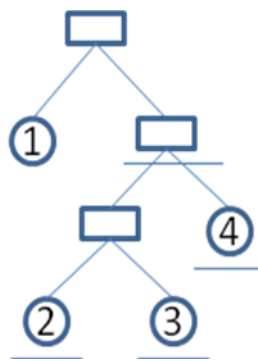


Figure 20: Valid discourse tree structure for the considered example

This is an optimal case because the text contains a sufficient number of markers. But let us consider Ex. 25 were the "and" marker is now missing:

Ex. 25

<John is determined to pass the NLP exam so,^[1]><because he had missed many courses,^[2]>
<was only vaguely implicated in the working sessions,^[3]><he will have a hard time until summer.^[4]>

In this case only two markers are identified within the text. The difference from the previous example is that the "and" marker is now missing (Table 13). One important aspect is that after the combinations are validated with the rules, there will be some cases from which no structure can be generated, because of the missing information.

So		because	
[1] – [2,3,4]	N_S	[2] – [3,4]	_NS



[1] – [2,3]		[2] – [3]	
[1] – [2]		[2,3] – [4]	
		[1] – [2,3,4]	N_S
		[1] – [2,3]	
		[1] – [2]	

Table 13: Markers with their patterns and computed argument lists missing the marker “and”

For these cases we will add two more rules:

- In cases were between any two *EDUs* placed in sequence the text contributes with no marker an empty marker it will be generated with the default argument pattern consisting in all of three combinations of nuclearity (N_N , N_S , S_N). For example, the combination [1] – [2,3,4] for *so* and [2,3] – [4] for *because* is validated against the two rules before, but between [2] and [3] we cannot find a relation, thus we cannot create the tree structure. So we will generate an empty marker gluing these two *EDUs* in order to be able to generate the tree structure. All three possibilities are shown in Figure 21.
- There is only one marker (belonging to the text or automatically generated) that covers the sequence of all *EDUs*; this expresses a natural condition of a well-formed tree with a single root node.

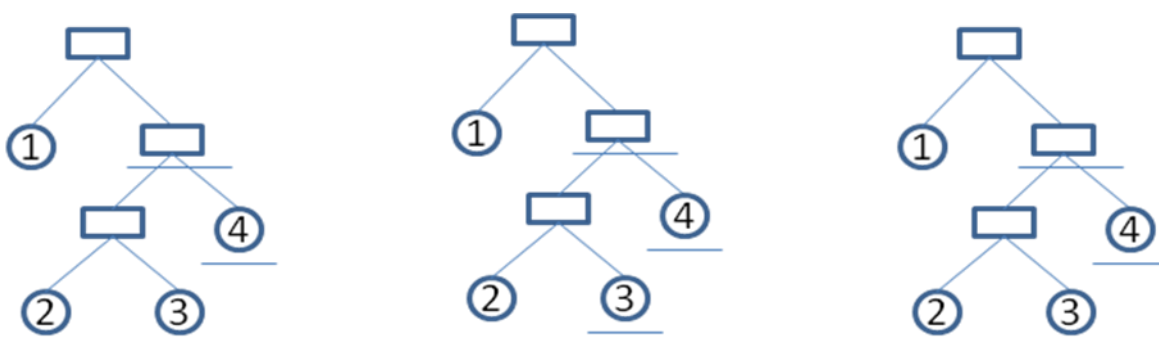


Figure 21: Automatically generated tree structures

Following the workflow for generating argument lists of markers and based on these rules the system generates all tree structures of the sentence. As was said at the beginning, this method cannot be applied on long sentences (more then 8-9 clauses) because it is time-consuming. In the best cases, a sentence containing ten clauses and nine markers, each of them



considering having only one unique pattern, the system will generate minimum $9^9 = 387420489$ combinations (considering each marker to have at least 9 lists), which must be validated against the rules. Although pinning out of inconsistent cases is done at each step, this method still remains extremely time consuming which makes it applicable only on sentences with a small number of clauses (<7). A threshold representing the maximum number of trees that can be generated in worst cases when the text contributes with no marker can be set to avoid running out of memory. In Table 14 we present the total number of trees that can be generated in relation with the number of component clauses.

<i># of clauses</i>	<i>possibilities</i>	<i>projections</i>	<i>total # of trees</i>
1	1	1	1
2	3	1	3
3	9	2	18
4	27	5	135
5	81	14	1.134
6	243	42	10.206
7	729	132	96.228
8	2.187	429	938.223
9	6.561	1.430	9.382.230
10	19.683	4.862	95.698.746
11	59.049	16.796	991.787.004
12	177.147	58.786	10.413.763.542
13	531.441	208.012	1.10546E+11

Table 14: Total number of possible trees relative to the number of clauses

The number of possibilities is a results of the fact that a sequence of two *EDUs* placed in sequence can give rise to 3 small structures: N_N , N_S and S_N . So the formula is:

$$\#possibilities = 3^{n-1} ; \text{ where } n \text{ represents the number of clauses.}$$

Let us take some examples. Let it be a sentence containing two clauses, *a* and *b*. In this case we can generate three different tree structures, like in Figure 22 (remember that we represent nuclear nodes as underlined and satellites – not underlined).

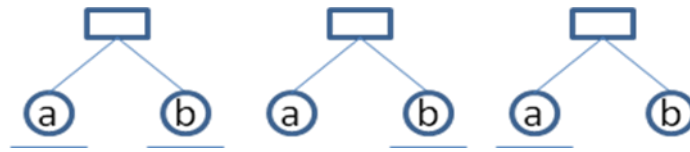


Figure 22: Total number of trees with two clauses

The following example presents a sentence containing three clauses, *a*, *b* and *c*. Only a part of the possible cases are depicted in Figure 23. The nine cases presented result only from one projection type of the tree structure.

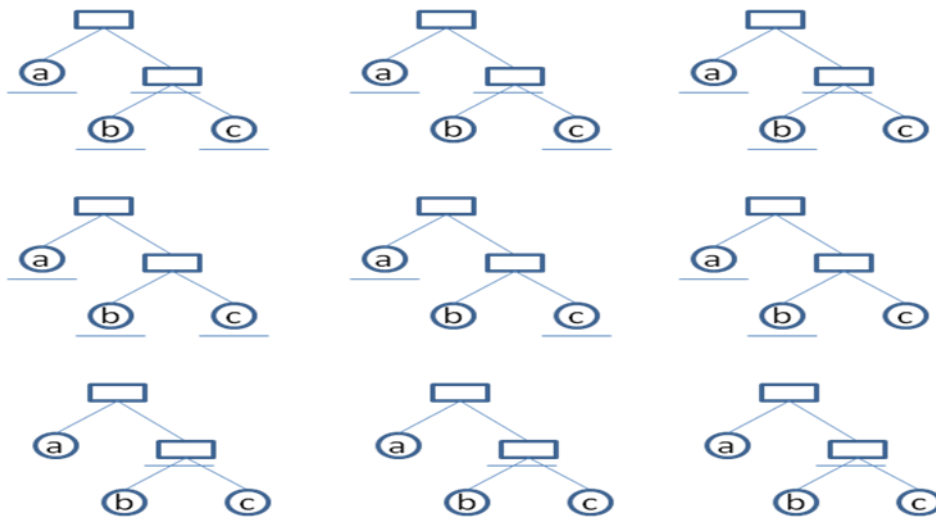


Figure 23: Total number of trees with three clauses and only one (and the same) projection type

The number of projections (Figure 24) is given by the Catalan numbers, counting the number of rooted binary trees with *n* leaves (Davis, 2010).

$$\#projections = \frac{(2*n-2)!}{n!(n-1)!} ; \text{ where } n \text{ represents the number of clauses.}$$

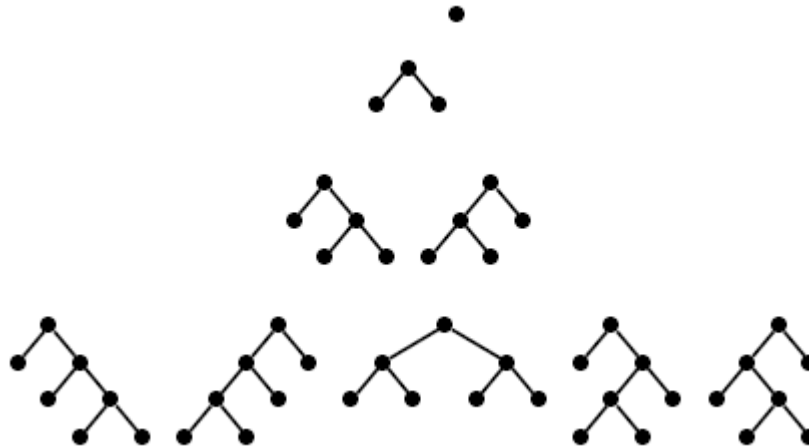


Figure 24: Rooted binary trees with 1, 2, 3 and 4 leaf nodes (Davis, 2010)

Space constraints do not permit an investigation across the whole spectrum of possibilities, such that for those sentences containing a big number of clauses an incremental parsing approach will be used, and for sentences with a smaller number of clauses (<7) we will apply this method (Table 15).

Algorithm 6 – Combinatorics at sentence level

```

1: initialize treshold
2: initialize listOfSentenceTrees = {0}
3: for each clause in sentence
4:   follow the combinatoricWorkflow and compute the lists markers
5:   together with their patterns as in table;
6: end for each clause
7: for each listOfCombinations
8:   verify the list against the rules and add to listOfValidCombinations;
9:   for each validList of combinations from listOfValidCombinations
10:    generate missing relations
11:    generate trees and add to listOfSentenceTrees
12:    if (listOfTrees.size > treshold ) break;
13:   end for each validList
14: end for each listOfCombinations
15: evaluate listOfSentenceTrees using heuristics and keep the best one

```

Table 15: A general overview of the combinatory system at sentence level

The evaluation process of the combinatory system takes into account also the number of cases of each of the markers patterns.



3.3.6 Heuristics: Centering

Centering theory (Grosz et. al, 1995) is a theory of local discourse structure which models the interaction of *cohesion* and *salience* in the internal organization of a text. CT involves speech to be divided in units called *utterances*. In our case an utterance may be considered, as in RST, an elementary discourse unit, mainly a clause. For each discourse unit a set of *forward-looking centers* $C_f(U_n)$ is assigned and for each discourse unit other than the initial one a single *backward-looking center* $C_b(U_n)$ is assigned. The elements of the C_f list are ranked according to discourse salience, and the highest-ranked member is called the *preferred center* $C_p(U_n)$. Pairs of successive discourse units (U_n, U_{n+1}) are classified into four different transition types: *continuation*, *retaining*, *smooth shifting* and *abrupt shifting* presented in Table 16:

	$C_b(U_{n+1}) = C_b(U_n)$ OR $C_b(U_n) = NULL$	$C_b(U_{n+1}) \neq C_b(U_n)$
$C_b(U_{n+1}) = C_p(U_{n+1})$	CONTINUATION	SMOOTH SHIFTING
$C_b(U_{n+1}) \neq C_p(U_{n+1})$	RETAINING	ABRUPT SHIFTING

Table 16: Types of transitions in Centering theory

The uses of different types of transitions are illustrated in Ex. 26, from (Grosz et. al, 1995), and are determined by two factors: whether $C_b(U_{n+1})$ is the same as $C_b(U_n)$ or not and whether $C_b(U_{n+1})$ is the same as $C_p(U_{n+1})$ or not:

Ex. 26

- a) **John** has been having a lot of trouble arranging his vacation.
- b) **He** cannot find anyone to take over his responsibilities.

$$C_f = \{\text{John}, \dots\};$$

$$C_b = \text{John};$$

$$\text{Transition} = \text{CONTINUATION};$$

- c) **He** called up **Mike** yesterday to work out a plan.

$$C_f = \{\text{John}, \text{Mike}, \dots\};$$

$$C_b = \text{John};$$



Transition = CONTINUATION;

d) **Mike** has annoyed **him** a lot recently.

$C_f = \{\text{Mike, John}\};$

$C_b = \text{John};$

Transition = RETAINING;

e) **He** called **John** at 5 am on Friday last week.

$C_f = \{\text{Mike, John}\}$

$C_b = \text{Mike};$

Transition = SMOOTH SHIFTING;

f) **Bill** told **John** **he** would rather look for help somewhere else

$C_f = \{\text{Bill, John}\};$

$C_b = \text{John};$

Transition = ABRUPT SHIFTING;

This heuristic favours adjunction positions that maximise the CT transitions on veins, which correspond to producing structures that reflect overall smoother understanding of the discourse.

3.3.7 Heuristics: Veins Theory

Veins Theory (VT) (Cristea et. al, 1998) extends the classical CT (Crosz et al., 1995) from local to global level of the discourse and makes two important claims: the first regards discourse cohesion, the second – discourse coherence.

VT has the following principles, some similar to RST:

- a discourse structure can be represented by a binary tree;
- a terminal node (leaf) of the tree represents an elementary discourse unit (clause) and is marked by a label;
- non terminal nodes represents a sequence of adjacent elementary discourse units which are labelled with rhetorical relations;
- among the children of a relation a polarity is established: at least one node must be of nuclear type;
- VT does not identify types of relations between tree nodes, unlike the 27 identified in RST.

To introduce the concept of vein, VT uses some notations:



- $mark(x)$ – is a function that takes a list of symbols and returns each symbol in x as being marked (for example between parenthesis);
- $simpl(x)$ – is a function that eliminates all the marked symbols from the string x ;
- $seq(x,y)$ – is a function that takes as arguments two non-intersecting list of symbols x and y , and returns the permutation of x concatenated with y corresponding to a left to right reading of the sequence of symbols in x and y on the terminal frontier of the tree. This function maintains the marked symbols if they exists.

The *head* expression of a node is an ordered list of the most important elementary discourse units (clauses) that the node covers and is computed *bottom-up* as follows:

- The *head* of a terminal node is its label.
- The *head* of a non-terminal node is the concatenation of the heads of its nuclear children.

The *vein* expression (Figure 25) of a node is an ordered list of elementary discourse units that are significant to understand the text covered by the node in the context of the whole discourse and is computed *top-down* as follows:

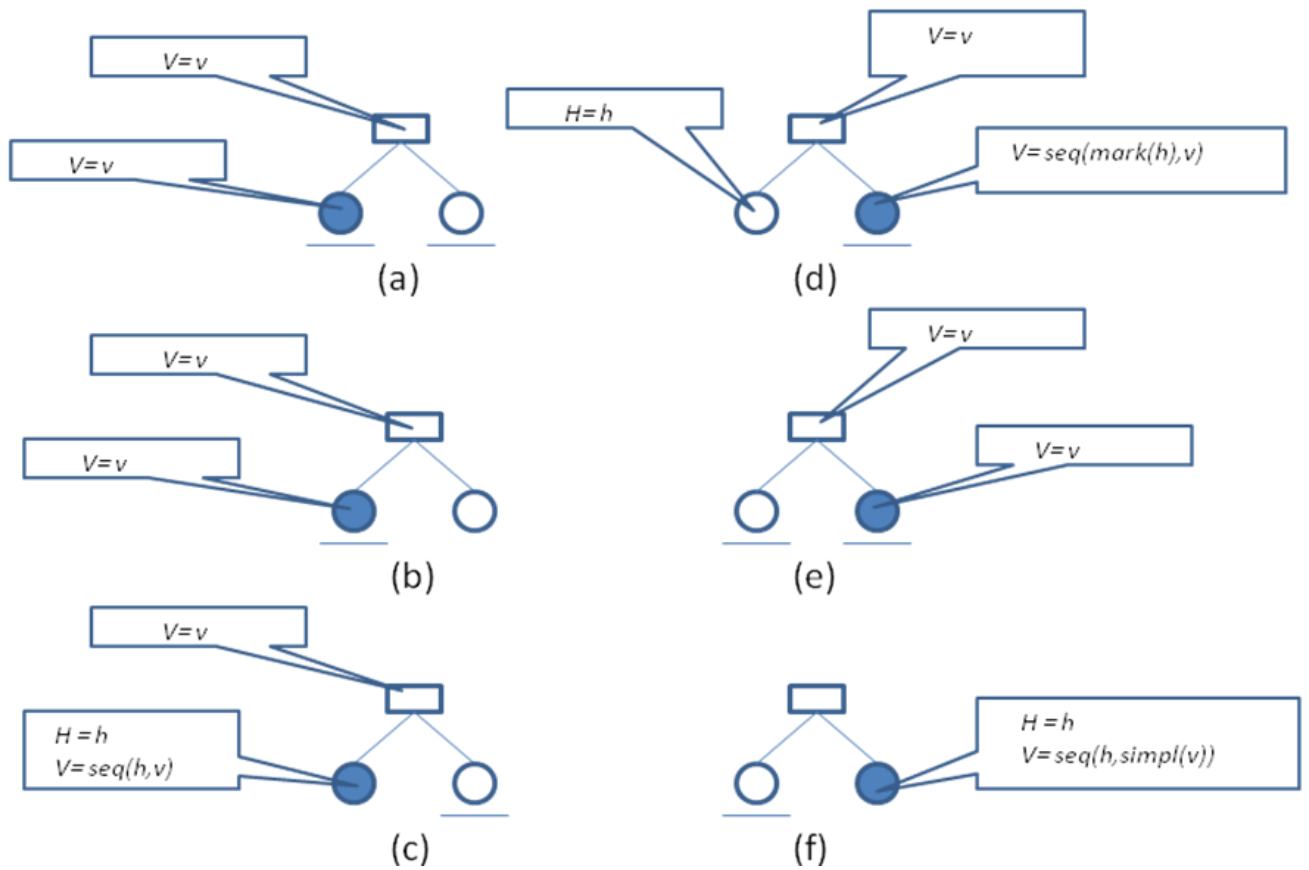


Figure 25: Veins expressions for all cases

- The *vein* expression of the root is its head.
- If the node is the right child of its parent then
 - If the node is of a nuclear type
 - If the node has a left non-nuclear sibling with head h then: $vein=seq(mark(h),v)$, where v is the vein of the parent node (d);
 - Else: $vein=v$, where v is the vein of the parent node (e);
 - Else (the node is of a non-nuclear type): $vein=seq(h, simpl(v))$, where h is the head of the node and v is the vein of the parent node (f);
- Else (the node is the left child of its parent):
 - If the node is of a nuclear type: $vein=v$, where v is the vein of the parent node (a), (b);
 - Else (the node is of a non-nuclear type): $vein=seq(h,v)$, where h is the head of the node and v is the vein of the parent node I;

An exemplification of how the veins are computed is illustrated in Ex. 27 and Figure 26.

**Ex. 27**

- a) Jack and Sue went to buy a new lawnmower
- b) because the old one has been stolen.
- c) Sue had seen the people who took it
- d) and pursued them for a while down the street,
- e) but they disappeared with a truck.
- f) After looking in a shop
- g) they understand they cannot afford a new one.

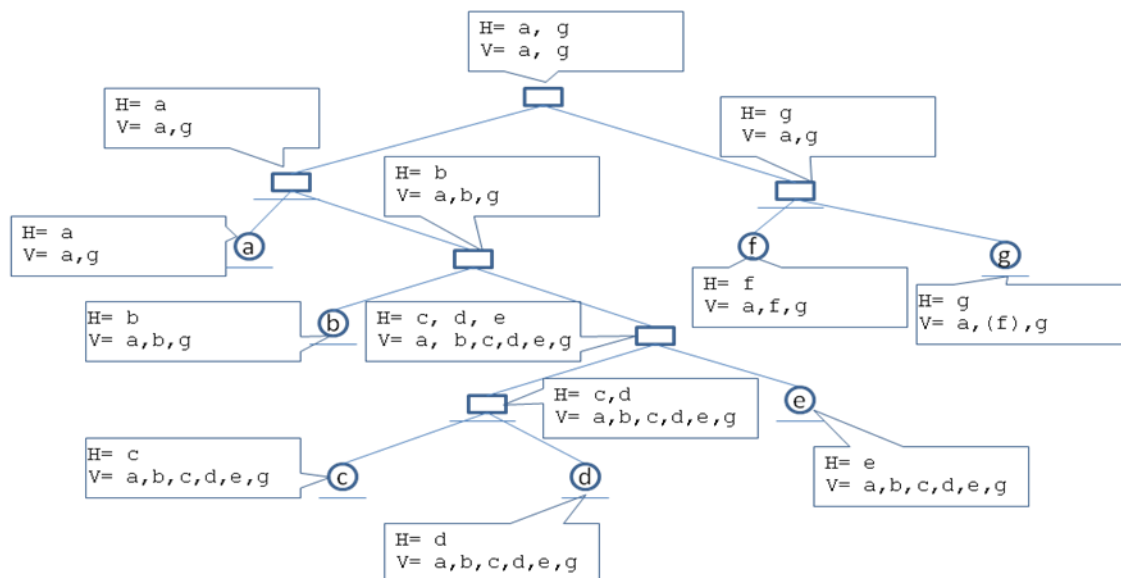


Figure 26: Computed vein expressions for the exemplification

3.3.8 Heuristics: favour lower adjunction levels

This heuristic favours adjunctions operated at the lower part of the RF (or inner-most Right Frontier), because the vein expressions on the lower part of the tree are longer than those on the upper part. The tree that develops when this heuristics is applied persistently will be predominantly balanced to the right and downward, which conforms to a discourse which most of the time develops the last idea, therefore always adding details to the lastly mentioned issue. On the contrary, a tree developing to the right and upward corresponds to a discourse that comes back to the original idea, completing it with new details.



3.3.9 Heuristics: opening the minimum referentiality domain

If the material node m contains a reference that can be satisfied by antecedents belonging to the VT domains D_1, \dots, D_k , give better scores to domains having less referents. This heuristic favours adjunctions on the upper levels (see below).

Supposing a predominantly left balanced tree (in which most of the hypotactic relations have the nuclear daughters on the left side; this is actually very common), an auxiliary tree also left balanced opens for the new material node an as longer domain of referentiality as the adjunction place is lower on the right frontier (for the reason mentioned in Section 3.3.8). In other words, if I go on adding details to the most recent topic I have access to the most part of what has been said until now. On the contrary, if I go on adding details to an old topic, I have access only to the old discourse⁵.

If we reverse the point of view from the accessibility issue, if an entity belonging to the new material node is linked to refer a mention (which will become its antecedent) belonging to the old discourse, the material node can be attached anywhere on the RF, but if the reference link is directed towards a new mention (antecedent), the adjunction cannot be made but on the lower part of the RF.

So, by favouring tree structures having minimal domains of referentiality, therefore adjunctions to the upper levels of the RF, this heuristic would counter-balance the tendency incurred by the previous heuristic.

3.3.10 Heuristics: maximum referentiality score

This heuristic favours adjunction positions where most referents of the material node find antecedents on the referentiality domains given by veins. In relation with the occurrence of the antecedent on the vein, the type of the referential expression counts also, because not all referential expressions are equal in evocation power. We have starting from the original experimental findings of (Ide and Cristea, 2000) where:

- zero pronoun: it is compulsory to have an antecedent on the vein,
- clitic: antecedent extremely wanted on the vein,
- overt pronoun: antecedent very much wanted on vein,

⁵ This is also conformant with the stack referentiality of the Attentional State Theory (Grosz and Sidner, 1986).



- common noun: good to have an antecedent on the vein,
- proper noun: it is not necessary to have an antecedent on the vein,

First, we have defined three cases of placing an antecedent:

A. the antecedent belongs to the previous unit (clause), where previous means in the sequence of the vein of the current unit,

B. the antecedent belongs to one unit of the vein of the current unit which is not the previous one,

C. the antecedent does not belong to any unit of the vein of the current unit,

A set of parameters (included in the `preferencesDiscourseParser.pref` file of the DP module) assign scores to anaphors, depending of the POS of their lexical heads and the positioning of their antecedents. A recommended set of values is given in Section 3.3.14.

3.3.11 Heuristic: if a substitution node is open, consume it!

Whenever possible solve first substitutions (by using gamma or delta A-trees) and propose no-name alpha or beta A-trees after (see Figure 13 in Section 3.3.3).

3.3.12 Heuristic: limit the length of the summary

If this heuristic (referred below as the *Limit Heuristic*) is selected, a summary roughly guided by a defined length is obtained. The correction is done at each step.

3.3.13 Heuristics: concluding remarks

At one step in the development of the discourse tree t , a forest of trees is obtained. For each tree, a global score is computed by summing up weighted scores of each individual heuristic, using the formula:

$$GS^t = \sum_i^N s_i * w_i$$

where s_i and w_i are the score of one heuristic and its corresponding weight, and N is the total number of heuristics applied.

Each heuristic triggers an evaluation procedure that produces a score, which is a number in the range 0 to 1.



3.3.14 Parameters for the DP module

The parameters of the DP module should be indicated in a preferences file, named `preferencesDiscourseParser.pref`. An example of such a file follows:

```
INPUT = "input.xml" – path towards to input file;
OUTPUTXML = "output.xml" – path towards to output file;
OUTPUTLISP = "" – path towards to output file in LISP format;
OUTPUTBIN = "" – path towards to output file that can be used as input by the DrawTrees
auxiliary module (see Section 4.2);
LOGFILE = "" – if needed, the name of a log file;
TAGSETFILE = "model/tagset.xml" – a file containing a list of tags for different part
of speeches (out of which the DP modules needs those for punctuation, pronouns, nouns and
proper nouns);
MARKERSTYPE = "model/markers_type.xml" – the place of the file containing
markers, built by the CS Training Module (see Section 3.2.3) and possibly updated manually
using the Markers Editor (see Section 4.1);
ADJREL = "N_N N_S S_N" – the configurations of nuclearity pairs for adjunction when
no marker is found;
CENTERING = "20.0" – weight of the centering heuristic;
V_CENTERING = "20.0" – weight for centering on the principal vein (general summary)
heuristic;
REFERENCES = "20.0" – weight for references heuristic;
// The Maximum Referentiality Score heuristic favours adjunction positions where most
referents find antecedents on the referentiality domains given by veins. The type of the
referential expression contributes in the computation of the score.
V_REFERENCES = "20.0" – weight for references on the principal vein (general
summary) heuristic;
// Applied by the Maximum Referentiality Score heuristic.
ADJUNCTION_SENT = "-10.0" – weight of the adjunction heuristic at the sentence level;
// In this case, the value -10 disfavors lower adjunction levels giving a low score (negative)
to adjunctions at the lower part of the RF (or inner-most RF).
```



ADJUNCTION_PAR = "-10.0" – weight for the adjunction heuristic at the paragraph level;
SUBSTITUTION = "20.0" – weight of the substitution heuristic: substitute first if no other needs;

// Whenever possible solve first substitutions and propose no-name alpha or beta A-trees after.

COMBINATORICS = "true" – if "true", the combinatorics algorithm is applied (see Section 3.3.5). But the parser switches this trigger automatically to "false" if the number of clauses is over 7 (to prevent the exponential explosion).

LIMIT = "10.0" – weight for the *Limit Heuristic*.

MIN_LIMIT = "0.2" – the minimal value for the interval of the summary length when applying the *Limit Heuristic* (see Section 3.3.12).

MAX_LIMIT = "0.4" – the maximal value for the interval of the summary length when applying the *Limit Heuristic* (see Section 3.3.12).

MAX_DISCOURSE_TREES = "1" – the number of trees retained at each step to continue the beam search with, in the incremental approach (at the discourse level);

MAX_PARAGRAPH_TREES = "1" – the number of trees retained at each step to continue the beam search with, in the incremental approach (at the paragraph level);

MAX_SENTENCE_TREES_INCR = "5" – the number of trees retained at each step to continue the beam search with, in the incremental approach (at the sentence level);

MAX_SENTENCE_TREES_COMB = "2000" – the maximum number of generated trees for one sentence when applying the combinatorics algorithm;

TAG = "POS" – the name of the tag containing the part of speech information (usually could be POS and MSD – see also Sections 3.2.3 and 3.2.4);

PROPERNOUN_DIR = "1.0" – if the anaphor is a PROPER NOUN and the antecedent is in case A (see Section 3.3.10);

PROPERNOUN_INDIR = "0.75" – if the anaphor is a PROPER NOUN and the antecedent is in case B (see Section 3.3.10);

PROPERNOUN_PRAGM = "0.5" – if the anaphor is a PROPER NOUN and the antecedent is in case C (see Section 3.3.10);

NOUN_DIR = "1.0" – if the anaphor is a NOUN and the antecedent is in case A;

NOUN_INDIR = "0.75" – if the anaphor is a NOUN and the antecedent is in case B;

NOUN_PRAGM = "0.5" – if the anaphor is a NOUN and the antecedent is in case C;



PRONOUN_DIR = "1.0" – if the anaphor is a PRONOUN and the antecedent is in case A;
PRONOUN_INDIR = "0.5" – if the anaphor is a PRONOUN and the antecedent is in case B;
PRONOUN_PRAGM = "-0.75" – if the anaphor is a PRONOUN and the antecedent is in case C;
LANGUAGE = "BG/DE/EL/EN/PL/RO" – the language of the input text⁶.

The DP module technical record

- Programming language: Java, platform independent.
- OS compatibility: Windows and Linux.
- Access conditions: free for non-commercial use.
- Copyright: UAIC for the core module; Each partner for their localised versions.
- Input:
 - for the stand-alone package (at the exit of the CS module): XML text containing TOK (with attributes: ID, POS, LEMMA), S (with the attribute ID), P (with the attribute ID), NP (with ID, HEADID and REF: ID of a DE element), DE (with attributes: ID, ReList), CLAUSE (with ID). Coding is UTF-8;
 - for the UIMA package: jCas containing types: RichToken (with attributes BEGIN, END, POS, BASE), Sentence (with attributes BEGIN, END), NpChunkWithHead (with attributes BEGIN, END, HeadToken, Parent, Children), CoreferenceChain (with attribute DE, containing itself a list of types NpChunkWithHead), Clause (with BEGIN, END).
- Preferences (one file, either created by the DP Calibration system, as described in Section 3.4.3, or filled in manually, see above):
 - preferencesDiscourseParser.pref.
- Resources:
 - tagset.xml (as defined by the TAGSETFILE parameter of the preferences file, see above);
 - markers_type.xml (as defined by the MARKERSTYPE parameter of the

⁶ This parameter is used by the DP module to overcome a particular feature of the Greek POS-tagger (hard-coded temporary solution). It is not compulsory for the other languages.



preferences file, see above)

– Output (only contributed by the DP module):

- an XML file, in UTF-8, if the `OUTPUTXML` parameter of the `preferencesDiscourseParser.pref` file is set to a non-null path, containing the new elements:
 - `NODE` (describing a node of the discourse structure), with the attributes: `ID`, `NUC`, `LEFT`, `RIGHT`, `TYPE`, `VEIN`, `HEAD`, `MARK`, where: `NUC` is the nuclearity pattern of the rhetorical relation of the node (`N_N`, `S_N` or `N_S`) – only for non-leaf nodes; `LEFT` and `RIGHT`, which are Ids of the left and right daughters of the current node – also only for non-leaf nodes; `TYPE`, which could have one of the values: “`PARAGRAPH`” – if the node is covering a paragraph, “`SENTENCE`” – if the node is covering a sentence, “`RELATION`” – for any non-terminal node under one of the types above, and “`LEAF`” – if the node is a leaf node; `VEIN`, which gives the vein expression of the node (as a sequence of clauses Ids); `HEAD`, which gives the head expression of the node (as a sequence of clauses Ids); `MARK`, which gives the marked clauses (see the Veins Theory – Section 3.3.7);
 - `ROOT` (unique, indicating the root node of the discourse structure): its format is identical to a `NODE` element.
- a LISP file, in UTF-8, if the `OUTPUTLISP` parameter of the `preferencesDiscourseParser.pref` file is set to a non-null path. The LISP format of a node describes recursively, in two similar sub-lists, the left and the right daughters of that node;
- a BIN file if the `OUTPUTBIN` parameter of the `preferencesDiscourseParser.pref` file is set to a non-null path. This format is compatible with the input of the `DrawTrees` module (Section 4.2);
- UIMA data type for the annotation of a node, namely `uaic.uimatypes.DiscourseNode` (with the attributes: `begin`, `end`, `nuc`, `leftChild` – giving, only for internal nodes, the `DiscourseNode` of the left child node of the current node, `rightChild` – giving, only for internal nodes, the `DiscourseNode` of the right child node of the current node, `clause`



- giving, only for terminal nodes, the `Clause` – see the Segmenter module technical record in Section 3.2.4, `veinClauses` – a list of `Clauses` forming the vein expression of the current node, `headClauses` – a list of `Clauses` forming the head expression of the current node, `markClauses` – the list of marked `Clauses` in the vein expression of the current node).
- Status: stable. Localised versions could be improved if more gold data is provided. Performances could still be improved by iterating extensively the whole ranges of the parameters in the Discourse Parser Calibration system (see Section 3.4.4).

3.4 The Summariser for short texts

In this document we will call *short text* a text spanning between $\frac{1}{2}$ a page up to 6 pages. Our summaries belong to the type usually known as *excerpt type summaries*⁷, which are summaries that copy contiguous sequences of tokens from the original text. Actually, in our case, such a summary should contain elementary discourse units (see Section 3.3.1) copy-pasted from the original text.

3.4.1 Types of summaries

The structure of a discourse as a complete tree gives more information than properly needed (at least for summarization purposes). Out of the discourse structure, three types of summaries can be extracted:

- general summary – The general summary of the discourse is given by the vein expression of the root node. All the clauses that appear in the vein expression of the root node represent the general summary.
- entity focused summary – For obtaining an entity focused summary we will look in the text for all the apparitions of the desired entity and all the references to it. All the clauses that appear in the reunion of the vein expressions of the elementary units in which the entity (or references) appears, give us the entity focused summary;
- clause focused summary – One of the conjectures of VT is that the vein expression of an elementary discourse unit provides a coherent “abstract” of the discourse

⁷ Contrary to an excerpt type summary is a *rephrase type summary*, which contains a reduced, freely produced, verbalisation of the original text.



fragment that contains that unit (Cristea et. al, 1999). To accomplish this kind of summarization we will analyze the vein expression of the clause we are interested in. The reunion of all the clauses that appears in the vein expression of clause represents the focused summary of the clause.

3.4.2 The Extract Summary module

The *Extract Summary* module takes the tree of a discourse structure on which vein expressions are noted and produces a general or a focussed summary of a certain length.

The general summaries, being vein expressions, inherit the coherence properties of veins (see Section 3.3.7). The same happens with summaries focussed on clauses. The vein of an *EDU* (clause), by definition, gives a summary of the whole discourse focused on that particular unit.

However, producing summaries focussed on entities is not that straightforward. Suppose one discourse entity is traced and a summary focused on that entity is wanted. If there is only one *EDU* in which the entity is mentioned, the vein expression of that *EDU* gives a very well-focused summary of the entity. A problem appears if the entity is mentioned in more than just one *EDU*. Because there is no a-priori reason to prefer one of the focused summaries obtained in this way to any of the others, it is clear that a combination of the vein expressions of each *EDU* in which the entity is mentioned should be considered. We proposed more methods of building a final summary from the collection of particular summaries. The first method takes the vein expression of the lowest node of the tree that covers all units in which the entity is mentioned. Since the length of a vein expression is proportional to the deepness of the node in the tree structure, this method results in shorter summaries. The second method considers that particular summary (vein expression) which includes most of the mentions of the entity. The third method simply takes the union of all vein expressions of the units that mention the central entity. Finally, the fourth method builds a histogram from all vein expressions of the units mentioning the central entity and selects all units above a certain threshold. The last two methods are not in themselves vein expressions, and therefore are more prone to incoherent summaries than the first two methods, the last one being the most exposed (Cristea et al., 2005).

The preferences file of the *Extract Summary* module is named `preferencesExtractSummary.pref` and contains the following parameters:



SUMMARY_NP = "ID" – if different than an empty string, the ID of an NP (an entity), on which a focussed summary will be generated;

SUMMARY_CLAUSE = "ID" – if different than an empty string, the ID of a clause, on which a focussed summary will be generated;

REDUCTION_RATE = "percent" – where percent applies to the original length of the document to give the length of the summary (not yet implemented);

INPUT = "string" – the path to the XML file, containing the discourse structure of the input text;

OUTPUT_CLAUSES = "file_name" – path and name of the output file which will contain the list of clause IDs representing the summary;

OUTPUT_WORDS = "file_name" – path and name of the output file which will contain the list of tokens IDs representing the summary;

TEXT_SUMMARY = "text_file" – path and name of an output file which will contain the summary in plain text.

TAG = "POS/MSD" – specifies the name of the TAG attribute identifying the pronouns (for replacing purpose).

The Extract Summary module technical record

- Programming language: Java, platform independent.
- OS compatibility: Windows and Linux.
- Access conditions: free for non-commercial.
- Copyright: UAIC.
- Input:
 - for the stand-alone package (at the exit of the DP module): an XML file, in UTF-8, containing minimally the elements: TOK (with attributes: ID, POS, LEMMA), NP (with ID, HEADID and REF: ID of a DE element), DE (with attributes: ID, ReList), CLAUSE (with ID), NODE (describing a node of the discourse structure, with the attributes: ID, VEIN, where VEIN gives the vein expression of the node as a sequence of clauses Ids), and ROOT (a unique element, indicating the root node of the discourse structure) whose format is identical to that of a



NODE element;

- for the UIMA package (only contributed by the DP module): the UIMA data type for the annotation of a node as outputted by the DP module, namely `uaic.uimatypes.DiscourseNode` (with the attributes: `begin`, `end`, `nuc`, `leftChild` – giving, only for internal nodes, the `DiscourseNode` of the left child node of the current node, `rightChild` – giving, only for internal nodes, the `DiscourseNode` of the right child node of the current node, `clause` – giving, only for terminal nodes, the `Clause` – see the Segmenter module technical record in Section 3.2.4, `veinClauses` – a list of `Clauses` forming the vein expression of the current node, `headClauses` – a list of `Clauses` forming the head expression of the current node, `markClauses` – the list of marked `Clauses` in the vein expression of the current node).
- Preferences (one file, see above):
 - `preferencesExtractSummary.pref`.
- Output:
 - Only if the `OUTPUT_CLAUSES` parameter is non-empty: the list of Ids of clauses, in UTF-8, belonging to the summary;
 - Only if the `OUTPUT_WORDS` parameter is non-empty: the list of Ids of tokens, in UTF-8, belonging to the summary;
 - Only if the `TEXT_SUMMARY` parameter is non-empty: the UTF-8 string of the summary;
 - UIMA data type for the `SUMMARY` annotation, namely `uaic.uimatypes.Summary` (with attributes `clausesSummary`, containing itself a list of `Clauses`, and `wordSummary`, containing itself a list of `RichTokens`).
- Status: beta version ready, heuristics for deciding the length under elaboration, in search for a parameterised definition of new heuristics.

3.4.3 The Evaluate Summary system

It is notoriously true that a gold corpus to be used in the evaluation of a discourse parser is very expensive, due at least to three factors. First, the determination of the discourse structure of a text involves choices which have not always only one solution, being therefore subjective.



Even very well trained human annotators could arrive at totally different structures for the same text. Secondly, the annotation process is extremely time consuming. And, thirdly, the cost of such a process is very high because of the complexity of the task and the high skills needed. Considering all these, the ATLAS consortium has decided to use an indirect method for evaluating the DP module, which skips a direct confrontation of a discourse tree parse against a discourse gold file and concentrates the evaluation of a discourse structure and the evaluation of a summary. The idea is that a good summary cannot be but the result of a good discourse structure. Inversely, a poor summary reflects defects in the discourse structure.

Therefore, we decided to compare the automatically produced summaries against gold summaries produced by human subjects. The Evaluator module computes Precision, Recall and F-measure by comparing tokens (words) in the test against the gold summarization files.

But since the summarization process is also extremely subjective, in building summarization gold corpora it is compulsory to have more than one annotator for the same text. It remains to decide how to evaluate an artificial summary if there are more than one human-produced summaries that count as being gold. Two measures can be proposed in this case:

- the **Histogram** measure: this is proposed in (Cristea et al., 2005). Partners have given each text, in which *EDUs* were already marked and numbered, to a number of human subjects and asked them to indicate 3 general summaries by writing down sequences of discourse unit numbers: displaying reduction rates of approximately 20%, 10% and 5% (counting as word reductions rates – see Section 5.2). Then, one different histogram is drawn for each reduction rate by counting the number of time each *EDU* of the original text is mentioned by the subjects as belonging to their summaries. In these histograms the sequence of *EDU* numbers is placed on the x-axis and the frequency of mentioning on the y-axis. A sliding horizontal line (threshold) is fixed in this histogram at a position such that the number of units above the line approximates the considered reduction rate. The respective golden summary is given by all units whose corresponding frequencies were above the threshold (see Figure 27).

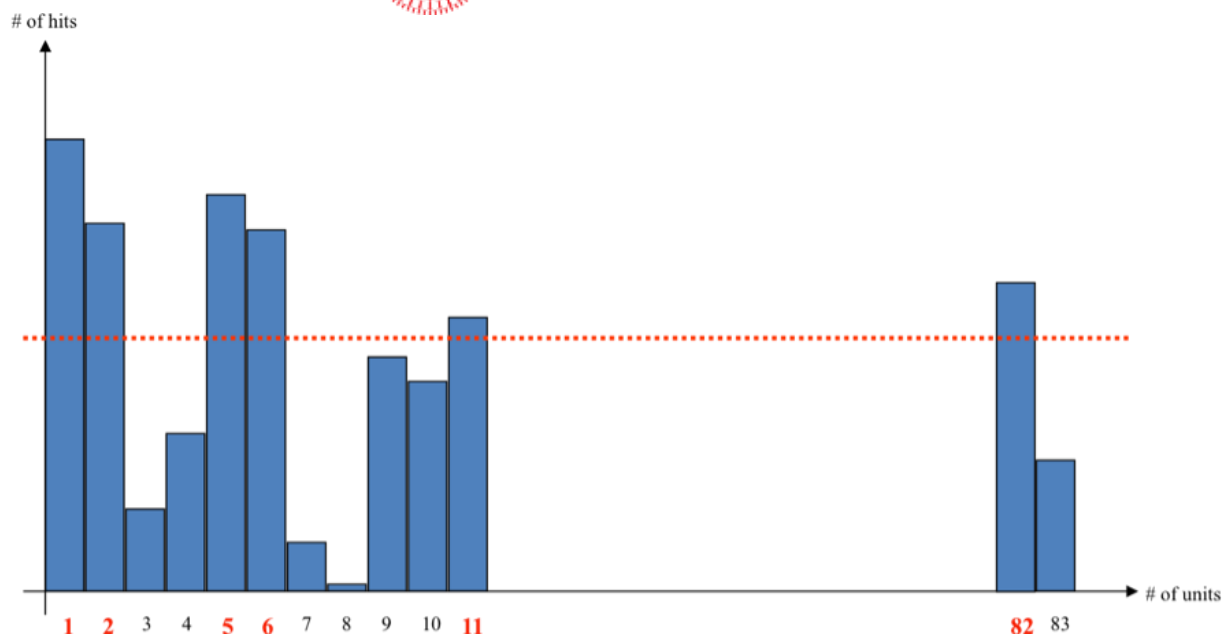


Figure 27: Approximating a gold summary out of a number of human - produced summaries

- the **Averaging** measure: compute P, R, F between the unique test file and the gold file of each subject and average the results.

As will be seen in the next section, the evaluation is involved in the calibration process.

The preference file, named `preferencesEvaluateSummary.pref` includes the following parameters:

`MEASURE` = "HISTOGRAM/AVERAGE" – the type of the measure, as above;

`GOLDFILE` = "path" – the path towards a folder containing the summary gold files, each of them containing a list of IDs of clauses or tokens;

`TESTFILE` = "path" – the path and name of the summary test file (as a list of IDs of clauses – if the `GOLDFILE` contains clauses, or tokens – if the `GOLDFILE` contains tokens);

`EVALFILE` = "path" – the path and name of a file where the results will be deposited.



The Evaluate Summary module technical record

- Programming language: Java, platform independent.
- OS compatibility: Windows and Linux.
- Access conditions: free for non-commercial.
- Copyright: UAIC.
- Input:
 - for the stand-alone package (at the exit of the DP module): an XML file, in UTF-8, containing minimally the elements: TOK (with attributes: ID, POS, LEMMA), NP (with ID, HEADID and REF: ID of a DE element), DE (with attributes: ID, ReList), CLAUSE (with ID), NODE (describing a node of the discourse structure, with the attributes: ID, VEIN, where VEIN gives the vein expression of the node as a sequence of clauses Ids), and ROOT (a unique element, indicating the root node of the discourse structure) whose format is identical to that of a NODE element.
- Preferences (one file, see above):
 - `preferencesEvaluateSummary.pref`.
- Output:
 - a text file in UTF-8 containing the results: precision, recall and F-measure.
- Status: beta version ready.

3.4.4 The DP calibration system

We used the same idea as in the case of the *Segmenter Calibration system* and built a *Discourse Parser Calibration System* (Figure 28). The difference is that the evaluation involved in the calibration process involves the summary, as explained in Section 3.4.3.

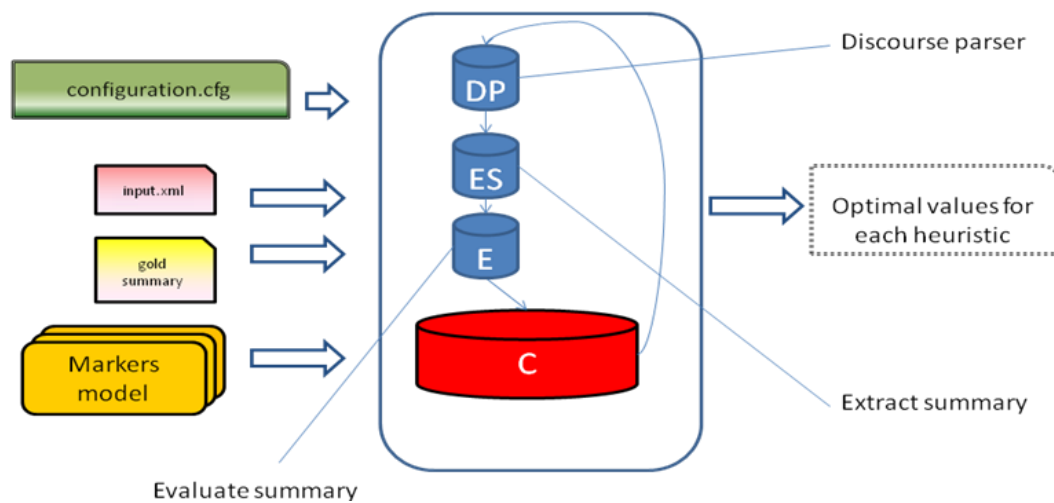


Figure 28: Discourse Parser Calibration System

The preferences of the DP Calibration system are defined in the file `configuration.cfg`. An example of this file follows. Most of the parameters have been described in Section 3.3.14. The rest are self-understood.

```

minimCentering = "20.0"
maximCentering = "30.0"
minimVCentering = "20.0"
maximVCentering = "30.0"
minimReferences = "20.0"
maximReferences = "30.0"
minimVReferences = "20.0"
maximVReferences = "30.0"
minimAdjunctionInt = "-10.0"
maximAdjunctionInt = "10.0"
minimAdjunctionExt = "-10.0"
maximAdjunctionExt = "10.0"
minimSubstitution = "20.0"
maximSubstitution = "30.0"
minimLimit = "10.0"
maximLimit = "20.0"
minimDiscourseTrees = "1"
maximDiscourseTrees = "3"
minimParagraphTrees = "1"
maximParagraphTrees = "3"
minimSentenceTreesIncr = "5"
maximSentenceTreesIncr = "10"
minimSentenceTreesComb = "2000"
maximSentenceTreesComb = "3000"
combinatorics = "true"
tag = "POS"
adjRel = "N_N N_S S_N"

```



```
testFile = "testFile.txt"  
goldFolder = "Folder"  
evalFile = "evalFile.txt"  
measure = "Fmeasure"  
outputFile = "outBG.txt"
```

The DP Calibration module technical record

- Programming language: Java, platform independent.
- OS compatibility: Windows.
- Access conditions: free for non-commercial use.
- Copyright: UAIC.
- Input: (the DP Calibration module runs only on XML formatted texts):
 - A set of GOLD summary files, defined in special folder, each of them containing a list of IDs of clauses or tokens;
 - The INPUT file: an XML text containing: TOK (with attributes: ID, POS, LEMMA), S (with the attribute ID), P (with the attribute ID), NP (with attributes ID, HEADID, REF: ID of a DE element), DE (with attributes: ID, ReList), CLAUSE (with ID). Coding is UTF-8;
- Preferences (1 text file, in UTF-8, as above):
 - configuration.cfg
- Resources (2 XML files, see in Section 3.3.14):
 - markers_type.xml;
 - tagset.xml.
- Output (as defined in Section 3.2.3 and Section 3.2.4):
 - preferencesDiscourseParser.pref
- Status: stable. Can be improved by iterating extensively the whole ranges of parameters.



3.5 The Summarization tools integrated in u-Compare

3.5.1 u-Compare

U-Compare is an integrated text mining/natural language processing system based on the UIMA Framework which provides access to a large collection of ready-to-use, interoperable, natural language processing components. U-Compare allows users to build complex NLP workflows from these components via an easy drag-and-drop interface, and makes visualization and comparison of the outputs of these workflows simple.

As the name implies, comparison of components and workflows is a central feature of the system. U-Compare allows sets of components to be run in parallel on the same inputs and then automatically generates statistics for all possible combinations of these components. Once a workflow has been created in U-Compare it can be exported and shared with other users or used with other UIMA compatible tools and so in addition to comparison, U-Compare also functions as a general purpose workflow creation tool.

U-Compare is a joint project between the University of Tokyo, the Center for Computational Pharmacology (CCP) at the University of Colorado Health Science Center, and the National Centre for Text Mining (NaCTeM) at the University of Manchester.

3.5.2 Integrating the summarization chain in u-Compare

The integration was done for the Romanian language and it contains the following engines:

- POS Tagger
- NP Chunker
- RARE
- Clause Splitter
- Discourse Parser

POS Tagger was integrated using the web service located at nlptools.info.uaic.ro/. It receives as input raw text and returns in output annotations for: Sentence, RichToken with RichToken:base, RichToken:posString, RichToken:end, RichToken:begin.

NP Chunker was integrated using the web service located at nlptools.info.uaic.ro/. It's



input is

The JCas object containing the output of the POS Tagger and the output annotations are:

```
NpChunkWithHead          with:          NpChunkWithHead:headToken,  
NpChunkWithHead:children,          NpChunkWithHead:parent,  
NpChunkWithHead:begin,NpChunkWithHead:end.
```

RARE was integrated using the source code. The description of their JCas objects can be found in the RARE Technical Record in Section 3.1.2. and it includes a description of the input and output of the POS Tagger and of the NP Chunker. The output annotation is `CoreferenceChain`. `CoreferenceChain` is a UIMA data type annotation namely `uaic.uimatypes.CoreferenceChain` with the attribute `DE`, containing itself a list of types `NpChunkWithHead`.

The Clause Splitter was integrated using the source code. The description of their JCas objects can be found in the CS module technical record in Section 3.2.4. It's input is the JCas object containing the output of the POS Tagger, of the NP Chunker and of RARE, and the output annotation is a `Clause`. `Clause` is a UIMA data type annotation, namely `uaic.uimatypes.Clause` with attributes `BEGIN` and `END`.

The Discourse Parser was integrated using the source code. The description of their JCas objects can be found in the DP module technical record in Section 3.3.14. It's input is the JCas object containing the output of the POS Tagger, of the NP Chunker, of RARE, and of the Clause Splitter, and the output annotation is `DiscourseNode`.

`DiscourseNode` is a UIMA data type annotation namely `uaic.uimatypes.DiscourseNode` (with the attributes: `begin`, `end`, `nuc`, `leftChild` – giving, only for internal nodes, the `DiscourseNode` of the left child node of the current node, `rightChild` – giving, only for internal nodes, the `DiscourseNode` of the right child node of the current node, `clause` – giving, only for terminal nodes, the `Clause`, `veinClauses` – a list of `Clauses` forming the vein expression of the current node, `headClauses` – a list of `Clauses` forming the head expression of the current node, `markClauses` – the list of marked `Clauses` in the vein expression of the current node.

RARE, Clause Splitter and Discourse Parser are stand alone tools, thus their integration into UIMA (actually, u-Compare) implied some changes. Because UIMA works with CAS type objects, and the tools were using XML for input/output a compromise was achieved. The input and output for each tool is a CAS, and a wrapper was developed, that reads the CAS, and



creates the XML file necessary for the tools.

4 AUXILIARY SOFTWARE

4.1 Graphical interface for viewing and editing discourse markers

Besides the actual implementation of the system we have created an application for viewing and editing discourse markers (Figure 29). This application is useful to review the correctness of markers annotated within the corpus.

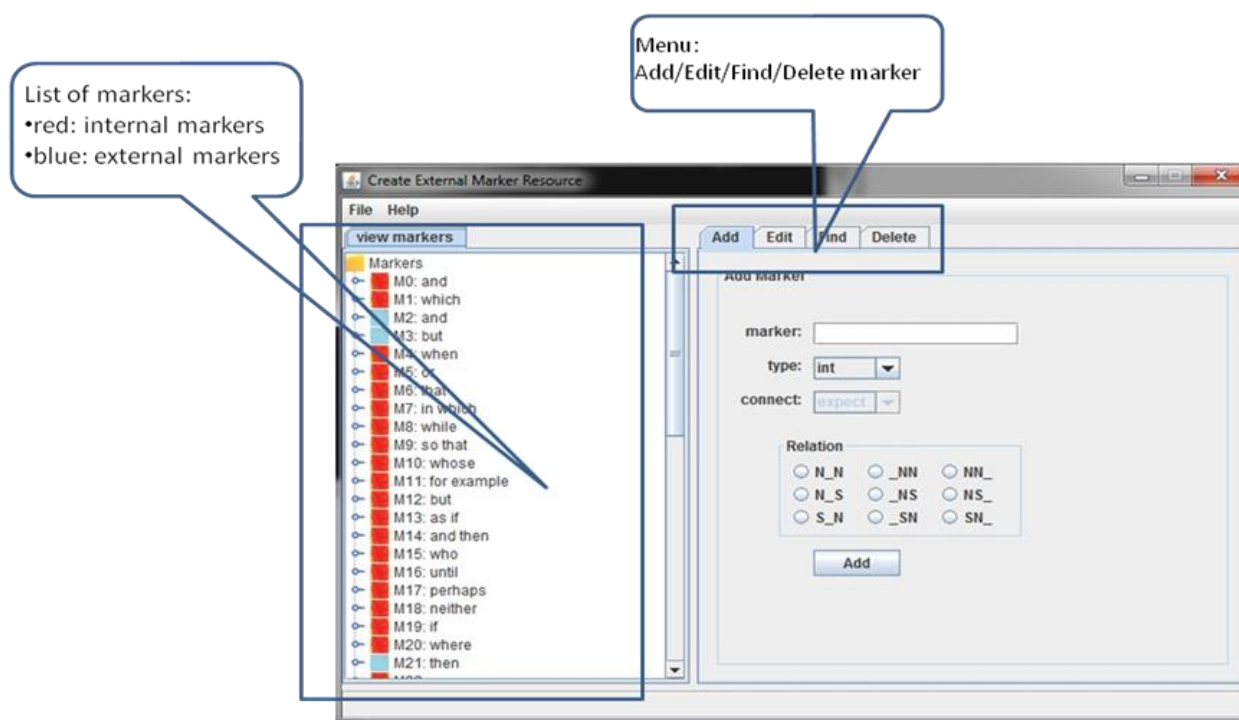


Figure 29: Markers' Editor

File menu (Figure 30) :

File/New – create a new empty file;

File/Open – open an XML file (markers_type.xml);

File/Save – saves the current file in XML format.

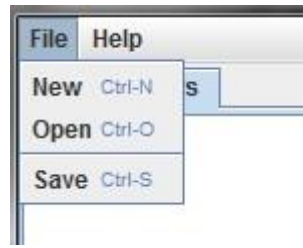


Figure 30: File menu of the Markers Editor

Add Menu (Figure 31) – Offers the signalling to add new markers that weren't annotated within the corpus.



Figure 31: Add menu

In order to add a new marker we must specify the text of the marker and if the marker is composed of more than one word, the words must be separated by spaces like in the next example: **marker:** . After the name of the marker is completed, the user must set also the values of the attributes: the *type* field, which may have one of the values “*int*” or “*ext*”.

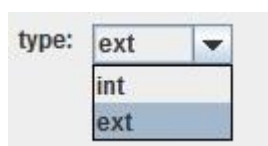


Figure 32: Add menu – marker type



If we select the type “*ext*” the user must complete also the *connect* attribute, which may have

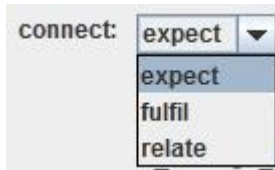


Figure 33: Add menu – marker type, connect attribute

one of the values: “*expect*”, “*fulfil*” or “*relate*”.. The last thing to be filled in is the relation type. Multiple relations can be selected for a marker, but at least one is mandatory. After all the information regarding a marker were completed, in order to add a new marker in the list the user must press the “*Add*” button. After adding a new marker the graphical view is automatically refreshed and the new marker appears in the “*view markers*” section. If the *marker* field was not filled in with the name of the marker, an error message will appear:

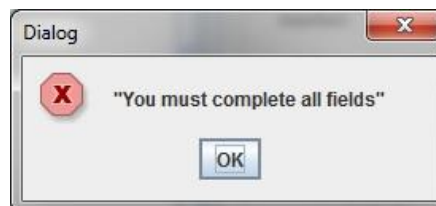


Figure 34: Error message – marker field

A similar error message will appear if no relation was selected:

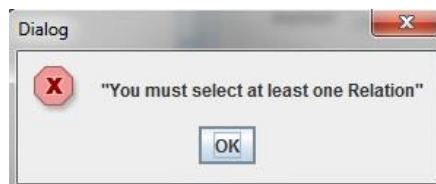


Figure 35: Error message – relation field

If the marker that the user wants to add already exists in the list of markers with the same type, an error message will appear signalling the id of the marker. A marker is considered as already existing in the list if both the marker text and the type of the marker are identically. We can have two “*and*” markers, one of them being of type *int*, meaning it is an internal marker, and the other one being of type *ext*, meaning that it is external. Thus, there are considered two different markers.

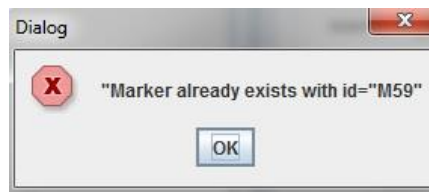


Figure 36: Error message – adding existing marker

The *Edit Menu* (Figure 37) offers the possibility to edit an existing marker. To edit a marker, first it must be selected. To select a marker we must click on the name of the marker from the “*view markers*” section and all the information regarding that marker will be automatically filled in the corresponding fields, then pressing the “*edit*” checkbox all the fields becomes enabled, offering the possibility to edit them. After the modifications are made, by pressing the “*Save*” button, all the modifications will be saved.

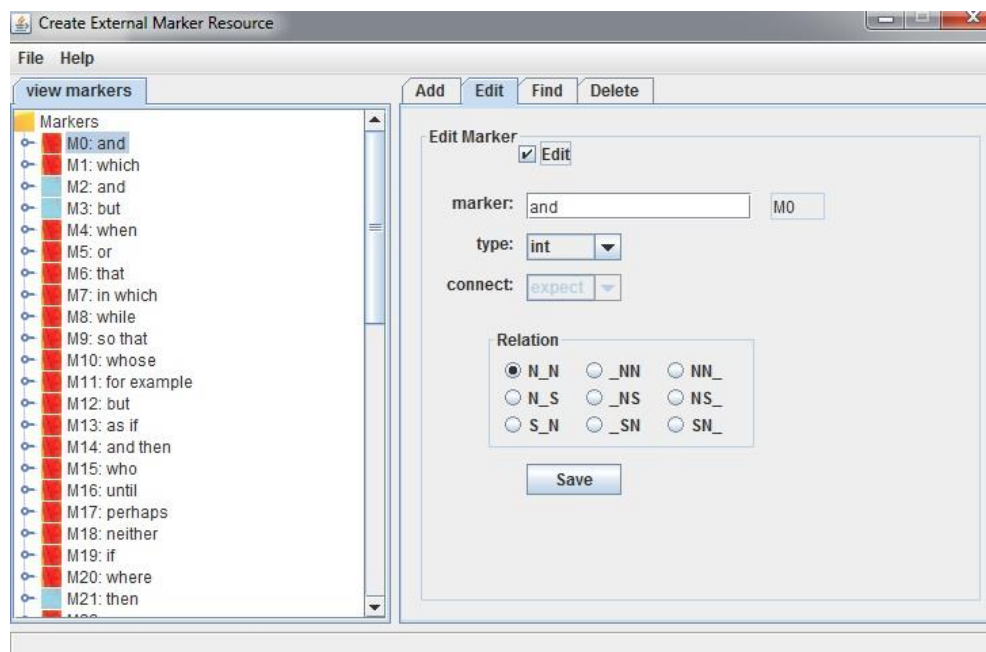
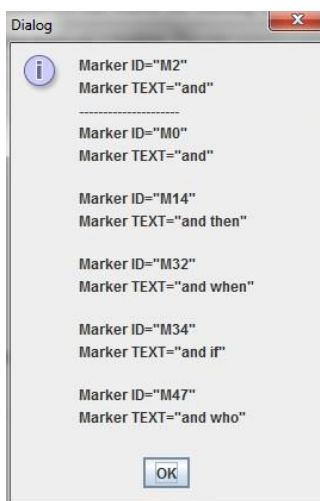


Figure 37: Edit Menu

Using the *Find Menu* we can search for existing markers. This is a useful operation when we want to verify the existence of a marker. The user can search a marker after its *id* or after the text. When the user searches a marker after its *id* he must complete the *id* field with the *id* he wants to search. If a marker with that *id* exists within the list of markers, a dialog will appear containing the name of the marker with that *id*:



Figure 38: Warning dialog



The user can also search a marker by text and the application will return both, the complete matching of the text and the partial matching. The searching function is case sensitive. For example if we want to verify if the marker “AND” exists in the list, the results returned contains all occurrences of markers which have in their composition the searched WORD.

Figure 39: Search marker results

If there is no marker, an error message will appear like in the next picture:



Figure 40: Error finding marker

Markers can also be removed by using the *Delete Menu* (Figure 41). We can delete a marker by selecting its *id* and after that pressing the “*Remove*” button. After this operation is performed, the graphical view is automatically refreshed and the new marker disappears from



the "view markers" section.

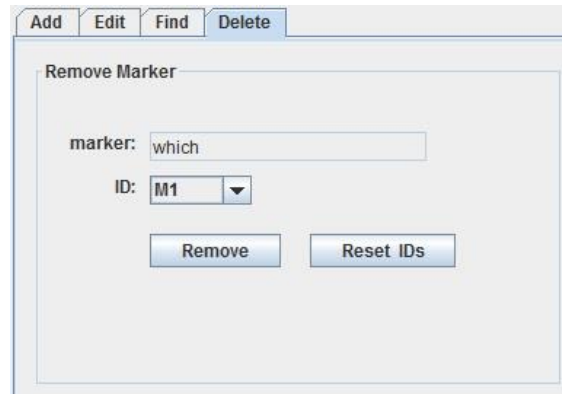


Figure 41: Delete Menu

The Discourse Markers Editor module technical record

- Programming language: Java, platform independent.
- OS compatibility: Windows and Linux.
- Access conditions: free for non-commercial.
- Copyright: UAIC.
- Input: markers_type.xml
- Output: the updated file markers_type.xml
- Status: stable, finalised.

4.2 Graphical interface for viewing and evaluating the discourse trees

Besides the actual implementation of the system UAIC has created a graphical interface (Figure 42) for viewing and evaluating the discourse trees. This application is useful for reviewing the adequacy of the generated trees and to make detailed analyses of the generated trees for discovering new heuristics to actually improve the system.

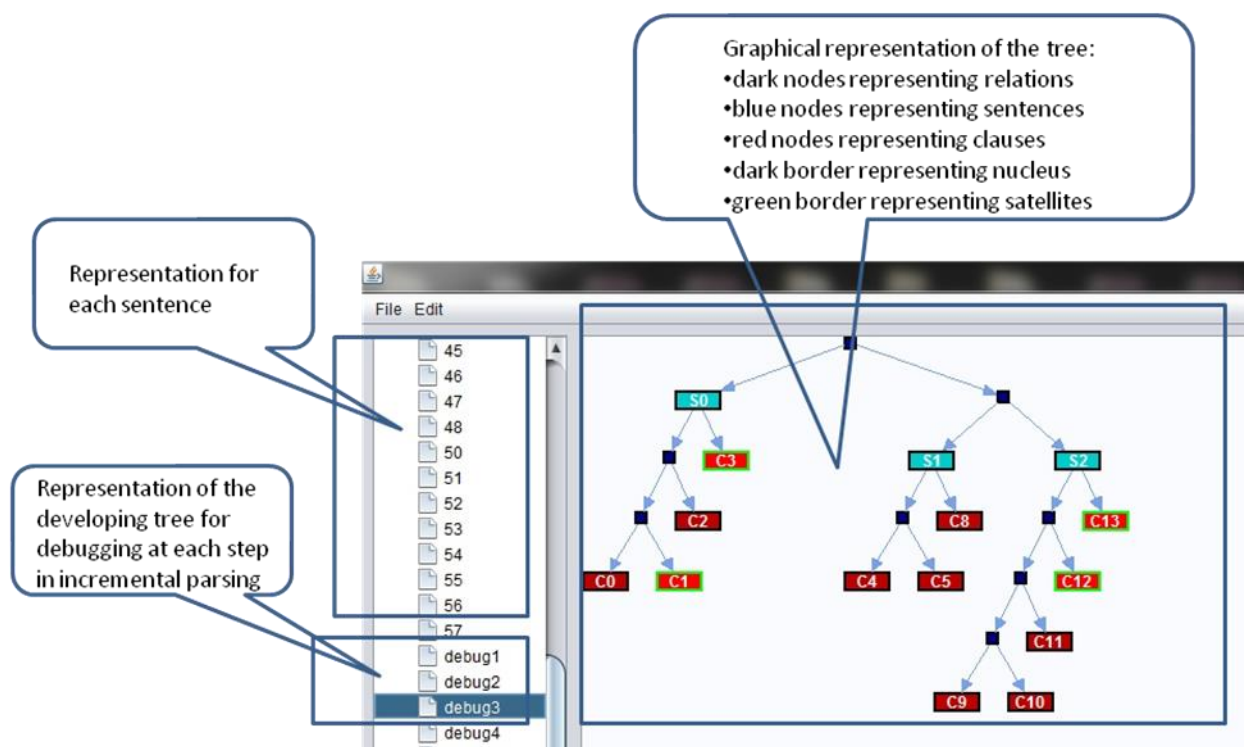


Figure 42: Draw Trees Graphical Interface

The application is created in Java using *jgraph*⁸ library in order to draw the structure of the trees. On the left side is presented the list of trees for each sentence, the final tree, and the entire process of building trees. In the right side is the graphical representation of the computed tree.

The application does not include yet a tree editing function.

The Discourse Trees Viewer module technical record

- Programming language: Java, platform independent.
- OS compatibility: Windows and Linux.
- Access conditions: free for non-commercial use
- Copyright: UAIC.
- Input: a BIN file (as described in the technical record of the DP module, Section 3.3.14).
- Output: none
- Status: stable, finalised.

⁸

<http://www.jgraph.com/>



5 BUILDING CORPORA

5.1 *Building the clause segmentation and markers corpora*

The summarisation model is described in (Cristea et al., 2005). The summariser exploits the discourse structure (a simplified variant of a rhetorical structure, see (Mann&Thompson, 1988), (Marcu, 2000)). In ATLAS this model is applied only for summarising short texts. The partners have been instructed to apply the following steps when building their language corpora (L stands for the respective *language*):

1. Look for translations in L of the short stories you can find in: svn://office.tetacom-bg.com/tetacom/documentation/atlas/WP5 – [Summarization/Summarisation package.zip](#) in the EN folder. We think that having parallel texts would ease the process of comparative evaluation among languages. Because parallel texts are difficult to obtain, we thought of a number of texts that we know should exist in each of our languages.
2. Collect other short texts to arrive to a total of 1000 sentences (approximately 20 short texts, each of 2 to 4 pages), in different domains than stories/fancies, for instance: financial news, political articles, geographical descriptions, etc. These should be annotated to clause boundaries and markers. Out of these, approximately a quarter (250 sentences, or 5 short texts) should also be annotated for summary.
3. Launch the WP4 processing chain of L on each of these texts, which should produce the following type of XML markers, added to the original text:
 1. `<S></S>` – indicating the sentence boundaries;
 2. `<W></W>` – marking tokens. The recommended attributes of tokens are: ID, POS and LEMMA. See examples in: “*example1.xml*”, “*example2.xml*”, “*example3.xml*”, “*example4.xml*”. These examples follow the XML dtd conventions in “*example.dtd*”.
4. Use a program that allows you to annotate texts by placing XML tags around words or sequences of words. The recommended tools are PALinkA⁹ and `clauseannotator.jar`¹⁰,

⁹ Perspicuous and Adjustable Links Annotator: <http://clg.wlv.ac.uk/projects/PALinkA/> and <http://clg.wlv.ac.uk/trac/palinka/wiki/Preferences>



or any program displaying a similar behaviour. With this annotation tool, mark boundaries of clauses, using the element tags `<CLAUSE></CLAUSE>`. A `CLAUSE` element has one compulsory attribute, `ID` – indicating a unique identifier of the clause, and one optional: `CONTINUE`. With this attribute you could mark interrupted clauses, like in Ex. 28:

Ex. 28

`<The youngest,>[0] <who was called simpleton,>[1] <set out to seek his brothers.>[2]`

Here, there are two clauses: [1] occurring inside the other, made out of the spans [0]+[2], as in the following annotation:

```
<S ID="S0"><CLAUSE ID="C0">The youngest,</CLAUSE> <CLAUSE ID="C1">who was
called Simpleton,</CLAUSE> <CLAUSE ID="C2" CONTINUE="C0">set out to seek his
brothers.</CLAUSE></S>
```

Other example can be found in “*example1-ann.xml*”, “*example2-ann.xml*”, “*example3-ann.xml*”, “*example4-ann.xml*” according to “*example1.dtd*”.

5. After the clauses are annotated, to facilitate the following step in your work, which deals with annotating discourse markers, you can run the `ListClauses.jar` file (found in the `ListClause` directory, a launch example being given in the `runListClauses.bat` file), which will output a text with each clause on one line, indicating also the `ID` and the `CONTINUE` attributes. Print this output and keep the sheet of paper by you while annotating markers (step 6). This will ease your annotation process.
6. Come back in the annotator tool, to annotate the markers occurring in the text. The rules to annotate markers are:
 - a marker is annotated with the `<MARKER></MARKER>` XML element;
 - the `ID` (compulsory) attribute defines a unique identifier for each marker;
 - the `NUC` (compulsory) attribute should have one of the following values: “`N_N`”, “`N_S`”, “`S_N`”, “`_NN`”, “`_NS`”, “`_SN`”, “`NN_`”, “`NS_`”, “`SN_`” where `N` represents a nucleus, `S` represents a satellite and the underline sign

¹⁰ Indicated by our Polish partners.



(_) represents the relative position of the marker with respect to the two arguments (spans of text) that it binds;

- the TYPE (optional) attribute should have one of the values: "int" or "ext". TYPE="int" means that the marker relates two clauses of a sentence (therefore is internal to a sentence).

TYPE="ext" means that the marker relates a sentence to another span of text outside itself (therefore is external to a sentence).

We interpret an implicit value of the TYPE attribute as being "int". This means that you are not obliged to specify this attribute in any of the internal type of markers (most numerous);

- only if TYPE="ext", you should (compulsorily) also fill in a CONNECT attribute, with one of the values: "expect", "fulfil" or "relate":

CONNECT="expect" means that the span unit the marker belongs to introduces an expectation, like the marker *on the one hand* in Ex. 29.

Ex. 29 (from (Cristea and Webber, 1997))

<On the one hand, John is very generous.>^[1] | <For example, suppose you needed some money.>^[2] | <Then, you would just have to ask him for it.>^[3] | <On the other hand, he is very difficult to find.>^[4]

You will easily recognise a marker raising an expectation by imagining that the discourse stops immediately after the sentence including the marker. If you perceive such an interrupted discourse as defective, then it means that the marker raises an expectation.

CONNECT="fulfil" means that the span of text the marker belongs to fulfils (satisfies) an open expectation. This is the case with *on the other hand* marker of Ex. 10, which announces the span [4] as fulfilling the expectation opened in [1].

CONNECT="relate" means that the span of text the marker belongs to relates in some way to the previous discourse, but neither raising nor fulfilling an expectation. This is the case of the *but* marker in Ex. 30 below, which announces that the span of text of the sentence [4]+[5] relates to the previous discourse: the span [2]+[3].

Ex. 30



<Suppose – just suppose, this guy was really what he said he was: a retired professional killer!>^[1] | <If he was just a nut,>^[2] <no harm was done.>^[3] | <But if he was the real thing,>^[4] <he could do something about Lolly.>^[5]

Out of all markers (underlined) in Ex. 30, only *suppose* and *but* are external. *Suppose* is raising an expectation (fulfilled by the sentence [2]+[5], while there is not a following CONNECT="fulfil" external marker). *But* is relating its sentence [4]+[5] to the previous discourse. The two *if* markers, in clauses [2] and [4], are internal markers, relating, respectively, the clause [2] with the clause [3], and the clause [4] with the clause [5]. Notice that the *suppose* marker has the NUC pattern "_NN" (because it prefixes its first argument and both arguments are equally important: if one is deleted, the text becomes incomprehensible); the *but* marker has NUC="N_N" (because it is positioned between the first and the second argument, and both arguments are equally important); finally, both *if* markers have NUC="_NN" (because they prefix their first argument and if one argument is deleted, the text loses the sense).

In the following, we show the desired annotation for Ex. 29:

```
<S ID="S1"><CLAUSE ID="1"><MARKER ID="m1" TYPE="ext" CONNECT="expect" NUC="_NN">On the one hand</MARKER>, John is very generous.</CLAUSE></S><S ID="S2"><CLAUSE ID="2"><MARKER ID="m2" TYPE="ext" CONNECT="relate" NUC="N_S">For example</MARKER>, <MARKER ID="m3" TYPE="ext" CONNECT="expect" NUC="_NN">suppose</MARKER>you needed some money.</CLAUSE></S><S ID="S3"><CLAUSE ID="3"><MARKER ID="m4" TYPE="ext" CONNECT="fulfil" NUC="N_N">Then</MARKER>, you would just have to ask him for it. </CLAUSE></S><S ID="S4"><CLAUSE ID="4"><MARKER ID="m5" TYPE="ext" CONNECT="fulfil" NUC="N_N">On the other hand</MARKER>, he is very difficult to find.</CLAUSE></S>
```

and for Ex. 30:

```
<S ID="S1"><CLAUSE ID="1"><MARKER ID="m1" TYPE="ext" CONNECT="expect" NUC="_NN">Suppose</MARKER> – just suppose, this guy was really what he said he was: a retired professional killer!</CLAUSE></S> <S ID="S2"><CLAUSE ID="2"><MARKER ID="m2" NUC="_NN">If</MARKER> he was just a nut,</CLAUSE> <CLAUSE ID="3">no harm was done.</CLAUSE></S> <S ID="S3"><CLAUSE ID="4"><MARKER ID="m3" TYPE="ext" CONNECT="relate" NUC="N_N">But</MARKER> <MARKER ID="m4" NUC="_NN">if</MARKER> he was the real thing,</CLAUSE> <CLAUSE ID="5">he could do something about Lolly.</CLAUSE></S>
```



In some languages some markers can be found inside the clauses, therefore are not interposed as separators in boundary positions, e.g., something like “He was clever, he therefore solved this problem”. If you fall on such a case, try to find the most natural rephrase of the sentence in which the marker is placed at one of the extreme positions inside its EDU and report the NUC pattern accordingly. As such, for the above case the nuclearity annotation of the marker therefore should be N_S.

Also, please be aware that the same word or sequence of words can function as a marker in some situations and not in others. For instance, in the following example:

Ex. 31

Both plans also prohibited common directors, or employees between Du Pont, Christiana, and Delaware, on the one hand, and General Motors, on the other.

Neither *on the one hand*, nor *on the other* should be considered discourse markers. The same happens in the example:

Ex. 32

<When you want to build a website of a company, for example,>^[1] <you need to know the company profile.>^[2]

Here, *for example* is not a discourse marker. It signals one possibility among many of what you would want to build and is incorporated in the first clause of this sentence. The only marker here is when, prefixing its first argument, and with the second argument coming after the first.

7. Run the training module that you will find at <svn://office.tetracom-bg.com/tetracom/documentation/atlas/WP5 – Summarization/>. Based on the annotated examples (training corpus) built at step 6, this module creates a file containing all annotated markers (`markers_type.xml`). With `MarkersEditor.jar`, that you will find at <svn://office.tetracom-bg.com/tetracom/documentation/atlas/WP5 – Summarization/>, you can view and edit this file, within a graphical interface. You will find there all the annotated markers with their properties. You can add new markers (you should do this if you know of markers that were not included in the training corpus) or edit existing ones, if you recognise mistakes (in this case we suggest that you



also correct the corresponding annotation errors in the training corpus and re-run the training module). You are free also to delete markers. In any case, however, it is hoped that at the end of this process the file `markers_type.xml` becomes synchronised with the corpus in which markers are annotated (for those markers for which there are occurrences in the corpus). This file will be used by the discourse parser in order to build the discourse trees of the texts.

The annotation process to build a corpus to be used to train the CS and the DP modules is divided in two stages:

The first phase consists of an automatic annotation made by a language processing chain, which, starting from the original file in “.txt”- *UTF-8*¹¹ format should produce the following type of *XML* markers (in this context “markers” refers to *XML* elements also called tags):

- `<S></S>` - indicating the sentence boundaries;
- `<W></W>` - indicating tokens with at least the following attributes: `ID`, `POS` and `LEMMA`.

In the second phase it should be used an annotator to annotate the files created in the first phase, by placing *XML* tags around words or sequences of words. We used *PALinkA*¹² (Figure 43) to annotate Romanian and English texts, but any program displaying a similar behaviour could be used.

With this annotation tool we annotate boundaries of clauses, using the element tags `<CLAUSE></CLAUSE>`. A `CLAUSE` element has one compulsory attribute, `ID` indicating a unique identifier of the clause, and one optional: `CONTINUE` with which we can mark interrupted clauses.

¹¹ Unicode Transformation Format-8: <http://www.utf8.com/>

¹² Perspicuous and Adjustable Links Annotator: <http://clg.wlv.ac.uk/projects/PALinkA/>

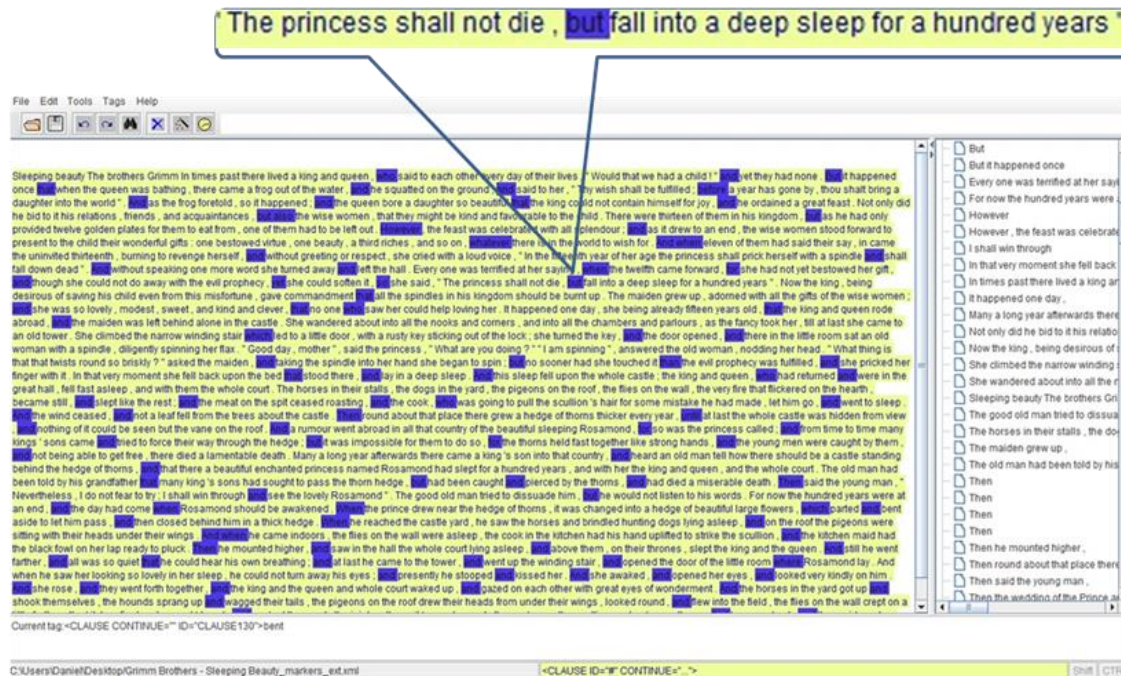


Figure 43: Example of a text annotated with PALinkA

After the clauses are annotated, the markers occurring in text should be annotated as well.

The rules to annotate markers are:

- a marker is annotated with the `<MARKER></MARKER>` XML element;
- the ID (compulsory) attribute defines a unique identifier for each marker;
- the NUC (compulsory) attribute should have one of the following values: "N_N", "N_S", "S_N", "_NN", "_NS", "_SN", "NN_", "NS_", "SN_" where N represents a nucleus, S represents a satellite and the underline sign (_) represents the relative position of the marker with respect to the two arguments (spans of text) that it binds; a nucleus represents a span of text (elementary discourse unit, *EDU*) which, if deleted from its context, makes the text incoherent; a satellite represents a span of text (*EDU*) which, if deleted from its context, does not harm the overall understandability, even if some details are now missing. In (Carlson et. al, 2001) the nucleus is defined as being more salient in the discourse structure, while the satellite represents supporting information.
- the TYPE (optional) attribute should have one of the values: "int" or "ext"; TYPE="int" means that the marker relates two clauses of a sentence (therefore is internal to a sentence) and TYPE="ext" means that the marker relates a sentence to



another span of text outside itself (therefore is external to a sentence).

- only if TYPE="ext", the annotator should (compulsorily) also fill in a CONNECT attribute, with one of the values: "expect", "fulfil" or "relate";
- CONNECT="expect" means that the span unit the marker belongs to introduces an expectation; a marker raising an expectation is recognized if it leaves the impression that the discourse cannot stop immediately after the sentence including the marker, as in the case of the marker *on the one hand* in the next example, from (Cristea and Webber, 1997):

Ex. 33

<On the one hand, John is very generous,^[1]> <For example, suppose you needed some money.^[2]> <Then, you would just have to ask him for it.^[3]> <On the other hand, he is very difficult to find.^[4]>

- CONNECT="fulfil" means that the span of text the marker belongs to fulfils (satisfies) an open expectation. This is the case with *on the other hand* marker of the example above, which announces the span [4] as fulfilling the expectation opened in [1].
- CONNECT="relate" means that the span of text the marker belongs to relates in some way to the previous discourse, but neither raising nor fulfilling an expectation. This is the case of the *but* marker in the following example, also from (Cristea and Webber, 1997), which announces that the span of text of the sentence [4] + [5] relates to the previous discourse: the span [2] + [3].

Ex. 34

<Suppose – just suppose, this guy was really what he said he was: a retired professional killer!^[1]> <If he was just a nut,^[2]> <no harm was done.^[3]> <But if he was the real thing,^[4]> <he could do something about Lolly.^[5]>

5.2 Building the corpora of summaries

The partners were requested to build a number of short texts summaries (produced by humans), to be used to evaluate the summariser for their languages. The minimum



recommended number of people-summarising-each-text was set to 2, but better was to have 3 subjects summarising the same text. The steps indicated for this task have been:

1. Chose 10 texts to be summarised. It is preferable that they be among the texts you have selected for building L corpus of segmented sentences and markers. As such, they have already clause level segmentation, including IDs of all clauses, which reduces considerably your amount of work.
2. Go to the svn server and copy onto your computer the Clause Annotator tool `summann.jar`¹³
3. Use this tool to annotate three summaries with different reduction rates for each input text.

Whether you use this tool or not, the output of your activity should be a list of clause IDs: those that you consider to be part of the summary. All summaries are excerpt type summaries. The three reduction rates are:

- a 20% summary – you should come up with a summary containing approximately 20% of the words in the original text;
- a 10% summary – only approximately 10% of the words in the original text should appear here. The strongly suggested strategy is to simply include only half of the *EDUs* you have selected for the 20%-summary;
- a 5% summary – only approximately 5% of the words in the original text should appear here. Again, the suggested strategy is to include only half of the *EDUs* you have selected for the 10%-summary.

¹³ Thanks to our Polish colleagues.



6 CORPORA AND LANGUAGE MODELS

In this section we will report on the work performed by each partner to build specific corpora that are used for the summarisation process as well as the related language data for diverse modules of the summarisation chain. The specific information will follow the patterns:

For the Anaphora Resolution (see Table 17 and Table 18)

- use of a localised version of RARE (yes/no), if yes then:
 1. `copulativeVerbs.txt` (yes/no), if yes then: # of entries
 2. `determiners.xml` (yes/no), if yes then: # of entries
 3. `familyNames.txt` (yes/no), if yes then: # of entries
 4. `femaleNames.txt` (yes/no), if yes then: # of entries
 5. `maleNames.txt` (yes/no), if yes then: # of entries
 6. # of rules, out of which: certifying, demolishing, promoting/demoting
- existence of an AR annotated corpus (yes/no), if yes then:
 - file name
 - # of tokens
 - # of Res (noun phrases), out of which: # pronouns, # proper nouns
 - # of Des (coreferential chains), out of which: # of distinct named entities

Totals for each of the numeric fields

General total
- evaluation of the AR module (yes/no), if yes then:
 - is it an announced MUC score for this language? (yes/no), if yes – reproduce it
 - the tested MUC score
 - other scores (yes/no), if yes name them and give values

<code>copulativeVerbs.txt</code>	
<code>determiners.xml</code>	
<code>familyNames.txt</code>	
<code>femaleNames.txt</code>	



maleNames.txt		
# of rules	certifying	
	demolishing	
	promoting	
Other resources (please specify)		

Table 17: RARE localisation files

File name		Total
# of tokens		
# of REs	# pronouns	
	# proper nouns	
# of DEs		
	# of NEs	
Evaluation	The best known MUC score	
	The ALTAS AR MUC score	

Table 18: The AR annotated corpus

For CS and DP corpora (see Table 19)

- file name
- # of sentences
- # of tokens
- # of clauses
- # of annotated occurrences of markers in the corpus (<MARKER></MARKER> parenthesis)
- # of distinct lexicals acting as markers in the corpus
- # of `int` markers (occurrences)
- # of `ext` markers (occurrences), out of which:
 - # of `CONNECT="expect"` `ext` markers (occurrences)
 - # of `CONNECT="fulfil"` `ext` markers (occurrences)
 - # of `CONNECT="relate"` `ext` markers (occurrences)
- existence of the `markers_type.xml` file (yes, no), if yes then # of entries

Totals for each of the numeric fields



File name		Total
# of sentences		
# of tokens		
# of clauses		
# of markers		
# of lexicals		
# of int markers		
# of ext markers	"expect"	
	"fulfil"	
	"relate"	
markers_type.xml		
Evaluation	P	
	R	
	F	

Table 19: The CS and DP corpus

The summaries corpora (see Table 20)

- file name
- # of sentences
- # of clauses
- 20% summary (yes/no), if yes # clauses
- 10% summary (yes/no), if yes # clauses
- 5% summary (yes/no), if yes # clauses
- evaluation against the 20% summary file, gives Precision, Recall, F-measure for one or both of the mentioned measures (as defined in Section 3.4.3): Histogram (H) and Averaging (A).

Totals for each of the numeric fields. The figures in this column referring to measures give averages of the percentages.

File name		Total
# of sentences		
# of clauses		
20% summary		
10% summary		
5% summary		
Evaluation against the	P (H)	
	R (H)	



20% summary file	F (H)		
	P (A)		
	R (A)		
	F (A)		

Table 20: The summaries corpus

In the *sumaries corpus* tables that are presented in the following sections, null values reflect errors in the processing chain (as signalled and analysed in the Conclusions – Section 7.1). It is our intentions to solve all these errors till the end of the ALTAS project.

6.1 Bulgarian

RARE was adopted for Bulgarian by means of several resources:

- A file `copulativeVerbs.txt` contains the Bulgarian copulative verb.
- A file `familyNames.txt` contains a list of commonly encountered family names in Bulgarian.
- A file `femaleNames.txt` contains a list of common Bulgarian female names.
- A file `maleNames.txt` consists of common Bulgarian male names.
- An XML file `constraints.xml` describes specific co-reference constraints in Bulgarian by means of a set of rules.

<code>copulativeVerbs.txt</code>	1
<code>determiners.xml</code>	13
<code>familyNames.txt</code>	1384
<code>femaleNames.txt</code>	645
<code>maleNames.txt</code>	761



Constraints.xml # of rules	certifying	0
	demolishing	3
	promoting	19
Other resources (please specify)		
stopwords.txt	71	
tagset.xml	871	

Table 21-BG: RARE localisation files for Bulgarian

The Bulgarian training corpus for clause splitting is relatively wide in terms of coverage of different text styles and categories – it includes fiction, law, scientific and journalistic texts, which are manually annotated. The annotation was performed with the tool ClauseChooser¹⁴, which was specially adopted for that type of manual annotation. The annotation conforms to the common principles accepted by all partners and is focused on sentence and clause borders, relations between clauses - coordinative or subordinative, as well as clause internal markers for clause connectivity. Each clause in a pair is identified as either main / independent, or subordinate / dependent according to its status in the sentence and with respect to the other clause in the pair. In this way, additional information about the ordering of the clauses and the position of the conjunctive words/phrases in the clauses, as well as language-specific clause-to-clause ordering constraints is obtained. Statistical information for the corpus can be found in the Table below.

¹⁴ <http://dcl.bas.bg/Tools/Chooser/Chooser.zip>



File name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
-----------	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----

Table 22-BG: The CS and DP corpus for Bulgarian

File name	1	3	6	12	13	17	18	19	20	23	24	25	26	27	Total
# of sentences	43	52	69	181	98	104	105	55	45	54	56	171	79	56	1168
# of clauses	90	161	175	383	192	279	243	234	176	141	109	318	238	216	2955
20% summary	10	21	27	73	38	47	43	35	27	29	20	56	51	43	520
10% summary	4	13	13	35	16	24	21	16	14	12	11	27	23	20	249

File name	16	17	18	19	20	21	22	23	24	25	26	27	28	Total
# of sentences	263	104	105	55	45	79	101	54	56	171	79	56	122	2749
# of tokens	2289	2343	1568	1059	724	1902	2605	1103	1486	1598	1761	1371	1823	51116
# of clauses	439	279	243	234	176	297	386	141	109	318	238	216	281	6468
# of markers	113	108	96	106	70	182	227	61	30	93	98	106	112	2507
# of lexicals	113	108	96	106	70	182	227	61	30	93	98	106	112	2507
# of int markers	113	108	96	106	70	182	227	61	30	93	98	106	112	2507
# of ext markers	"expect"	0	0	0	0	0	0	0	0	0	0	0	0	0
	"fulfil"	0	0	0	0	0	0	0	0	0	0	0	0	0
	"relate"	0	0	0	0	0	0	0	0	0	0	0	0	0
markers_type.xml														

Evaluation	P	0.9717													
	R	0.7666													
	F	0.8571													
5% summary	3	5	7	17	7	12	8	9	9	6	7	14	16	9	129
Evaluation against the 20% summary File	P	0.215				0.27	0.238		0.18	0.15	0.216				0.212
	R	0.48				0.55	0.469		0.289	0.189	0.413				0.316
	F	0.297				0.365	0.316		0.223	0.167	0.284				0.254

Table 23-BG: The Bulgarian summaries corpus. File names are given by code and code – file name correspondence is presented in the table below.



Code	Filename	Code	Filename
1	00000764tAAL.xml	15	00002194bBTZ.xml
2	00000854bBJC.xml	16	00002279bDDE.xml
3	00001079bBEA.xml	17	Allan_Edgar_Poe-The_Masque_of_the_Red_Death_bg.xml
4	00001173bBQJ.xml	18	ArthurClarke.xml
5	00001370bCSG.xml	19	Grimm-SleepingBeauty.xml
6	00001650bCJI.xml	20	Grimm-TheQueenBee.xml
7	00002022bBDZ.xml	21	H. P. Lovecraft - The other gods_bg.xml
8	00002023bCFZ.xml	22	H.P._Lovecraft-_The_quest_of_Iranon_bg.xml
9	00002027bBUZ.xml	23	L00002537bDBZ.xml
10	00002030bBHZ.xml	24	L00002763bBIG.xml
11	00002034bBBK.xml	25	L00003210bDJB.xml
12	00002057bDMZ.xml	26	L00003984tDLE.xml
13	00002161bBCZ.xml	27	Polaris.xml
14	00002193bBFZ.xml	28	Washington_Irving-The_adventure_of_the_German_student_BG.xml

Table 24-BG: Code – file name correspondence for Table 22-BG and Table 23-BG

For Bulgarian, 14 Gold summaries were developed, each one by two human experts. The Gold corpus includes text from different styles and categories – some of the texts are part of the clause splitting training corpus. For each text, 3 summaries were produced with different length: 20%, 10% and 5% of the clauses from the original text. **Table 23** presents the statistical data.



6.2 English

1.copulativeVerbs.txt	yes	1
2.determiners.xml	yes	7
3.familyNames.txt	yes	26866
4.femaleNames.txt	yes	4275
5.maleNames.txt	yes	1224
# of rules	certifying	2
	demolishing	2
	promoting	19
Other resources (please specify)		N/A

Table 25-EN: RARE localisation files for English

File name	1000 place s befo re you die	Adir onda ck park	Afric a afar depr essio ns short	Afric a alber tine rift	A. E. Poe - The Masq ue of the Red Death	Ame rica wild river s	Anima l domest ication loong	Anth ropo cene	A. C. Clar ke forg et thee, oh Eart h	Asta na city long	Aust ralia slot cany ons	Braz il dune s	Cam bodg ia heali ng land s	Stran ge room	Can chin a go gree n long	
# of sentences	55	70	22	94	102	72	219	93	76	136	98	42	172	41	153	
# of tokens	1100	1715	520	2297	2741	1649	5175	2210	1992	3271	2086	982	3500	895	3547	
# of clauses	131	156	54	202	236	153	422	180	181	326	200	79	312	81	295	
# of markers	46	43	18	72	118	55	126	64	82	112	46	27	83	17	109	
# of lexicals																
# of int markers	29	29	14	49	88	39	75	47	73	88	32	14	53	13	67	
# of ext markers	"expect"	7	6	0	8	0	5	15	5	0	4	2	6	8	1	15
	"fulfil"	7	0	2	2	0	1	4	2	0	2	5	2	3	1	15
	"relate"	3	8	2	13	30	10	32	10	9	18	7	5	19	2	11

Table 26-EN: The CS and DP corpus for English



File name	C. Dickens life	Cleopatra loong6	Coral barrier long	A.E.Poe The tell tale	Grimm Brothers Sleeping Beauty	Grimm Brothers The queen Bee	H.P. Lovecraft Polaris	H.P. Lovecraft The other gods	W. Irving The Adventure of the German student	Total
# of sentences	34	192	111	153	38	26	52	72	123	2246
# of tokens	1075	4908	2579	2457	1504	936	1706	2246	2413	53504
# of clauses	68	376	222	286	173	117	172	239	219	4880
# of markers	26	136	64	101	113	71	99	147	57	1832
# of lexicals										0
# of int markers	17	101	38	76	96	66	88	134	57	1383
# of ext markers	"expect"	4	15	7	9	0	0	0	0	117
	"fulfil"	1	7	3	4	0	0	0	0	61
	"relate"	4	13	16	12	17	5	11	13	270
markers_type.xml										
Evaluation	P									0.9818
	R									0.9364
	F									0.9585

Table 27-EN: The CS and DP corpus for English (continued)

File name	A.E. Poe The Masque of the Red Death	A.C. Clarke If I forget thee, oh Earth	Grimm Brothers Sleeping beauty	Grimm Brothers The queen bee	H.P. Lovecraft Polaris	H.P. Lovecraft The other gods	W. Irving The adventure of the German student	Total
# of sentences	102	76	38	26	52	72	123	489
# of clauses	90	161	175	383	192	279	219	1499
20% summary	10	21	35	23	37	48	44	218
10% summary	4	13	17	11	18	24	22	109
5% summary	3	5	8	5	9	12	11	53
Evaluation against the 20% summary file	P	0.224	0.199	0.163	0.323	0.4	0.286	0.265
	R	0.382	0.274	0.347	0.484	0.530	0.456	0.412
	F	0.282	0.230	0.222	0.387	0.456	0.351	0.321

Table 28-EN: The English summaries corpus



6.3 German

1.copulativeVerbs.txt	yes	7
2.determiners.xml	yes	52
3.familyNames.txt	yes	100
4.femaleNames.txt	yes	305
5.maleNames.txt	yes	255
# of rules	certifying	-
	demolishing	2
	promoting	12
Other resources (please specify)		N/A

Table 29-DE: RARE localisation files

File name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
# of sentences	86	105	43	101	41	98	71	87	25	217	44	66	30	72	82	51	54	102	
# of tokens	2164	1739	886	1964	858	2848	1738	1540	546	4366	602	1228	853	1821	2102	1186	1262	4136	
# of clauses	178	235	64	163	65	189	135	148	41	491	63	89	125	104	139	80	78	339	
# of markers	54	37	12	25	8	29	22	17	8	67	7	10	31	12	13	14	11	19	
# of lexicals																			
# of int markers	39	33	4	15	6	17	15	11	2	54	3	5	30	5	3	7	5	10	
# of ext markers	"expect	4	1	3	1	1	3	2	2	2	5	1	1	0	2	4	1	1	1
	"fulfi"	1	1	2	2	0	0	1	1	0	0	0	0	2	1	2	1	2	
	"relat"	12	2	3	7	1	9	4	3	4	8	3	4	1	3	5	4	4	6



markers_type.xml		
Evaluation	P	0.9347826
	R	0.6615385
	F	0.7747748

Table 30-DE: CS and DP corpus for German. File names are given by code and code – file name correspondence is presented in the table below.

File name	1	2	3	4	5	6	7	8	9	Total
# of sentences	124	82	86	105	72	71	52	66	123	489
# of clauses	225	139	178	235	104	135	491	89	219	1499
20% summary	40	22	32	38	20	29	80	17	44	218
10% summary	20	11	17	17	9	17	42	9	22	109
5% summary	10	7	10	8	4	9	20	5	11	53
Evaluation 20%	P									
	R									
	F									

Table 31-DE: The German summaries corpus

Code	Filename, in Table 30-	Filename, in Table 31-DE
1	22a79e82-7c04-45e6-b7e6-0e9dd3	1edac4ec-d9d5-4a20-9a30-d4a872b875ac
2	26bdf0f7-eec1-44ce-86f1-8bfa7b	d7e0ad14-0942-4486-8c49-517f1060937b
3	27350145-c855-4fed-87cd-eeac49	22a79e82-7c04-45e6-b7e6-0e9dd34ca7b6
4	274f0093-bd9c-4078-acfa-6ee1a2	26bdf0f7-eec1-44ce-86f1-8bfa7b2e18ef
5	2b909b7c-1556-4555-b04c-15fcaa	cef2f2c5-5ab1-4243-bd10-021d32890a22
6	5047fda9-f908-4105-91a2-f006dc	58bb81e0-1908-4e87-88ae-b32afbb82538
7	58bb81e0-1908-4e87-88ae-b32afb	8c647524-39f0-47ce-aa1a-af22c9a983fe
8	63b2fd52-1266-4094-a6ad-de91a6	c789c65c-8eb4-4141-a46d-f43dd89aea86
9	876732eb-e431-418d-b55b-19e191	ce771460-2470-454f-87be-8be9bbb2173a
10	8c647524-39f0-47ce-aa1a-af22c9	--
11	96ee1cbc-c838-4fea-a286-597caa	--
12	c789c65c-8eb4-4141-a46d-f43dd8	--
13	ce771460-2470-454f-87be-8be9bb	--
14	cef2f2c5-5ab1-4243-bd10-021d32	--



15	d7e0ad14-0942-4486-8c49-517f10	--
16	ecf78269-0a1f-47b2-af64-a39bfc	--
17	f28329f6-6086-44ed-a7fa-01f0ea	--
18	f8cd839c-f107-42f7-b4a1-c5bf7e	--

Table 32-DE: Code – file name correspondence for Table 30-DE and Table 31-DE

6.4 Greek

In order to make RARE work for the Greek language two type of resources have been prepared: a set of internal files to be used in a pre-RARE phase, and a set of files accessed by RARE. More specifically:

- A file called `copulativeVerbs.txt`, has been build that contains the list of Greek copulative verbs that are placed in the `Resources` directory. This is a text file containing one lemma of a copulative verb on each line.
- A file called `determiners.xml`, has been build that contains the list of Greek determiners (articles). This is a text XML file including 2 types of articles: definite and indefinite. The tags should be: `<defDET>`, `<undefDET>`.
- A file called `familyNames.txt`, that contains a list of commonly encountered family names in the Greek language.
- A file called `femaleNames.txt`, that contains a list of common female names.
- A file called `maleNames.txt`, containing a list of common male names.
- Finally an XML file called `constraints.xml` describes specific co-reference constraints applied in modern Greek language. This file contains a set of `rule` XML elements.

All these files have been placed in the `Resources` directory.

A detailed description is given in the following tables concerning the existence of the aforementioned files and the number of entries in the Greek version of RARE.

<code>1.copulativeVerbs.txt</code>	yes	39
<code>2.determiners.xml</code>	yes	22



3.familyNames.txt	yes	454
4.femaleNames.txt	yes	630
5.maleNames.txt	yes	927
# of rules	certifying	-
	demolishing	1
	promoting	13
7.stopwords.txt		59

Table 33-EL: RARE localisation files

The Greek corpus is relatively wide in terms of coverage; it mainly contains different text categories such as novels, political, financial and geographical texts, as well as law texts that have been manually annotated. The summarisation of short texts of Modern Greek Language draws on the whole language processing chain and in addition, makes use of specific modules such as: clause splitting, anaphora resolution, discourse parsing and summarization. The strategy for short texts is based on identification of the discourse structure and produces a summary that can be classified as a type of excerpt, thus it is possible to indicate the length of the summary as a percentage of the original text.

Summarisation tools of Modern Greek Language consider methods such as, anaphora resolution, clause splitting and discourse parsing method. Depending on the difficulty of the task, some annotation activities have used specialised tools such as PALinka.

Concerning Clause Splitting (CS) and Discourse Parsing (DP), these have been applied in 10 short texts. For the CS and DP of these texts a number of statistical data is quoted below:

- File name
- Number of sentences
- Number of tokens
- Number of clauses
- Number of annotated occurrences of markers in the corpus (<MARKER></MARKER> parenthesis)
- Number of distinct lexicals acting as markers in the corpus
- Number of `int` markers (occurrences)
- Number of `ext` markers (occurrences), out of which:
 - Number of `CONNECT="expect"` `ext` markers (occurrences)
 - Number of `CONNECT="fulfil"` `ext` markers (occurrences)
 - Number of `CONNECT="relate"` `ext` markers (occurrences)



- Existence of the markers_type.xml file (yes, no), if yes then # of entriesTotals for each of the numeric fields
- General total

Remark for the Greek Language:

External and internal markers bind different periods or sentences in Modern Greek Language. The most important external markers are the following: “ωστόσο, έπειτα, μολαταύτα, και, αλλά, αε άλλα λόγια,αντίθετα, άρα, επομένως, ώστε, τώρα, τότε, έπειτα, ύστερα., για αυτό, δηλαδή, θέλω να πω, συμπερασματικά, πράγματι, παρόλα αυτά”. Moreover, the morphology of Modern Greek language is concerned with the internal binding of clauses through internal markers such as: “Και, όπου, αλλά, ο οποίος, η οποίος, το οποίο, όταν, όπως, ενώ, αν, εκτός αν, μέχρι να, παρόλο που, είτε”. Greek language has a relatively free word order. The same sentence may be expressed in several different word orders. Therefore, markers play a significant role in the structure and coherence of Greek sentences.

A numeric overview of this manually annotated corpus is provided in the following table.



File name:	Grimm Brothers "Sleeping Beauty"	Grimm Brothers "The Queen Bee"	Lovercraft "Polaris"	Lovercraft "The Other Gods"	Edgar Allan Poe "Manuscript in a Bottle"	Lovercraft "The White Ship"	Raptopoulos "The Friends"	Political Article 1	Political Article 2	Financial Article 1	Financial Article 2	Financial Article 3	Geographical Article 1	Geographical Article 2	Law Text (100 sentences)	Science Article 1	Science Article 2	Total	
Number of sentences	39	29	56	84	42	83	73	68	42	42	46	26	79	73	102	86	85	1055	
Number of tokens	1452	886	1589	1599	985	2412	1622	2006	1018	1885	1796	664	1486	1730	4924	1895	2258	30207	
Number of clauses	189	187	165	181	91	226	148	140	93	137	157	52	154	166	302	186	204	2778	
Number of markers	123	101	103	121	32	129	74	76	57	77	89	25	58	84	154	87	103	1493	
Number of lexicals																			
Number of int markers	110	75	93	105	32	116	62	66	47	63	85	20	49	76	149	74	98	1320	
Number of ext markers	"expect"	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	
	"fulfill"	-	-	-	4	-	3	5	-	2	5	2	3	1	-	1	4	-	30
	"relate"	11	12	10	12	5	10	8	10	8	9	2	2	8	8	4	9	5	133
markers_type.xml	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	
Evaluation	P	0.903																	
	R	0.842																	
	F	0.871																	

Table 34-EL: the CS and DP corpus for Greek



6.5 Polish

- 1 use of a localised version of RARE – yes (see Table 36-PL)
- 2 existence of an AR annotated corpus – no
- 3 evaluation of the AR module – no

copulativeVerbs.txt	Yes, 1 entry	
determiners.xml	Yes, 2 entries	
familyNames.txt	Yes, 399529 entries	
femaleNames.txt	Yes, 337 entries	
maleNames.txt	Yes, 416 entries	
# of rules	certifying	0
	demolishing	0
	promoting	10
Other resources (please specify)		stopwords.txt, 37 entries tagset.xml

Table 36-PL: RARE localisation files

File number	# of sentences	# of tokens	# of clauses	# of markers	# of lexicals	# of int markers	# of ext markers		Evaluation		
							expect	fulfil	P	R	F
1	120	1782	252	-		-	-	-			
2	96	1997	293	-		-	-	-			
3	79	1229	165	-		-	-	-			
4	45	713	110	-		-	-	-			
5	112	1842	263	-		-	-	-			
6	120	1888	250	104		98	0	0			
7	50	859	100	35		33	0	0			
8	121	1758	241	68		67	0	0			
9	99	2098	236	135		116	1	1			
10	34	695	108	63		58	0	0			
11	74	1520	189	99		92	0	0			
12	103	3857	188	88		80	0	1			
13	43	1139	179	106		99	0	0			
Total	1096	21377	2574	698		643	1	2	0.890	0.973	0.821

Table 37-PL: The CS and DP corpus for Polish



Filenames:

1. onet_pl_Spisek_przeciw_konsumentom_PL.txt
2. paradowska_Katastrofa_PL.txt
3. onet_pl_Wyspa_kary_dla_ladacznic_PL.txt
4. polityka_pl_Tym_PL.txt
5. polityka_pl_Chomikuj_PL.txt
6. onet_pl_Mamo_tato_zostaje_z_Wami_PL.txt
7. onet_pl_Katastrofa_kolejowa_PL.txt
8. onet_pl_Rugbysci_jedza_zmarlych_PL.txt
9. Allan_Edgar_Poe_-_The_Masque_of_the_Red_Death_PL.txt
10. Grimm_Brothers_-_The_Queen_Bee_PL.txt
11. Arthur_C_Clarke_-_If_I_forget_thee,_oh_Earth_PL.txt
12. 100_sents_from_32011R0305_PL.txt
13. Grimm_Brothers_-_Sleeping_Beauty_PL.txt

File name	1	2	3	4	5	6	7	Total	
# of sentences	120	50	121	99	34	74	43	541	
# of clauses	250	100	241	236	108	189	179	1303	
20% summary	35,30, 46,32*	16,17, 17,20	44,41, 44,42	48,44, 47,47	19,22, 24,21	32,35, 36,35	35,33, 33,33	229	
10% summary	17,13, 22,13	8, 8, 7, 11	18,22, 21,22	24,21, 24,23	10,10, 11,12	16,18, 19,17	19,17, 15,15		
5% summary	7, 7, 12, 6	4, 4, 6, 5	11,11, 9, 12	10,10, 11,11	5, 5, 6, 3	8, 8, 9, 8	11, 8, 8, 8		
Evaluation	P	0.175	0.105	0.157	0.211	0.125	0.258	0.164	0.171
	R	0.376	0.223	0.330	0.449	0.287	0.53	0.336	0.361
	F	0.239	0.143	0.213	0.287	0.174	0.347	0.220	0.232

Table 38-PL: The summaries corpus for Polish

* The cells including 4 values denote the numer of clauses as left by 4 different annotators. The evaluation has been done against the longest gold summary in these cases.

Filenames:

1. onet_pl_Mamo_tato_zostaje_z_Wami_PL.txt
2. onet_pl_Katastrofa_kolejowa_PL.txt
3. onet_pl_Rugbysci_jedza_zmarlych_PL.txt
4. Allan_Edgar_Poe_-_The_Masque_of_the_Red_Death_PL.txt
5. Grimm_Brothers_-_The_Queen_Bee_PL.txt
6. Arthur_C_Clarke_-_If_I_forget_thee,_oh_Earth_PL.txt
7. Grimm_Brothers_-_Sleeping_Beauty_PL.txt



6.6 Romanian

1.copulativeVerbs.txt	yes	11
2.determiners.xml	yes	4
3.familyNames.txt	yes	41732
4.femaleNames.txt	yes	4798
5.maleNames.txt	yes	1542
# of rules	certifying	2
	demolishing	2
	promoting	9
Other resources (please specify)		N/A

Table 39-RO: RARE localisation files for Romanian

File name	Bun simt	Cetateanul si literatura lui	Clasicii II	Clasicii I	Cocorii	Comediile DN Caragiale	Conspiratiile esentiale	Dropi	Feminism I	Feminism II	Geni agitatie	Geniul neprihanit	Geniul organizator	Huhurezi barbosii	
# of sentences	49	53	26	33	51	40	28	55	29	33	45	32	48	38796	
# of tokens	1444	1093	712	1019	1973	1393	1046	1075	891	1210	1162	1128	1318	85	
# of clauses	108	131	54	61	119	125	93	140	65	70	96	78	95	18	
# of markers	40	55	17	23	26	28	28	20	21	21	20	25	32		
# of lexicals														14	
# of int markers	28	46	11	15	23	25	16	16	15	12	16	21	19	2	
# of ext markers	"expect"	2	3	2	1	0	1	0	0	1	2	3	0	3	2
	"fulfil"	3	2	0	1	1	0	3	0	2	1	0	1	2	0
	"relate"	7	4	4	6	2	2	1	4	3	6	1	3	8	2

Table 40-RO: The CS and DP corpus for Romanian



File name	Influențe si curente I	Influențe si curente II	Influențe si curente III	Intelectualul	Introducere I	Introducere II	In procesul intelectualilor	Irlanda subacvatică	Libertatea gustului	Literatură populară I	Literatură populară II	Mecanizarea scrisului	Minul povestitor I	Minul povestitor II	
# of sentences	34	26	27	57	35	51	48	33	48	30	40	25	25	50	
# of tokens	902	839	869	1734	112	1316	1128	927	1373	986	1177	967	967	1509	
# of clauses	90	90	78	157	100	117	70	79	123	81	90	81	74	112	
# of markers	36	21	0	20	31	24	24	12	33	32	24	21	25	11	
# of lexicals															
# of int markers	27	19	0	18	27	22	13	10	22	30	20	18	8	18	
# of ext markers	"expect"	2	1	0	0	1	0	1	2	2	0	1	1	2	2
	"fulfil"	1	0	0	0	1	0	0	0	2	0	2	1	0	1
	"relate"	6	1	0	2	2	2	10	0	6	2	1	1	1	4

Table 41-RO: The CS and DP corpus for Romanian (continued)

Table 42-RO: The CS and DP corpus for Romanian (continued)

File name	Motivul scriiturii	Neintelegeri între public și artiști	Ocnas	Paradisul impreciziei	Perioada critică II	Poeti și critici II	Prefața	Prefața zăfăpol	Romanul din minulescu	Stil clasic I	Stil clasic II	Tainele artei	toleranta	Totals
# of sentences	32	47	28	36	39	44	33	44	35	30	40	37	39	1571
# of tokens	975	1307	815	962	1465	1155	1250	1175	1015	885	1223	982	1028	47016
# of clauses	85	95	79	76	99	84	93	93	67	50	79	78	80	3720
# of markers	24	27	14	16	39	18	19	23	11	11	22	12	23	947
# of lexicals														
# of int markers	20	21	12	13	36	14	17	20	10	8	13	8	16	745
# of ext markers	"expect"	1	2	0	1	0	1	2	1	0	1	2	1	48
	"fulfil"	1	0	0	0	0	0	0	0	0	1	0	2	28
	"relate"	2	4	2	2	3	3	0	2	1	6	3	4	125
markers_type.xml	87													
Evaluation	P	0.9071												
	R	0.8819												
	F	0.8943												



File name	Arthur C. Clarke Daca te voi uita vreodata a Pamant	Fratii Grimm Frumoasa adormita	Fratii Grimm regina albinelor	Franz Kafka unsper eze fecioiri	Lovecraft zeii ceilalti	Edgar Allan Poe Masca mortii rosii	Edgar Allan Poe Un vis	Total	
# of sentences	70	55	57	74	94	110	66	526	
# of clauses	176	212	144	201	232	224	128	1317	
20% summary	40	43	29	40	46	30	26	254	
10% summary									
5% summary									
Evaluation	P	0.302	0.243	0.241	0.160	0.221	0.198	0.157	0.217
	R	0.434	0.406	0.175	0.169	0.323	0.531	0.179	0.316
	F	0.356	0.304	0.203	0.164	0.262	0.288	0.167	0.249

Table 43-RO: The summaries corpus for Romanian



7 CONCLUSIONS

7.1 *Critical analysis of the ATLAS Summarisation system*

The quality of the summaries depends on many factors. We have identified the following types of errors:

- Type 1 errors – due to the quality of the prerequisite chains (correctness of the annotation received in input by the proper summarisation chain – see Section 1.3);
- Type 2 errors – due to a technical defect in some component module of the proper summarisation chain;
- Type 3 errors – due to the quantity and quality of the corpora used for training different modules in different languages;
- Type 4 errors – due to the improper fixing of the parameters of different modules.

In the following we will give examples from all these types of errors.

Analysis of Type 1 Errors

The example below displays a POS error in the RO processing chain: *merse* is tagged as an adjective instead of a verb and this may induce clause segmentations errors.

```
<W ID="67.13" LEMMA="merge" OFFSET="82" POS="Afpfprn">merse</W>
```

Analysis of Type 2 Errors

The example below shows an anaphora resolution error: RARE is not able to find the correct antecedent of an anaphor. The error is induced by the rules in the constraints file.



```

<NP ID="1" DBID="0" HEADID="2">
  <W ID="1" DBID="0" LEMMA="ο" POS="article"
MSD="article:sin:nom:m:def" NUM="singular" GEN="masculine">ο</W>
  <W ID="2" DBID="0" LEMMA="πρίγκιπας" POS="noun"
MSD="noun:sin:nom:m" NUM="singular" GEN="masculine">πρίγκιπας</W>
</NP>
....
<NP ID="0" DBID="0" HEADID="10">
  <W ID="10" DBID="0" LEMMA="εκείνος" POS="pronoun"
MSD="pronoun:sin:nom:m" NUM="singular" GEN="masculine">εκείνος</W>
</NP>
<W ID="11" DBID="0" LEMMA="μετατρέπομαι" POS="verb"
MSD="verb:sin:pas:pas" NUM="singular">μετατράπηκε</W>
<W ID="12" DBID="0" LEMMA="σε" POS="preposition"
MSD="preposition">σε</W>
<W ID="13" DBID="0" LEMMA="φράχτη" POS="adjective"
MSD="adjective:sin:acc:m" NUM="singular" GEN="masculine">φράχτη</W>
....
<NP ID="4" DBID="0" HEADID="28">
  <W ID="28" DBID="0" LEMMA="αυτός" POS="pronoun"
MSD="pronoun:sin:nom:m" NUM="singular" GEN="masculine">αυτός</W>
</NP>

```

In the above example, the correct reference chain should be:

“NP-ID-1 πρίγκιπας, NP-ID-4 αυτός”.

However, due to the rule “*numberGenderPersAgreementPronoun*” (a pronoun matches a previous pronoun if they match in Number and Gender), the pronouns “εκείνος” and “αυτός” were matched. As a consequence, the RARE engine produced the wrong reference chain

reList=" NP-ID-1 πρίγκιπας, NP-ID-0 εκείνος, NP-ID-7 αυτός”.

Obviously, the correct correlation should have been between the pronoun «εκείνος» and the noun «φράχτης».

Figure 44: RARE error caused by a matching of 2 pronouns

Analysis of Type 3 Errors

The example below shows a clause segmentation error, which should be caused by insufficient data in the training corpus:



```

<CLAUSE ID="12" TYPE="COMB_">
  <W ID="W99" LEMMA="and" POS="CC" NUM="" MSD="CC">and</W>
  <NP ID="12" HEADID="W101" ref="0">
    <W ID="W100" LEMMA="the" POS="DET" NUM="" MSD="DT">the</W>
    <W ID="W101" LEMMA="queen" POS="N" NUM="SG" PERSONTYPE="FEMALE_PERSON"
MSD="DT">queen</W>
  </NP>
  <W ID="W102" LEMMA="bear" POS="V" NUM="" MSD="VBD">bore</W>
  <NP ID="13" HEADID="W104" ref="8">
    <W ID="W103" LEMMA="a" POS="DET" NUM="" MSD="DT">a</W>
    <W ID="W104" LEMMA="daughter" POS="N" NUM="SG"
PERSONTYPE="FEMALE_PERSON" MSD="DT">daughter</W>
  </NP>
</CLAUSE>
<CLAUSE ID="13" TYPE="MARKER_">
  <W ID="W105" LEMMA="so" POS="ADV" NUM="" MSD="RB">so</W>
  <W ID="W106" LEMMA="beautiful" POS="A" NUM="" MSD="JJ">beautiful</W>
  <W ID="W107" LEMMA="that" POS="PREP" NUM="SG" MSD="IN">that</W>
  <NP ID="14" HEADID="W109" ref="11">
    <W ID="W108" LEMMA="the" POS="DET" NUM="" MSD="DT">the</W>
    <W ID="W109" LEMMA="king" POS="N" NUM="SG" PERSONTYPE="MALE_PERSON"
MSD="DT">king</W>
  </NP>
  <W ID="W110" LEMMA="can" POS="V" NUM="" MSD="MD">could</W>
  <W ID="W111" LEMMA="not" POS="ADV" NUM="" MSD="RB">not</W>
  <W ID="W112" LEMMA="contain" POS="V" NUM="" MSD="VB">contain</W>
  <W ID="W113" LEMMA="himself" POS="PRON" NUM="SG" MSD="PRP">himself</W>
  <W ID="W114" LEMMA="for" POS="PREP" NUM="" MSD="IN">for</W>
  <W ID="W115" LEMMA="joy" POS="N" NUM="SG" PERSONTYPE="OBJECT"
MSD="NN">joy</W>
  <W ID="W116" LEMMA="," POS="," NUM="" MSD=",">,</W>
</CLAUSE>

```

Figure 45: Segmenter error

As can be seen, the segmentation between the clauses 12 and 13 was marked before the group *so beautiful*, this being improperly attached to the second clause.

Analysis of Type 4 Errors

In principle, the richer the set of parameters, the finer a calibration process can be designed and therefore, the did not get their values as a result of a calibration process, but rather have been fixed manually, are those corresponding to the Maximum Referentiality heuristic (in Section 3.3.10).



7.2 *Further improvements and developments*

The actual tests have shown that, in general, the system produces useful summaries. However, till the end of the project we envisage a number of enhancements, that are mentioned below.

A set of **improvements intended to overcome Type 2 Errors:**

- Do CS before AR and include a demolishing rule that does not allow to co-refer pronouns in different persons if they are in the same clause. The actual constraints definition language does not allow this kind of constraint. The same happens in the case of a rule that would have to interdict coreference between two nested REs, and which has now been cabled in the RARE code. Difficulty: high; placing CS before AR is straightforward, but modifying the constraints definition language of RARE and its interpreter is difficult;

- At present, pairs of verbs are classified as composed or not based on a fixed set of features (which do not depend on the language and have been inherited for all languages after a training organised with the English clauses-and-markers corpus) – see Section 3.2.3. This is clear a source of errors for languages other than English, which could be avoided if a training would be organised based on language specific corpora. Difficulty: low, because corpora exist;

Improvements intended to overcome Type 3 Errors:

- It is clear that much more gold material is needed at each level of processing in order to improve the quality of the component parts of the summariser. Difficulty: high, because it is time consuming and costly.

- The editing tool (presented in Section 4.1) can also be used to improve the quality of the resources used by the CS and the DP modules. Difficulty: medium. Although the editing tool is very handy, a proper editing supposes a rather high competence, and much practice.

Improvements intended to overcome Type 4 Errors:

- The problem mentioned above in connection with the Maximum Referentiality heuristic of Section 3.3.10 will be solved in the next version of the DP module by including in the Calibration cycle also the parameters connected to the Maximum Referentiality heuristic (see Section 3.3.10).

**Further improvements:**

- The DP module evaluates thousands of trees, before establishing which is the best candidate for a discourse structure of a text. This process is extremely time consuming and the DP module uses a multi-threading launch of the evaluation procedure in order to speed up the computations. A great part of this process involves the computation of veins. However vein expressions could be inferred without being computed exhaustively, by applying a set of formulas, which uses the vein expressions already existent in the developing tree and takes into consideration the place of the adjunction/substitution on the RF and other parameters. Difficulty: high, because it involves a number of theoretical developments, which are now under scrutiny.

- Resolving dangling pronouns in the summaries: due to the use of veins at the base of summaries, the appearance of dangling pronouns should be extremely rare. However, if this still happens, the summarisation module can solve this by replacing it with an antecedent which is not a pronoun. Difficulty: low. It presupposes a simple replace in the output string of the pronoun with an antecedent non-pronoun string taken from its chain of corefering expressions.

- Fixing the punctuation. As can be noticed, the summaries present some punctuation problems. For instance, a final dot is missing if the dot belonged to a clause that has been abandoned in the summary.

- The present version of the ATLAS summariser works only on short texts. A major development will refer to extending the ATLAS summarisation system to long texts. Difficulty: very high, as a completely different method should be thought. UAIC has started to investigate methods for this extension and a theoretical solution has been proposed.

REFERENCES

- Carlson L., Marcu D., Okurowski M. (2001). Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In *Proceedings of the 2nd SIGDIAL Workshop on Discourse and Dialogue*, Denmark.
- Cristea,D., Dima,G.E. (2001). [An integrating framework for anaphora resolution](#). In *Information Science and Technology*, Romanian Academy Publishing House, Bucharest, vol. 4, no. 3-4, p 273-291.
- Cristea D., Ide N., Romary L. (1998). Veins Theory: A Model of Global Discourse Cohesion and Coherence. In *Proceedings of the 17th international conference on Computational Linguistics*.



- Dan Cristea, Nancy Ide, Daniel Marcu, Valentin Tablan (1999). Discourse Structure and Co-Reference: An Empirical Study, In *Proceedings of the Workshop on the Relation Between Discourse Structure and Reference*, Maryland, June 1999, workshop in conjunction with *The 37th Annual Meeting of the Association for Computational Linguistics - ACL'99*.
- Cristea, D., Postolache, O., Pistol, I. (2005). Summarisation through Discourse Structure. In Alexander Gelbukh (Ed.): *Computational Linguistics and Intelligent Text Processing, 6th International Conference CICLing 2005*, Mexico City, Mexico, February 2005, Proceedings, Springer LNSC, vol. 3406, ISBN 3-540-24523-5, pp. 632-644.
- Cristea, D., Webber, B. L. (1997). Expectations in Incremental Discourse Processing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, Madrid
- Davis T. (2010). Catalan Numbers, <http://www.geometer.org/mathcircles/catalan.pdf>
- Feng, V.W., Hirst, G. (2012). Text-level Discourse Parsing with Rich Linguistic Features. In *Proceedings of ACL-2012*.
- Grosz, B.J. Joshi, A.K. and Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 12(2), 203-225.
- Grosz, B. and Sidner, C. (1986). Attention, Intention and the Structure of Discourse. *Computational Linguistics*, 12, 175-204.
- Hernault, H., Prendinger, H., duVerle, D.A. and Ishizuka, M (2010). HILDA: A discourse parser using support vector machine classification. *Dialogue and Discourse*, 1(3):1–33.
- Ide, N., Cristea, D. (2000). A Hierarchical Account of Referential Accessibility. In *Proceedings of The 38th Annual Meeting of the Association for Computational Linguistics*, ACL'2000, Hong Kong, China.
- Joshi A., Schabes Y. (1997). Tree-Adjoining Grammars, in G. Rozenberg and A.Salomaa, editors, *Handbook of Formal languages*.
- Leffa V. (1988). Clause processing in complex sentences. In *Proceedings of the First International Conference on Language Resource and Evaluation*, volume 1, pages 937 – 943, May 1998.
- Mann, W.C., Thompson, S.A. (1988). Rhetorical Structure Theory: A theory of Text Organization. *Text*, 8:3, 243–281.
- Marcu, D. (2000). *The Theory and Practice of Discourse Parsing and Summarization*. The MIT Press.
- Nguyen, V.V., Nguyen, M.L., Shimazu, A. (2009). Clause Splitting with Conditional Random Fields, *Information and Media Technologies* 4(1): 57-75, reprinted from: *Journal of Natural Language Processing* 16(1): 47-65 (2009) © The Association for Natural Language Processing.
- Orăsan, C. (2000). A hybrid method for clause splitting in unrestricted English texts. In *Proceedings of ACIDCA'2000*, Monastir, Tunisia.
- Orăsan, C., Cristea, D., Mitkov, R., Branco, A. (2008). Anaphora Resolution Exercise – An Overview. In *Proceedings of LREC-2008*, Marrakech, Morocco.
- Parveen D., Sanyal R., Ansari A. (2011). Clause Boundary Identification using Classifier and



- Clause Markers in Urdu Language, *Polibits Research Journal on Computer Science*, 43, pp. 61-65.
- Puşcaşu G. (2004). A Multilingual Method for Clause Splitting. In *Proceedings of the 7th Annual Colloquium for the UK Special Interest Group for Computational Linguistics (CLUK 2004)*, Birmingham, UK.
- Subba R., Di Eugenio B. (2007). Automatic discourse segmentation using neural networks. In *Proceedings of 11th Workshop on the Semantics and Pragmatics of Dialogue*, Trento, Italy.
- Şoricuţ, R. and Marcu, D. (2003). Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, Volume 1, pages 149–156.



ANNEX I: EXAMPLE OF RULES TO BE USED BY RARE FOR ENGLISH

```

<?xml version="1.0" encoding="UTF-8"?>
<document lan="EN">
  <rule ID="numberGenderPersAgreementPronoun" type="promoting"
score="4.5">
    <constraint on="RE" attr="POS" val="PRON"/>
    <constraint on="DE" attr="POS" val="PRON"/>
    <relation pred="equal" attrList="NUM GEN PERS"/>
    <!-- A pronoun will corefer (with a score of 4.5) a previous pronoun
if they match in NUM AND GEN AND PERS. -->
  </rule>

  <rule ID="NumGenPersNotMatchPronPron" type="demolishing">
    <constraint on="RE" attr="POS" val="PRON"/>
    <constraint on="DE" attr="POS" val="PRON"/>
    <relation pred="notEqual" attrList="NUM GEN PERS"/>
    <!-- A pronoun cannot corefer a previous pronoun if they do not
match in one of the attributes NUM OR GEN OR PERS. -->
    <!-- If one of the attributes NUM GEN PERS does not exist in one of
RE/DE, the test passes to the next. -->
    <!-- The test succeeds (and the demolishing rule applies) if one of
the attributes NUM GEN PERS in RE does not match with its values in DE. -->
  </rule>

  <rule ID="NumGenPersNotMatchPronPron" type="certifying">
    <constraint on="RE" attr="MSD" val="Np*"/>
    <constraint on="DE" attr="MSD" val="Np*"/>
    <relation pred="equal" attrList="LEMMA"/>
    <!-- Two proper nouns having equal lemmas should be coreferential. -
->
  </rule>

  <rule ID="PronounNounNumberAgreement" type="promoting" score="4">
    <constraint on="RE" attr="POS" val="PRON"/>
    <constraint on="DE" attr="POS" val="N"/>
    <relation pred="equal" attrList="NUM"/>
    <!-- A pronoun will corefer (with a score of 4.) a previous noun if
they match in NUM. -->
  </rule>

  <rule ID="NounPronounNumberAgreement" type="promoting" score="3.5">
    <constraint on="RE" attr="POS" val="N"/>

```



```

    <constraint on="DE" attr="POS" val="PRON"/>
    <relation pred="equal" attrList="NUM"/>
    <!-- The reverse rule: a noun will corefer a previous pronoun if
they match in NUM (according to some authors a noun referring a pronoun is
not as frequent as in the other direction). -->
  </rule>

<rule ID="LemmaEqualityNounNoun" type="promoting" score="5">
  <constraint on="RE" attr="POS" val="N"/>
  <constraint on="DE" attr="POS" val="N"/>
  <relation pred="lemmaEquality" attrList="LEMMA NUM"/>
  <!-- A noun will corefer a previous noun if they match in LEMMA AND
NUM. -->
</rule>

<rule ID="NumFemaleAgreementPronNoun" type="promoting" score="4.5">
  <constraint on="RE" attr="POS" val="PRON" />
  <constraint on="RE" attr="GEN" val="FEMALE" />
  <constraint on="DE" attr="POS" val="N" />
  <constraint on="DE" attr="PERSONTYPE" val="FEMALE-NAME" />
  <relation pred="equal" attrList="NUM"/>
  <!-- A feminine pronoun will corefer a previous feminine noun name
if they match in NUM. -->
</rule>

<rule ID="NumMaleAgreementPronNoun" type="promoting" score="4.5">
  <constraint on="RE" attr="POS" val="PRON" />
  <constraint on="RE" attr="GEN" val="MALE" />
  <constraint on="DE" attr="POS" val="N" />
  <constraint on="DE" attr="PERSONTYPE" val="MALE-NAME" />
  <relation pred="equal" attrList="NUM"/>
  <!-- A masculine pronoun will corefer a previous masculine noun name
if they match in NUM. -->
</rule>

<rule ID="ItIsObject" type="promoting" score="4.0">
  <constraint on="RE" attr="POS" val="PRON"/>
  <constraint on="RE" attr="LEMMA" valList="it its itself"/>
  <constraint on="DE" attr="PERSONTYPE" val="OBJECT"/>
  <relation pred="equal" attrList="NUM"/>
  <!-- An "it" pronoun will corefer a previous object type noun if
they match in NUM. -->
</rule>

<rule ID="ThatIsSingular" type="promoting" score="3.0">
  <constraint on="RE" attr="POS" val="PRONREL"/>
  <constraint on="RE" attr="LEMMA" valList="that"/>
  <constraint on="DE" attr="POS" val="N" />

```



```
<relation pred="equal" attrList="NUM"/>
<!-- A "that" relative pronoun will corefer a previous noun if they
match in NUM. -->
</rule>

<rule ID="SheIsShip" type="promoting" score="3.0">
  <constraint on="RE" attr="POS" val="PRON"/>
  <constraint on="RE" attr="GEN" val="FEMALE"/>
  <constraint on="DE" attr="LEMMA" valList="ship vessel destroyer"/>
  <relation pred="equal" attrList="NUM"/>
  <!-- A "she" pronoun could refer a "ship" lemma if they match in
NUM. -->
</rule>

<rule ID="ItIsBabySing" type="promoting" score="3.0">
  <constraint on="RE" attr="POS" val="PRON"/>
  <constraint on="RE" attr="GEN" valList="NEUTER"/>
  <constraint on="DE" attr="PERSONTYPE" val="PERSON"/>
  <constraint on="DE" attr="LEMMA" valList="kid baby child"/>
  <relation pred="equal" attrList="NUM"/>
  <!-- An "it" pronoun could refer a "kid" lemma if they match in NUM.
-->
</rule>

<rule ID="antecedentCollectiveNoun" type="promoting" score="2.5">
  <constraint on="RE" attr="POS" val="PRON"/>
  <constraint on="RE" attr="NUM" val="PL"/>
  <constraint on="RE" attr="PER" val="3"/>
  <constraint on="DE" attr="LEMMA" valList="government parliament
senate party gathering group commission army patrol"/>
  <constraint on="DE" attr="NUM" valList="SG"/>
  <!-- A collective SG noun (e.g. "goverment") could be refered by a
plural pronoun in EN. -->
</rule>

<rule ID="antecedentExquisitePerson" type="promoting" score="2.5">
  <constraint on="RE" attr="POS" val="PRON"/>
  <constraint on="RE" attr="NUM" val="PL"/>
  <constraint on="RE" attr="PER" val="1"/>
  <constraint on="DE" attr="POS" val="N"/>
  <constraint on="DE" attr="LEMMA" valList="king queen emperor"/>
  <constraint on="DE" attr="NUM" valList="SG"/>
  <!-- The pronoun "we" could refer an exquisite person in SG. -->
</rule>

<rule ID="ExquisitePersonAntecedentPronoun" type="promoting"
score="2.5">
  <constraint on="RE" attr="POS" val="N"/>
```




```
<constraint on="RE" attr="LEMMA" valList="king queen emperor"/>
<constraint on="RE" attr="NUM" valList="SG"/>
<constraint on="DE" attr="POS" val="PRON"/>
<constraint on="DE" attr="NUM" val="PL"/>
<constraint on="DE" attr="PER" val="1"/>
<!-- The reverse rule of above: "We, the king, ..." -->
</rule>
</document>
```



ANNEX II: SUMMARIES OF SHORT TEXTS

Bulgarian

1) Original text:

МАСКАТА НА АЛЕНАТА СМЪРТ

Дълго време «Алената смърт» опустошавала страната.

Нито една епидемия дотогава не била толкова страшна и толкова гибелна.

Кръвта била нейният Аватар, а нейният печат — червенината и ужасът на кръвта.

Остри болки, внезапно виене на свят, обилен кръвоизлив от всички пори и смърт.

Алените петна по тялото и особено по лицето на жертвата представлявали проклятието на чумата, която лишавала осъдения от помощта и съчувствието на неговите близки.

И всичко траяло само половин час: заразата, развоят и развързката на болестта.

Ала принц Просперо бил жизнерадостен, неустрашим и умен.

Когато народът в неговите владения наполовина измрял, той призовал хиляда здрави и жизнерадостни приятели измежду рицарите и дамите на своя двор и се оттеглил заедно с тях в дълбокото уединение на едно от укрепените си и построено като замък абатство.

Просторен и величествен замък, изграден според причудливия и все пак царствен вкус на принца.

Яка и висока стена го опасвала.

В стената имало железни врати.

След като влезли в абатството, придворните докарали горнило и тежки чукове и занитили вратите.

Решили в случай на неочаквани пристъпи на отчаяние или безумие да не оставят никакви пътища нито за влизане, нито за излизане.

Абатството снабдили с пребогат запас провизии.

С подобни предпазни мерки придворните можели вече смело да хвърлят предизвикателство на заразата.



Външният свят трябвало вече да се погрижи сам за себе си.

А междуременно да тъгуваш или да се отдаваш на размисли, било истинско неблагоразумие.

Принцът се погрижил за всякакви удоволствия.

Имало шутове, имало импровизатори, имало танцьорки, имало музиканти, имало Красота, имало каквото поискаш.

И всичко това заедно с безопасността било тук, вътре.

Отвън останала само «Алената смърт».

Към края на петия или шести месец от уединението, когато чумата вилнеела най-свирепо отвън, за да развесели своите хиляди приятели, принц Просперо им устроил маскараден бал с невидано великолепиe.

Разгулна гледка представлявал този маскарад.

Но нека първо да ви разкажа за балните зали.

Седем на брой — същински царствени покои.

В много дворци обаче тези покои имат вид на дълга и права зала, а когато крилата на плъзгащите се врати се дръпнат до самите стени, погледът не среща почти никакво препятствие.

Но тук случаят бил съвършено различен.

Както и можело да се очаква от пристрастието на принца към bizzare.

Залите били толкова неправилно разположени, че погледът обгръщал наведнъж само една от тях.

На всеки двадесет-тридесет метра имало остър завой, а зад всеки завой — нов ефект.

По средата на всяка стена, отляво и отдясно, към закрит проход гледал висок и тесен готически прозорец, който следвал криволиците на самите зали.

Прозорците били от цветни стъкла, чиито краски се менели в зависимост от тона на подредбата в залата, към която се разтваряли.

Така например в най-източната преобладавало синьото – и прозорците били в яркосиньо.

Във втората зала украсата и gobлените били пурпурни.

Третата била цялата в зелено.

Четвъртата била мебелирана и греела в оранжево, петата – в бяло, шестата – във виолетово.



Седмата зала била плътно обвита в черни кадифени драперии, които покривали целия таван и всички стени и се спускали на тежки дипли върху килим от същата материя и в същия цвят.

Единствено в тази зала цветът на прозорците не съответствувал на украсата.

Стъклата били алени, кървавочервени.

Сред изобилието на златни украшения, пръснати тук и там или провесени от тавана, в нито една от седемте зали нямало ни лампа, ни канделабър.

Ни лъч от лампа или от свещ в целия низ от зали.

Ала в коридорите, край залите, срещу всеки прозорец имало по тежка тринога поставка с пламенееща жаравя – тя хвърляла отблясъци през многоцветните стъкла и по такъв начин озарявала ярко залата.

И създавала безброй пъстри и фантастични сенки.

В най-западната или черната зала въздействието от отблясъците на пламенеещата жаравя, падащо през кърваво обагрените стъкла върху черните завеси, било крайно кошмарно и извиквало по лицата на гостите такова безумно изражение, че малцина от тях намирали достатъчно смелост да прекратят нейния праг.

В тази зала, на западната страна, висял окачен гигантски абаносов часовник.

Махалото се люлеело наляво-надясно с унило, сънно, монотонно тиктакане; когато стрелката на минутите склучела кръга край циферблата и часовникът трябвало да удари, от медните му дробове се изтръгвал ясен, силен, дълбок и извънредно мелодичен звън, но с такъв особен тон и удар, че при изтичането на всеки час музикантите от оркестъра се принуждавали да спрат за момент изпълнението си и да се вслушат в звъна, а валсиращите прекъсвали по неволя танца и за миг във веселите гости се прокрадвал смут; и докато напевният звън на часовника отеквал, най-лекомислените пребледнявали, а по-улегналите и по-сдържаните прокаравали длани по челата си, сякаш някой ги е изтръгнал от унеса и размислите им.

Загълхнели ли отзвучите, весел смях пак заливал внезапно цялата компания; музикантите се споглеждали и се усмихвали като че на собствената си възбуда и безразсъдство и се кълнели шепкешком един на друг, че следващият звън от часовника няма да предизвика подобни чувства; а после, след изтичането на шестдесетте минути (заклучили в себе си три хиляди и шестстотин секунди от летящото Време), от часовника отново се разнасял звън и отново, както и преди, настъпвало смущение, трепет и размисъл.

Независимо от всичко пиршеството било весело и пишно.

Вкусвете на принца били особени.

Той имал око за багрите и за тяхното въздействие.

Не признавал обикновената decora.

Дръзки и пламенни били плановете му, в замислите му искрял див блясък.

Други биха го взели за луд.

Приближените му обаче знаели, че не е.



Трябвало да го чуете, да го видите, за да се уверите, че наистина не е такъв.

Седемте зали били подредени в значителна степен по негово разпореждане по случай големия фете, а неговият вкус подсказал кой участник с каква маска да се маскира.

И естествено всички били гротески.

Имало много великолепиe и блясък, много пикантност и фантазия, много повече от онова, което сме виждали по-късно в «Ернани».

Арабески – фигури с несъответстващи и по предназначение, и по място крайници.

Безумни капризи, породени като че от разстроено въображение.

Много красота, много поквара, доста ужаси и немалко от онова, което би извикало отвращение.

Напред и назад в седемте зали вилнеели в същност рой съновидения.

И те, съновиденията, се кривели на всички страни, по тях трепкаха багрите на самите зали, караха необузданата музика на оркестъра да отеква като ехо на стъпките им.

И отново ударите на абаносовия часовник откъм залата с кадифе.

И отново всичко се умълчавало за момент, всичко утихвало освен гласа на часовника.

Съновиденията замръзвали по местата си.

Но отзвукът от напевния звън замирал ... само едно мигновение траел той ... и весел, тих смях го сподирял.

И отново гръмвала музиката, съновиденията оживявали, почвали да се кривят на всички страни, само че по-весело от преди пламенеели багрите на многоцветните прозорци, през които струели отблясъците от триногите.

В най-западната от седемте зали обаче не се осмелявала да прекрачи никоя маска, защото нощта била дълбока и през кърваво обогрение стъкла се изливали още по-ярко алени лъчи; чернотата на тъмните драперии навявала ужас; до онзи, чиито крака дръзнали да стъпят върху черния килим, откъм близкия абаносов часовник достигал задавеният ек, подчертан много по-внушително от звъна, който долитал до ушите на останалите, отдали се на веселие в другите зали.

Останалите зали били претъпкани, в тях туптяло трескаво сърцето на живота.

Празненството продължило вихрено чак до полунощ, когато часовникът най-сетне започнал да отброява звънко ударите.

Тогавата, така ми разказаха, музиката спряла, валсиращите престанали да се въртят, всички замрели смутено.

Ала сега часовникът трябвало да отрони дванадесет удара, имало повече време и сигурно повече мисли се промъкнали крадешком в съзнанието на по-разсъдливите измежду пируващите.

И тъкмо в този миг, навярно преди още мълчанието да е погълнало завинаги последните отзвуци на последния звън, много



от участващите в карнавала забелязали за първи път присъствието на една маска, която до този момент не била привлякла вниманието на никого.

Слухът за новодошлия обходил веднага шепнешком залите и всички зашумели, засуетили се, зашушукали – надигнала се вълна от неодобрение и почуда...

накрая тя прераснала в страх, ужас и отвращение.

Лесно е да се допусне, че в общество от привидения, каквито току-що описах, не е било възможно обикновен образ да причини преживяване от подобно естество.

Маскарадното своеволие в онази нощ наистина почти минавало пределите на допустимото, а въпросната фигура със своя Иродов вид надминавала дори самия Ирод и прекрачвала далеч-далеч отвъд представите за благоприличие и на самия принц.

В сърцата и на най-безразсъдните има струни, които не можеш да докоснеш, без да затрептят.

Дори и най-безпътните, за които и животът, и смъртта са шега, дори за тях има неща, с които не се шегуват.

Всички присъстващи се смаяли от облеклото и поведението на непознатия, в които не личало ни остроумие, ни благопристойност.

Фигурата била висока и мършава и от главата до петите била увита в саван.

Маската, скриваща лицето, почти изцяло наподобявала образа на вкочаняващ труп и дори най-внимателното вглеждане отблизо разсейвало с голям труд тази заблуда.

Обезумелите пируващи въпреки всичко щели да понесат и това, а нямало да бъде чудно дори да го одобрят.

Но маскираният стигнал дотам, че дръзнал да се предреши като Алената смърт.

Наметалото му било оросено с кръв, а високото чело и цялото лице – белязани с аления ужас.

Когато погледът на принц Просперо попаднал върху призрачния образ (който с бавна и тържествена стъпка крачел напред-назад сред спрелите танцьори, като че да изпълни по-правдоподобно ролята си), всички забелязали, че принцът се сгърчил конвулсивно; най-напред го разлюляла неудържима тръпка от ужас или от погнуса, а в следващия миг лицето му почервняло от гняв.

— Кой посмя...

– попитал дрезгаво той застаналите край него придворни – кой посмя да ни оскърби с тази богохулна подигравка?

Хванете го и му свалете маската ... за да разберем кого да обесим на бойниците при изгрев слънце!

Принц Просперо произнесъл тези думи в най-източната, в синята зала.

Звучно и отчетливо проехтели те из седемте зали, защото принцът бил смел и силен, а музиката секнала веднага, щом той дал знак с ръка.



Просперо стоял в синята зала, заобиколен от група пребледнели придворни.

В началото, когато изрекъл тези думи, придворните като че понечили да се спуснат към самозванеца, който в този миг също бил наблизко, но в следващия, с твърда и царствена стъпка, той сам пристъпил към принца.

Всички присъстващи били обзети от безумен ужас, вдъхнат от наглото високомерие на маскирания, и никой не посмял да протегне ръка и да го хване; невъзпрепятствуван, странникът минал на по-малко от метър край принца, а в това време стълпеното множество, подтикнато сякаш от един и същ импулс, се отдръпнало от центъра на залите и се прилепило към стените; непознатият, с предишната царствена и отмерена крачка, с която направил от първия миг впечатление, продължил безпрепятствено пътя си от синята зала в пурпурната ... от пурпурната в зелената ... от зелената в оранжевата... от нея в бялата ... оттам във виолетовата, преди някой да направи какъвто и да е опит да го задържи.

Точно тогава принц Просперо, побеснял от гняв и срам поради собственото си моментно малодушие, се понесъл като вихър през шестте зали, но никой не го последвал, понеже всички били сковани от смъртен ужас.

Просперо вдигнал високо голата си кама, доближил бързо на три-четири крачки движещата се фигура, която била стигнала вече до дъното на кадифената зала; неочаквано тя се обърнала и се намерила лице срещу лице със своя преследвач.

Разнесъл се остър вик, проблясваща, камата паднала върху черния килим, върху който миг по-късно се проснал мъртъв и принц Просперо.

Тогава, набрали безумната смелост на отчаянието, неколцина от пируващите тозчас нахълтали в черната зала и се нахвърлили върху високата маскирана фигура, застанала права и неподвижна в сянката на абаносовия часовник, обаче дъхът им секнал от неизказан ужас, когато под мъртвешкия саван и под маската на труп, които сграбчили с такава необуздана ярост, не намерили нищо осезаемо.

Едва тогава те усетили присъствието на Алената смърт.

Тя се била промъкнала подобно на крадец в нощта.

Един подир друг почнали да се сгромолясват пируващите из опръсканите с кръви зали, където само до преди миг се веселели, и издъхвали — всеки в окаяното положение, в което паднал повален.

И животът на абаносовия часовник помръкнал заедно с последния проблясък на веселие.

И пламъците в триногите угаснали.

И над всичко навеки се възцарили Мракът, Тлението и Алената смърт.

2) Automatic Summary:

Нито една епидемия дотогава не била толкова страшна и толкова гибелна . Кръвта била нейният Аватар , а нейният червенината и ужасът на кръвта . Остри болки , внезапно виене на свят , обилен кръвоизлив от всички пори и смърт . И всичко траяло само половин час : заразата , развоят и развързката на болестта . Когато народът в неговите владения наполовина измрял , той призовал хиляда здрави и жизнерадостни приятели измежду рицарите и дамите на своя двор и се оттеглил заедно с тях в дълбокото уединение на едно от укрепените си и построено като замък абатство . Решили в случай на неочаквани пристъпи на отчаяние или безумие да не оставят никакви пътища нито за влизане , нито за излизане . да се погрижи сам за себе си . Принцът се погрижил за всякакви удоволствия . И всичко това заедно с безопасността било тук , вътре . Отвън останала само Към края на петия или шести месец от уединението , когато чумата вилнеела най-свирепо отвън , за да развесели своите хиляди приятели , принц Просперо им устроил маскараден бал с невиджано великолепие . Но нека първо да ви разкажа за балните зали . В много дворци обаче тези покои имат вид на дълга и права зала , а когато крилата на плъзгащите се врати се дръпнат до самите стени , погледът не среща почти никакво препятствие . Както и можело да се очаква от пристрастието на принца към bizzare . На всеки двадесет-тридесет метра имало остър завой , а зад всеки нов ефект . По средата на всяка стена , отляво и отдясно , към закрит проход гледал висок и тесен готически прозорец , който следвал



криволиците на самите зали . Прозорците били от цветни стъкла , чиито краски се менели в зависимост от тона на подредбата в залата , към която се разтваряли . Ала в коридорите , край залите , срещу всеки прозорец имало по тежка тринога поставка с пламенееща жаравя тя хвърляла отблясъци през многоцветните стъкла и по такъв начин озарявала ярко залата . В най-западната или черната зала въздействието от отблясъците на пламенеещата жаравя , падащо през кърваво обогрените стъкла върху черните завеси , било крайно кошмарно и извиквало по лицата на гостите такова безумно изражение , че малцина от тях намирали достатъчно смелост да прекрачат нейния праг . В тази зала , на западната страна , висял окачен гигантски абаносов часовник . Махалото се люлеело наляво-надясно с унило , сънно , монотонно тиктакане ; когато стрелката на минутите склочела кръга край циферблата и часовникът трябвало да удари , от медните му дробове се изтръгвал ясен , силен , дълбок и извънредно мелодичен звън , но с такъв особен тон и удар , че при изтичането на всеки час музикантите от оркестъра се принуждавали да спрат за момент изпълнението си и да се вслушат в звъна , а валсиращите прекъсвали по неволя тан и за миг във веселите гости се прокрадвал смут ; и докато напевният звън на часовника отеквал , най-лекомислените пребледнявали , а по-улегналите и по-сдържаните прокарвали длани по челата си , сякаш някой ги е изтръгнал от унеса и размислите им . И те , съновиденията , се кривели на всички страни , по тях трепкали багрите на самите зали , И отново ударите на абаносовия часовник откъм залата с кадифе . И отново всичко се умълчавало за момент , Но отзвукът от напевния звън замирал . . В най-западната от седемте зали обаче не се осмелявала да прекрачи никоя маска , И тъкмо в този миг , навярно преди още мълчанието да е погълнало завинаги последните отзвущи на последния звън , Маскарадното своеволие в онази нощ наистина почти минавало пределите на допустимото , а въпросната фигура със своя Иродов вид надминавала дори самия Ирод и прекрачвала далеч-далеч отвъд представите за благоприличие и на самия принц . Но маскираният стигнал дотам , че дръзнал да се предреша като Алената смърт . Наметалото му било оросено с кръв . . . а високото чело и цялото лице белязани с аления ужас . Принц Просперо произнесъл тези думи в най-източната , в синята зала . В началото , когато изрекъл тези думи , придворните като че понечили да се спуснат към самозванеца , Всички присъстващи били обзети от безумен ужас , вдъхнат от наглото високомерие на маскирания , и никой не посмял да протегне ръка и да го хване ; невъзпрепятствуван , странникът минал на по-малко от метър край принца , а в това време стълпеното множество , подтикнато сякаш от един и същ импулс , се отдръпнало от центъра на залите и се прилепило към стените ; непознатият , с предишната царствена и отмерена крачка , с която направил от първия миг впечатление , продължил безпрепятствено пътя си от синята зала в пурпурната . . . от пурпурната в зелената . . . от зелената в оранжевата . . . от нея в бялата . .

English

I

1) Original text:

Two kings' sons once went out in search of adventures, and fell into a wild, disorderly way of living, so that they never came home again.

The youngest, who was called simpleton, set out to seek his brothers, but when at length he found them they mocked him for thinking that he with his simplicity could get through the world, when they two could not make their way, and yet were so much cleverer.

They all three traveled away together, and came to an ant-hill.

The two elder wanted to destroy it, to see the little ants creeping about in their terror, and carrying their eggs away, but simpleton said, leave the creatures in peace, I will not allow you to disturb them.

Then they went onwards and came to a lake, on which a great number of ducks were swimming.

The two brothers wanted to catch a couple and roast them, but simpleton would not permit it, and said, leave the creatures in peace, I will not suffer you to kill them.

At length they came to a bee's nest, in which there was so much honey that it ran out of the trunk of the tree where it was.

The two wanted to make a fire beneath the tree, and suffocate the bees in order to take away the honey, but simpleton again stopped them and said, leave the creatures in peace, I will not allow you to burn them.

At length the three brothers arrived at a castle where stone horses were standing in the stables, and no human being was to be seen, and they went through all the halls until, quite at the end, they came to a door in which were three locks.

In the middle of the door, however, there was a little pane, through which they could see into the room.

There they saw a little grey man, who was sitting at a table.

They called him, once, twice, but he did not hear, at last they called him for the third time, when he got up, opened the locks, and came out.

He said nothing, however, but conducted them to a handsomely-spread table, and when they had eaten and drunk, he took each of them to a bedroom.

Next morning the little grey man came to the eldest, beckoned to him, and conducted him to a stone table, on which were inscribed three tasks,



by the performance of which the castle could be delivered from enchantment.

The first was that in the forest, beneath the moss, lay the princess's pearls, a thousand in number, which must be picked up, and if by sunset one single pearl was missing, he who had looked for them would be turned into stone.

The eldest went thither, and sought the whole day, but when it came to an end, he had only found one hundred, and what was written on the table came true, and he was turned into stone.

Next day, the second brother undertook the adventure, but it did not fare much better with him than with the eldest, he did not find more than two hundred pearls, and was changed to stone.

At last it was simpleton's turn to seek in the moss, but it was so difficult for him to find the pearls, and he got on so slowly, that he seated himself on a stone, and wept.

And while he was thus sitting, the king of the ants whose life he had once saved, came with five thousand ants, and before long the little creatures had got all the pearls together, and laid them in a heap.

The second task, however, was to fetch out of the lake the key of the king's daughter's bed-chamber.

When simpleton came to the lake, the ducks which he had saved, swam up to him, dived down, and brought the key out of the water.

But the third task was the most difficult, from amongst the three sleeping daughters of the king was the youngest and dearest to be sought out.

They, however, resembled each other exactly, and were only to be distinguished by their having eaten different sweetmeats before they fell asleep, the eldest a bit of sugar, the second a little syrup, and the youngest a spoonful of honey.

Then the queen of the bees, whom simpleton had protected from the fire, came and tasted the lips of all three, and at last she remained sitting on the mouth which had eaten honey, and thus the king's son recognized the right princess.

Then the enchantment was at an end, everything was delivered from sleep, and those who had been turned to stone received once more their natural forms.

Simpleton married the youngest and sweetest princess, and after her father's death became king, and his two brothers received the two other sisters.

2) Automatic Summary:

Two kings' sons once went out in search of adventures, so that they never came home again. The youngest, who was called simpleton, set out to seek his brothers. The two elder wanted to destroy it, to see the little ants creeping about in their terror, and carrying their eggs away, but simpleton said, leave the creatures in peace, I will not allow you to disturb them. Then they went onwards and came to a lake. The two brothers wanted to catch a couple and roast them. The two wanted to make a fire beneath the tree, and suffocate the bees. At length the three brothers arrived at a castle and no human being was to be seen, and they went through all the halls. There they saw a little grey man, who was sitting at a table. They called him, once, twice. He said nothing, however. Next morning the little grey man came to the eldest, beckoned to him, and conducted him to a stone table. The first was and if by sunset one single pearl was missing, he for them would be turned into stone. The eldest went thither, and sought the whole day. Next day, the second brother undertook the adventure. And while he was thus sitting, the king of the ants. The second task, however, was to fetch out of the lake the key of the king daughter bed-chamber. But the third task was the most difficult, from amongst the three sleeping daughters of the king was the youngest and dearest to be sought out.

II

1) Original text:

Into the north window of my chamber glows the Pole Star with uncanny light. All through the long hellish hours of blackness it shines there. And in the autumn of the year, when the winds from the north curse and whine, and the red-leaved trees of the swamp mutter things to one another in the small hours of the morning under the horned waning moon, I sit by the casement and watch that star. Down from the heights reels the glittering Cassiopeia as the hours wear on, while Charles' Wain lumbers up from behind the vapour-soaked swamp trees that sway in the night-wind. Just before dawn Arcturus winks ruddily from above the cemetery on the low hillock, and Coma Berenices shimmers weirdly afar off in the mysterious east; but still the Pole Star leers down from the same place in the black vault, winking hideously like an insane watching eye which strives to convey some strange message, yet recalls nothing save that it once had a message to convey. Sometimes, when it is cloudy, I can sleep.

Well do I remember the night of the great Aurora, when over the swamp played the shocking coruscations of the daemon-light. After the beams came clouds, and then I slept.

And it was under a horned waning moon that I saw the city for the first time. Still and somnolent did it lie, on a strange plateau in a hollow



betwixt strange peaks. Of ghastly marble were its walls and its towers, its columns, domes, and pavements. In the marble streets were marble pillars, the upper parts of which were carven into the images of grave bearded men. The air was warm and stirred not. And overhead, scarce ten degrees from the zenith, glowed that watching Pole Star. Long did I gaze on the city, but the day came not. When the red Aldebaran, which blinked low in the sky but never set, had crawled a quarter of the way around the horizon, I saw light and motion in the houses and the streets. Forms strangely robed, but at once noble and familiar, walked abroad, and under the horned waning moon men talked wisdom in a tongue which I understood, though it was unlike any language I had ever known. And when the red Aldebaran had crawled more than half way around the horizon, there were again darkness and silence.

When I awaked, I was not as I had been. Upon my memory was graven the vision of the city, and within my soul had arisen another and vaguer recollection, of whose nature I was not then certain. Thereafter, on the cloudy nights when I could sleep, I saw the city often; sometimes under that horned waning moon, and sometimes under the hot yellow rays of a sun which did not set, but which wheeled low around the horizon. And on the clear nights the Pole Star leered as never before.

Gradually I came to wonder what might be my place in that city on the strange plateau betwixt strange peaks. At first content to view the scene as an all-observant uncorporeal presence, I now desired to define my relation to it, and to speak my mind amongst the grave men who conversed each day in the public squares. I said to myself, "This is no dream, for by what means can I prove the greater reality of that other life in the house of stone and brick south of the sinister swamp and the cemetery on the low hillock, where the Pole Star peers into my north window each night?"

One night as I listened to the discourse in the large square containing many statues, I felt a change; and perceived that I had at last a bodily form. Nor was I a stranger in the streets of Olathoë, which lies on the plateau of Sarkis, betwixt the peaks Noton and Kadiphonek. It was my friend Alos who spoke, and his speech was one that pleased my soul, for it was the speech of a true man and patriot. That night had the news come of Daikos' fall, and of the advance of the Inutos; squat, hellish, yellow fiends who five years ago had appeared out of the unknown west to ravage the confines of our kingdom, and finally to besiege our towns. Having taken the fortified places at the foot of the mountains, their way now lay open to the plateau, unless every citizen could resist with the strength of ten men. For the squat creatures were mighty in the arts of war, and knew not the scruples of honour which held back our tall, grey-eyed men of Lomar from ruthless conquest.

Alos, my friend, was commander of all the forces on the plateau, and in him lay the last hope of our country. On this occasion he spoke of the perils to be faced, and exhorted the men of Olathoë, bravest of the Lomarrians, to sustain the traditions of their ancestors, who when forced to move southward from Zobna before the advance of the great ice-sheet (even as our descendants must some day flee from the land of Lomar), valiantly and victoriously swept aside the hairy, long-armed, cannibal Gnophkehs that stood in their way. To me Alos denied a warrior's part, for I was feeble and given to strange faintings when subjected to stress and hardships. But my eyes were the keenest in the city, despite the long hours I gave each day to the study of the Pnakotic manuscripts and the wisdom of the Zobnarian Fathers; so my friend, desiring not to doom me to inaction, rewarded me with that duty which was second to nothing in importance. To the watch-tower of Thapnen he sent me, there to serve as the eyes of our army. Should the Inutos attempt to gain the citadel by the narrow pass behind the peak Noton, and thereby surprise the garrison, I was to give the signal of fire which would warn the waiting soldiers and save the town from immediate disaster.

Alone I mounted the tower, for every man of stout body was needed in the passes below. My brain was sore dazed with excitement and fatigue, for I had not slept in many days; yet was my purpose firm, for I loved my native land of Lomar, and the marble city of Olathoë that lies betwixt the peaks of Noton and Kadiphonek.

But as I stood in the tower's topmost chamber, I beheld the horned waning moon, red and sinister, quivering through the vapours that hovered over the distant valley of Banof. And through an opening in the roof glittered the pale Pole Star, fluttering as if alive, and leering like a fiend and tempter. Methought its spirit whispered evil counsel, soothing me to traitorous somnolence with a damnable rhythmical promise which it repeated over and over:

"Slumber, watcher, till the spheres

Six and twenty thousand years

Have revolv'd, and I return

To the spot where now I burn.

Other stars anon shall rise

To the axis of the skies;

Stars that soothe and stars that bless

With a sweet forgetfulness:

Only when my round is o'er

Shall the past disturb thy door."



Vainly did I struggle with my drowsiness, seeking to connect these strange words with some lore of the skies which I had learnt from the Pnakotic manuscripts. My head, heavy and reeling, drooped to my breast, and when next I looked up it was in a dream; with the Pole Star grinning at me through a window from over the horrible swaying trees of a dream-swamp. And I am still dreaming.

In my shame and despair I sometimes scream frantically, begging the dream-creatures around me to waken me ere the Inutos steal up the pass behind the peak Noton and take the citadel by surprise; but these creatures are daemons, for they laugh at me and tell me I am not dreaming. They mock me whilst I sleep, and whilst the squat yellow foe may be creeping silently upon us. I have failed in my duty and betrayed the marble city of Olathoë; I have proven false to Alos, my friend and commander. But still these shadows of my dream deride me. They say there is no land of Lomar, save in my nocturnal imaginings; that in those realms where the Pole Star shines high and red Aldebaran crawls low around the horizon, there has been naught save ice and snow for thousands of years, and never a man save squat yellow creatures, blighted by the cold, whom they call "Esquimaux".

And as I writhe in my guilty agony, frantic to save the city whose peril every moment grows, and vainly striving to shake off this unnatural dream of a house of stone and brick south of a sinister swamp and a cemetery on a low hillock; the Pole Star, evil and monstrous, leers down from the black vault, winking hideously like an insane watching eye which strives to convey some strange message, yet recalls nothing save that it once had a message to convey.

2) Automatic Summary:

Into the north window of my chamber glows the Pole Star with uncanny light . All through the long hellish hours of blackness it shines there . And in the autumn of the year , when the winds from the north curse and whine , and the red-leaved - trees of the swamp mutter things to one another in the small hours of the morning under the horned waning moon , I sit by the casement and watch that star . Just before dawn Arcturus winks ruddily from above the cemetery on the low hillock , and Coma Berenices shimmers weirdly afar off in the mysterious east ; but still the Pole Star leers down from the same place in the black vault , winking hideously like an insane watching eye Well do I remember the night of the great Aurora , when over the swamp played the shocking coruscations of the daemon-light - . And it was under a horned waning moon Still and somnolent did it lie , on a strange plateau in a hollow betwixt strange peaks . In the marble streets were marble pillars , the upper parts of And overhead , scarce ten degrees from the zenith , glowed that watching Pole Star . than half way around the horizon , there were again darkness and silence . And on the clear nights the Pole Star leered as never before . One night as I listened to the discourse in the large square containing many statues , I felt a change ; and perceived It was my friend Alos That night had the news come of Daikos fall , and of the advance of the Inutos ; squat , hellish , yellow fiends who five years ago had appeared out of the unknown west to ravage the confines of our kingdom , and finally to besiege our towns . Alos , my friend , was commander of all the forces on the plateau , and in him lay the last hope of our country . On this occasion he spoke of the perils to be faced , and exhorted the men of Olatho , bravest of the Lomarrians , to sustain the traditions of their ancestors , and victoriously swept aside the hairy , long-armed - , cannibal Gnophkehs To the watch-tower - of Thapnen he sent me , there to serve as the eyes of our army . Should the Inutos attempt to gain the citadel by the narrow pass behind the peak Noton , and thereby surprise the garrison , I was to give the signal of fire Alone I mounted the tower , But as I stood in the tower s topmost chamber , I beheld the horned waning moon , red and sinister , quivering through the vapours And through an opening in the roof glittered the pale Pole Star , fluttering as if alive , and leering like a fiend and tempter . Methought its spirit whispered evil counsel , soothing me to traitorous somnolence with a damnable rhythmical promise And I am still dreaming . But still these shadows of my dream deride me . And as I writhe in my guilty agony , frantic to save the city.

German

1) Original text:

Eine Teilung stand am Beginn des langen Weges der Deutschen zu einer Nation . Ludwig der Fromme , der schwache Sohn Karls des Großen , hatte das mächtige Reich des Vaters nur noch mühsam zusammenhalten können , und unter den Enkeln des großen Kaisers brach es dann morsch auseinander . einigten sich diese drei nach langen Streitereien in Verden , dem heutigen Verdun : Lothar , der Älte , erhielt die Kaiserwürde , dazu Italien und einen verhältnismäßig schmalen Streifen , der von den Alpen im Süden bis nach Friesland im Norden reichte . Das Land westlich davon fiel an Karl den Kahlen . Aus ihm ging das spätere Frankreich hervor . Östlich des Rheins aber , bis zur Elbe , und Donau Saale erstreckte sich das ostfränkische Reich Ludwigs des Deutschen . Ein Vierteljahrhundert später wurde die Grenze dieses Reiches noch einmal geringfügig nach Westen verschoben . Dann blieb sie als Grenze zwischen Frankreich und Deutschland unverändert das ganze Mittelalter hindurch bestehen , nicht nur als politische Trennungslinie , sondern mit kleinen Abweichungen auch als Sprachgrenze . Der Westen sprach das aus dem Latein hervorgegangene Französisch , der Osten dagegen die " Volkssprache " , die " Lingua theodisca " , wie sie im Gegensatz zum Latein des gelehrten Klerus genannt wurde . Das neue ostfränkische Reich hatte damit eine eigene Sprache un seine eigenen Könige . Schwache Könige allerdings , die sich weder im Innern durchsetzen , noch die drohenden Gefahren an den Grenzen abwenden konnten . Normannen im Norden , Sarazenen im Süden , Mährer und Ungarn im Osten und Südosten bedrängten die Grenzen und fielen mit fester Regelmäßigkeit in das Reich ein und plünderten . Die Uneinigkeit im Innern erleichterte noch diese Raubzüge ; denn zwischen den adeligen Geschlechtern to Fehden und Blutrache . Niemand trauerte daher dem Aussterben des karolingischen Herrscherhauses nach , das mit dem Tod des achtzehnjährigen Ludwig IV . im Jahr 911 erlosch . Noch einmal wählten die Großen des Reiches mit dem Frankenherzog Konrad einen nahen Verwandten der Karolinger , aber der neue König hatte es in den wenigen Jahren seiner Regierung nicht leicht ; denn zu mächtig waren inzwischen die Stammesherzöge geworden . Konrad I . war ein glückloser Herrscher , aber ein weitschauender Politiker . Noch auf dem Sterbebett empfahl er , den mächtigsten seiner früheren Gegner , den Sachsenherzog Heinrich , zu seinem Nachfolger zu wählen . Die Sage erzählt , dieser habe die Nachricht von seiner Wahl beim Vogelfang erhalten , doch sah die Wirklichkeit weit prosaischer aus . Nur der Adel aus Sachsen und Franken wählte ihn , während sich die Süddeutschen abseits hielten . Fast schien es , als sollten die Stammesherzöge mächtiger bleiben als der König . Trotzdem gilt 919 , das Jahr der Thronbesteigung Heinrichs , zu Recht als das des ersten deutschen Reiches Geburtsjahr . Karolingische Reiterei Karolingische Reiterei Trotz verschiedener Schwächen bleibt dieses Reich auch über



Jahrhunderte hinweg bestehen und wird zu einem entsprechenden politischen Faktor in Mitteleuropa . Das Königtum gewann nicht nur an Ansehen , sondern half mit bei der Formung des Reiches und damit der deutschen Nationen . Nur hundert Jahre regierten die Könige aus dem sächsischen Stamm , aber diese Zeit bildete eine wichtige Phase mittelalterlicher deutscher Geschichte . Heinrich I . erreichte , was seinem Vorgänger nicht gelungen war : er setzte die königliche Gewalt gegenüber den Herzögen durch und fügte die deutschen Stämme in dem neuen " Regnum Teutonicum " - dem Reich der Deutschen - zusammen . Auf dieser Leistung konnte sein Sohn Otto I . weiterbauen , der einzige deutsche König , dem die Geschichte den Beinamen " der Große " verlieh . Energisches Vorgehen im Innern wie an den Grenzen des Reiches kennzeichnen seine Regierung . So bannte er 955 in der Schlacht am Lechfeld bei Augsburg für immer die Bedrohung des jungen Reiches durch die Ungarn . Gegenüber den Herzögen festigte er die königliche Gewalt , und 962 ließ er sich in Rom zum Kaiser krönen . Es war die Geburtsstunde des " Heiligen Römischen Reiches Deutscher Nationen " ; denn von nun an blieb die römische Kaiserkrone den deutschen Königen vorbehalten , zugleich aber verbanden sich deren Schicksale eng mit Italien und dem Papsttum . Ein wechselvolles Auf und Ab , ein gelegentliches Miteinander und weit häufigeres Gegeneinander von Kaiser und Papst kennzeichnen die Geschichte des Mittelalters , und noch heute schwelt der alte Streit um Wert und Unwert deutscher Italienpolitik . Otto I . hatte , wie es schien , durch sein energisches Vorgehen in Rom ein klares Nebeneinander von Papst und Kaiser geschaffen . Aber schon ein Jahrhundert später sollte sich zeigen , daß diese Ordnung keineswegs klar genug abgegrenzt war , sondern sich in das Gegenteil verkehrte . Den sächsischen Königen folgte das fränkisch-salische Geschlecht . Die beiden ersten Herrscher aus diesem Haus führten Deutschland zu machtvoller Größe . gewann Konrad II . im Westen das Königreich Burgund , das nun zusammen mit dem Reich und Italien den dritten Teil des römisch-deutschen Imperiums bildete . An Konrads Sattel hängen Karls Steigbügel " sagten die Zeitgenossen und verglichen ihren Kaiser mit Karl dem Großen . Sein Sohn und Nachfolger Heinrich III . vollendete das Werk des Vaters und hielt vor allem in Italien energisch Ordnung . So ließ er gleich drei rivalisierende Päpste absetzen und den Bamberger Bischof Suidger zu ihrem rechtmäßigen Nachfolger wählen . Auch die beiden nächsten Päpste kamen aus Deutschland und waren Schützlinge Heinrichs . Als dieser aber , erst neununddreißigjährig , 1056 starb , änderte sich die politische Lage schlagartig . Sein Sohn Heinrich IV . war ein sechsjähriges Kind und abhängig von den deutschen Fürsten . In Italien schien für das Papsttum nun die Stunde gekommen , sich aus der kaiserlichen Bevormundung zu lösen . Als 1073 Gregor VII . ein ungemein energischer ehemaliger Mönch , in Rom den Stuhl Petri bestieg , bahnte sich die entscheidende Kraftprobe zwischen dem römischen Papst und dem deutschen Kaiser an . Gregor wollte das Papsttum von jedem Einfluß weltlicher Gewalt befreien . Der offene Streit entzündete sich an dem alten Recht des deutschen Königs , von sich aus Bischöfe einzusetzen , das nun der Papst allein für sich beanspruchte . Heinrich verfügte über keinen Rückhalt bei den deutschen Fürsten . Als ihn der Papst bannte , beschlossen diese sogar seine Absetzung , wenn er sich nicht binnen Jahresfrist von der im Mittelalter so bedeutsamen Kirchenstrafe würde lösen können . Daraufhin zog Heinrich 1077 im strengsten Winter mit nur wenigen Getreuen nach Oberitalien , wo Gregor in der Burg Canossa weilte . Durch demütige Buße erreichte zwar der junge Herrscher die Lösung vom Bann , doch hatte die Auseinandersetzung zwischen ihm und dem Papst damit keineswegs ihr Ende errei . Bald war der Streit in neuer Heftigkeit entflammt , und erst unter Heinrich V . kam es zu einem Kompromiß , der weder Kaiser noch Papst einen vollen Erfolg brachte . Gewinner waren die deutschen Fürsten und die italienischen Städte , die in diesen unruhigen Jahrzehnten ihre Macht stärken konnten . Als 1137 das schwäbische Geschlecht der Staufer die deutsche Königswürde erlangte , machte sich diese doppelte Belastung unheilvoll bemerkbar . Trotzdem war es eine glanzvolle Zeit , die ihren Höhepunkt mit der Friedrichs I Regierung . Barbarossa erlebte . Er galt als das große ritterliche Vorbild , ein gleichermaßen hervorragender Staatsmann wie Feldherr , unter dessen Herrschaft auch die Literatur und die bildende Kunst einen Höhepunkt erlebten und von deren einstiger Größe heute noch zahlreiche Burgen und stattliche Ruine zeugen . Aber diese hohe Zeit des Rittertums hatte auch ihre dunklen Seiten . Wie seine Vorgänger mußte sich Barbarossa mit den Machtansprüchen der Fürsten auseinandersetzen . Der Welfe Heinrich der Löwe , gleichzeitig Herzog von Bayern und Sachsen , war stark genug , um eigenmächtige selbständige Politik treiben zu können . Barbarossa seinerseits suchte zuerst die Position des Kaisertums im Süden zu stärken und unternahm sechs Italienzüge . Auch er geriet dabei in einen schweren Konflikt mit dem Papst , auf dessen Seite sich die lombardischen Städte stellten . Aber klüger als sein Vorgänger Heinrich IV . suchte er den Ausgleich , der ihm freie Hand gewährte , im Reich gegen den Welfenherzog vorzugehen und dessen Macht zu brechen . Als es Barbarossa gelang , seinen Sohn und Nachfolger mit der Erbin des Normannenreiches in Unteritalien- Sizilien zu vermählen , stand er auf dem Höhepunkt seiner Macht und seines Ansehens . Das Reich war befriedet , dem Selbstständigkeitsstreben der Fürsten eine deutliche Grenze gesetzt , und die staufische Herrschaft reichte von der Nordsee bis nach Sizilien , von der Provence bis nach Pommern . Mit Recht konnte sich Barbarossa als das Haupt des Abendlandes sehen . Doch schneller als vermutet , brach diese Vormachtstellung des Staufischen Kaisertums wieder zusammen . Friedrichs Sohn Heinrich der VI . regierte nur acht Jahre , dann raffte ihn ein tückisches Fieber hinweg . Nie mehr sollte in den folgenden Jahrhunderten das Reich eine ähnliche Machtfülle nach außen , nie mehr die gleiche Geschlossenheit im Innern erlangen . Fortan blieb es nur ein Glied in der Kette der neu entstehenden europäischen Nationalstaaten . Im Innern wurde die kaiserliche Zentralgewalt abgelöst von erneuten partikularistischen Bestrebung der Fürsten . Friedrich II . der Enkel Barbarossas , sah Deutschland nicht mehr als den seines Imperiums Kern , sondern nur als ein Nebenland , das er vorwiegend von Sizilien aus regierte und in dem er wichtige königliche Rechte preisgab , um seine Kräfte für eine erneute Auseinandersetzung mit dem Papsttum frei zu haben . Den deutschen Fürsten kam solche Entwicklung nur zugute , sie wurden die wahren Landesherrn und damit die entscheidenden Träger der innerpolitischen Entwicklung . Damit änderte sich aber in den folgenden Jahrzehnten rasch die innere Struktur des Reiches . war mit Konrad IV . der letzte staufische König gestorben . Nach zwei Jahrzehnten der Unruhe , die Schiller so treffend als " die kaiserlose , die schreckliche Zeit " bezeichnete , folgten von 1273 bis 1347 fünf Könige aus verschiedenen Geschlechtern . Gewählt wurden sie nun vo den " Kurfürsten " , dem ursprünglich auf sieben Mitglieder beschränktem Kollegium der mächtigsten geistlichen und weltlichen Reichsfürsten . Sie waren aber nie nur " die ersten in des kysere kore " , sondern auch die ersten , die sich bei jeder neuen Wahl schamlos bereicherten . Um sich ihrerseits gegenüber den Fürsten behaupten zu können , versuchten die Könige ihre eigene Hausmacht auszubauen , d . h . sie suchten für ihre Familien ein entsprechend großes Territorium zu erwerben . Der Habsburger Rudolf I . begann dieses Spiel und begründete seine Hausmacht , indem er die österreichischen Besitzungen der ausgestorbenen Babenberger erwarb . Eine andere große Hausmacht im östlichen Teil des Reiches begründeten die Luxemburger , die Böhmen und Mähren für sich gewannen . In den letzten zwei Jahrhunderten des Mittelalters wandelte sich das innere Bild des Reiches . Die aus den alten Stammesherrschaften hervorgegangenen Herzogtümer waren längst zerfallen . An ihre Stelle traten eine Vielzahl von Landesherrschaften . Deutlich verschob sich dabei das politische Schwergewicht des Reiches von Rhein und Main nach Osten , wo seit dem hohen Mittelalter im Zuge einer stetigen Kolonisations- und Siedlerbewegung Neuland gewonnen worden war . Drei der sieben Kurfürsten , nämlich der Herzog von Sachsen (Obersachsen) , der Markgraf von Brandenburg und der König von Böhmen , stammten aus solchen Neusiedelgebieten . Das Rittertum , einst die Stütze staufischer Kaisermacht , verlor immer mehr an Bedeutung , dafür gewannen die Städte und ihre Bürger an Ansehen . Zwar lebte immer noch der weitaus größte Teil der Bevölkerung auf dem Lande , aber die Zahl der Städte wuchs rasch , und gegen Ende des Mittelalters waren es im Reich schon an die zweitausend , die meisten von ihnen hatten allerdings kaum mehr als tausend Einwohner . Trotzdem wurden sie zu kulturellen und wirtschaftlichen Zentren und bildeten die neu aufkommenden Städtebünde , politische Machtfaktoren , mit denen die Fürsten ernstlich rechnen mußten . So beeinflusste der große der Hanse Politik Städtebund und Wirtschaft im weit über die Grenzen des Reiches Norden hinaus und griff sogar in die inneren Angelegenheiten der skandinavischen Länder ein . Erstaunlicherweise fehlten äußere Bedrohungen des Reiches über Jahrhunderte hinweg . Otto I . hatte 955 die Ungarngefahr gebannt . standen dann , wie aus dem Nichts auftauchend , die Mongolen im Osten an den Grenzen . Vergebens traten ihnen deutsche und polnische Ritter in gemeinsamen Kampf



entgegen . Nur der Tod des Großkhans in der fernen Mongolei verhinderte die Überflutung Deutschlands und dann ganz Westeuropas durch die mongolischen Reiterheeren . Nach 1426 unternahmen die böhmischen Hussiten zehn Jahre lang regelmäßig blutige Raubzüge in die mittel- und ostdeutschen Gebiete . Trotzdem war das Mittelalter keine Zeit der Ruhe und des Friedens . Fast regelmäßig kämpften die deutschen Kaiser mit ihren Heeren in Ober- und Mittelitalien ; deutsche Ritter nahmen an den Kreuzzügen teil oder kämpften , als es in Palästina nichts mehr zu kämpfen gab , mit den Rittern des deutschen Ordens gegen deren heidnische und christlich-slavische Gegner , und gegen Ende des 14. Jahrhunderts zogen kaiserliche Heere durch Ungarn gegen die Türken , die neuen Feinde der Christenheit . Im Reich kam es immer und immer wieder zu Bürgerkriegen , wie man sie nicht anders bezeichnen kann ; denn die Fürsten befehdelten oft genug den König , bekämpften sich untereinander oder kämpften im späten Mittelalter gegen die Städtebünde . Und wenn die Großen schon einmal Ruhe gaben , sorgten die Kleinen für Unruhe . Jeder glaubte , von seiner Burg aus irgendwelchen Nachbarn und vor allem den " Pfeffersäcken " in den Städten den Kampf ansagen zu können . Symbole allgemeiner Verrottung waren die Fehdebücher der Städte , die sorgfältig geführt werden mußten , um überhaupt noch zu wissen , welche Herren den Bürgern den Kampf angesagt hatten . Das ausgehende 15. und beginnende 16. Jahrhundert in Deutschland hat ein Historiker einmal als die " Zeit der großen Wende " bezeichnet . Tatsächlich erfaßte ein rascher Wandel alle Bereiche des politischen , wirtschaftlichen , kulturellen und sozialen Lebens . Im Innern des Reiches zerfiel die alte ständische Ordnung . Das im Niedergang begriffene Rittertum verlor durch das Aufkommen der Landsknechte als Fußtruppen endgültig seine militärische Bedeutung . Umgekehrt wuchs das Ansehen der Bürger in den Städten . Handelsgesellschaften und einzelne Familien wie etwa die Fugger und Welser in Augsburg leiteten eine Epoche des Frühkapitalismus in Deutschland ein . Sehr rasch fanden sie Anschluß an den internationalen Markt , der seinerseits wieder durch die großen Entdeckungen zu Beginn des 16. Jahrhunderts einen ungeheuren Aufschwung erlebte . Die Erfindung des Buchdrucks mit beweglichen Lettern durch Gutenberg förderte seit der Mitte des 15. Jahrhunderts den geistigen Gedankenaustausch und damit die von Italien her sich ausbreitende Geistesströmung des Humanismus . So ist es nur verständlich , wenn der streitbare Humanist und Dichter Ulrich von Hutten in den Ruf ausbrach : " O Jahrhundert , o Wissenschaft , es ist eine Lust zu leben ! Auch der Thesenanschlag Luthers 1517 , die Geburtsstunde der Reformation , darf mit unter dem Gesichtspunkt dieser Lust am Disputieren , an der geistigen und geistvollen Auseinandersetzung mit den anstehenden Problemen gesehen werden , wenn er au rasch über das ursprüngliche Ziel hinauswuchs und zum Funken wurde , der einen gewaltigen Brand entzündete , dessen Wirkung niemand vorausahnen konnte . Aber die " Lust am Leben " war nicht für alle gleichermaßen gegeben . Politische Rechtlosigkeit , in einigen Teilen Deutschlands verbunden mit bedrückenden wirtschaftlichen Lasten , führte zu jenen Unruhen bei den Bauern , aus denen sich der Bauernkrieg entwickelte , de wir besser als eine erste politische und soziale Revolutionsbewegung sehen sollten . Aber die Bauern waren zu schwach , um sich gegen Fürsten und Städte durchsetzen zu können . Es sollten noch einmal rund dreihundert Jahre vergehen , bis einige der neuen freiheitlichen Gedanken wieder aufgenommen und verwirklicht werden konnten .

2) Automatic Summary:

Eine Teilung stand am Beginn des langen Weges der Deutschen zu einer Nation . Aus ihm ging das spätere Frankreich hervor . Dann blieb sie als Grenze zwischen Frankreich und Deutschland unverändert das ganze Mittelalter hindurch bestehen , nicht nur als politische Trennungslinie , sondern mit kleinen Abweichungen auch als Sprachgrenze . Das neue ostfränkische Reich hatte damit eine eigene Sprache und seine eigenen Könige . Normannen im Norden , Sarazenen im Süden , Mährer und Ungarn im Osten und Südosten bedrängten die Grenzen und fielen mit fester Regelmäßigkeit in das Reich ein und plünderten . Niemand trauerte daher dem Aussterben des karolingischen Herrscherhauses nach , das mit dem Tod des achtzehnjährigen Ludwig IV . denn zu mächtig waren inzwischen die Stammesherzöge geworden . war ein glückloser Herrscher , aber ein weitschauender Politiker . Nur der Adel aus Sachsen und Franken wählte ihn , während sich die Süddeutschen abseits hielten . Trotzdem gilt 919 , das Jahr der Thronbesteigung Heinrichs , zu Recht als das des ersten deutschen Reiches Geburtsjahr . Nur hundert Jahre regierten die Könige aus dem sächsischen Stamm , aber diese Zeit bildete eine wichtige Phase mittelalterlicher deutscher Geschichte . erreichte , was seinem Vorgänger nicht gelungen war : er setzte die königliche Gewalt gegenüber den Herzögen durch und fügte die deutschen Stämme in dem neuen " Regnum Teutonicum " - dem Reich der Deutschen - zusammen . Energisches Vorgehen im Innern wie an den Grenzen des Reiches kennzeichnen seine Regierung . Gegenüber den Herzögen festigte er die königliche Gewalt , und 962 ließ er sich in Rom zum Kaiser krönen . Ein wechselvolles Auf und Ab , ein gelegentliches Miteinander und weit häufigeres Gegeneinander von Kaiser und Papst kennzeichnen die Geschichte des Mittelalters , und noch heute schwelt der alte Streit um Wert und Unwert deutscher Italienpolitik . hatte , Aber schon ein Jahrhundert später sollte sich zeigen , Die beiden ersten Herrscher aus diesem Haus führten Deutschland zu machtvoller Größe . An Konrads Sattel hängen Karls Steigbügel " sagten die Zeitgenossen und verglichen ihren Kaiser mit Karl dem Großen . vollendete das Werk des Vaters und hielt vor allem in Italien energisch Ordnung . Als dieser aber , erst neununddreißigjährig , 1056 starb , änderte sich die politische Lage schlagartig . war ein sechsjähriges Kind und abhängig von den deutschen Fürsten . ein ungemein energischer ehemaliger Mönch , in Rom den Stuhl Petri bestieg , bahnte sich die entscheidende Kraftprobe zwischen dem römischen Papst und dem deutschen Kaiser an . Der offene Streit entzündete sich an dem alten Recht des deutschen Königs , von sich aus Bischöfe einzusetzen , das nun der Papst allein für sich beanspruchte . Daraufhin zog Heinrich 1077 im strengsten Winter mit nur wenigen Getreuen nach Oberitalien , wo Gregor in der Burg Canossa weilte . Bald war der Streit in neuer Heftigkeit entflammt , und erst unter Heinrich V . Gewinner waren die deutschen Fürsten . Trotzdem war es eine glanzvolle Zeit , die ihren Höhepunkt mit der Friedrichs I Regierung . Aber diese hohe Zeit des Rittertums hatte auch ihre dunklen Seiten . Der Welfe Heinrich der Löwe , gleichzeitig Herzog von Bayern und Sachsen , war stark genug ,

Greek

I

1) Original text:

Η αντίληψη για τις σχέσεις της κυβέρνησης και του κοινοβουλίου, της νομοθετικής και της εκτελεστικής εξουσίας καθορίζεται από ένα πρότυπο που δεν υπάρχει πια.



Τα σημερινά κοινοβούλια έχουν πάψει προ πολλού να είναι κοινοβούλια με σταθερές πλειοψηφίες που κρίνουν κατά περίπτωση την αποδοχή μιας πολιτικής ή ενός νομοσχεδίου με βάση τη συζήτηση στην αίθουσα. Αλλά και οι βουλευτές έπαψαν να αλλάζουν στρατόπεδα ή απόψεις επικαλούμενοι τη συνείδησή τους από την εποχή που τα εργατικά κόμματα στην Ευρώπη άρχισαν να έχουν μαζική παρουσία και να επιβάλλουν ενιαία στάση στους εκπροσώπους τους.

Η εικόνα έτσι σε κάθε κόμμα είναι μία και ενιαία για όλα τα θέματα. Η πλειοψηφία ακολουθεί τις κυβερνητικές εισηγήσεις, σε αντίθετη δε περίπτωση δημιουργείται μείζον πρόβλημα, το οποίο όμως κυβέρνηση και κυβερνητική πλειοψηφία αποφεύγουν κατά κανόνα πάση θυσία. Το αποτέλεσμα των διαδικασιών εντός του κοινοβουλίου είναι από την πρώτη κιόλας μέρα σχεδόν προδιαγεγραμμένο για όλη την τετραετία. Ένα διαφορετικό αποτέλεσμα προϋποθέτει τη δημιουργία μιας διαφορετικής πλειοψηφίας. Προκύπτει, όπως το 1993, από διεργασίες εκτός της αίθουσας συνεδριάσεων. Τότε η κυβέρνηση της Νέας Δημοκρατίας έχασε την πλειοψηφία μετά από κρίση στο εσωτερικό της που οδήγησε και στη δημιουργία και νέου κόμματος.

Το σύνταγμα ορίζει ότι δεν μπορούν να επιβληθούν περιορισμοί στη γνώμη του βουλευτή, ο οποίος σε κάθε περίπτωση ψηφίζει κατά συνείδηση. Βουλευτές επικαλούνται κατά καιρούς τη συνταγματική ρύθμιση διαμαρτυρόμενοι για την «κομματική πειθαρχία». Συνήθως η έκρηξη αυτή συμβαίνει σε θέματα που αφορούν προσωπικά τους ζητήματα, όπως το επαγγελματικό ασυμβίβαστο ή ο τρόπος εκλογής τους. Η ψήφος κατά συνείδηση σωστά τονίζεται από το σύνταγμα, διότι έτσι κάθε βουλευτής είναι υπεύθυνος των αποφάσεών του και οι αποφάσεις της Βουλής δεν μπορούν να αμφισβητηθούν ως προϊόν πίεσης. Ο βουλευτής πρέπει να είναι ελεύθερος στην έκφραση της γνώμης του, όπως και τα κόμματα από τη μεριά τους πρέπει να είναι ελεύθερα να επιβάλλουν την κομματική πειθαρχία, μια που οι ψηφοφόροι εκλέγουν κόμματα και σε αυτά δίνουν πρωτίστως την εντολή να διαμορφώνουν πολιτική. Και οι δύο αρχές μπορούν και πρέπει να συνυπάρχουν.

Ο βουλευτής και το κόμμα έχουν έτσι την ευθύνη να δείξουν ποιες αρχές και ποιες πολιτικές υπερασπίζονται, όταν η ελευθερία γνώμης του βουλευτή συγκρούεται με την κομματική πειθαρχία. Αυτή η ευθύνη καθιστά τη Βουλή χώρο πολιτικής αντιπαράθεσης και θέτει φραγμό στην προβολή προσωπικών επιδιώξεων.

Η σημασία των κοινοβουλευτικών συζητήσεων έχει επηρεαστεί επίσης αρνητικά από τις νέες μορφές δημόσιου διαλόγου που αναπτύχθηκαν με τη βοήθεια των μέσων ενημέρωσης. Όταν πλέον οι ειδήσεις κυκλοφορούν ταχύτατα και οι επιπτώσεις στην οικονομία και στην πολιτική ζωή προκαλούνται άμεσα, η κυβέρνηση και τα κόμματα δεν μπορούν να περιμένουν τη σύγκληση της Βουλής για να πάρουν θέση.

Πρέπει να δρουν ταχύτατα και με τρόπο ώστε το μήνυμά τους να μεταδίδεται αμέσως. Η κοινοβουλευτική διαδικασία δεν καλύπτει αυτές τις προϋποθέσεις.

Μερικοί θέλουν τη Βουλή πεδίο ρητορικών αγώνων, κι ο κανονισμός της Βουλής έχει επηρεαστεί από την αντίληψη αυτή. Παρέχει επίσης στους αρχηγούς των κομμάτων και στους κοινοβουλευτικούς εκπροσώπους υπερβολικά πολύ χρόνο ομιλίας σε σχέση με αυτό που ισχύει στην Ευρωβουλή και σε άλλα κοινοβούλια της Ευρωπαϊκής Ένωσης. Η εποχή των ρητορειών έχει ευτυχώς παρέλθει. Η πληροφόρηση που διαθέτουν οι βουλευτές από πριν, ο δημόσιος διάλογος στα ΜΜΕ, η δυνατότητα των πολιτών να αντλήσουν και από άλλες πηγές πέραν των ομιλιών την αναγκαία γνώση πιέζουν για τεκμηριωμένες και συνοπτικές ομιλίες. Αυτό περιορίζει το θέμα που παρείχαν τα κοινοβούλια σε παλαιότερες εποχές και βοηθά την ουσία.

Ο κανονισμός της Βουλής προβλέπει τη διεξαγωγή συζητήσεων μεταξύ του πρωθυπουργού και των προέδρων των κομμάτων της αντιπολίτευσης για τα θέματα της εξωτερικής πολιτικής, για τις σχέσεις με την Ευρωπαϊκή Ένωση αλλά και για άλλα θέματα γενικότερου ενδιαφέροντος. Οι συζητήσεις αυτές αποτέλεσαν τόσο κατά την οκταετία όσο και σε προηγούμενες περιόδους το αποκορύφωμα της σύγκρουσης μεταξύ κυβέρνησης και αντιπολίτευσης. Είχαν ως εκ τούτου ένα στοιχείο δραματικής κορύφωσης σε σχέση με τη συνήθη εργασία της Βουλής. Ωστόσο για μένα υπήρξαν ευπρόσδεκτη ευκαιρία για να εξηγήσω τους στόχους μας και να αναλύσω την πορεία της οικονομίας και των εξωτερικών μας σχέσεων.

Όμως, την προσοχή των μέσων ενημέρωσης και ως εκ τούτου και του κοινού αποσπούσαν μόνο ορισμένα λεπτά, όταν ο πρωθυπουργός σχολίαζε την ομιλία του αρχηγού της αξιωματικής αντιπολίτευσης και ο τελευταίος απαντούσε, όταν δηλαδή υπήρχε άμεση προσωπική αντιπαράθεση, συνήθως με ένταση και υψηλούς τόνους.

Εάν ο καβγάς ήταν μεγάλος υπήρχε θέμα και τα μέσα έμεναν ικανοποιημένα από τη συζήτηση, εάν αντίθετα δεν είχε την αναμενόμενη ένταση, η συζήτηση δεν είχε ενδιαφέρον. Το μεγάλο κοινό, που δεν παρακολουθούσε ολόκληρη τη συζήτηση, αποκτούσε ως εκ τούτου περιορισμένη εικόνα. Συνολικά δηλαδή η συζήτηση υποβαθμιζόταν και ενισχύονταν η αρνητική εντύπωση της πολιτικής ως θεάτρου. Προσπάθησα για αυτό σε συνεργασία με τον πρόεδρο της Βουλής κ. Α. Κακλαμάνη να τροποποιηθεί ο κανονισμός της Βουλής, ώστε να περιοριστούν οι απανωτές δευτερολογίες και τριτολογίες μεταξύ των δύο πρωταγωνιστών και να παρεμβάλλονται οι επικεφαλής των άλλων κομμάτων καθώς και ο αρμόδιος για το θέμα υπουργός. Και πράγματι η αντιπαράθεση μεταξύ του πρωθυπουργού και του αρχηγού της αξιωματικής αντιπολίτευσης περιορίστηκε σε ένα βαθμό, δεν έπαψε όμως να αποτελεί το κεντρικό γεγονός της συζήτησης καλύπτοντας την ουσία των θεμάτων.

Επιπλέον για να αναδειχθεί η κοινοβουλευτική συζήτηση τροποποιήθηκε από την αρχή της οκταετίας, μετά από πρότασή μου, η διαδικασία των επίκαιρων ερωτήσεων προς τον πρωθυπουργό. Η νέα ρύθμιση προέβλεπε ο πρωθυπουργός να απαντά πολύ πιο συχνά, κάθε εβδομάδα, σε ερωτήσεις οιοδήποτε βουλευτή και όχι μόνο των κοινοβουλευτικών εκπροσώπων. Η «ώρα του πρωθυπουργού», όπως ονομάστηκε, ήταν στην πραγματικότητα πεντάλεπτες τοποθετήσεις σε δύο ερωτήσεις. Οι απαντήσεις απαιτούσαν συνήθως τεκμηρίωση και αρκετή δουλειά. Η ώρα του πρωθυπουργού γνώρισε διαφορετικές τύχες κατά τη διάρκεια της οκταετίας. Η αρχική περίοδος ήταν περίοδος πολλών ερωτήσεων, είτε από βουλευτές για θέματα της περιφέρειά τους είτε από βουλευτές που ήθελαν να αναδειχθούν κονταροχτυπούμενοι με τον πρωθυπουργό.

2) Automatic Summary:

Αλλά και οι βουλευτές έπαψαν να αλλάζουν στρατόπεδα ή απόψεις επικαλούμενοι τη συνείδησή τους από την εποχή που τα εργατικά κόμματα στην Ευρώπη άρχισαν να έχουν μαζική παρουσία και να επιβάλλουν ενιαία στάση στους εκπροσώπους τους. Η εικόνα έτσι σε κάθε κόμμα



είναι μία και ενιαία για όλα τα θέματα . Η πλειοψηφία ακολουθεί τις κυβερνητικές εισηγήσεις , σε αντίθεση δε περίπτωση δημιουργείται μείζον πρόβλημα , το οποίο όμως κυβέρνηση και κυβερνητική πλειοψηφία αποφεύγουν κατά κανόνα πάση θυσία . Ένα διαφορετικό αποτέλεσμα προϋποθέτει τη δημιουργία μιας διαφορετικής πλειοψηφίας . Συνήθως η έκρηξη αυτή συμβαίνει σε θέματα που αφορούν προσωπικά τους ζητήματα , όπως το επαγγελματικό ασυμβίβαστο ή ο τρόπος εκλογής τους . Η ψήφος κατά συνείδηση σωστά τονίζεται από το σύνταγμα , διότι έτσι κάθε βουλευτής είναι υπεύθυνος των αποφάσεών του και οι αποφάσεις της Βουλής δεν μπορούν να αμφισβητηθούν ως προϊόν πίεσης . Ο βουλευτής και το κόμμα έχουν έτσι την ευθύνη να δείξουν ποιες αρχές και ποιες πολιτικές υπερασπίζονται , όταν η ελευθερία γνώμης του βουλευτή συγκρούεται με την κομματική πειθαρχία . Πρέπει να δρουν ταχύτητα και με τρόπο ώστε το μήνυμά τους να μεταδίδεται αμέσως . Παρέχει επίσης στους αρχηγούς των κομμάτων και στους κοινοβουλευτικούς εκπροσώπους υπερβολικά πολύ χρόνο ομιλίας σε σχέση με αυτό που ισχύει στην Ευρωβουλή και σε άλλα κοινοβούλια της Ευρωπαϊκής Ένωσης . Η πληροφόρηση που διαθέτουν οι βουλευτές από πριν , ο δημόσιος διάλογος στα ΜΜΕ , η δυνατότητα των πολιτών να αντλήσουν και από άλλες πηγές πέραν των ομιλιών την αναγκαία γνώση πιέζουν για τεκμηριωμένες και συνοπτικές ομιλίες . Ο κανονισμός της Βουλής προβλέπει τη διεξαγωγή συζητήσεων μεταξύ του πρωθυπουργού και των προέδρων των κομμάτων της αντιπολίτευσης για τα θέματα της εξωτερικής πολιτικής , για τις σχέσεις με την Ευρωπαϊκή Ένωση αλλά και για άλλα θέματα γενικότερου ενδιαφέροντος . Ωστόσο για μένα υπήρξαν ευπρόσδεκτη ευκαιρία για να εξηγήσω τους στόχους μας και να αναλύσω την πορεία της οικονομίας και των εξωτερικών μας σχέσεων . Το μεγάλο κοινό , που δεν παρακολουθούσε ολόκληρη τη συζήτηση , αποκτούσε ως εκ τούτου περιορισμένη εικόνα . Προσπάθησα για αυτό σε συνεργασία με τον πρόεδρο της Βουλής κ. Α. Κακλαμάνη να τροποποιηθεί ο κανονισμός της Βουλής , ώστε να περιοριστούν οι απανωτές δευτερολογίες και τριτολογίες μεταξύ των δύο πρωταγωνιστών και να παρεμβάλλονται οι επικεφαλής των άλλων κομμάτων καθώς και ο αρμόδιος για το θέμα υπουργός . Η νέα ρύθμιση προέβλεπε ο πρωθυπουργός να απαντά πολύ πιο συχνά , κάθε εβδομάδα , σε ερωτήσεις οιοδήποτε βουλευτή και όχι μόνο των κοινοβουλευτικών εκπροσώπων . Οι απαντήσεις απαιτούσαν συνήθως τεκμηρίωση και αρκετή δουλειά .

II

1) Original text:

Λέγομαι Μπάζιλ Έλτον και είμαι φύλακας του φάρου του Βόρειου Σημείου που ο πατέρας και ο παππούς μου φυλούσαν πριν από μένα . Ο γκριζός φάρος υψώνεται μακριά από την ακτή , πάνω σε βυθισμένους γλιστερούς βράχους που φαίνονται όταν η παλίρροια χαμηλώνει και εξαφανίζονται όταν τα νερά ανεβαίνουν . Εδώ και έναν αιώνα , τα μεγαλόπρεπα καράβια των επτά θαλασσών περνάνε μπροστά από αυτό το φανάρι . Στις μέρες του παππού μου ήταν πάρα πολλά , στις μέρες του πατέρα μου ήταν λιγότερα και τώρα είναι τόσο λίγα που μερικές φορές αισθάνομαι τόσο μόνος σαν να ήμουν ο τελευταίος άνθρωπος που έχει απομείνει στον πλανήτη μας .

Από μακρινές ακτές έφταναν τα παλιά μεγάλα εμπορικά ιστιοφόρα με τα άσπρα πανιά , έρχονταν από τις μακρινές ακτές της Ανατολής , όπου έλαμπαν ζεστοί ήλιοι και γλυκά αρώματα περιπλανιόνταν πάνω από παράξενους κήτους και χαρούμενους ναούς . Οι παλιοί καπετάνιοι της θάλασσας έρχονταν συχνά στον παππού μου και του μιλούσαν για αυτά τα πράγματα , αυτός με τη σειρά του τα έλεγε στον πατέρα μου και ο πατέρας μου σε εμένα τα απογεύματα του φθινοπώρου όταν ο άνεμος σύρλιαζε αλλόκοτα από την Ανατολή . Στα βιβλία που μου έδιναν οι άνθρωποι όταν ήμουν νέος και γεμάτος απορίες , έχω διαβάσει για αυτά και για άλλα ακόμη .

Αλλά πιο θαυμαστή από τη γνώση των παλαιών ανθρώπων και από τη γνώση των βιβλίων είναι η μυστική γνώση του ωκεανού . Μπλε , πράσινο , γκριζό , άσπρο ή μαύρο , λειός , αναστατωμένος ή με κύματα ψηλά σαν βουνά , αυτός ο ωκεανός δεν είναι σιωπηλός . Όλες μου τις μέρες τον παρατηρώ και τον αφογκράζομαι και τον ξέρω καλά . Στην αρχή μου έλεγε μόνο απλές μικρές ιστορίες για ήρεμες ακρογιαλιές και κοντινά λιμάνια . Αλλά όσο περνούσαν τα χρόνια έγινε πιο φιλικός και μιλούσε για άλλα πράγματα , πράγματα πιο παράξενα σε χώρο και χρόνο . Μερικές φορές στο αμυδρό φως της αυγής οι γκριζοί ατμίοι του ορίζοντα υποχωρούσαν για να με ανταμείψουν με φευγαλέες ματιές σε αυτά που ήταν πιο πέρα , και μερικές φορές τη νύχτα τα βαθιά νερά γίνονταν καθαρά και φωσφορίζοντα για να με ανταμείψουν με φευγαλέες ματιές σε αυτά που ήταν από κάτω .

Αυτές οι φευγαλέες ματιές αφορούσαν τόσο συχνά τα πράγματα που υπήρξαν όσο και τα πράγματα που θα μπορούσαν να υπάρχουν , γιατί ο ωκεανός είναι πιο αρχαίος από τα βουνά και είναι φορτωμένος με τις μνήμες των ονείρων και του Χρόνου .

Το άσπρο καράβι συνήθιζε να έρχεται από το Νότο , όταν το φεγγάρι ήταν γεμάτο και ψηλά στον ουρανό . Γλιστρούσε πολύ απαλά και σιωπηλά μέσα στη θάλασσα από τη μεριά του Νότου . Είτε η θάλασσα ήταν αργιμεμένη είτε ήρεμη , είτε ο άνεμος ήταν φιλικός είτε ανάποδος , πάντα γλιστρούσε απαλά και σιωπηλά με τα πανιά του να αχνοφαινούνται και τις μακριές , παράξενες σειρές των κουπιών του να κινούνται ρυθμικά . Μια νύχτα διέκρινα πάνω στο κατάστρωμα έναν άνδρα με γενειάδα και μανδύα , που φαινόταν να μου έγνεφε να επιβιβαστώ για μακρινές και άγνωστες ακτές . Αργότερα τον είδα πολλές φορές κάτω από το γεμάτο φεγγάρι , και πάντοτε μου έγνεφε .

Το φεγγάρι ήταν πολύ λαμπερό τη νύχτα που απάντησα στο κάλεσμα και περπάτησα πάνω από τα νερά σε μια γέφυρα από φεγγαροακτίνες προς το άσπρο καράβι . Ο άνδρας που μου είχε γνήσει με καλωσόρισε τώρα σε μια απαλή γλώσσα που καταλάβαινα πολύ καλά . Καθώς γλιστρούσαμε μακριά προς τον μυστηριώδη Νότο , χρυσαφένιο από τη λάμψη του ολόγιμου μελωμένου φεγγαριού , οι ώρες κυλούσαν απαλά σαν τραγούδια των κωπηλατών .

Όταν η μέρα ξημέρωσε ροδαλή ή φωτεινή , είδα τις πράσινες ακτές μακρινών χωρών , λαμπρών και υπέροχων , άγνωστων σε μένα . Προβλήτες για τους εμπόρους διακοσμημένες με δέντρα υψώνονταν αγέρωχα πάνω από τη θάλασσα και εδώ και εκεί φαινόταν η ανταύγεια από τις άσπρες σκεπές και τις κιονοστοιχίες παράξενων ναών . Καθώς πλησιάζαμε προς την πράσινη ακτή , ο άνδρας με τη γενειάδα μου μίλησε για αυτή τη χώρα , τη χώρα της Ζαρ , όπου κατοικούν όλα τα όνειρα και οι σκέψεις για την ομορφιά που έρχονται για μια φορά στους ανθρώπους και μετά ξεχνιούνται . Όπως κοιτούσα πάλι τις προβλήτες , είδα ότι αυτά που μου έλεγε ήταν αληθινά , γιατί ανάμεσα σε αυτά που έβλεπα



μπροστά μου ήταν πολλά πράγματα που είχα δει κάποτε μέσα από την ομίχλη, πέρα από τον ορίζοντα και μέσα στα φωσφορίζοντα βάθη του ωκεανού. Εκεί, επίσης, υπήρχαν μορφές και φαντασίες πιο θαυμαστές από σιδηρόπυραυλο είχα γνωρίσει ποτέ. Τα οράματα νεαρών ποιητών που πέθαναν από ανάγκη προτού ο κόσμος να μπορέσει να μάθει αυτά που είχαν δει και ονειρευτεί. Αλλά δεν πατήσαμε το πόδι μας πάνω στα κεκλιμένα λιβάδια της Ζαρ, γιατί λένε πως όποιος περπατήσει σε αυτά μπορεί να μην ξαναγυρίσει ποτέ στην πατρική του γη.

Καθώς το άσπρο καράβι έπλεε σιωπηλά, μακριά από τις γεμάτες ναούς πεζούλες της Ζαρ, είδαμε μπροστά μας στον μακρινό ορίζοντα τις κορυφές μιας μεγαλόπρεπης πόλης και ο άνδρας με τη γενειάδα μου είπε ότι αυτή είναι η Θαλαριόν, η Πόλη των Χιλιών Θαυμάτων, όπου κατοικούν όλα τα μυστήρια που ο άνθρωπος αγωνίζεται μάταια να κατανοήσει.

Κοίταξα πάλι από κοντά και είδα ότι η πόλη ήταν μεγαλύτερη από κάθε άλλη πόλη που είχα δει και είχα ονειρευτεί. Οι κορυφές των ναών της έφθαναν μέχρι τον ουρανό, έτσι που κανένας άνθρωπος δεν μπορούσε να δει την άκρη τους, και μακριά πέρα από τον ορίζοντα απλώνονταν τα σοβαρά γκρίζα τείχη της, πάνω από τα οποία μπορούσε κανείς να διακρίνει μόνο μερικές σκεπές, στολισμένες με πλούσιες κορνίζες και θελκτικά γλυπτά, αλλά παρ'όλα αυτά και δυσόσιονες.

Επιθυμούσα σφοδρά να πάω σε αυτή τη γοητευτική αλλά και αποθητική πόλη και παρακάλεσα τον άνδρα με τη γενειάδα να με αποβιβάσει στην προκουμία κοντά στην τεράστια σκαλιστή πύλη Ακαριέλ. Αλλά αρνήθηκε ευγενικά στην επιθυμία μου λέγοντας ότι «πολλοί έχουν πάει στην Θαλαριόν, την Πόλη των Χιλιών Θαυμάτων, αλλά κανένας δεν έχει γυρίσει από εκεί. Εκεί περπατούν δαίμονες και τρελά πράγματα που δεν είναι πια άνθρωποι και οι δρόμοι είναι άσπροι από τα άθραφα κόκαλα αυτών που κοίταζαν την εικόνα του Λάθι που βασιλεύει πάνω στην πόλη».

Έτσι το άσπρο καράβι συνέχισε τον πλου του πέρα από τα τείχη της Θαλαριόν και ακολουθούσε για πολλές μέρες ένα πουλί που πετούσε προς το Νότο του οποίου τα γυαλιστερά φτερά έμοιαζαν με τον ουρανό μέσα από όπου είχε εμφανιστεί. Μετά φτάσαμε σε μια ευχάριστη ακτή με πολύχρωμα λουλούδια όπου όσο έφτανε το μάτι μας βλέπαμε χαριτωμένα ηλιόλουστα άλση και αστραφτερά δέντρα κάτω από έναν ήλιο στη μεσημβρία. Από κοιλάδες που ήταν έξω από την όρασή μας έρχονταν ξεσπάσματα τραγουδιών και αποσπάσματα λυρικής αρμονίας, ανακατεμένα με σιγανά γέλια τόσο γλυκά που παρακινούσα τους κοπηλάτες να κάνουν γρήγορα από τη σφοδρή μου επιθυμία να φτάσω εκεί. Ο άνδρας με τη γενειάδα δεν έλεγε τίποτα, αλλά με παρατηρούσε καθώς πλησιάζαμε τη στρωμένη με κρίνους ακτή. Ξαφνικά, ένας άνεμος που φυσούσε από τα ανθισμένα λιβάδια και τα γεμάτα δέντρα δάση έφερε μια μυρωδιά που με έκανε να τρέμω. Ο άνεμος δυνάμωσε και ο αέρας γέμισε από αυτή τη θανατερή μυρωδιά τάφου, που έμοιαζε να έρχεται από τα ξεσκέπαστα νεκροταφεία και τις πόλεις που τις είχε χτυπήσει η πανούκλα. Καθώς πλέαμε σαν τρελοί μακριά από αυτή την καταραμένη ακτή, ο άνδρας με τη γενειάδα μίλησε τελικά και είπε ότι «αυτή είναι η Ξάρα, η γη των Απραγματοποίητων Απολαύσεων».

Έτσι για μια ακόμη φορά το άσπρο καράβι ακολουθούσε το πουλί του ουρανού σε ζεστές ευλογημένες θάλασσες όπου φυσούσαν χαϊδευτικά αρωματικές αύρες. Ημέρα με την ημέρα και νύχτα, ολοένα ταξιδεύαμε και, όταν το φεγγάρι ήταν γεμάτο, ακούγαμε το απαλό τραγούδι των κοπηλατών, τόσο γλυκό όσο εκείνη τη μακρινή νύχτα όταν φεύγαμε από την πατρική μου γη. Ήταν κάτω από το φως του φεγγαριού όταν τελικά αγκυροβόλησαμε στο λιμάνι της Σόνα Νιλ, το οποίο φυλάει δύο δίδυμα κρυστάλλινα ακρωτήρια που υψώνονται πάνω από την θάλασσα και συναντιούνται σε ένα υπέρλαμπρο τόξο. Αυτή είναι η γη της Φαντασίας. Αποβιβαστήκαμε στην κατάφυτη ακτή πάνω σε μια γέφυρα από φεγγαροακτίνες.

Στη γη της Σόνα Νιλ δεν υπάρχει ούτε χρόνος ούτε χώρος, ούτε πόνος, ούτε θάνατος. Εκεί έμεινα για πολλούς αιώνες. Στη Σόνα Νιλ τα δάση και τα λιβάδια είναι πράσινα, τα λουλούδια λαμπερά και ευωδιαστά, τα ρυάκια γάργαρο και γαλανά, οι πηγές καθαρές και δροσερές και μεγαλόπρεποι και θαυμαστοί οι ναοί, τα κάστρα και οι πόλεις. Σε αυτήν τη γη δεν υπάρχουν σύνορα πέρα από κάθε επίπεδο ομορφιάς υψώνεται ένα άλλο ακόμη πιο όμορφο. Ο ευτυχισμένος λαός κινείται ελεύθερα στην εξοχή και ανάμεσα στη λαμψη των πόλεων, και όλοι οι άνθρωποι είναι προικισμένοι με αδιατάρακτη χάρη και ανόθευτη ευτυχία. Αιώνες κατοικούσα εκεί και περιπλανιόμουν μακάρια σε κήπους και σε άσπρα μονοπάτια από συνόρευν με διακριτικές μυρωδιές. Ανέβαινα πάνω σε ήρεμους λόφους από όπου μπορούσα να δω γοητευτικά πανοράμια πόλεων με καμπαναριά που φάλιαζαν αναπαικτικά σε καταπράσινες κοιλάδες και χρυσαφένιους θόλους από γιγάντιες πολιτείες που έλαμπαν στον άπειρο μακρινό ορίζοντα. Στο φως του φεγγαριού έβλεπα την αστραφτερή θάλασσα, τα κρυστάλλινα ακρωτήρια και το ήρεμο λιμάνι όπου ήταν αγκυροβολημένο το άσπρο καράβι.

Μια νύχτα με γεμάτο φεγγάρι, του αμνημόνευτου χρόνου Θαρπ, είδα τη σιλουέτα του ουρανού πουλιού που μου έγνεφε και αισθάνθηκα τα πρώτα σκιρτήματα της ανησυχίας. Μίλησα στον άνδρα με τη γενειάδα και του είπα για την καινούρια μου επιθυμία να ξεκινήσω για τη μακρινή Καθουρία που κανένας άνθρωπος δεν έχει δει, αλλά που όλοι πιστεύουν ότι βρίσκεται πάνω από τις βασάλτινες κολόνες της Δύσης.

Είναι η Γη της Ελπίδας και εκεί λάμπουν τα τέλεια ιδανικά όλων όσα γνωρίζουμε ή τουλάχιστον έτσι λένε: «Να είσαι προσεκτικός με αυτές τις επικίνδυνες θάλασσες που λένε ότι βρίσκεται η Καθουρία. Στη Σόνα Νιλ δεν υπάρχει πόνος ή θάνατος αλλά ποιος μπορεί να πει τι βρίσκεται πίσω από τις βασάλτινες κολόνες της Δύσης;». Ωστόσο επιβιβάστηκα στο άσπρο καράβι και μαζί με τον απρόθυμο άνδρα με τη γενειάδα άφησα το ευτυχισμένο λιμάνι για καινούριες, αταξίδευτες θάλασσες.

Το ουράνιο πουλί πετούσε μπροστά και μας οδηγούσε προς τις βασάλτινες κολόνες της Δύσης, αλλά αυτή τη φορά οι κοπηλάτες δεν τραγουδούσαν απαλά τραγούδια κάτω από το γεμάτο φεγγάρι. Συχνά έφτιαχνα με το μυαλό μου την εικόνα της άγνωστης γης της Καθουρίας με θαυμαστά δάση και παλάτια και αναρωτιόμουν τι καινούργιες απολαύσεις με περιμέναν εκεί. «Η Καθουρία», έλεγα στον εαυτό μου «είναι η κατοικία των θεών και η γη με τις αμέτρητες πόλεις του χρυσού. Τα δάση της είναι από αλόη και σάνδαλο, όμορφα σαν τα ευωδιαστά δάση του Καμορίν, και ανάμεσα στα δέντρα πετάνε χαρούμενα γλυκόλαλα πουλιά. Στα πράσινα και ανθισμένα βουνά της Καθουρίας υψώνονται ναοί από ροζ μάρμαρο, πλούσιοι σε σκαλισμάτα και στολίδια και έχουν στις αυλές τους δροσερές πηγές από ασήμι, όπου τα αρωματισμένα νερά που έρχονται από τον ποταμό Ναργκ κελαρύζουν με συναρπαστικές μελωδίες. Οι πόλεις της Καθουρίας είναι περικυκλωμένες από χρυσαφένια τείχη και τα πεζοδρόμιά τους είναι επίσης από χρυσό. Στους κήπους τους υπάρχουν παράξενες ορχιδέες και αρωματισμένες λίμνες που οι όχθες τους είναι περικυκλωμένες από κοράλλι και κεχρμπάρι. Τη νύχτα οι δρόμοι και οι κήποι φωτίζονται από χαρούμενα φανάρια φτιαγμένα από τρίχρωμο δστρακο της χελώνας και αντηχούν οι απαλές νότες του τραγουδιού και του λαγουτάρι. Τα σπίτια στις πόλεις της Καθουρίας είναι όλα παλάτια και είναι χτισμένα πάνω από ευωδιαστά κανάλια στα οποία τρέχουν τα νερά του ιερού Ναργκ. Είναι όλα χτισμένα από μάρμαρο και πορφύριτη και οι σκεπές τους είναι από λαμπερό χρυσάφι που αντανακλά τις ακτίνες του ήλιου και επαυξάνει το μεγαλείο των πόλεων καθώς οι ευδαίμονες θεοί τις κοιτάζουν από τις μακρινές τους κορυφές. Ωραιότερο από όλα είναι το παλάτι του μεγάλου



μονάρχη Ντόριεμπ και έχει πολλούς πυργίσκους από μάρμαρο πάνω στους τοίχους του. Μεγάλα πλήθη μαζεύονται στις τεράστιες αίθουσές του όπου κρέμονται τα τρόπαια όλων των εποχών. Η σκεπή είναι από καθαρό χρυσάφι στηριγμένη σε ψηλές κολόνες από ρουμίνη και αζουρίτη, γεμάτες με σκαλισμένες φηγούρες θεών και ηρώων, έτσι που όποιος κοιτάζει νομίζει ότι αντικρίζει το ζωντανό Όλυμπο. Και τα πατώματα του παλατιού είναι από γυαλί, κάτω από το οποίο κυλάνε χαρούμενα τα έντεχνα φωτισμένα νερά του Ναργκ, γεμάτα πλουμιστά ψάρια που δεν είναι γνωστά πέρα από τα σύνορα της αξιολάμπης Καθουρίας. Με αυτόν τον τρόπο μιλούσα στον εαυτό μου για την Καθουρία αλλά ο άνδρας με τη γενειάδα συνέχεια με προειδοποιούσε να γυρίσουμε πίσω στις ευτυχισμένες ακτές της Σόνα Νιλ γιατί η Σόνα Νιλ είναι γνωστή στους ανθρώπους, ενώ την Καθουρία δεν την έχει ποτέ κανένας αντικρίσει.

Την τριακοστή πρώτη μέρα που ακολουθούσαμε το πουλί είδαμε τις βασάλτινες κολόνες της Δύσης. Ήταν τυλιγμένες σε ομίχλη, έτσι που κανένας άνθρωπος δεν μπορούσε να κοιτάξει πέρα από αυτές ή να δει τις κορυφές τους, οι οποίες μερικοί λένε ότι φθάνουν μέχρι τους ουρανούς. Ο άνδρας με την γενειάδα με εκλιπαρούσε να γυρίσουμε πίσω, αλλά αρνήθηκα γιατί από την ομίχλη πέρα από τις κολόνες μου φάνηκε ότι άκουσα τις νότες των τραγουδιών και των λαγουτάρηδων, γλυκύτερες και από τα γλυκύτερα τραγούδια της Σόνα Νιλ, που τραγουδούσαν τα δικά μου εγκώμια, τα δικά μου, εμένα που είχα ταξιδέψει από πολύ μακριά, από το γεμάτο φεγγάρι και κατοικούσα στη γη της Φαντασίας. Έτσι το άσπρο καράβι ακολούθησε τους μελωδικούς ήχους και αρμένιζε προς την ομίχλη ανάμεσα στις βασάλτινες κολόνες της Δύσης. Και όταν σταμάτησε η μουσική και διαλύθηκε η ομίχλη, δεν είδαμε τη Γη της Καθουρίας αλλά μια ακατανίκητη θάλασσα πάνω στην οποία το αβοήθητο σκάφος μας πήγαινε προς κάποιον άγνωστο σκοπό. Σύντομα ήρθε στα αυτιά μας ο μακρινός ήχος των νερών που πέφτουν και στα μάτια μας εμφανίστηκε στο μακρινό ορίζοντα μπροστά ο τιτάνιος αφρός ενός τερατώδους καταρράκτη, εκεί όπου οι ωκεανοί του κόσμου χύνονται στην άβυσσο του τίποτα. Τότε ο άνδρας με τη γενειάδα με δάκρυα στα μάτια μου είπε: «Αρνήθηκαμε την όμορφη γη της Σόνα Νιλ που δε θα ξαναδούμε ποτέ. Οι θεοί είναι μεγαλύτεροι από τον άνθρωπο και νίκησαν».

2) Automatic Summary:

Λέγομαι Μπάζιλ Έλτον και είμαι φύλακας του φάρου του Βόρειου Σημείου που ο πατέρας και ο παππούς μου φυλούσαν πριν από μένα. Εδώ και έναν αιώνα, τα μεγαλόπρεπα καράβια των επτά θαλασσών περνάνε μπροστά από αυτό το φανάρι. Στις μέρες του παππού μου ήταν πάρα πολλά, στις μέρες του πατέρα μου ήταν λιγότερα και τώρα είναι τόσο λίγα που μερικές φορές αισθάνομαι τόσο μόνος σαν να ήμουν ο τελευταίος άνθρωπος που έχει απομείνει στον πλανήτη μας. Όλες μου τις μέρες τον παρατηρώ και τον απουγκράζομαι και τον ξέρω καλά. Μερικές φορές στο αμυδρό φως της αυγής οι γκριζοί ατμοί του ορίζοντα υποχωρούσαν για να με ανταμείψουν με φευγαλέες ματιές σε αυτά που ήταν πιο πέρα, και μερικές φορές τη νύχτα τα βαθιά νερά γίνονταν καθαρά και φωσφορίζοντα για να με αν με φευγαλέες ματιές σε αυτά που ήταν από κάτω. Το άσπρο καράβι συνήθιζε να έρχεται από το Νότο, όταν το φεγγάρι ήταν γεμάτο και ψηλά στον ουρανό. Μια νύχτα διέκρινα πάνω στο κατάστρωμα έναν άνδρα με γενειάδα και μανδύα, που φαινόταν να μου έγνεφε να επιβιβαστώ για ακρινές και άγνωστες ακτές. Το φεγγάρι ήταν πολύ λαμπερό τη νύχτα που απάντησα στο κάλεσμα και περπάτησα πάνω από τα νερά σε μια γέφυρα από φεγγαροακτίνες προς το άσπρο καράβι. Όταν η μέρα ξημέρωσε ροδαλή ή φωτεινή, είδα τις πράσινες ακτές μακρινών χωρών, λαμπρά και υπέροχων, άγνωστων σε μένα. Καθώς πλησιάζαμε προς την πράσινη ακτή, ο άνδρας με τη γενειάδα μου μίλησε για αυτή τη χώρα, τη χώρα της Ζαρ, όπου κατοικούν όλα τα όνειρα και οι σκέψεις για την ομορφιά που έρχονται για μια φορά στους ανθρώπους και μετά ξεχνιούνται. Εκεί, επίσης, υπήρχαν μορφές και φαντασίες πιο θαυμαστές από οτιδήποτε είχα γνωρίσει ποτέ. Καθώς το άσπρο καράβι έπλεε σιωπηλά, μακριά από τις γεμάτες ναούς πεζούλες της Ζαρ, είδαμε μπροστά μας στον μακρινό ορίζοντα τις κορυφές μιας μεγαλόπρεπης πόλης και ο άνδρας με τη γενειάδα μου είπε ότι αυτή είναι η Θαλαριόν, η Πόλη των Χιλίων Θαυμάτων, όπου κατοικούν όλα τα μυστήρια που ο άνθρωπος αγωνίζεται μάταια να κατανοήσει. Οι κορυφές των ναών της έφθαναν μέχρι τον ουρανό, έτσι που κανένας άνθρωπος δεν μπορούσε να δει την άκρη τους, και μακριά πέρα από τον ορίζοντα απλώνονταν τα σοβαρά γκριζά τείχη της, πάνω από τα οποία μπορούσε κανείς να διακρίνει μόνο μερικές σκεπές, στολισμένες με πλούσιες κορνίζες και θεατρικά γλυπτά, αλλά παράξενες και δυσόσιωνες. Εκεί περπατούν δαιμόνες και τρελά πράγματα που δεν είναι πια άνθρωποι και οι δρόμοι είναι άσπροι από τα άθραφα κόκαλα αυτών που κοίταζαν την εικόνα του Λάθι που βασιλεύει πάνω στην πόλη». Από κοιλάδες που ήταν έξω από την όρασή μας έρχονταν ξεσπάσματα τραγουδιών και αποσπάσματα λυρικής αρμονίας, ανακατεμένα με σιγανά γέλια τόσο γλυκά που παρακινούσα τους κωπηλάτες να κάνουν γρήγορα από τη σφοδρή μου επιθυμία να φτάσω εκεί. Ξαφνικά, ένας άνεμος που φυσούσε από τα ανθισμένα λιβάδια και τα γεμάτα δέντρα δάση έφερε μια μυρωδιά που με έκανε να τρέμω. Έτσι για μια ακόμη φορά το άσπρο καράβι ακολούθησε το πουλί του ουρανού σε ζεστές ευλογημένες θάλασσες όπου φυσούσαν χαϊδετικά αρωματικές αύρες. Ήταν κάτω από το φως του φεγγαριού όταν τελικά αγκυροβολήσαμε στο λιμάνι της Σόνα Νιλ, το οποίο φυλάνε δύο δίδυμα κρυστάλλινα ακρωτήρια που υψώνονται πάνω από την θάλασσα και συναντιούνται σε ένα υπέρλαμπρο τόξο. Στη γη της Σόνα Νιλ δεν υπάρχει ούτε χρόνος ούτε χώρος, ούτε πόνος, ούτε θάνατος. Στη Σόνα Νιλ τα δάση και τα λιβάδια είναι πράσινα, τα λουλούδια λαμπερά και ευωδιαστά, τα ρυάκια γάργαρα και γαλανά, οι πηγές καθαρές και δροσερές και μεγαλόπρεποι και θαυμαστοί οι ναοί, τα κάστρα και οι πόλεις. Αιώνες κατοικούσα εκεί και περιπλανιόμουν μακάρια σε κήπους και σε άσπρα μονοπάτια που συνόρευαν με διακριτικές μυρωδιές. Στο φως του φεγγαριού έβλεπα την αστραφτερή θάλασσα, τα κρυστάλλινα ακρωτήρια και το ήρεμο λιμάνι όπου ήταν αγκυροβολημένο το άσπρο καράβι. Μίλησα στο άνδρα με τη γενειάδα και του είπα για την κανούρια μου επιθυμία να ξεκινήσω για τη μακρινή Καθουρία που κανένας άνθρωπος δεν έχει δει, αλλά που όλοι πιστεύουν ότι βρίσκεται πάνω από τις βασάλτινες κολόνες της Δύσης. Είναι η Γη της Ελπίδας και εκεί λάμπουν τα τέλεια ιδανικά όλων όσα γνωρίζουμε ή τουλάχιστον έτσι λένε: «Να είσαι προσεκτικός με αυτές τις επικίνδυνες θάλασσες που λένε ότι βρίσκεται η Καθουρία. Συχνά έφτιαχνα με το μυαλό μου την εικόνα της άγνωστης γης της Καθουρίας με θαυμαστά δάση και παλάτια και αναρωτιόμουν τι κανούργιες απολαύσεις με περιέμεναν εκεί. Στα πράσινα και ανθισμένα βουνά της Καθουρίας υψώνονται ναοί από ροζ μάρμαρο, πλούσιοι σε σκαλισμένα και στολίσια και έχουν στις αυλές τους δρ πηγές από ασήμι, όπου τα αρωματισμένα νερά που έρχονται από τον ποταμό Ναργκ κελαρύζουν με συναρπαστικές μελωδίες. Στους κήπους τους υπάρχουν παράξενες ορχιδέες και αρωματισμένες λίμνες που οι όχθες τους είναι περικυκλωμένες από κοράλλι και κεχρμπάρι. Τα σπίτια στις πόλεις της Καθουρίας είναι όλα παλάτια και είναι χτισμένα πάνω από ευωδιαστά κανάλια των οποία τρέχουν τα νερά του ιερού Ναργκ. Μεγάλα πλήθη μαζεύονται στις τεράστιες αίθουσές του όπου κρέμονται τα τρόπαια όλων των εποχών. Και τα πατώματα του παλατιού είναι από γυαλί, κάτω από το οποίο κυλάνε χαρούμενα τα έντεχνα φωτισμένα νερά του Ναργκ, γεμάτα πλουμιστά ψάρια που δεν είναι γνωστά πέρα από τα σύνορα της αξιολάμπης Καθουρίας». Ήταν τυλιγμένες σε ομίχλη, έτσι που κανένας άνθρωπος δεν μπορούσε να κοιτάξει πέρα από αυτές ή να δει τις κορυφές τους, οι οποίες μερικοί λένε ότι φθάνουν μέχρι τους ουρανούς. Έτσι το άσπρο καράβι ακολούθησε τους μελωδικούς ήχους και αρμένιζε προς την ομίχλη ανάμεσα στις βασάλτινες κολόνες της Δύσης. Τότε ο άνδρας με τη γενειάδα με δάκρυα στα μάτια μου είπε: «Αρνήθηκαμε την όμορφη γη της Σόνα Νιλ που δε θα ξαναδούμε ποτέ».



Polish

1) Original text:

Przed laty żyli sobie król i królowa, którzy powtarzali codziennie: - Ach, gdybyśmy mieli dziecko! - a dziecka wciąż nie mieli. Zdarzyło się pewnego razu, kiedy królowa zażywała kąpieli, że z wody wyskoczyła na brzeg żaba i w te słowa do niej przemówiła: - Twoje pragnienie spełni się, nim rok minie, wydasz na świat córeczkę.

I stało się, jak przepowiedziała żaba, królowa urodziła dziewczynkę tak urodziwą, że król nie posiadał się z radości i wydał wspaniałą ucztę. Zaprosił nie tylko krewnych, przyjaciół i znajomych, ale również wróżki, aby sprzyjały dziecku i otaczały je opieką. Było ich trzynaście w całym królestwie, ale że w pałacu posiadano tylko dwanaście złotych talerzy, na których miały jeść, jedna z nich musiała zostać w domu.

Po niezwykle wystawnej uczcie wróżki poczęły składać noworodkowi dary: jedna obdarzyła królową cnotami, druga urodą, trzecia bogactwem, tak że w końcu otrzymała ona wszystko, co tylko można sobie wymarzyć na tym świecie. Kiedy jedenaście wróżek poczyniło już swoje zaklęcia, wkroczyła nagle do komnaty trzynasta. Chciała się zemścić za to, że nie została zaproszona, i nie witając się z nikim ani na nikogo nie patrząc, zawołała na cały głos: - Niechaj królowa w piętnastym roku życia ukuje się wrzecionem i umrze. Powiedziawszy to zawróciła i opuściła salę. Wszyscy obecni stali przerażeni, wtedy wystąpiła dwunasta wróżka, która nie wymieniła jeszcze swego życzenia, a że nie mogła odwrócić złowrogiego zaklęcia, tylko je złagodzić, rzekła: - Niech to nie będzie śmierć, lecz stuletni głęboki sen, w jaki królowa zapadnie.

Król, który pragnął uchronić ukochane dziecko przed nieszczęściem, wydał rozkaz, aby wszystkie wrzeciona w całym jego państwie zostały spalone. Dary, jakimi wróżki obsypały dziewczeczkę, spełniły się co do joty: królowa była taka śliczna, cnotliwa, miłutka i roztropna, że wystarczyło na nią spojrzeć, by ją pokochać. Tak się złożyło, że w dniu, kiedy królowa kończyła piętnaście lat, ani króla, ani królowej nie było w domu i dziewczynka została w pałacu zupełnie sama. Chodziła sobie po całym gmachu, zaglądała swobodnie do wszystkich komnat i pokoi, aż stanęła u stóp starej wieży. Wspięła się po kręconych schodach na górę i znalazła się przed maleńkimi drzwiczkami. W zamku tkwił zardzewiały klucz, a kiedy królowa go przekreśliła, drzwi otworzyły się gwałtownie, w izdebce zaś siedziała stara kobieta z wrzecionem i przędła pilnie len.

- Dzień dobry, babciu - rzekła królowa - co ty tu robisz? - Przędę - odparła staruszka, kiwając głową. - A co to tak śmiesznie podskakuje? - spytała dziewczynka, wzięła do ręki wrzeciono i chciała także prząść. Ledwie dotknęła wrzeciona, spełniła się zła przepowiednia i ukuła się nim w palec. A w chwili gdy poczuła ukłucie, padła na łóżko stojące obok i pograżyła się w głębokim śnie. Ten sen ogarnął również cały pałac: król i królowa, którzy wrócili właśnie do domu i wkraczali na salę, zasnęli wraz z całym swym dworem. Zasnęły konie w stajni, psy na podwórzu, gołębie na dachu, muchy na ścianie, a nawet ogień płonący w kuchennym piecu zgasł i usnął, a pieczeń przestała się dusić, i kucharz, który chciał pociągnąć za włosy kuchcika, bo coś przeskrobał, puścił jego czuprynę i zasnął. I wiatr ucichł, a liście drzew na dziedzińcu znieruchomiały.

Wokół pałacu zaczął rosnąć cierniowy żywopłot, który z każdym rokiem stawał się coraz wyższy i wreszcie zarósł cały pałac, tak że nie było go już wcale widać, znikła nawet chorągiewka na dachu. Po kraju zaś rozeszła się wieść o śpiącej królowej Różyczce, bo tak ją nazywano, i coraz to jakiś królewicz przybywał i chciał przedrzeć się przez gęsty żywopłot. Żadnemu się to jednak nie udało, bowiem ciernie powstrzymywały każdego, jakby miały ręce, i młodzieńcy, wczepieni w nie, nie mogli się już wyzwolić i umierali żalną śmiercią. Po wielu, wielu latach przybył znów do kraju tego młody królewicz. Słyszał on, jak jakiś starzec opowiada o cierniowym żywopłocie, za którym ukryty jest pałac, a w nim od stu lat śpi prześliczna królowa, zwana Różyczką, wraz z nią zaś śpi król i królowa, i wszyscy dworzanie. Od swego dziadka wiedział również, że wielu już królewiczów próbowało przedrzeć się przez ów żywopłot, ale zawiśli na nim i zginęli smutną śmiercią. - Ja się niczego nie boję - rzekł młodzieniec - udam się do zaczarowanego pałacu i zobaczę śliczną Różyczkę! Poczciwy starzec odradzał mu, jak mógł, ale on wcale go nie słuchał.

A wtedy właśnie minęło sto lat i nadszedł dzień, w którym śpiąca królowa miała się zbudzić. Kiedy królewicz zbliżył się do żywopłotu, było on pokryty wielkimi wspaniałymi kwiatami, które same się przed nim rozstały, wpuszczając go do środka, a potem zwały się znowu w gęsty żywopłot. Na dziedzińcu pałacowym królewicz ujrzał uśpione konie i także psy myśliwskie, na dachu siedziały gołębie i każdy miał lepek schowany pod skrzydełko. A kiedy przekroczył próg pałacu, muchy spały na ścianach, kucharz stał z wyciągniętą ręką, jakby chciał chwycić kuchcika za czuprynę, a dziewczka kuchenna siedziała, trzymając na kolanach czarną kurę, którą chciała oskubać. Poszedł dalej i zobaczył wszystkich dworzaków śpiących na sali, a obok tronu leżeli uśpieni król i królowa. Królewicz szedł coraz dalej, a cisza dokoła była tak głęboka, że słyszał własny oddech, i wreszcie dotarł do wieży i otworzył drzwi izdebki, gdzie spała królowa. I ujrzał dziewczeczkę tak piękną, że nie mógł oderwać od niej oczu, pochylił się i pocałował ją. Po tym pocałunku królowa otworzyła oczy, zbudziła się i spojrzała na niego z uroczym uśmiechem. Zeszli potem razem na dół, a wtedy zbudził się król, królowa i cały dwór i jedni patrzyli na drugich ze zdumieniem. Konie na podwórku podniosły się potrząsając grzywami, psy myśliwskie jęły skakać i machać ogonami, gołębie na dachu wyciągnęły główki spod skrzydełek, rozejrzały się i pofrunęły w pole, muchy na ścianach poruszyły się, ogień pod kuchnią zapłonął i zaczął gotować jedzenie, pieczeń zaskwierczała, kucharz trzepnął kuchcika w ucho, a dziewczka kuchenna oskubała do końca kurę.

Po czym odbyło się huczne wesele królewicza z królową Różyczką i oboje żyli długo i szczęśliwie.

2) Automatic Summary:

Zdarzyło się pewnego razu, kiedy królowa zażywała kąpieli, że z wody wyskoczyła na brzeg żaba i w te słowa do niej przemówiła: - Twoje pragnienie spełni się, nim rok minie, wydasz na świat córeczkę. Zaprosił nie tylko krewnych, przyjaciół i znajomych, ale również wróżki,



aby sprzyjały dziecku i otaczały je opieką. Było ich trzynaście w całym królestwie, ale że w pałacu posiadano tylko dwanaście złotych talerzy, na których miały jeść, jedna z nich musiała zostać w domu. Król, który pragnął uchronić ukochane dziecko przed nieszczęściem, wydał rozkaz, aby wszystkie wrzeczona w całym jego państwie zostały spalone. Tak się złożyło, że w dniu, kiedy królowna kończyła piętnaście lat, ani króla, ani królowej nie było w domu i dziewczynka została w pałacu zupełnie sama. Wspięła się po kręconych schodach na górę i znalazła się przed maleńkimi drzwiczkami. A w chwili gdy poczuła ukłucie, padła na łóżko stojące obok i pogrążyła się w głębokim śnie. Zasnęły konie w stajni, psy na podwórzu, gołębie na dachu, muchy na ścianie, a nawet ogień płonący w kuchennym piecu zgasł i usnął, a pieczeń przestała się dusić, i kucharz, który chciał pociągnąć za włosy kuchcika, bo coś przeszkrobał, puścił jego czuprynę i zasnął. Żadnemu się to jednak nie udało, bowiem ciernie powstrzymywały każdego, jakby miały ręce, i młodzieńcy, wczepieni w nie, nie mogli się już wyzwolić i umierali żalosną śmiercią. Słyszał on, jak jakiś starzec opowiada o cierniowym żywopłocie, za którym ukryty jest pałac, a w nim od stu lat śpi prześliczna królowna, zwana Różyczką, wraz z nią zaś śpią król i królowa, i wszyscy dworzanie. A wtedy właśnie minęło sto lat i nadszedł dzień, w którym śpiąca królowna miała się zbudzić. Kiedy królewicz zbliżył się do żywopłotu, było on pokryty wielkimi wspaniałymi kwiatami, które same się przed nim rozstały, wpuszczając go do środka, a potem zwarły się znowu w gęsty żywopłot. Poszedł dalej i zobaczył wszystkich dworzan śpiących na sali, a obok tronu leżeli uśpieni król i królowa. I ujrzał dziewczeczkę tak piękną, że nie mógł oderwać od niej oczu, pochylił się i pocałował ją. Konie na podwórku podniosły się potrząsając grzywami, psy myśliwskie jęły skakać i machać ogonami, gołębie na dachu wyciągnęły główki spod skrzydełek, rozejrzały się i pofrunęły w pole, muchy na ścianach poruszyły się, ogień pod kuchnią zapłonął i zaczął gotować jedzenie, pieczeń zaskwierczała, kucharz trzepnął kuchcika w ucho, a dziewczka kuchenna oskubała do końca kurę.

Romanian

I

1) Original text:

Când Marvin a împlinit zece ani, tatăl l-a dus prin coridoarele lungi și răsunătoare care urcau prin Administrativ și Energetic, până au ajuns în sfârșit la nivelurile cele mai de sus, prin vegetația dezvoltându-se cu rapiditate, a Fermelor. Lui Marvin îi plăcea aici; era amuzant să urmărească plantele uriașe și subțiri cum se târau cu o ardoare aproape vizibilă spre lumina soarelui care, filtrată prin cupolele de plastic, cobora pentru a le întâlni. Mireasma vieții se simțea la tot pasul, deșteptându-i în inimă doruri inexprimabile; acum nu mai respira aerul uscat și răcoros specific nivelurilor rezidențiale, purificat de orice arome, cu excepția izului ușor înțepător de ozon. Dorea să mai poată întârzia aici, dar tata nu l-a lăsat. Au mers mai departe, până la intrarea Observatorului, pe care nu-l vizitase niciodată; totuși nu s-au oprit și Marvin a știut, cu o senzație de excitație nestăpânită, că mai rămăsesse o singură destinație posibilă. Pentru prima dată în viață, avea să iasă Afară.

În sala imensă de întreținere, existau o duzină de vehicule de teren, înzestrate cu anvelope-balon late și cabine presurizate. Probabil că tata era așteptat, fiindcă au fost conduși de îndată la mașinuța de cercetare de lângă gigantică ușă circulară a ecluzei. Încordat de nerăbdare, Marvin a luat loc în cabina strămtă, în vreme ce tata a pornit motorul și a verificat comenzile. Ușa interioară a ecluzei a glisat deschizându-se, apoi s-a închis înapoia lor; a auzit reducându-se treptat vuietul pompelor de aer, pe măsură ce presiunea scădea spre zero. După aceea, s-a aprins indicatorul "Vid", ușa exterioară s-a deschis, și înaintea lui Marvin a apărut tărâmul în care nu pătrunsesse niciodată.

Îl văzuse, desigur, în fotografii: privise de o sută de ori astfel de imagini pe ecranele de televiziune. Acum, însă, realitatea se întindea în jurul său, arzând sub soarele aprig care se târa atât de lent peste cerul mai negru decât cărbunele. S-a uitat spre vest, cât mai departe de splendoarea orbitoare a soarelui... și a zărit stelele, așa cum i se spusese, dar n-o crezuse niciodată cu adevărat. Le-a cercetat mult timp, minunându-se că puteau fi atât de luminoase și totuși atât de mici. Erau niște punctulețe nemișcate, ce nu păl-pâiau, și brusc și-a reamintit câteva versuri citite într-o carte a tatei:

Steaua sus răsare

Ca o taină mare.

Ei bine, el cunoștea taina stelelor. Cel care făcuse comparația aceea nu fusese prea inteligent... A abandonat nedumeririle și și-a concentrat din nou atenția asupra peisajului din jur.

Goneau cu aproape o sută șazececi de kilometri pe oră, peste un șes neted, anvelopele-balon ridicând înapoia lor norișori de praf. Colonia nu se mai zărea nicăieri; în cele câteva minute cât privise stelele, domurile și turnurile radio dispăruseră sub orizont. Existau totuși alte semne ale prezenței umane, deoarece, la mai puțin de doi kilometri în față, Marvin putea distinge structurile cu forme bizare, adunate în jurul intrării unei mine. La răstimpuri, un fuier de vânturi se ridica dintr-un coș scund, dispersându-se instantaneu.

Au trecut pe lângă mină în câteva secunde; tata conducea cu o îndemănare îndrăzneată și părea preocupat, de parcă — era un gând straniu



pentru un copil — dorea să scape de ceva care-l săcâia. Peste alte câteva minute, au ajuns la marginea platoului pe care fusese construită Colonia. Terenul cobora abrupt sub ei, într-o pantă amețitoare al cărei capăt se pierdea în umbră. În față, cât vedeai cu ochii, se întindea un ținut pustiu în care se învâlmășeau cratere, lanțuri muntoase și râpe. Crestele munților, reflectând soarele aflat jos pe boltă, ardeau aidoma unor insule de foc într-un ocean de întuneric, iar deasupra lor, stelele continuau să sclipească la fel de imobile.

Mai departe nu se putea înainta... și totuși, exista un drum. Marvin și-a încheștat pumnii când vehiculul s-a aplecat peste muchie și a început lunga coborâre. După aceea, a zărit poteca și s-a destins puțin. Se părea că și alți oameni mai trecuseră pe aici.

Noaptea s-a lăsat șocant de brusc, atunci când au traversat marginea umbrei și soarele a dispărut înapoia coamei platoului. Perechea de faruri s-a aprins imediat, proiectând dungi alb-albastre pe stâncile din față, așa încât nu mai era nevoie să verifice viteza cu care mergeau. Au călătorit ore întregi prin văioage și pe lângă poalele unor munți ale căror piscuri păreau că ating stelele, ieșind uneori, pentru câteva clipe, în lumina soarelui, atunci când terenul suia.

Acum, în dreapta, se zărea un șes cutat și colbuit, iar în stânga, cu meterezele și terasele ridicându-se preț de mulți kilometri spre cer, se întindea un perete de munți ce mășăluiau în depărtare, până ce vârfurile lor se pierdeau sub marginea lumii. Nu se vedea nici un semn care să ateste că oamenii exploraseră cândva aceste ținuturi, totuși vehiculul a trecut la un moment dat pe lângă epava unei rachete, alături de care se afla o movilă de pietre având deasupra o cruce din metal.

Băiatului i se părea că munții își continuau marșul la nesfârșit, dar, în cele din urmă, după multe ore, aceștia se terminară printr-un promontoriu uriaș și prăpăstios, ce se înălța abrupt din mijlocul unui pâlcc de dealuri. Mașina a coborât într-o vale puțin adâncă, ce se arcuia într-o curbă prelungă spre versantul opus al lanțului muntos, și atunci Marvin a început să înțeleagă că în fața lor se petrecea ceva foarte ciudat.

Soarele se găsea acum jos, înapoia dealurilor din dreapta, așadar valea ce se afla drept înainte ar fi trebuit să fie cufundată în beznă. Cu toate acestea, ea era scăldată într-o strălucire rece și albă, ce se revărsa peste țarcurile pe sub care treceau. Apoi, pe neașteptate, au ieșit în șes deschis și sursa luminii s-a ivit în toată splendoarea ei.

Acum, după oprirea motoarelor, în micuța cabină pogorâse liniștea. Unicele zgomote erau susurul slab al alimentatorului cu oxigen și ocazionala trepidație metalică produsă de pereții exteriori ai vehiculului, care-și împrăștia fluxul termic. Nici un pic de căldură nu venea dinspre imensa seceră argintie care plutea imediat deasupra orizontului îndepărtat, inundând ținutul cu o lumină perlată. Era atât de strălucitoare încât, abia după câteva minute, Marvin i-a putut accepta provocarea și a privit direct în sclipirea ei; totuși, în cele din urmă, a reușit să distingă contururile continentelor, limita cețoasă a atmosferei și insulele albe ale norilor. Și chiar de la acea depărtare, a văzut scânteierea razelor de soare pe banchizele polare.

Spectacolul era minunat și-i atrăgea sufletul peste abisul spațiului. Acolo, în semiluna luminoasă, se găseau toate minunile pe care el nu le cunoscuse niciodată: nuanțele amurgurilor, oftatul oceanului pe nisipul plajelor, răpăitul ploii, binecuvântarea lină a ninsorii. Ele și altele o mie ar fi trebuit să reprezinte moștenirea sa legitimă, însă băiatul le cunoștea doar din cărți și din înregistrări vechi, iar gândul acesta îi umplea inima cu tristețea unui surghiunit.

De ce nu se mai puteau întoarce? Totul părea atât de pașnic sub bancurile de nori călători... Apoi, ochii lui Marvin s-au acomodat cu lumina orbitoare și băiatul a observat că porțiunea discului aflată în beznă sclipea palid, cu o fosforescență malignă... și și-a amintit: privea rugul funerar al unei planete, epilogul radioactiv al Apocalipsei. La patru sute de mii de kilometri distanță, lucirea atomilor muri-bunzi continua să fie vizibilă, aidoma unui memento etern al trecutului funest. Aveau să treacă multe secole înainte ca strălucirea ucigașă să dispară din piatră, iar viața să poată invada din nou lumea tăcută și pustie.

După aceea, tata a început să vorbească, depănând istoria care până atunci nu însemnase pentru Marvin mai mult decât basmele auzite în copilărie. Multe lucruri nu izbutea să le înțeleagă; îi era imposibil să-și imagineze configurația multicoloră a vieții de pe planeta pe care n-o văzuse nicicând. Nu putea pricepe nici forțele care o distruseră în final, lăsând în urma ei un singur supraviețuitor, Colonia, care scăpase grație izolării ei. Cu toate acestea, a reușit să-și imagineze agonia acelor zile, când "coloniștii" aflaseră că navele de aprovizionare n-aveau să mai coboare niciodată pe cozi de foc printre stele, aducând daruri de acasă. Una câte una, stațiile radio își încetaseră apelurile; pe globul adumbrat, luminițele orașelor păliseră și muriseră, iar ei rămăseseră singuri în cele din urmă, mai singuri decât fusese vreodată un grup de oameni, purtând în mâini-le lor viitorul rasei.

Urmaseră apoi anii de disperare și necontenita bătălie pentru a supraviețui pe o lume ostilă și neimblânzită. Bătălia aceea fusese câștigată, deși cu greutate; micuța oază de viață se afla în siguranță, înfruntând cele mai rele lucruri pe care le putea face Natura. Totuși, dacă nu exista un țel, un viitor căruia să i se dedice, Colonia avea să-și piardă voința de a supraviețui, și atunci nici o mașină, nici o artă și nici o știință n-ar mai putea-o salva.

Abia atunci Marvin a priceput rolul pelerinajului. El n-avea să se plimbe niciodată pe malurile râurilor planetei legendare și pierdute, nici n-avea să audă tunetele rostogolindu-se peste coamele rotunjite ale dealurilor ei. Într-o bună zi — cât de îndepărtată însă? — copiii copiilor lui urmau să se întoarcă pentru a-și revendica moștenirea. Până atunci, vânturile și ploile vor fi curățat otrăvurile din solurile părjolite, purtându-le către oceane, iar în adâncurile apelor veninul se va fi diluat, nemaifiind în stare să facă nici un rău vieții. Abia în acea zi, uriașele nave care continuau să aștepte aici, pe șesurile tăcute și colbuite, aveau să se înalțe iarăși în spațiu, pe drumul spre casă.

Acesta era visul; iar în altă zi, ceva mai apropiată — Marvin a știut-o printr-o străfulgerare de intuiție — el urma să-l transmită fiului său, aici, exact în același loc, cu munții înapoia lor și cu lumina argintie a cerului scâldându-le chipurile.

Nu a privit îndărăt când s-au întors și au pornit către Colonie, îndreptându-se spre cei de-un neam cu el, aflați în lungul exil, băiatul nu rezista să vadă încă o dată măreția înghețată a secerii Pământului pâlind pe stâncile din jur...



2) Automatic Summary:

Când Marvin a împlinit zece ani , tatăl l- a dus prin coridoarele lungi și răsunătoare care urcau prin Administrativ și Energetic , până au ajuns în sfârșit la nivelurile cele mai de sus , prin vegetația dezvoltându- se cu rapiditate , a Fermelor . Lui Marvin îi plăcea aici ; Dorea să mai poată întârzia Au mers mai departe , până la intrarea Observatorului , Pentru prima dată în viață , avea să iasă Afară . Pro - babil Încordat de nerăbdare , Marvin a luat loc în cabina strâmtă , în vreme ce tata a pornit motorul și a verificat comenzile . După aceea , s- a aprins indicatorul " Vid " , ușa exterioară s- a deschis , și înaintea lui Marvin a apărut tărâmul Îl văzuse , desigur , S- a uitat spre vest , cât mai departe de splendoarea orbitoare a soarelui ... Le- a cer - cetat mult timp , minunându- se că puteau fi atât Ei bine , el cunoștea taina stelelor . A abandonat nedumeririle și și- a concentrat din nou atenția asupra peisajului din jur . Goneau cu aproape o sută șazeci de kilometri pe oră , peste un șes neted , anvelopele- balon ridicând înapoia lor norișori de praf . La răstimpuri , un fuier de va - pori se ridica dintr- un coș scund , dispersându- se instantaneu . Au trecut pe lângă mină în câteva secunde ; tata conducea cu o îndemănare îndrăzneată și părea preocupat , de parcă — era un gând straniu pentru un copil — dorea să scape În față , cât vedeai cu ochii , se întindea un ținut pustiu Mai departe nu se putea înainta ... și totuși , exista un drum . Noaptea s- a lăsat șocant Perechea de faruri s- a aprins imediat , proiectând dungi alb-albastre pe stâncile din față , așa încât nu mai era nevoie să verifice viteza Au călătorit ore întregi prin văioage și pe lângă poa - lele unor munți Acum , în dreapta , se zărea un șes cutat și colbuit , iar în stânga , cu meterezele și terasele ridicându- se preț de mulți kilometri spre cer , se întindea un pe de munți ce mărsăluiau Nu se vedea nici un semn care să ateste Băiatului i se părea că munții își continuau marșul la , Mașina a coborât într- o vale puțin adâncă , Soarele se găsea acum jos , înapoia dealurilor din dreapta , Apoi , pe neaștept - tate , au ieșit Acum , după oprirea motoarelor , în micuța cabină pogorâse liniștea . Unicele zgomote erau susurul slab al alimentatorului cu oxigen și ocazionala trepidație metalică produsă de pereții exteriori ai vehiculului , Nici un pic de căldu - ră nu venea dinspre imensa seceră argintie Era atât totuși , în cele din urmă , a reușit să distingă contururile continentelor , limita cețoasă a atmosferei și insulele albe ale norilor . Spectacolul era minunat și- i atrăgea sufletul peste abisul spațiului . Acolo , Ele și altele o mie ar fi trebuit să reprezinte moștenirea sa legitimă , De ce nu se mai puteau întoarce ? și și- a amintit : privea rugul fune al unei planete , epilogul radioactiv al Apocalip - sei . La patru sute de mii de kilometri distanță , lucirea atomilor muri - bunzi continua să fie vizibilă , aidoma unui memento etern al trecutului funest . După aceea , tata a început să vorbească , depănând istoria Multe lucruri nu izbutea să le înțeleagă ; îi era imposibil să -și imagineze configurația multicoloră a vieții de pe pla - neta Cu toate acestea , a reușit să -și imagineze agonia acelor zile , Una câte una , stațiile radio își încetaseră apelurile ; Urmaseră apoi anii de disperare și neconținută bătălie pentru a supraviețui pe o lume ostilă și neîmblanzită . Bătălia aceea fusese câștigată , deși cu greutate ; micuța oază de viață se afla în siguranță , înfruntând cele mai rele lucruri Abia atunci Marvin a priceput rolul pelerinajului . Într- o bună zi — cât iar în adâncurile apelor veninul se va fi diluat , nemaifiind Acesta era visul ; iar în altă zi , ceva mai apropiată — Marvin a știut -o printr- o străfulgerare de intuiție — el urm să -l transmită fiului său , Nu a privit îndărăt .

II

1) Original text:

A fost odată , în vremurile de demult , a fost un împărat și- o împărăteasă , și cum nu se îndurase soarta să le hărăzească un urmaș , nu trecea zi în care să nu se tânguie amândoi : "Cât de goală ne e casa și ce fericiți am fi de- am avea și noi un copil!" Dar coconul cel mult dorit se lăsa așteptat de multă vreme...

Și iată că- ntr- un sfârșit , pe când împărăteasa se scălda în apa unui râu , o broască sări din unde pe prundișul malului și- i grăi astfel :

- Află , Măria ta , că-ți va fi îndeplinită dorința ; căci până a nu trece anul , vei aduce pe lume o fetiță .

Și se împlini întocmai ceea ce spusese broasca ; împărăteasa născu o fetiță , și atâta era de frumoasă fetița , că împăratul nu- și mai încăpea în piele de bucurie . Cum voia să se împărtășească și alții din bucuria lui , Măria sa dădu o mare petrecere , la care pofți nu numai rudele , prietenii și cunoscuții , ci și ursitoarele , ca să le câștige toată bunăvoința și grija față de soarta copilei .

În toată țara aceea erau treisprezece ursitoare , dar pentru că împăratul nu avea decât douăsprezece talere de aur în care să le servească bucatele , una din ele nu fu pofțită la ospăț . Petrecerea se prăznu cu mare strălucire și , când fu să se încheie , ursitoarele înzestrară copila cu cele mai alese daruri : una îi hărăzi să se bucure de virtute , a doua , de frumusețe , a treia , de bogăție și așa , pe rând , ursitoarele îi dăruiră tot ce- și poate dori omul pe lumea asta . În clipa când cea de- a unsprezecea ursitoare tocmai își sfârșea urarea , numai ce intră valvărtej cea de- a treisprezecea . Pasămite , venise să se răzbune pentru că nu fusese pofțită și ea la serbare . Și , fără ca măcar să arunce cuiva vreo privire , rostii cu un glas tunător :

- Când va împlini fata cincisprezece ani , să se- nșepe c- un fus și să moară !

Apoi nu mai spuse nici un cuvânt și , întorcând spatele la cei de față , părăsi sala tronului . Toți încremeniseră de spaimă , dar cea de- a douăsprezecea ursitoare , care nu- i rostise încă urarea , se apropie de leagăn și , fiindcă nu- i sta în putință să ridice blestemul , ci doar să- l mai îndulcească , glăsui așa :

- Dar să nu fie moartă , ci să cadă într- un somn adânc , care să țină o sută de ani !

Împăratul , care dorea din tot sufletul s- o ferească pe iubita sa copilă de năpasta blestemului , dădu poruncă să se pună pe foc toate fusele din împărăție .

În ăst timp , fetița creștea și darurile cu care o înzestraseră ursitoarele își arătau roadele cu prisosință : era atât de frumoasă , de cuminte , de prietenoasă și de înțeleaptă , că oricine o vedea o îndrăgea pe dată .



Se întâmplă însă ca tocmai în ziua în care domnița împlinea cincisprezece ani împăratul și împărăteasa să fie duși de-acasă și fata să rămână singură în palat. Ca să-i treacă de urât, luă la rând încăperile; zăbovi ea prin tot felul de săli și odăi și colindă așa, după pofta inimii, ceasuri întregi, până ce ajunse și la un turn vechi. Urcă scara îngustă și întortocheată ce ducea sus și se pomeni în dreptul unei uși micuțe. În broască era vărâtă o cheie ruginită și când o răsuci, ușa sări în lături și fetei îi fu dat să vadă într-o cămăruță o femeie bătrână, bătrână, care sta și torcea cu sânguință un fuior de in.

- Bună ziua, bătrânico, spuse domnița, da' ce faci aici?

- Iaca, torc și eu! răspunse bătrâna, dând din cap.

- Da' ce naiba-i aia de se răsuțește așa de repede? întrebă fata și, luând fusul în mână, încercă și ea să toarcă.

Dar de-abia-l atinse, că se și împlini blestemul ursitoarei și domnița se înțepă la un deget.

În clipa în care simți înțepătura, fata căzu pe patul care se afla acolo și se cufundă într-un somn adânc. Și somnul ăsta cuprinse întreg palatul: împăratul și împărăteasa, care tocmai atunci se întorseseră acasă și intraseră în sala tronului, adormiră pe loc și o dată cu ei adormi întreaga curte. Adormiră și caii în grajd, și câinii prin cotloanele curții, și hulubii pe acoperiș, și muștele pe pereți; ba chiar și focul care ardea în vatră își potoli vâpaia și se stinse cu totul. Friptura încetă de-a mai sfârâi, iar bucătarul, care tocmai se pregătea să-și apuce de chică ucenicul, fiindcă făcuse o boroboată, îi dădu drumul și adormi și el. Chiar și vântul își opri suflarea și nici o frunzuliță nu se mai mișcă în copacii din preajma palatului.

Și ce să vezi: de jur-împrejurul palatului începu să crească un tufiș de mărăcini, ca un gard viu! An de an, mărăcinișul se înălța tot mai mult și, în cele din urmă, cuprinse toți pereții, până la căpriori, ba se mai întinse și pe deasupra, de nu mai puteai vedea palatul defel, nici chiar steagul de pe acoperiș.

A mers vestea-n lume despre frumoasa adormită, căci așa o numeau toți pe tânăra domniță; și la câte-un răstimp se găsea câte-un fiu de crai care încerca să străbată prin tufișul de mărăcini și să pătrundă până la palat.

Dar de răzbit n-a răzbit nici unul, din pricină că mărăcinii se prindeau laolaltă ca și când ar fi avut niște mâini ghimpoase și nu-i lăsau să înainteze măcar un pas.

După amar de ani, prin acele meleaguri iată că mai veni un fecior de împărat. Și întâlnindu-se el cu un bătrân, acesta-i povesti despre gardul de mărăcini care împrejmuia palatul de amar de ani... Și-i mai zise că-n palat o fiică de crai, frumoasă cum nu se mai află alta pe lume, dormea somn adânc, de vreo sută de ani. Și că tot de atunci dorm și împăratul și împărăteasa și toți curtenii... Și mai știa bătrânul, de la bunicu-său, că se perindaseră pe-acolo mulți feciori de împărat care încercaseră să răzbească la palat prin tufărișul de mărăcini, dar că toți s-au pierdut fără urmă, agățați printre spini, murind de-o moarte cumplită.

Atunci tânărul îi vorbi cu semeție:

- Mie nu mi-e teamă defel și află c-am să mă duc s-o văd pe frumoasa adormită, chiar de-ar fi să-mi pierd viața!

Și oricât încercă bătrânul cel omenos să-l abată pe flăcău de la acest gând, nu izbuti, căci acesta era neînduplecat în hotărârea lui.

Se nimeri însă că tocmai atunci se împlinea cei o sută de ani și sosise ziua în care frumoasa adormită trebuia să se deștepte din somnul ei lung. Când feciorul de împărat se apropie de tufărișul de mărăcini, mai-mai nu-i veni să-și creadă ochilor, căci mărăcinii se prefăcuseră pe dată în puzderie de flori mari și frumoase, ce se dădeau la o parte din fața lui, lăsându-l să treacă nevătămat. Și după ce trecea el, cărarea se strângea din nou și florile se preschimbau iarăși într-un zid ghimpos, de nestrăbătut. În curtea palatului văzu o mulțime de cai și niște ogari care dormeau pe lespezile de piatră. Pe acoperiș, hulubii dormeau cu căpșoarele vârate sub aripi... De-ndată ce intră în palat, văzu același lucru: muștele dormeau pe pereți, iar în bucătărie bucătarul ținea mâna întinsă, ca și cum ar fi vrut să apuce de chică pe ajutorul său, în timp ce o slujnică ședea pe-un scăunel, având în față o găină neagră, pe care abia apucase s-o jumulească de câteva pene, cu o sută de ani în urmă.

Și feciorul de crai merse mai departe, până ce ajunse în sala tronului. Aici îi găsi pe toți curtenii dormind un somn ca de plumb iar mai în fund, sus, pe tron, dormeau împăratul și împărăteasa. Colindă el mereu alte și alte încăperi, fără să se oprească măcar o clipă. Și tot palatul era cufundat într-o tăcere așa de adâncă, că-ți puteai auzi răsuflarea. În cele din urmă, flăcăul ajunse în turn și deschise ușa cămăruței în care dormea tânăra domniță. Și era atât de frumoasă în somnul ei, că feciorul de împărat nu-și mai putu lua ochii de la ea și, aplecându-se, o sărută.

Dar de îndată ce-o atinse cu buzele, frumoasa adormită clipi din gene, deschise ochii și-l privi cu drag. Coborâră apoi amândoi din turn și, ca la un semn, împăratul și împărăteasa se treziră și ei și tot atunci se treziră toți curtenii. Se priveau unii pe alții, mirați, și mai-mai că nu le venea să-și creadă ochilor. Caii din curte se ridicară și începură să-și scuture coama; ogarii săriră sprinten în sus, dând vesel din coadă și gudurându-se; hulubii de pe acoperiș scoaseră căpșoarele de sub aripi și, rotindu-și privirile în zare, își luară zborul spre câmpie; muștele începură să se miște pe pereți bătăind întruna; focul din bucătărie se dezmoști și-și întepi vâlvătaia; mâncarea de pe plită porni să clocotească; friptura prinse iar să sfârâie; bucătarul trase o palmă zdravănă ucenicului său, care prinse să țipe ca din gură de șarpe; iar slujnica se apucă de jumulit găina și n-o mai lăsă din mână până nu-i smulse toate penele...

Apoi se prăznuie cu mare strălucire și alai nunta feciorului de împărat cu frumoasa adormită și amândoi trăiră fericiți până la sfârșitul zilelor.



2) Automatic Summary:

A fost odată , și- o împărăteasă , și cum nu se îndurase soarta să le hărăzească un urmaș , nu trecea zi și ce fericiți am fi de- am avea și noi un copil ! Și iată că -ntr-un sfârșit , pe când împărăteasa se scălda Și se împlini întocmai ceea ce spusese broasca ; Cum voia să se împărtășească și alții din bucuria lui , Măria sa dădu o mare petrecere , În clipa când cea de- a unsprezecea ursitoare tocmai își sfârșea urarea , numai ce intră valvărteje cea de- a treisprezecea . Pasămite , venise să se răzbune pentru că nu fusese poflită și ea la serbare . Și , fără ca măcar să arunce cuiva vreo privire , rosti cu un glas tunător : - Apoi nu mai spuse nici un cuvânt și , întorcând spatele la cei de față , părăsi sala tronului . Toți încremeniseră Împăratul , care dorea din tot sufletul s- o ferească pe iubita sa copilă de năpasta blestemului , dădu poruncă să se pună pe foc toate fusele din împărăție . În ăst timp , fetița creștea și darurile cu care o înzestraseră ursitoarele își arătau roadele cu prisosință : era atât Se întâmplă Ca să -i treacă de urât , luă la rând încăperile ; Urcă scara îngustă și întortocheată - Bună ziua , bătrânico , spuse domnița , - Iaca , torc și eu ! răspuse bătrâna , dând din cap . - Da' ce naiba -i aia întrebă fata și , luând fusul în mână , încercă și ea să toarcă . Dar de- abia -l atinse , Adormiră și caii în grajd , și câinii prin cotloanele curții , și hulubii pe acoperiș , și muștele pe pereți ; ba chiar și focul care ardea în vatră își potoli văpaia și se stinse cu totul . Chiar și vântul își opri suflarea și nici o frunzuliță nu se mai mișcă în copacii din preajma palatului . A mers vestea -n lume despre frumoasa adormită , căci așa o numeau toți pe tânăra domniță ; și la câte -un răstimp se găsea câte -un fiu de crai Dar de răzbit n- a răzbit nici unul , din pricină că mărăcinii se prindeau laolaltă ca și când ar fi avut niște mâini ghimpoase și nu -i lăsau să înainteze măcar un pas . Și că tot de atunci dorm și împăratul și împărăteasa și toți curtenii ... Și mai știa bătrânul , de la bunicu -său , Atunci tânărul îi vorbea cu semeție : - Mie nu mi- e teamă defel și află c- am să mă duc s- o văd pe frumoasa adormită , chiar de- ar fi să -mi pierd viața ! Și oricât încercă bătrânul cel omenos să -l abată pe flăcău de la acest gând , nu izbuti , căci acesta era neînduplecat în hotărârea lui . Se nimeri Când feciorul de împărat se apropie Și după ce trecea el , cărarea se strângea din nou și florile se preschimbau iarăși într- un zid ghimpos , de nestrăbătut . În curtea palatului văzu o mulțime de cai și niște ogari Pe acoperiș , hulubii dormeau cu căpșoarele vârate sub aripi ... De -ndată ce intră Și feciorul de crai merse mai departe , până ce ajunse în sala tronului . Aici îi găsi pe toți curtenii dormind un somn Și tot palatul era cufundat într- o tăcere așa Coborâră apoi amândoi din turn și , ca la un semn , împăratul și împărăteasa se treziră și ei și tot atunci se treziră toți curtenii . Se priveau unii pe alții , mirați , și mai- mai că nu le venea să -și creadă ochilor . muștele începură să se miște pe pereți bâzâind întruna ; focul din bucătărie se dezmoști și- și înteți valvătaia ; mâncarea de pe plită porni să clocotească ; friptura prinse iar să sfârșie ; bucătarul trase o palmă zdravănă ucenicului său , Apoi se prăznui cu mare strălucire și alai nunta feciorului de împărat cu frumoasa adormită și amândoi trăiră fericiți până la sfârșitul zilelor .