

Defining and Classifying Space Builders for Information Extraction

Barbara Gawronska, Björn Erlendsson, Niklas Torstensson

Dept. of Humanities and Information, University of Skövde, Box 408, 541 28 Skövde, Sweden

Abstract. The paper addresses the question of Information Extraction aimed at multilingual text generation, or text re-writing. This method provides an alternative to traditional Machine Translation, but is also related to text summarization. Given a source text, a re-writing system selects and structures the textual information in order to generate a “content report”. The present approach is inspired by recent IE-research, classical speech act theory, and Cognitive Semantics, especially the Theory of Mental Spaces and employed in an experimental system for understanding of news reports. The authors focus on the problem of identification and interpretation of ‘space builders’, i.e. linguistic signals for establishing mental spaces.

1 Introduction

The problem of finding relevant information in large number of texts has been attracting the attention of more and more NLP-researchers during the last decade. This field of research is most commonly referred to as Information Retrieval (IR) and Information Extraction (IE), but, as often is the case with quickly emerging and growing research areas, there is no exact consensus as to the terminology. An additional complication is the fact that certain terms are defined and used in different ways by computer scientists, computational linguists, and information researchers. Terms like Information Retrieval vs. Document/Text Retrieval, Information Extraction vs. Information Refinement, or Summarization vs. Abstracting, are in some contexts used as nearly synonymous, in other contexts – as related by subsumption. Moreover, terms referring to certain subfields of research are sometimes differentiated in respect

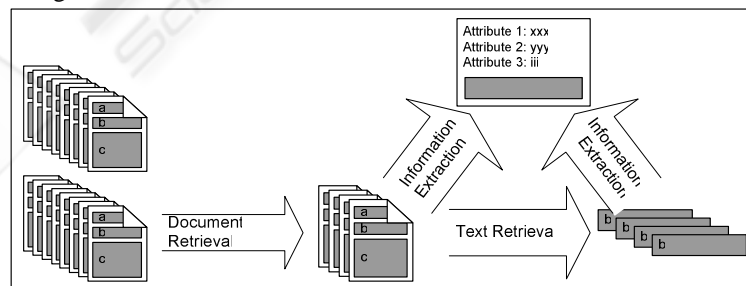


Fig. 1. Document and Text Retrieval in relation to Information Extraction

to methods and techniques: work on text summarization based on stochastic methods is often referred to as research on “extracting”, while work involving semantic and pragmatic representations is called “abstracting” or “reading comprehension” research.

It is not our ambition to clarify all the terminological discrepancies in the field, but we would like to shed some light on our use of the central terms, before we proceed with our approach to Information Extraction (IE) and Reading Comprehension. We limit ourselves to notions related to NLP.

We adhere to Cowie’s and Wilks’ definition of IE, stating that IE “is the name given to any process which selectively structures and combines data which is found, explicitly stated or implied, in one or more texts” [2].

This definition implies that some process aimed at finding the data to be structured and combined must precede the IE process. Practically, it means that, in most cases, a system traditionally labeled as an “IR-system” selects and clusters potentially relevant documents without investigating the semantic structure of the documents in depth. Since we (agreeing with Harabagiu et al. [10]) find the term Information Retrieval misleading, we prefer to call the preliminary selection process Document Retrieval and Text/Paragraph Retrieval (see Figure 1). The term IE is reserved for the process of selecting and structuring information found in the chosen documents or paragraphs. The terms ‘question answering’, ‘summarization’ and ‘re-writing’ (the latter to be discussed in Section 2) refer to the different ways of utilizing an IE process. The intended use has of course impact on the construction of a particular IE system.

2 ‘Re-writing’ or ‘Re-creation’ – Combining IE with Multilingual Text Generation

An application of IE-outputs that has not been mentioned to any greater extent in the NLP-literature – as far we know – is re-writing, or re-creation of a text. The idea comes from the theories of literature and translation ([28], [27]). Translation of a literary text normally requires some degree of re-creation, or “re-writing”: metaphors may need an adaptation to the target language (TL) culture, certain semantic changes may be necessary in order to preserve the rhythmic structure of the text, and so on.

A re-writing process may also apply to non-literary texts –. We define re-writing as a process aimed at understanding an input text and rendering the most important part of the text content in a way that is more comprehensible for the reader than the original text. The output text may, for example, be formulated in a syntactically less complex way than the input. If the input text and the output texts are written in different languages, re-writing may serve as an alternative to Machine Translation.

It can be of more use to obtain a comprehensible and pragmatically correct target language report that renders the content of a source language (SL) text without following the SL-texts syntactic and textual structure in detail, than to get a poor syntax-based translation. The re-writing process includes Information Extraction, and is very similar to summarization, but the output does not have to fulfill the formal demands on summaries formulated by evaluation fora of today [13]. The output may be considerably longer than 30% of the original text, depending on the nature of the found information. In other words, the Compression Ratio (the relation of the length of the target text to the length of the source text) is not central for evaluation of a re-written text, while the Retention Ratio (the relation of the information in the target text to the information of the source text) and readability are of crucial importance.

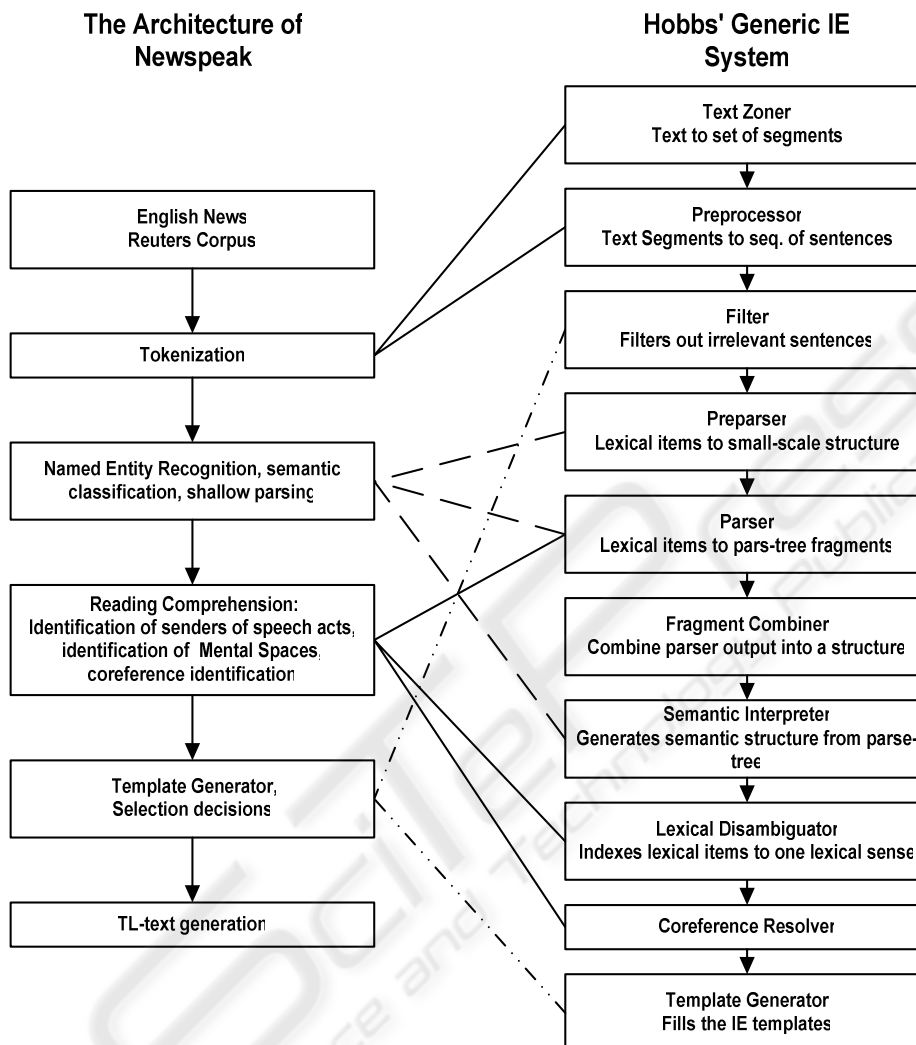


Fig. 2. The experimental system aimed at multilingual re-writing presented in relation to Hobbs' Generic IE System

The project presented below aims at multilingual re-writing of English news reports. In Figure 2, we show the general architecture of the system called Newspeak and relate it to the modules of a "Generic IE System" enumerated in Hobbs [12].

The main difference between Newspeak and the generic IE system lies in the fact that in Newspeak, the filtering process applies after the text has been analyzed by the Reading Comprehension module. Furthermore, the syntactic and semantic analysis is concentrated to one module.

3 The Theoretical Framework

The Reading Comprehension module in the Newspeak system is, as follows from the schema in Figure 2, based on Cognitive Semantics, in particular on Fauconnier's ([3], [4], [5]). The Mental Space Theory is of course not the only possible model of text comprehension, but, as it will be argued for below, it seems very suitable for processing the domain of news reports.

The current approaches to reading comprehension aimed at IE differ with regard to the terminology and to the details of their theoretical background ([19], [29]), mainly because of the characteristics of the texts to be summarized. 'Plot units' [18] and Rhetorical Structure Theory [20] inspired approaches aimed at discovering coherence relations ([21], [22], [8], [9]) are no doubt very successful when analyzing long narratives, where temporal, causative, and resemblance relations play a crucial role for the text structure. When extracting information from human-human dialogues, the game-theoretical approach appears to be fruitful – as in the Verbmobil system. For summarization of longer monologues, like politicians' speeches, more robust approaches, like 'squeezing', or 'compressing' [15] strategies, i.e. omitting optional syntactic phrases (mostly attributes), is quite efficient. The differences in approaches to summarization are no doubt psycholinguistically motivated, since the focus of human memorization strategies differ in different cognitive tasks: understanding a short message requires different neurolinguistic activity than understanding a long written narrative.

A striking characteristics of today's news reports is the presence of different versions of a certain event, different attitudes to the event and/or different hypotheses regarding the cause or the epistemic status of this event. These are normally encoded in the same, often quite short text, for example:

1. different versions of an event:

The Palestinians said the Israel Defense Forces had staged incursions into Hebron (...) and Tulkarem, killing one and injuring 10. *The Israel Defense Forces (IDF) had no immediate comment* on the accusation that troops had entered Tulkarem, and *strongly denied* there was an incursion at Hebron.

2. different actors express different attitudes

U.S. intelligence officials have received threats that terrorists will strike a U.S. nuclear power plant July 4. The government is taking the threats seriously, though officials have preliminary determined that the information is not credible enough to act upon...

Fauconnier's Mental Space Theory with its focus on counterfactuals, different epistemic modalities, and propositional attitudes, provides a very pertinent tool for analysis of this kind of structures. A good example of how this theory can be used for structuring information in news reports is given by Sanders and Redeker, although their analysis does not aim at any computerized application [30].

According to Fauconnier [3], Natural Language communication involves establishing different mental spaces, where the base space corresponds, roughly speaking, to the reality, perceived from the sender's perspective. Other, embedded, or 'dependent' mental spaces are set up for different "time periods, possible and impossible worlds, intentional states and propositional attitudes, epistemic and deontic modalities" [4]. Objects present in the base space may have counterparts in other spaces. The original objects and their counterparts do not have to share all features; the mapping between them can be partial. This claim suits the domain of news reports very well: it allows e.g. that coreference links can be established even when a certain

person is presented as a murderer in one version of an event, and as innocent person in another version.

Explicit signals for establishing a new mental space – so-called space-builders – are, according to the classical Mental Space Theory, time and place adverbials (*in 1963, in Canada* as opposed to *in England*), noun phrases referring to pictures and narratives (*in this story, in this movie*), conditional constructions, verbs denoting speech acts and verbs denoting mental activities, tense markers, and modality markers.

Implementing a computerized text comprehension module based on Fauconnier's theory requires, however, a more stringent definition of 'space-builders' and a strategy for discovering such words and phrases in the process of semantic analysis. The following section is devoted to this problem.

4 Defining and Classifying 'Space-builders'

The goals of the process we term 'Identification of Mental Spaces' in text processing are:

- to distinguish between the very news and the historical background of the reported event (the latter does not need to be rendered in the re-written version), as in: *India's defense minister Wenesday blamed Pakistan for the violence a day earlier that left at least 33 people dead and 40 wounded. (...) Authorities say that around 30,000 people have been killed during the campaign in the Muslim majority state.*

- to distinguish between different versions of an event (see example 1. above)

- to distinguish between 'real' events and hypothetical events (is the text about a terrorist attack that has happened or about a warning of an attack? Is the text about the results of an election, or only about a result prognosis?)

- to facilitate coreference resolution.

When designing a module that should divide and re-structure the input text into parts corresponding to the mental spaces, the definition of space-builders must be formulated in more detail, at least operationally. A crucial question to be answered is whether, and under which circumstances the categories enumerated by Fauconnier as possible space-builders actually introduce a new mental space. This returns us to the definition of a mental space itself.

4.1 Spatiotemporal Dimensions and Mental Spaces

Harder [11, p. 94] criticizes Fauconnier's claim about place and time adverbials like *in Canada* or *in 1963* as builders of mental spaces. Harder argues that "mental spaces, like the real world, can be assumed to have spatial and temporal dimensions inside them" and points out that it would be close to absurd to set up a new mental space for, let say, each birth date when compiling birth dates for the past fifty years [11]. He proposes that the main factor prompting space building is "potential contradiction". Harder is no doubt right in observing that the original definition of mental spaces is too generous and that it, in fact, formally allows computing a new mental space for every second. However, it seems intuitively wrong to banish all time and place adverbials from the category of space builders. A reader of a news report certainly draws a cognitive borderline between what happened yesterday in a given country, and what belongs to the historical background. The problem can be solved by taking the sender's and reader's perspective into account ([16], [32], [17]). In Harder's example, the

person interested in birth rates for the last fifty years sets up a mental space covering those 50 years, while for a news reporter, the base space normally is restricted to the last two or perhaps three days. We thus propose that a new mental space is set up either by virtue of potential contradiction, or when an event takes place outside the spatiotemporal scope of the base mental space. These limits depend on the sender's perspective and are, as a consequence, genre- and domain-dependent. For the purpose of processing news reports, we assume a temporal limit of maximum two days before the date of the report and a spatial limit corresponding to the country in question.

The next and quite complicated problem is to decide which words and phrases denoting acts of communication should be regarded as space-builders. This theoretical question is related to a practical one, namely, whether the existing lexical resources (here, we concentrate on WordNet – [24], [6]) can be used for identifying this subcategory of space-builders.

4.2 Speech Act Verbs as Space Builders

The starting point for the analysis of speech-act related space-builders was Searle's classification of speech acts. Table 1 is based on Searle's original classification [31] and its interpretation in Coulthard [1].

Table 1. Searle's speech act classification

Macro-class	Words-world relation	The psychological state of the sender	Sample verbs
Representatives	The speaker fits his words to the world	Belief that p	claim, announce, forecast, predict
Directives	Attempt to achieve a situation where the world fits to the words	Wanting that p	ask, beg, order, forbid, instruct
Commissives	Commit the speaker to act in order to fit the world to the words	Intending p	promise, offer, swear, threaten
Declarations	Alter the world		wed, baptize, name, call, dub
Expressives	No dynamic world-words relationship	Specified in the sincerity condition expressed by the prepositional content	thank, apologize, congratulate, regret, pardon

Our goal required a certain reformulation of Searle's model. The 'directives' and the 'commissives' could be regarded as one macro-class: 'builders' of hypothetical mental spaces. The 'representatives' required some further divisions. The criterion for our classification was not 'the psychological state of the sender' (since we cannot have access to the real psychological state or the real beliefs of the information sources), but 'the intended psychological state of the receiver (R)'. By this, we mean 1) whether the sender wants the receiver to believe that certain states-of affairs are true, false, or hypothetical, 2) whether the sender wants to impose a certain

evaluation of some event or person on the receiver. If at least one of these two criteria gets a positive value, the word/phrase is regarded as a potential space-builder.

It follows from criterions 1) and 2) that “declarations” that do not include any evaluation component (like typical Austinian performatives: *baptize*, *wed*, *dub*) are not treated as space-builders in our approach, while declarations that may involve subjective evaluation (*X called Y Z*) are regarded as openers of new mental spaces. This is of course disputable. Our argument in favor of this decision is the fact that subjective evaluation opens the possibility for potential contradiction between mental spaces, while typical “neutral” declaratives, like *wed* or *baptize* merely introduce performative speech events within a certain mental space.

A problematic group are phrases referring to utterance refusal, like *X declined to say*, *X neither confirmed nor denied*, *X had no comments on...* which are very frequent in news reports. We will return to this problem in Section 6, in the context of an authentic example.

5 Identification of ‘Space-builders’

The identification of spatiotemporal space-builders in news texts is relatively straightforward given access to a module for interpretation of time expressions (i.e. computing the names of the day of the week given a date etc.; a good tool for this task is available in the Delphi programming language, used for preprocessing in our project) and an ontology module representing the main geographical facts.. Interpreting spatiotemporal conditions is of course not entirely free from complications, but definitely less intricate than identifying and interpreting speech act related space-builders.

The main lexical resource for English used in the system is, as mentioned, WordNet (version 1.6). The general problem with the use of WordNet in NLP is, however, its frequently discussed fine-grainedness ([14], [33], [26], [23], [7], and many others). Inspired by the work of Mendes and Chaves, and also by the ideas by Montoyo et al., who propose labeling WordNet synsets by terms used in standard news agencies classification system, we decided to enrich the WordNet noun structures by identifying the telic hypernyms that were most salient in the domain of news reports. This has significantly reduced the lexical ambiguity with respect to concrete nouns ([26], [7]), but the problem of ambiguous interpretation of abstract nouns and verbs referring to communication acts still required a solution.

Using both the verb and the noun part of WordNet for semantic tagging generates extremely polysemous solutions, since the average number of synsets per noun is ca 4.3, and per verb – between 7 and 8 (e.g. *say* – the most frequent space-builder in news reports – is present under seven different top nodes in WordNet verb hierarchy).

Since about 80% of all nouns in WordNet are homonymous with verbs, we investigated whether it would be useful to use only the noun part of WordNet and a restricted list of most frequent speech act verbs which are not homonymous with nouns (*say*, *deny*, *confirm*, *inform*, *tell*, *declare*).

Two main nodes in the noun hierarchy seemed to be suitable for our purpose: one connected to the word speech-act and defined as “The use of language to perform some act” and the other one representing the noun statement and defined: “a message that is stated or declared; a communication (oral or written) setting forth particulars or facts etc”. When expanded, these two nodes cover a very large area of the vocabulary. To test the usefulness of this, and to find out if this could be of use for finding ‘space-builders’, all sub-nodes found under these two were expanded, and the classification was examined. The texts used for this task were retrieved from a part of Reuters’ corpus (from the categories disaster and accidents and war). The texts (totally

203.900 words) were semantically tagged using WordNet and the short verb list mentioned above. A total of 5771 words (234 lexemes) were classified as speech acts. In order to identify words which function as mental space openers, all instances tagged as speech-acts have been viewed by a human informant, and classified as either being or not being mental space openers, according to the criteria formulated in Section 4.2. Out of all word instances automatically tagged as speech acts, only 55.2%, were classified by human judges as possible space builders; 79% of the instances judged as potential space builders obtained the tag from the short additional verb list.

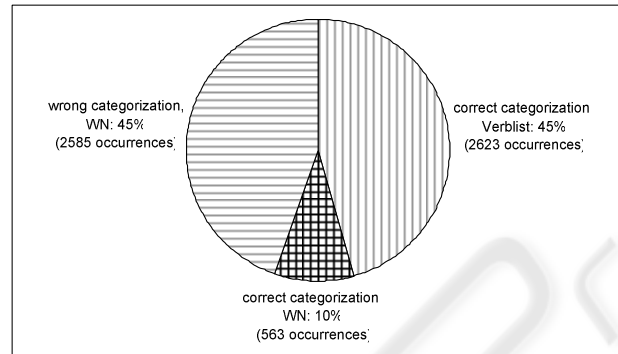


Fig. 3. Speech act classification based on WordNet (WN) and the system-internal verb lexicon with morphological rules

The conclusion to be drawn from this investigation was that WordNet hierarchies are not suitable for direct automatic identification and classification of speech act based space-builders. A better way to provide a base for space-builder identification is to develop a list of potential space-builders and to enrich it with information about contextual patterns. WordNet can be used as a tool for identification of words and phrases denoting speech acts and speech events in general, but with respect to space-builder identification WordNet structures over-generate heavily. The system-internal verb lexicon has generated an overwhelming majority (79%) of the correct answers, while all instances judged as “impossible as space openers” have been generated by WordNet.

For the time being, the “Verblist” used as a complement to WordNet is a database table organized around 43 lexical stems. The stems are connected to inflectional rules, derivational rules (e.g. rules relating confirmation to confirm, or refusal to refuse), contextual patterns and interpretation rules. The latter deal with classification and interpretation of mental spaces connected to a given lexical entry (e.g. promise normally introduces a hypothetical, future mental space, while inform opens a mental space which normally is ‘real’ from the sender’s perspective). In Table 2, we show a fragment of the verb lexicon. Column 2 contains the codes indicating the epistemic status of the mental space normally introduced by the verb (1=true from the sender’s perspective, 2 = true from the perspective of at least 2 senders, -1=untrue, 0 = hypothetical). These values may undergo changes in interplay with negation and/or modal verbs. The third column indicates whether an evaluation component is involved, and the last column deals with the senders attitude to the object (as in *X called Y a hero/a terrorist*). Figure 4 shows a sample derivational rule, relating certain nouns to verbs (e.g. *denial* to *deny*, and *proposal* to *propose*).

Table 2. A fragment of the verb lexicon (Verblast)

Verb	Epistem. status	Evaluation	Object alias
announce	1		
call X1 - X2	1		X1 = X2
claim	1		
condemn	1	negative	
confirm	2		
deny	-1	negative	
order	0	positive	
predict	0		

```

if pos('al',VerbStr)=length(VerbStr)-1 then
begin
  delete(VerbStr,length(VerbStr)-1,2);
  if VerbStr[length(VerbStr)]='i' then
  begin
    VerbStr[length(VerbStr)] := 'y';
  end
  else
  begin
    VerbStr:= VerbStr+'e';
  end;
end;

```

Fig. 4. A sample derivational rule

DUBAI, March 31st

Saudi Arabia's interior minister **has refused to comment** on a Saudi dissident held in **Canada** for his **alleged role** in a blast that killed 19 U.S. airmen in the kingdom, but **hinted** he **might say** something **next week**. The English-language Arab News daily **reported on Monday** that Prince Nayef **refused to answer** reporters' questions on the arrest of Hani Abdel-Rahim Hussein al-Sayegh. Prince Nayef **hinted** he **might give answers** at a press conference he is scheduled to give **on April 8** at the holy Moslem site of Mina after inspecting arrangements for the annual Moslem haj pilgrimage, which this year falls in the **middle of April**. Sayegh, who **denies** any role in the bombing and **has said** he was in Syria last June, was arrested **in Ottawa on March 18**.

Fig. 5. A sample input text

6 Identification of Mental Spaces

Our approach can be illustrated by processing a sample text from Reuters' corpus. The text fragment is cited below. Space-builders are highlighted.

Figure 6 shows the general structure of mental spaces identified in the text and the space builders that allowed this interpretation. The 'utterance-refusal' phrase *has refused to comment* in the base space opens a new mental space Ma 1, where the main referent is the arrested

dissident. Spaces Ma 2 and Ma 3 contain the different versions regarding the dissident’s alleged role in the bomb attack in June. Ma 4 is identified on the basis of the time adverbial March 18 and contains the background of Ma 1 (the event of arresting the dissident). The spaces on the right branch correspond to the future press conference (Mb 1) and the future Muslim pilgrimage (Mb 2). In Figure 7, the elements of the left branch of the structure are presented in detail.

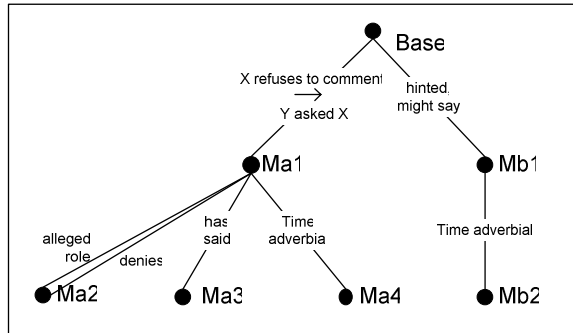


Fig. 6. The structure of mental spaces in the sample text

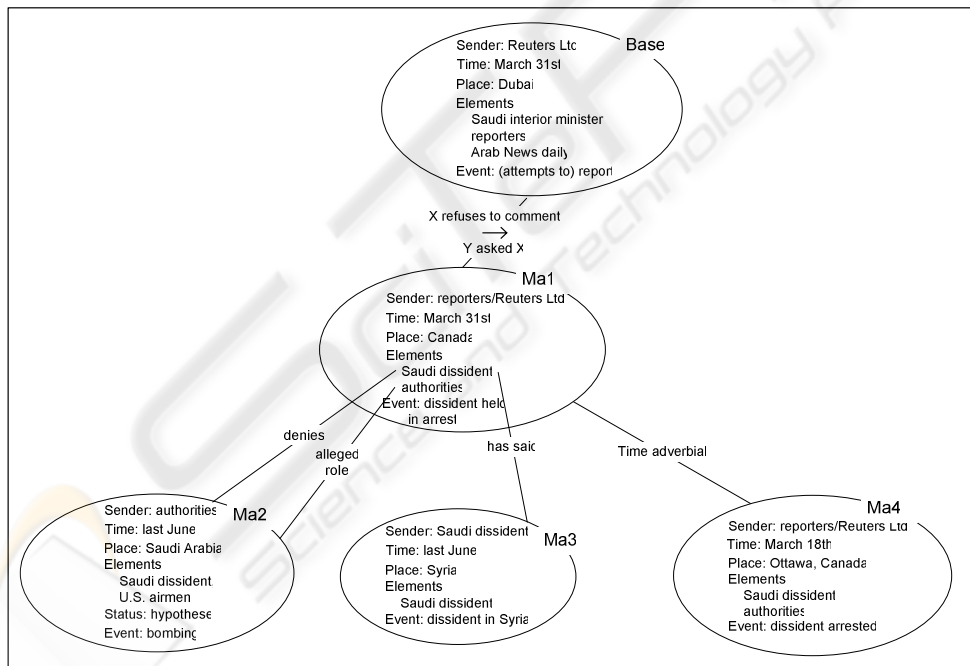


Fig. 7. The structure and content of mental spaces in the sample text (the dissident theme)

In the implemented system, the different mental spaces correspond to filled templates (Prolog structures), where variables connected to the attributes Sender, Time, Place, Event etc. are filled with values extracted from the text. These templates serve as input (a kind of interlingua) to the text generation module. The main output language is for the time being Swedish; we also started to implement a Polish module. In Figure 8, we show the output (an English translation of the news text re-written into Swedish; although it does not show the exact lexical choice, it gives a good picture of the textual structure of the output, its advantages and shortcomings).

Output:

Saudi Arabia's interior minister, Prince Nayef has refused to comment on a Saudi dissident, Hani Abdel-Rahim Hussein al-Sayegh, held in arrest in Canada. The dissident was arrested in Ottawa on March 18. He is suspected for participation in a bombing in Saudi Arabia last June. 19 U.S. airmen were killed in the attack. The dissident says he was in Syria last June. He denies his participation in the attack. The interior minister said he might give some answers on a press conference on April 8. The annual Moslem pilgrimage takes place in the middle of April.

Fig. 8. The re-written text

The text generation module contains sentence frames with variables, to which the TL equivalents of the attribute values are transferred from the filled templates. These frames are formulated on the basis of studies of news texts in the target language. The main advantage of this approach, compared to sentence-by-sentence Machine Translation, is a higher degree of cohesion and the possibility of a more idiomatic phrase and word choice. For example, in Swedish, the equivalent of He is suspected for participation in bombing is formulated with an infinite clause instead for the noun meaning 'participation':

Han misstänks för att ha deltagit i bombattacken
 He suspect+pass for to have participated in bomb+attack+def

which sounds much more idiomatic than

??? Han är misstänkt för deltagande i bombande
 he is suspected for participation in bombing

which is syntactically correct, but sounds very "un-Swedish".

7 Conclusions

The experimental system outlined here is still under development, so all components could not have been evaluated in a systematic way yet. The most extensive evaluation was based on news reports (ca 300 000 words) randomly chosen from the domains of politics, war, accidents, and disasters. This text set did not overlap with the texts mentioned in section 5. The results concerned the following modules:

- Named Entity Recognition: recall 98%, precision 86%
- Named Entity Classification (i.e. ascribing semantic categories to strings recognized as proper names): recall 70%, precision 87%

- Identification of speech acts as space-builders: recall 97%, precision 86%
- Identification of phrases referring to senders of speech acts: recall 82%, precision 98%.

Coreference resolution and readability of outputs are for the time being objects of internal evaluation, and any reliable results cannot be reported yet. The informal internal evaluation, though, indicates a continuous increase of precision. This is due to the fact that the coreference resolution process benefits from the information inferred from space-builders. Referents belonging to the same mental space are checked for coreference in the first hand, before the search for possible counterparts in other mental spaces begins. This converts many anaphora cases from “non-trivial” [25] to quite trivial ones and diminishes the need of extralinguistic knowledge (within one mental space, there is normally a considerably more limited number of possible referents than in a full text). However, the coreference resolution module requires further elaboration and more extensive testing.

The outputs evaluated hitherto show an acceptable readability level (the majority is judged as 4 on a 1 to 5 scale), but the recall is still not satisfactory and heavily domain limited. The template filling module needs to be extended and completed with a more robust technique (like Knight’s and Marcu’s compression – [15]) as an alternative. At its current development stage, the system is no doubt more precision- than recall-oriented, something that was expected given the Mental Space approach.

An interesting direction for further development of the approach applied here are applications aimed at Information Fusion (integration of information received from different sources) within e.g. bioinformatics and automated production processes. Our preliminary investigations indicate that there is a growing interest of integrating Information Extraction from Natural Language in those domains and that precision-oriented NLP is more required than systems achieving high recall and low precision rates.

References

1. Coulthard, M., *An Introduction to Discourse Analysis*. London: Longman (1985)
2. Cowie, J., Wilks, Y., *Information Extraction*. In Dale, R., Moisl, H., and Somers, H. (eds.): *Handbook of Natural Language Processing*. New York: Marcel Dekker (2000)
3. Fauconnier, G., *Mental Spaces. Aspects of Meaning Construction in Natural Language*. Cambridge, MA: MIT Press (1985)
4. Fauconnier, G., and Sweetser, E., (eds.) *Spaces, Worlds, and Grammar*. Chicago University Press (1996)
5. Fauconnier, G., and Turner, M., *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books (2002)
6. Fellbaum, C, *WordNet. An Electronic Lexical Database*. MIT Press (1998)
7. Gawronska, B., *Employing Cognitive Notions in Multilingual Summarization of News Reports*. In *Proceedings of NLULP 2002*. Copenhagen (2002) 103 - 119.
8. Harabagiu, S., *WordNet-Based Inference of Textual Cohesion and Coherence*. In *Proceedings of FLAIRS-98*, May 1998, Sanibel Island, FL (1998) 265-269.
9. Harabagiu, S, *WordNet-based inference of textual context, cohesion and coherence*. Ph.D. thesis, University of Southern California, Los Angeles, CA (1997)
10. Harabagiu, S., Maiorano, S., and Pasca, M., *Open-domain textual question answering techniques*. *Natural Language Engineering* 9 (3) (2003) 231 - 267.
11. Harder, P., *Mental spaces: Exactly when do we need them?* *Cognitive Linguistics* 14-1, 2003 (2003) 91 - 96

12. Hobbs, J., Sketch of an ontology underlying the way we talk about the world. *International Journal of Human-Computer Studies*, 43 (1995) 819-830.
13. Hovy, E., Text Summarization. In Mitkov, R., (ed.) *The Oxford Handbook of Computational Linguistics*. Oxford University Press (2003)
14. Ide, N., and Véronis, J., Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art. *Computational Linguistics* 24 (1) (1998) 1-40.
15. Knight, K., and Marcu, D., Statistics-based summarization - step one: sentence compression. In *Proceedings of the conference of the American Association for Artificial Intelligence. AAAI. Austin, Texas (2000)* 703 - 710
16. Langacker, R., *Concept, Image, and Symbol. The Cognitive Basis of Grammar*. Berlin/New York: Mouton de Gruyter (1991)
17. Lee, M. and Wilks, Y., An ascription-based approach to speech acts. *Proceedings of COLING '94, Kyoto (1996)* 344-348.
18. Lehnert, W., Plot Units: A Narrative Summarization Strategy. In Mani, I., and Maybury, M.T., (eds.) *Advances in Automatic Text Summarization*. Cambridge, Massachusetts, London, England: The MIT Press (1999) 177-213.
19. Mani, I., and Maybury, M.T., (eds.), *Advances in Automatic Text Summarization*. Cambridge, Massachusetts, London, England: The MIT Press (1999)
20. Mann, W. C., and Thompson, S., Rhetorical Structure Theory: Toward a functional theory of text organization. *Text* 8 (3) (1988) 243-281.
21. Marcu, D., *The rhetorical parsing, summarization and generation of natural language texts*. Ph.D. thesis, University of Toronto (1997)
22. Marcu, D., Building Up Rhetorical Structure Trees. *The Proceedings of the Thirteenth National Conference on Artificial Intelligence*, vol 2, Portland, Oregon, August 1996 (1996) 1069-1074.
23. Mendes, S., and Chaves, R. P., Enriching WordNet with Qualia Information. In *Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations at NAACL 2001 (2001)* 107 – 112.
24. Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K.J., Introduction to WordNet: an on-line lexical database. *International Journal of Lexicography* 3 (4) (1990) 235 - 244
25. Mitkov, R., Towards a more consistent and comprehensive evaluation of anaphora resolution algorithms and systems. *Applied Artificial Intelligence: An International Journal*, 15 (2001) 253 - 276
26. Montoyo, A., and Palomar, M., WordNet Enrichment with Classification Systems. In *Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations at NAACL 2001 (2001)* 101-106.
27. Newmark, P., *A Textbook of Translation*. Prentice Hall International (UK) Ltd (1988)
28. Nida, E. A., and Taber, C. R., *The Theory and Practice of Translation*. Leiden: E. J. Brill (1969)
29. Nirenburg, S., Mahesh, K., *Knowledge-Based Systems for Natural Language Processing. The Computer Science and Engineering Handbook (1997)* 637-653.
30. Sanders, J., and Redeker, G., Perspective and the Representation of Speech and Thought in Narrative Discourse. In: Fauconnier, G., and Sweetser, E., (eds.): *Spaces, worlds and grammar*. The University of Chicago Press, Chicago and London (1996) 290-317.
31. Searle, J.R., *Speech acts*. Cambridge: Cambridge University Press (1969)
32. Wilks, Y., *Relevance, points of view and speech acts: An artificial intelligence view*. Technical Report MCCS-85-25, New Mexico State University (1985)
33. Vossen, P., *EuroWordNet. A Multilingual Database with Lexical Semantic Networks*. Dordrecht: Kluwer Academic Publishers (1998)