INFORMATION EXTRACTION AND TEXT GENERATION OF NEWS REPORTS FOR A SWEDISH-ENGLISH BILINGUAL SPOKEN DIALOGUE SYSTEM

Barbara Gawronska and David House[1]

Department of Languages, University of Skövde, Sweden
[1]Currently at Department of Speech, Music and Hearing, KTH, Stockholm, Sweden

ABSTRACT

This paper describes an experimental dialog system designed to retrieve information and generate summaries of internet news reports related to user queries in Swedish and English. The extraction component is based on parsing and on matching the parsing output against stereotypic event templates. Bilingual text generation is accomplished by filling the templates after which grammar components generate the final text. The interfaces between the templates and the language-specific text generators are marked for prosodic information resulting in a text output where deaccentuation, accentuation, levels of focal accentuation, and phrasing are specified. These prosodic markers modify the default prosody rules of the text-to-speech system which then reads the text with subsequent improvement in intonation.

1. INTRODUCTION

There has been much emphasis recently on spoken dialogue systems both for real applications and for use as research tools. The need of a continuous interplay between the grammar-discourse component and the speech recognition - speech synthesis components has stimulated investigations across the borderlines of the traditional linguistic subdomains [9, 11]. The mutual dependence between prosody and information structure in discourse has become one of the central issues in language processing. It is also the main research question dealt with in our experimental system for interactive bilingual news reading, "Newspeak". More specific questions concern semantic aspects of parsing, information extraction from news texts with emphasis on referent tracking, bilingual text generation, and labelling of accentuation and phrasing to control and improve prosody in the speech synthesis output.

The experimental system prototype uses CNN's internet-based Quick News service in English and Swedish (all examples presented here are therefore taken from authentic Quick News texts). The user first selects the language whereby the system reads a summary of the current news generated by the system. The user can then query the system for more information about a particular event including the background and development of the event. This feature is somewhat similar to news browsing as described in [6]. Restricted domains are currently politics (subdomains: state visits, elections, strikes and riots), and disasters (subdomains: natural disasters, transportation accidents and terrorist incidents).

2. EXTRACTION COMPONENT

One may ask why a dialogue system that utilises human produced on-line news demands an information extraction component. Would it not be sufficient to let the system read the headlines and then ask the user what news he/she is particularly interested in? Our initial Wizard-of-Oz experiments convinced us that the system should be designed in another way. Headlines are often not informative enough, since their primary function is to focus the user's attention (frequently, by means of puns or metaphors), while a dialogue system should diagnose the user's intention. Moreover, in a speech-to-speech system, the speaker must be able to keep the last dialogue move of the interlocutor in short-term memory. This is even more difficult with speech synthesis using default prosody. In our experiments, five of six subjects were not able to remember the headlines when not simultaneously provided with a written text on the screen. Thus, a move containing four or five contextually unrelated headlines (*Netanyahu accepts offer to hold Mideast talks in London, Octavio Paz, Mexicos foremost poet, dies at 84, Leaders set date to start W. hemisphere trade talks, Sinn Fein delays vote on peace accord*) is not an appropriate starting point for a system-user initiative change. Furthermore, the user may be interested in certain aspects of an event development. An intelligent dialogue system can in such cases not rely on static stereotypes of the user's knowledge and intentions.

Another reason for processing the whole news text before starting the dialogue is the syntactic ambiguity of many headlines. Articles and finite verbs are frequently omitted in headlines, which may lead to the incorrect identification of phrase boundaries (e.g. *hits* in *Asia hits European stocks* may be interpreted as a noun). The risk of faulty interpretations is in most cases eliminated by extracting information from the whole text.

The search for the most relevant information must not rely on key words only. Statistic based cues without connections to syntactico-semantic parsing easily become misleading. If the system is too eager to draw quick conclusions from the appearance of certain lexical items, a headline like *Blair moves to educate an army of computer engineers* may be interpreted as referring to a military operation.

The extraction component is therefore based on lexicon-governed parsing and matching the parsing output against stereotypic event templates. The output of the parsing component is a structure containing event-type labels combined

with domain-specific semantic roles. Parsing involves the following subprocedures.

2.1. Identification of main event-types

The main cues used here are verbal nouns denoting events (*explosion, execution, visit*), predicative adjectives denoting change of state (*Air show pilots **dead** after collision*) and verbs that are not combined with verbal subjects. E.g. the verb *hit* in a headline like (1) *Explosion hits Bosnia Serb TV transmitter* is not regarded as salient since its subject is a verbal noun; the headline gets a representation similar to the one of (2) *Bomb damages Bosnia Serb TV transmitter*. This is achieved by means of a continuous interplay between phrase structure rules and the lexicon. The lexical information is to a certain extent inspired by the work of Boguraev and Pustejovsky [1].

In example (1), the noun *explosion* matches the following lexical entry, containing a description of the default interpretation (physical explosions) and of the abstract sense of the noun (as in *a population explosion*). The semantic information is connected to the most frequent syntactic patterns.

```
elex(explosion,noun,ref(explode),sg,count,
% a bomb explosion
[domain(default([ attack, accident,concrete])),
result(default(destruction_of(X))),
patterns([X, explosion],[explosion,of,X],[explosion,vt,X])],
% population explosion, explosion of drug abuse etc.
[domain([quantity,change]),
result(increase_of(X)),
patterns([X,explosion],[explosion_of(X)])]).
```

Since example (1) is syntactically structured according to the pattern {explosion, vt, X}, i.e. explosion+transitive verb+an NP denoting an object X, the default (concrete) reading is tested first. The noun explosion (by the code "ref(explode)") connected to the valency frame of the verb *explode*, structured as shown below:

```
elex(explode,verb,inf,regular,
% a bomb exploded
[domain([attack, concrete]),
cause([concrete,explosive]),result(destruction_of(X)),
pattern([cause,explode])],
% a chemical factory exploded
[domain([attack,accident,concrete]),
cause([concrete,explosive]),result(destruction_of(X)),
pattern([X,explode])],
% the army took the bomb to a safe place and exploded it
% or: the research has exploded the myth that…
[domain([action,concrete,abstract]),cause(_),
 result(destruction_of(X)),
 pattern([cause,explode,X])],
% he exploded into/with laughter
[domain([emotion_sign,concrete,abstract]),
cause([emotion,strong]),
 pattern([X,explode,pp([in,with,into],emotion) ])]).
```

The pattern labelled [attack,accident,concrete] is then chosen as the most probable one, since the cause of the explosion is not mentioned in the example. The next step is to transfer the

information from the headline to the template provided with the same label. The template, formulated in a slightly simplified Prolog notation is shown below. The variable Index corresponds to the date of the day; Time is computed as "the day before Index", if the text does not contain another time specification.

Template representation of example (1):

```
event(Index, event_type([ attack,accident,concrete], explosion),
       place(_),  time(default(Index,Time)),
       damage('TVtransmitter',attr('Bosnia Serb')),
       injuries(_),cause(_)).
```

Example (2) - *Bomb damages Bosnia TV transmitter* differs from example (1) with respect to the cause-slot, and, consequently, the description of the event type. In the case of (2), the label of the appropriate template is also found via the representation of the verb *explode*. The parser identifies the noun *bomb* as belonging to the category "explosive", the verb *damage* as referring to a destructive process; subsequently, the lexical connections to *explode* become strong and worth testing.

Template representation of example (2):

```
event(Index,event_type( [attack,concrete], explosion),
       place(_), time(default(Index,Time)),
       damage(´TVtransmitter',
       attr('BosniaSerb')),injuries(_),
       cause(bomb, suspected_agent(_))).
```

In the course of text parsing, the unfilled slots may get constant values, and the default values, such as Time, may be changed.

2.2 Identification of background information
-less salient mental spaces-

The parsing output may consist not only of partially or fully specified stereotypic subdomain templates but also of semantic representations that do not fit any pre-defined domain patterns. Those 'untypical' representations are (by means of valency frames picked up from the lexicon) formulated in terms of traditional semantic roles like agent, patient, goal, etc., provided by an index meaning "background information connected to the main index" and left in a temporary data base as potential starting points for bilingual text generation. But it can even be the case that the parser finds pieces of information that seemingly contradict each other. E.g. in a CNN report about an ETA attack, the sentence, *There were no injuries and little damage,* is followed by, *ETA has killed some 800 people in a nearly 30-year campaign of violence.*

In such cases, tense markers, conditional markers, modal verbs, and time and space adverbs are used as cues for distinguishing the information about the current event from less relevant temporal and hypothetical mental spaces [8] that may contain some background information or speculations about the possible consequences of the event reported. The distinction is coded by appropriate values of the Index, Time and Place variables. This part of the parsing procedure includes attempts to identify metaphorical expressions (as in *the millennium problem is a ticking time bomb*).

2.3 Identification of the most salient referent within the relevant mental space

Referent identification is crucial for building up the given-new information structure representation [10]. In most cases, principles based on the neo-Gricean approach [2,5] may apply to the news texts. However, the general coreference principle (the referent is normally introduced by an expression that is semantically more specific than the following anaphoric expression) is sometimes not obeyed: in news texts, a more specific description may point back to a less specific one *(A bomb damaged the home of an official at Tokyo's international airport Wednesday...No one was injured in the pre-dawn blast at the suburban Tokyo home of Tadonori Yamaguchi).* In most narratives, this would lead to an effect of lack of coreference *(An official left the room. Tadonori Yamaguchi was angry).* The domain of news requires some re-interpretation of the Gricean quantity principle: if the first description of a referent is less specific than the second one, then the information given by the more specific description is not salient. In fact, important referents are always introduced by highly specific descriptions *(Russian President Boris Yeltsin).*

3. TEXT GENERATION

Bilingual text generation is accomplished by filling the stereotypical templates after which the Swedish and English grammar components generate the final text. Referent structure contained in the templates is converted to basic prosodic information where focal accents, degree of accentuation, deaccentuation and phrase boundaries can be specified [3,7]. The template-based method enables marking the giveness/newness of the most salient referents in a quite simple way. Since the main discourse object (in most cases, the event mentioned in the headline) is already identified, the program introduces the event as "new" in the very first sentence of the generated text and then marks all NPs denoting the same as "given" (in the program code, the constant "topic" is used). The NPs denoting cause and place are by default marked as "new"; if the place information contains both the name of the country and the city, the name of the country is treated as prosodically more prominent. In the phrases referring to results (in the example below, damages and injuries), quantifiers and attributes (**no** injuries, **little** damage) are in focus, since the existence of damages and injuries is presupposed in the template; the new information concerns the extent of damages and injuries. Time adverbs in summaries of the news of the day are regarded as containing new, but less salient information (the time of the event is quite predictable for the user). In summaries of longer courses of events time adverbs may be more stressed depending on textual relations.

The filled template functions as an interlingua representation in the generation module (implemented in Definite Clause Grammar). The structures on the right of the arrow send the information from the template to the English or the Swedish grammar. Phrases with the prefix "e-" interact with the English syntax and lexicon, and phrases with the prefix "s-" with the corresponding Swedish modules.

Input to the text generators - an example:

```
event(march22,
event_type([attack,concrete], explosion),
    place(country('Spain'),city('Bilbao')),
        time(['Friday',morning,early]),
            damage(little),
                injuries(none),
    cause(bomb,suspected_agent('ETA'))) →

{language(english)},

eintro(main(explosion,
    cause(bomb)),country('Spain'),city('Bilbao'),
        time(['Friday',morning,early])),
eresult(topic(explosion),
    damage(little),
        injuries(none)),
esuspected_agent(topic(explosion),'ETA');

{language(swedish)},

sintro(main(explosion,
    cause(bomb)), country('Spain'),city('Bilbao'),
        time(['Friday',morning,early])),
sresult(topic(explosion),
    damage(little),
        injuries(none)),
ssuspected_agent(topic(explosion),'ETA').
```

Sample sentence generation rule for Swedish (for explanation of prosodic markers see section 4):

```
sintro(main(explosion, cause(Y)), country(Country),city(City),
        time(Timeinfo)) →
snp(head(Y),indef,foc(7)),
sv(V,fin,past),{slex(_,explosion,ref(V),_,_,_,_,_,_)},
snp(attr(Country,foc(6)),head(City),def),
sadvp(Timeinfo,time,foc(5)).
```

Generated output:

English:

A bomb exploded in Bilbao, Spain, early Friday morning. The explosion caused only little damage. There were no injuries. ETA is suspected of being responsible for the attack.

Swedish:

En bomb exploderade i den spanska staden Bilbao tidigt på fredagmorgonen. Explosionen förorsakade enbart små materiella skador. Inga personskador rapporterades. Förmodligen ligger ETA bakom bombdådet.

The advantage of bilingual text generation, compared with automatic translation from English to Swedish is the possibility of achieving a more idiomatic lexical and phrasal choice, as the Swedish module has direct access to the standardised ways of introducing certain types of objects and event in Swedish news. E.g. *den spanska staden Bilbao*, lit. *the Spanish town Bilbao*, sounds more natural in Swedish than *Bilbao, Spain*. The correct output is here achieved by structuring the semantic codes according to language-specific syntactic rules (in the example above, the variable Country is placed in the slot used for

adjective attributes, and the feature "def" controls the morphologically marked definiteness). The Swedish output thus does not suffer from syntactic and lexical interference, something that is difficult to achieve in automatic translation.

4. TEXT-TO-SPEECH OUTPUT

The system currently uses the Infovox multilingual text-to-speech system for British English and Swedish speech output [4]. Levels of prominence are specified for individual words using a numerical scale from 0 to 9 where 0 represents deaccentuation of the word, 1 is normal stress in English and word accent I or II in Swedish, and 2 to 9 are increasing levels of focal accent. Focal accent is realised in English by increasing fundamental frequency and duration and in Swedish by the addition of a high tone (H) following the preceding word-accent low tone [3]. The prosodically marked text is read by the text-to-speech system which results in a modification of the default prosody rules. Improvement in intonation is particularly evident concerning accentuation reflecting referent relationships and phrasing related to accentuation and information structure. Phrase boundaries can also be controlled by using commas in the text, and by sentence length in the generated text.

In the example of generated text presented above, Event Type, Place, Time, Damage, Injuries and Cause are each given a focal accent in the template representation as providing new information. Levels of focal accentuation are then specified for each language. Thus in the first sentence of the English example, Cause is given level 7, Country level 6 and Morning level 5. In the second sentence Topic does not to receive focal accent as Event has already been mentioned. The generated output with prosodic markings is as follows (capitals indicate focal accent, level of focus is given in parenthesis):

A (7)BOMB exploded in Bilbao, (6)SPAIN, early Friday (5)MORNING. The explosion caused only (5)LITTLE damage. There were (5)NO injuries. (7)ETA is suspected of being responsible for the attack.

In this example focal accent assignment in Swedish coincides with that of English and is given the following marking:

En (7)BOMB exploderade i den (6)SPANSKA staden Bilbao tidigt på fredag(5)MORGONEN. Explosionen förorsakade enbart (5)SMÅ materiella skador. (5)INGA personskador rapporterades. Förmodligen ligger (7)ETA bakom bombdådet.

The assignment and generation of focal accents and the resulting prosodic phrasing clearly corresponds more closely to semantic structure and information relationships than to syntactic structure. Thus information relationships in the generated text are reflected in the speech synthesis output.

5. SYSTEM DEVELOPMENT

The system is currently being developed using a Wizard of Oz human recogniser and the Infovox text-to-speech system. Experimentation to date has shown that within the domain of politics, the subdomains state visits and elections cause the most difficulty for extraction and text generation as these are the least stereotypical. The system is more successful at extracting and generating text concerning the more stereotypical subdomains of politics (strikes and riots) and the subdomains of disasters (natural disasters, transportation accidents and terrorist incidents). Within these subdomains, single templates containing information about the development of a certain event can quite easily be combined into a "course of events" template that may be used for generating summaries of several days news concerning the same topic. Further testing will involve a more extensive news database and more fine-grained manipulation of the accentuation component of the text-to-speech system with a view towards improving the interaction of accentuation, phrasing and information structure.

6. REFERENCES

1. Boguraev, B. and Pustejovsky, J. "Issues in text-based lexicon acquisition," In. Boguraev, B. & J. Pustejovsky (eds.): *Corpus processing for lexical acquisition,* Cambridge, Massachusets, London, England: The MIT Press, 3-20, 1996.

2. Brennan, S.E. 'Lexical entrainment in spontaneous dialog'. *ISSD 96, Philadelphia*, 41-44, 1996.

3. Bruce, G., Granström, B., Gustafson, K., Horne, M., House, D. and Touati, P. "Towards an enhanced prosodic model adapted to dialogue applications," *Proceedings of ESCA Workshop on Spoken Dialogue Systems, Vigsø*, 201-204, Aalborg, 1995

4. Carlson, R. Granström, B. and Hunnicutt, S. "Multilingual text-to-speech developoment and applications," in A.W. Ainsworth (ed), *Advances in speech, hearing and language processing,* JAI Press, London, 1990.

5. Gawronska, B. *An MT oriented model of aspect and article semantics,* Lund: Lund University Press, 1993.

6. Hauptmann, A.G. and Witbrock M.J. "Informedia: News-on-demand Multimedia Information Acquisition and Retrieval," in M.T. Maybury (ed), *Intelligent Multimedia Information Retrieval,* AAAI Press, 213-239, 1997.

7. Horne, M. and Filipsson, M. "Implementation and evaluation of a model for synthesis of Swedish intonation," *ICSLP 96, Philadelphia*, 1848-1851, 1996.

8. Fauconnier, G. *Mental Spaces: Aspects of meaning construction in natural language*. Cambridge, Mass.:MIT Press, 1985.

9. Lee, M. and Wilks, Y. "An ascription-based approach to speech acts*," Proceedings of CoLing'96, Copenhagen, Denmark,* 699-704, 1996.

10. Steedman, M. "Reprsenting discourse information for spoken dialogue generation," *ISSD 96, Philadelphia*, 89-92, 1996.

11. Zue, V. "Conversational interfaces: Advances and challenges," *Proceedings Eurospeech '97. Rhodes, Greece*, KN-KN18, 1997.