Crosslinguistic Analysis and Modeling of Prosody

Antonis Botinis[†], Barbara Gawronska[‡], Robert Bannert[††] and Elisabet Sandblom[‡]

† University of Athens, Greece

‡ University of Skövde, Sweden

†† University of Umeå, Sweden

E-mail: abotinis@phil.uoa.gr, barbara.gawronska@isp.his.se, Robert.Bannert@ling.umu.se, elisabet.sandblom@isp.his.se

ABSTRACT

This study concerns analysis and modeling of prosody in a computational context. It aims at an exploration of the interaction between prosodic markers on the one hand, and the lexical, the syntactic and discourse level of the utterances on the other. The analysis is a step towards a computational discourse-grammar-prosody model. We concentrate on the pattern of distribution and duration of silent pauses in German and Swedish. The possibility of integrating the results in a multilingual text generation system is then discussed.

1. INTRODUCTION

The present study is an attempt to integrate experimental phonetics and computational linguistics methodology in the analysis and modeling of prosody. Our background of prosody stems from earlier research (e.g. [1]), according to which prosodic categories were associated with linguistic levels. This research paradigm has been followed by further studies and research paradigms on crosslinguistic prosody (see [1], [2],[3]).

The ultimate goal of the research is a speech production model where prosodic representation is the outcome of a multiple interaction with reference to segmental, lexical, syntactic, and discourse levels. The present study focuses on one type of prosodic markers – silent pauses; their distribution and duration patterns are examined from a crosslinguistic perspective.

The distribution of silent pauses and other prosodic boundary markers in spontaneous speech in Swedish has been investigated in, among others, in [4], [5], [6]. These studies show that there is a very strong correlation between pause duration and perceived prosodic boundary strength in Swedish, while no strong relationships can be found between F0 reset, or reduction in articulation rate on the one hand, and perceived boundary strength on the other. Furthermore, on the basis of preceding pause duration, the syntactic vs. discourse status of a frequent Swedish conjunction (*men*-'but/and/so') may be predicted. These, and others studies provide sufficient basis for treating silent pauses as a prominent boundary marker on discourse level. It seems therefore relevant to investigate whether there is any significant correlation between pauses and syntactic patterns, and whether the patterns of interaction differ in different languages.

2. METHODOLOGICAL PROCEDURES

An exemplification of our methodology with reference to text structure, in accordance to category and functional representations, and prosodic silent pause correlation follows below.

Material. The speech material consists of Swedish and German broadcast news obtained from the Swedish State Sound and Picture Archive and from the German Television, which were analysed with the wavesurfer (http://www.speech.kth.se/wavesurfer/,2003-02-20) software package and transcribed by hand for the purpose of lexical and syntactic analysis.

The present study concentrates on the analysis of fragments of broadcast news, presented by both female and male speakers and dealing with similar topics. We chose fragments consisting of read texts in order to eliminate, or at least to minimize pragmatic factors such as turntaking (TRP markers), speech overlap, external disturbance, content planning pauses etc ([6]). What we wanted was to focus on the linguistic and textual structure of the utterances in relation to pure speech planning.

Syntactic analysis. The text was tagged into traditional lexical and inflectional categories (word classes, number, definiteness, tense), and then analysed in terms of syntactic phrase categories and syntactic functions. The syntactic analysis was performed partly by humans, partly automatically by means of a computerized grammar formalism, called Referent Grammar ([7],[8]), implemented in Prolog (Definite Clause Grammar). Referent Grammar (RG) is an eclectic formalism, inspired by GPSG and LFG. The output of the parsing procedure consists of two kinds of structures: the so-called C(ategory) representation and the F(unctional) representation. The latter representation makes use of traditional syntactic functions like subject, direct object, indirect object, adverbial phrase etc. Table 1 offerss a sample C- and F-representation. The words in the C- and F-structures are coded in Machinese English (a kind of interlingua). The variables R1, R2 and R3 mark discourse referents and can be used for coreference identification ([8], [9]).

ISBN 1-876346-48-5

Table 1. A sample result of the automatic RG-analysis.

Input (Swedish):

Tankfartyget Prestige var på väg från Lettland till Gibraltar

Oil tanker-def Prestige was on its way from Latvia to Gibraltar

C-representation:

```
sent(np(app(n(oil_tanker,sg,def),n(prop('Prestige')))),
      v1(v(be,past),vparticle(on_way)),
            pp(p(from),np(n(prop('Latvia')))),
            pp(p(to),np(n(prop('Gibraltar'))))
```

F-representation:

```
f(focus(,np(R1,
app(n(oil_tanker,sg,def),n(prop('Prestige')))),
      subject (R1),
pred(v(be,past),vparticle(on_way)),
      advl(p(from),np(R2,n(prop('Latvia')))),
            pp(p(to),np(R3,n(prop('Gibraltar'))))
```

Text units. In addition to the syntactic analysis exemplified in Table 1 above, the texts were segmented into discourse units. With reference to the present materials, four main discourse units were used: (1) topic introduction, (2) topic development, (3) transition towards dialogue and (4) topic reset. Table 2 offers an example of these categories. The unit labelled "topic development" was further segmented into smaller text units, i.e. "paragraphs", dealing with different aspects or different phases of the main course of events described in the news text.

Table 2. Discourse units in a typical broadcast news fragment (English translation of a Swedish news fragment).

Speaker 1:

Topic introduction: After being adrift and leaking oil since last Wednesday,, the vessel broke apart today, approximately 170 km of the Spanish coast.

Topic development: The tanker Prestige was on its way from Latvia to Gibraltar when it lauded in distress and received a large rent across the hull.

Transition toward dialogue: How serious is this oil spill?

(Speaker 2, spontaneous speech)

Topic reset: Exactly what it was that the vessel collided with is so far unclear.

Text, syntax and pauses. The pauses observed in the spoken material were related to discourse units (category A below), phrasal units (B and D) as well as to functional units (C and E). For Swedish, the following categories were established:

A. Discourse-related pauses, i.e. pauses after topic introduction, topic reset and before the beginning of a paragraph. The category "transition towards dialogue" was not taken into consideration because of the scarceness of data (only two occurrences in the material).

B. Pauses between full sentences (i.e. maximal units on the category level).

C. Topicalization related pauses (after heavy topicalized constituents).

D. Pauses between coordinated clauses.

E. "Additions": optional adverbial phrases (PP:s and subordinated adverbial clauses) occurring as the final functional constituent in a sentence, after obligatory verbal complements.

In the German material we found, as expected, silent pauses of the types A and B (discourse unit markers and intersentential pauses). The third frequent category was pauses after topicalized constituents; however, the syntactic functions of the topicalized constituents in German was not identical to these of topicalized constituents in Swedish. In Swedish, pauses of type D occurred only after heavy subjects (i.e. subjects with a more complicated structure than Det+A+N) and topicalized adverbial clauses. In German, the topicalized constituent followed by a pause could also function as a direct or indirect object.

Another pause category present in German, but not in Swedish, comprised pauses marking the beginning of not-topicalized subordinated finite and infinite clauses. This fact is probably due to the differences between the German and Swedish syntactic patterns. German, unlike Swedish, requires verb-final constructions in subordinated clauses and clauses with modal verbs. This, in turn, enables central embeddings. It is probable that this causes the need of strong prosodic markers that signal which constituents belong to the embedded clause, and which ones to the main clause.

Yet another difference in pause distribution between the two languages was the low frequency of pauses before final "additions" (category E in Swedish). Only one instance of this category was found in the German, compared to 13 instances in Swedish. Again, the final position of the verb in many German constructions may serve as an explanation.

Instead of pauses before final "additions", in German we found a category that we will call for "functional parallelism markers". It comprises pauses between constituents having the same syntactic function, mostly between two attributes or two adverbial phrases, like in

Am Nachmittag / dann
'in the afternoon'/ then
or

ISBN 1-876346-48-5

schwefelhaltigen / giftigen Öls
sulphurous / poisonous oil+gen.

The pause categories in German, labeled A-E, are thus as follows:

A. Like in Swedish and even called "major silence pauses".

B. Like in Swedish and even called "major silence pauses".

C. As topicalization related pauses.

D. As pauses marking transitions between main clause and embedded clause.

E. As pauses marking functional parallelism.

The distribution of silent pauses in the material was analyzed as a function of (a) language, (b) gender, (c) pause category.

3. Results

The results, presented in figures, were subjected to statistical analysis with the statistical package StatView. Analyses of variance were used for the evaluation of variable effects in a factorial model.

Figure 1 shows the duration of silent pauses as a function of language, i.e. German and Swedish. Although the overall duration is considerably in Swedish, it did not reach significance (at 0.05 level).

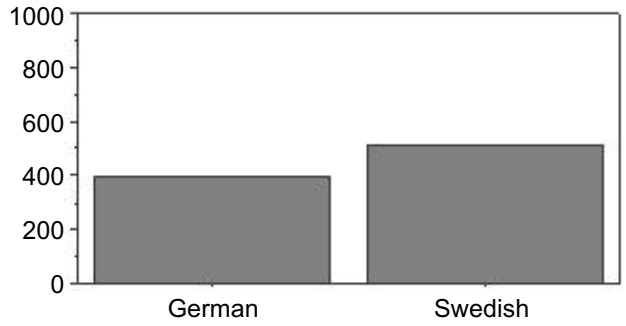

Figure 1. Prosodic silent pause duration in msecs as a function of language (German and Swedish).

Figure 2 shows the duration of silent pauses as a function of gender, i.e. female and male. Gender has hardy any effect on silent pause durations.

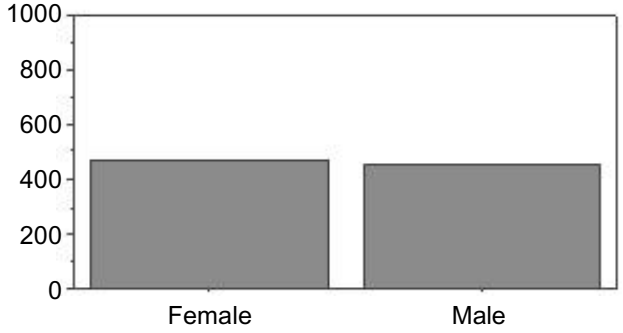

Figure 2. Prosodic silent pause duration in msecs as a function of gender (female and male).

Figure 3 shows the duration of silent pauses as a function of context category. The effect of category was highly significant ($F=40.6$, $p<0.0001$). In a scheffe's post-hoc test categories A and B were significant with regard to all other categories whereas categories C, D and E did not differ significantly. Figure 4, on the other hand, shows the silent pause durations of context categories splitted by German and Swedish.

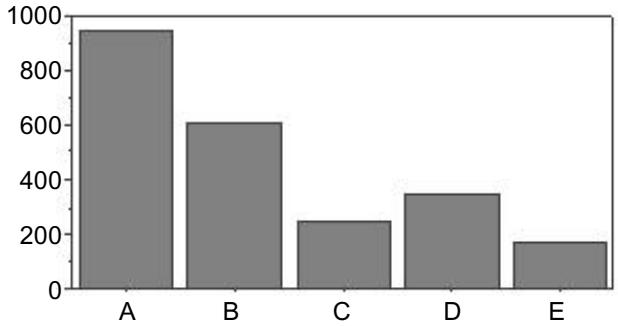

Figure 3. Prosodic silent pause duration in msecs as a function of context category (see text).

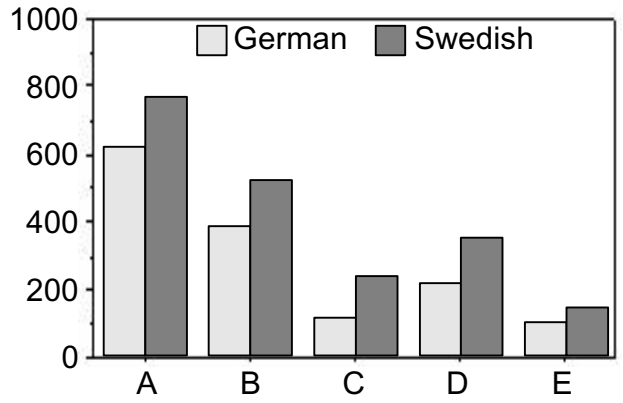

Figure 4. Prosodic silent pause duration in msecs as a function of context category (see text) as well as language (German and Swedish).

Figure 5 shows the duration of silent pauses as a function of pause category, i.e. major and minor. The effect of pause category was highly significant ($F=95.4$, $p<0.0001$).

ISBN 1-876346-48-5

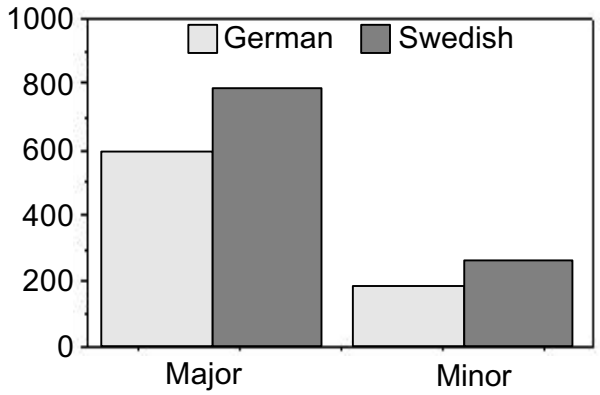

Figure 5. Prosodic silent pause duration in msecs as a function of type of prosodic silence pause (major and minor) as well as language (German and Swedish).

4. CONCLUSIONS

The results of the present study indicate the following: first, pauses may be prosodic cues of syntactic and text boundaries; in case of pause absence, we assume that other prosodic cues as well as lexical and other linguistic cues may segment speech production in the corresponding units. Second, there seem to be two types of pauses; minor pauses, i.e. up to approximately 300 msecs. and major pauses, i.e. considerably more than 300 msecs. German and Swedish seem to have fairly the same structure with regards to major pause distribution and structure whereas, with regards to minor pauses, these two languages have a divergent structure. In both languages we observed topicalization-related pauses, but the constitutens preceding these pauses differ with regard to their syntactic functions. In Swedish, pauses marking beginnings or ends of subordinated clauses and infinitive clauses were observed only in connection to topicalisation or "addition", whereas German speakers very frequently marked the borderlines between sper- and subordinated clauses by silent pauses.

Although the duration differences between the different categories of minor pauses did not prove to be statistically significant in our study, there seems to be a tendency for longer pauses on clause boundary (sub- and subordinated clauses in German, coordinated clauses with subject ellipse in Swedish) then between constituents on a lower syntactic level. This should be verified by studying a larger material.

Some aspects of the present study can be tested in the context of multilingual automatic text generation of the type "concept-to-syntax-to-speech", as sketched out in [10]. In such systems, text a generated from language independent templates, structured into topics and subtopics (paragraphs). This structure is mapped onto language specific syntactic (functional and category – se table 1) rules, which can serve as input to a text-to-speech component. Providing the syntactic input with estimated duration of major and minor pauses with respect to the facts observed in the present investigation, and evaluating the output, will be the next step in developing the prosody model. Furthermore, the analysis will be extended to include changes in the intonation contour.

REFERENCES

[1] A. Botinis, *Stress and Prosodic Structure in Greek*, Lund University Press, 1989.

[2] A. Botinis, R. Bannert and M. Tatham, "Contrastive tonal analysis of focus perception in Greek and Swedish", in A. Botinis (ed.), *Intonation: Analysis, Modelling and Technology*, Dordrecht: Kluwer Academic Publishers, 2000.

[3] A. Botinis, B. Granström and B. Möbius, "Developments and paradigms in intonation research", *Speech Communication* 33, 263-296.

[4] P. Hansson, "Perceived Boundary Strength", in *Proceedings of 7th International Conference on Spoken Language Processing. ICSLP 2002. INTERSPEECH 2002*. September 16-20, 2002. Denver, Colorado USA, 2277-2280.

[5] M. Horne, P. Hanson, G.. Bruce, J. Frid, and M. Filipson, "Discourse markers and the segmentation of spontaneous speech. The case of Swedish "men" – "but/and/so". *Working Papers* 47, 1999, Lund University, Dept. of Linguistics, 123-139.

[6] K. Lundholm, "Pausering i spontana dialoger: En undersökning av olika paustypers längd". Ms, Dept. of Linguistics, Lund University, 2000.

[7] B. Sigurd, "Referent Grammar (RG). A generalized phrase structure grammar with built-in referents", *Studia Linguistica* 1987:2, 115-135.

[8] B. Gawronska, "Translation great problem". On the problem of inserting articles when translating from Russian into Germanic languages, in *Proceedings of Coling 90*, Volume 2, Helsinki, Finland, pp. 133-138

[9] B. Gawronska, *An MT oriented model of aspect and article semantics* (=*Travaux de l´institut de linguistique de Lund* 28). Lund University Press, 1993.

[10] B. Gawronska. and D. House, "Information Extraction and text generation of news reports for a Swedish-English bilingual spoken dialogue system." In: R.H. Mannell & J. Robert-Ribes (eds): *ICSLP (5th International Conference on Spoken Language Processing) '98 Proceedings*, Sydney, Australia, volume 4, 1998, pp. 1139-1142.