Stress, pauses, pronominal types and pronominal functions in Danish spoken data

Costanza Navarretta
Centre for Language Technology
University of Copenhagen

Abstract

In this paper we present a study of the relation between types of third personal singular neuter pronoun and their functions in Danish spoken data where stress information is marked so that personal and demonstrative occurrences of the pronouns can be distinguished. This study confirms that there are language specific differences in the way various types of pronoun are used to refer to abstract entities in Danish compared to English data. In the second part of the paper we describe supervised machine learning experiments with the purpose of investigating the role of stress and pause information on the automatic recognition of the pronominal functions of third personal singular neuter pronouns in spoken corpora. The results of these experiments show that the inclusion of stress and pause information in the data improves classification.

1 Introduction

This paper is about the relation between types of third person singular neuter pronoun (sn-pronoun henceforth), their function and their referent types in Danish spoken data. The paper also addresses the role of stress and pause information on the automatic recognition of the pronominal function in these data.

The most frequently used sn-pronoun in Danish is *det* (it/this/that). In texts it is ambiguous with respect to its pronominal type while in spoken language it is possible to distinguish the unstressed personal *det* and the stressed demonstrative *det*. *Det* has many functions in discourse such as non-referential, anaphoric, cataphoric and deictic. When it is used anaphorically it can have nominal phrases antecedents in all genders and numbers or it can have non-nominal antecedents such as predicates in copula constructions, verbal phrases, clauses and discourse segments of various sizes. In the first case we call it an *individual anaphor* while in the second case we call it an *abstract anaphor*[1]. The Danish sn-pronouns *det* and *dette* (these) are used in more and different contexts

---

[1] The antecedents of sn-pronouns can also be substantives which refer to abstract entity, such as *accident* and *situation*, but in the present study we only account for the distinction between nominal and non-nominal antecedents.

than the corresponding English sn-pronouns: *it*, *this* and *that* (Navarretta 2002). Because theories accounting for the use of English sn-pronouns have been the backbone of algorithms for resolving individual and abstract pronominal anaphora, these algorithms cannot be directly applied to Danish data (Navarretta 2004a). It is thus important both from a theoretical and an applied point of view to investigate and account for the language specific uses of the Danish sn-pronouns. The present study is a contribution to this effort and focuses on two aspects: a) the relation between pronominal types, anaphoric types and antecedent types in spoken data and b) the role of stress and pause information on the automatic recognition of the pronominal functions of sn-pronouns using supervised machine learning algorithms.

The present work was done under the DAD project (The abstract *det*) which was funded by the Danish Research Councils between January 2007 and June 2009. The main aims of the DAD project were the following: a) to provide parallel and comparable corpora of texts and spoken data in Danish and Italian annotated with information about the occurrences of sn-pronouns with focus on abstract anaphora; b) to analyze the use of these pronouns in these two languages and compare their uses to the use of English sn-pronouns as described i.a. in (Webber 1988, Gundel et al. 2003); c) to explore the automatic treatment of these pronouns in the annotated corpora.

The paper is organized as follows. In section 2 we discuss the background literature then we present the Danish pronouns, the corpora and the annotation which we have used in the present work (section 3). In section 4 we describe the relation between pronominal types and referred entities in the data. In section 5 we discuss related work about the recognition of the function of sn-pronouns and in section 6 we present our machine learning experiments and discuss the results of these experiments. Finally, we make some concluding remarks and discuss future work.

2 Background

The first problematic aspects when processing sn-pronouns automatically is the recognition of their functions in discourse. The second obstacle is the identification of the antecedents of their anaphoric occurrences, especially when the anaphors are abstract. Most theories and empirical studies on abstract anaphora are based on English data, i.a. (Asher 1993, Hegarty 2003, Gundel et al. 2003). Exceptions are studies of abstract anaphora in Scandinavian and Romance languages by

i.a. (Fraurud 1992, Borthen et al. 1997, Navarretta 2000, Navarretta 2007), which indicate that there are differences in the way abstract pronominal reference is done in these languages and in English.

All cognitive theories of referring nominal expressions suggest that there is a relation between the type of referring expression and the assumed salience status of the referred entities in the hearer's model of the on-going discourse, see i.a. (Givón 1983, Prince 1981, Ariel 1988, Gundel et al. 1993). According to these theories personal pronouns refer to the most prominent (central, salient) entities in discourse, while demonstrative pronouns refer to less prominent entities. Using Gundel et al.'s (1993) terminology in the *Givenness Hierarchy*, these entities are only activated in the hearer's cognitive model, while entities referred to by personal pronouns are in focus.

Webber (1988) notices that English personal pronouns cannot often refer to abstract entities when the antecedent is a clause, because the clause is not accessible to the pronoun. Based on corpus studies, Gundel et al. (2003, 2004, 2005) and Hegarty (2003) confirm Webber's observation and explain the use of demonstrative pronouns to co-refer with clausal antecedents in terms of the *Givenness Hierarchy* (Gundel et al. 1993). Entities introduced in discourse by clauses are only activated in the hearer's cognitive status. Their referents are in most cases facts, situation and propositions thus these entities are usually referred to by demonstrative pronouns. Eventualities which are introduced in discourse by verbal phrases and entities which are introduced by nominal phrases occurring in central syntactic positions in the current or in the preceding utterance are in focus and are in most cases referred to by personal pronouns. Hegarty notices also that facts and propositions are the most abstract objects in Asher's (1993) hierarchy of abstract objects and thus facts and propositions are usually referred   to by demonstrative pronouns.

The fact that the personal pronoun *it* and the demonstrative pronouns *this* and *that* have usually different types of antecedents has been used in anaphora resolution algorithms to identify individual and abstract occurrences of sn-pronouns (Eckert and Strube 2001, Byron 2002, Strube and Müller 2005, Müller 2008). All these algorithms rely on linguistic and semantic knowledge and on pre-annotated information that allows sorting out non anaphoric occurrences of sn-pronouns. All algorithms are built upon and tested on domain specific corpora and Eckert and Strube's algorithm has only been manually tested by the two authors. An adaptation of this algorithm to Danish is

described in (Navarretta 2002, 2004a). Also this version of the algorithm has only been tested manually.

The observation that personal pronouns usually refer to non-clausal antecedents in English does not account for the use of abstract anaphora in other languages. Fraurud (1992) notices that in written Swedish it is not possible to determine the pronominal type of the pronoun *det* (it/this/that) and thus the pronominal type cannot be indispensable to the recognition of abstract anaphors in this language. Navarretta (2002, 2005) makes the same observation for the Danish pronoun *det* (it/this/that). She also shows that there are a number of differences in the way proximal demonstrative pronouns are used as anaphors in Danish and English. Finally, Navarretta (2007) presents a study of abstract anaphora in Danish, English and Italian parallel and comparable texts which indicates that there are systematic differences in the way various expressions are used in abstract reference in the three languages. Studying the use of various pronominal types in spoken Danish is one of the aims of our work.

3 The Danish data

3.1 The Danish sn-pronouns

As previously mentioned, *det* (it/this/that) is the most frequently used sn-pronoun in Danish. The demonstrative pronoun *dette* (this) is mostly used in formal texts and it seldom occurs in spoken language. Sn-pronouns in spoken language include the unstressed personal *det* and the stressed demonstrative pronouns *d'et*[2] (this/that), *d'et h'er* (this) and *d'et d'er* (that).

3.2 The spoken corpora

We have worked with three spoken corpora. The first two corpora were collected, transcribed and annotated with i.a. prosodic information by phoneticians at the university of Copenhagen in the DanPASS project (Grønnum 2006) which was a Danish version of the MAPTASK project. In the DAD project we worked with all the DanPASS monologues consisting of 21,224 running words and with 16 of the 32 DanPASS two-party dialogues (33,971 running words). The third spoken corpus we have worked with, consists of extracts of multi-party free conversations from the LANCHART corpus (Gregersen 2007), which has been collected by sociolinguistic researchers at the university of Copenhagen and of TV interviews (LANCHART+TV henceforth). The LANCHART+TV data consist of 26,304 running words.

3.3 The annotation

---

[2] Stress is henceforth indicated by an apostrophe.

The DanPASS and LANCHART+TV corpora contain various types of annotation. In the present study we only look at stress information on sn-pronouns[3] and at the annotations added to the corpora under the DAD project. The original format of the transcriptions and annotations in the three corpora was the PRAAT[4]'s TextGrid-format . This format has been transformed into the XML format required by the PALinkA tool's (Orasan 2003). The data have been annotated manually with information about sn-pronouns. The annotation scheme used in the project is an extension of the MATE/GNOME scheme (Poesio 2004) which is described in (Navarretta and Olsen 2008). In the DAD annotation all sn-pronouns are marked together with information about their function, such as non-referential, deictic and individual anaphoric. The antecedents of all anaphoric occurrences have also been marked and their syntactic types have been coded. Two referential link types between anaphors and antecedents have been distinguished: identity and non-identity. The anaphoric distance (the distance between the anaphors and antecedents in terms of utterances) and the semantic type of abstract referents have also been marked. The semantic types which have been recognized are partly inspired by Asher's (1993) hierarchy of abstract objects and comprise the following categories: eventuality, fact-like, proposition-like, property and speech act.

Most of the corpora were annotated independently by two expert annotators and then the two annotations were compared. The remaining data were only coded by one annotator and revised by the other. In case of disagreement the two annotators decided together which annotation to adopt. In difficult cases a third linguist was consulted to choose one annotation. Inter-coder agreement was measured in terms of *kappa* scores (Cohen 1960, Carletta 1996) on the first subset of the annotated data (most of the text corpora and the DanPASS dialogues). The *kappa*-scores for most categories were between 0.7 and 1 (Navarretta and Olsen 2008).

4 Pronominal types, pronominal functions and referent types in spoken Danish

In table 1 the types of pronoun and their anaphoric uses in the three annotated corpora are shown, while in table 2 the pronominal types and their distribution as individual and abstract anaphors in the same corpora are given.

---

[3] Stress information in the LANCHART-TV corpus was annotated in the DAD project.
[4] http://www.fon.hum.uva.nl/praat/.

C o r p u s	P r o n o u n	A n a p h o r	O t h e r	T o t a l
DanPASS	Det	146	61	210
Monologues	d'et	105	25	130
	Total	254	86	340
	det	311	123	434
DanPASS	dette	1	0	1
Dialogues	d'et	263	70	333
	Total	575	193	768
	det	577	398	975
LANCHART+TV	dette	1	0	1
Dialogues	d'et	180	79	259
	Total	758	477	1235

Table 1: Anaphoric and non-anaphoric uses of sn-pronouns in spoken corpora

Corpus	**Pronoun**	**Individual**	**Abstract**	**Total**
DanPASS	det	131 (82.4%)	28 (17.6%)	159
monologues	d'et	82 (78.1%)	23 (21.9%)	105
	Total	213 (80.7%)	51 (19.3%)	264
	det	202 (65%)	109 (35%)	311
DanPASS	dette	1 (100%)	0 (0%)	1
dialogues	d'et	142 (54%)	121 (46%)	263
	Total	345 (60%)	total 230 (40%)	575
	det	357 (61.9%)	220 (38.1%)	577
LANCHART+TV	dette	1 (100%)	0 (0%)	1
dialogues	d'et	78 (43.3%)	102 (56.7%)	180
	Total	436 (57.5%)	322 (42.5%)	758

Table 2: Sn-pronouns with individual and abstract anaphoric uses in spoken corpora

The data in table 2 show that the DanPASS monologues contain fewer abstract anaphors than the
dialogue corpora. The data also indicate that stressed and unstressed *det* are abstract anaphora

nearly with the same frequency in the DanPASS data, while the unstressed *det* is the most frequently used abstract anaphor in the LANCHART+TV corpus.

In table 3 the relation between types of pronoun and types of referent is given. Propositions, situations, fact-like entities, speech acts and some eventualities are introduced in discourse by clauses, while eventualities and properties are introduced in discourse by verbal phrases and predicates in copula constructions respectively.

C o r p u s	R e f e r e n t t y p e	**Det**	d ' e t	T o t a l
DanPASS monologues	property	2	2	4
	eventuality	3	8	11
	fact-like	3	4	7
	proposition	20	9	29
DanPASS dialogue	property	3	4	7
	eventuality	26	45	71
	fact-like	36	53	89
	proposition	40	16	56
	speech act	4	3	7
LANCHART+TV dialogues	property	21	6	27
	eventuality	115	48	163
	fact-like	61	41	102
	proposition	18	4	22
	other	5	2	7

Table 3: Pronominal types and referent types

The data in table 3 indicate that Danish demonstrative pronouns are not the most frequently used pronouns in abstract reference differing from English, see among others (Gundel et al. 2004, 2005). In the DanPASS dialogues, however, abstract demonstrative pronouns are slightly more frequent than abstract personal pronouns. The data show also that personal pronouns refer often to facts and propositions (their antecedents are clauses) and thus their use differs from that of the English *it*. Concluding the Danish spoken data does not indicate that demonstrative pronouns are the preferred abstract anaphora when referring to entities introduced in discourse by clauses.

Borthen et al. (1997) notice that in Norwegian there are more occurrences of personal pronouns referring to entities introduced by clauses than it is the case in English. They explain these occurrences by extralinguistic factors. Although we agree with Borthen et al. that many factors influence salience of entities in discourse, such as information structure and world knowledge, see among others (Hajicovà et al. 1990, Kaiser 2000, Gundel et al. 2003, Navarretta 2002) we believe that the differences in the use of various pronominal types in abstract reference in Danish and English are in many cases systematic. They are also so frequent that they cannot only be explained in terms of extralinguistic factors. In (Navarretta 2008) we suggest that they can often be accounted for by the languages' pronominal systems and their syntactic structure. We especially point out that in Danish constructions such as left dislocations and clefts are much more frequent than in English and as a consequence of this, clauses are more often in focus in Danish than in English.

5 The role of stress and pause and hesitation information on the recognition of the function of sn-pronouns

5.1 Related studies

There are few studies for automatically recognizing the main functions of sn-pronouns. Evans (2001), and Boyd et al. (2007) use a machine learning approach for recognising non-referential occurrences of the English *it* in text corpora. Boyd et al. generalize some of the rules implemented in rule-based systems via word patterns which they add to the system as features. They also use external knowledge sources in the form of two word lists containing weather verbs and idioms. Boyd et al. achieve the best results using 25 features (precision was 82% and recall 72% on the given corpus). Müller (2006) runs machine learning experiments for recognizing non-referential uses of *it* on multi-party dialogues using a number of shallow knowledge annotations in the data and achieves 80% precision and 60.9% recall. Machine learning experiments for recognizing non-noun-referential and noun-referential uses of the pronoun *it* were done by Bergsma et al. (2008) using distributions in large text corpora, that is looking at which words can substitute the pronoun in similar contexts . They achieve better results than the other algorithms which rely on annotations of various types.

The classification of referential and non-referential uses of the Dutch pronoun *het* (it) in two text corpora is tested by Hoste et al. (2007). Hoste et al. recognize the following uses of this pronoun: individual and abstract anaphoric, non-referential, anticipatory subject and anticipatory object. The reported results of the classification give an improvement of approx. 30% for all distinctions with

respect to the baseline (the most frequent class). The authors also measure the effects of the classification on a machine learning based co-reference resolution system.

Navarretta (2009) presents unsupervised and supervised machine learning experiments on the DAD Danish corpora with the purpose of recognizing the functions of sn-pronouns. The experiments were run on both written and spoken data and the classification of pronominal functions which she uses is more fine-grained than in the experiments conducted on other languages. The following function classes were marked in the data:

- expletive (all non-referential uses);

- cataphoric (the pronoun precedes the linguistic expression necessary to its interpretation);

- deictic (the pronoun refers to something in the physical word);

- individual anaphoric;

- individual vague anaphoric (the individual antecedents are implicit in discourse);

- abstract anaphoric;

- abstract vague anaphoric (the abstract antecedents are implicit in discourse);

- textual deictic (pure textual deictic, that is the anaphors refer to, but are not co-referential with, preceding linguistic expressions (Lyons, 1977);

- abandoned (the pronouns occur in unfinished and abandoned utterances).

Navarretta run her experiments with different machine algorithms and on various datasets: a) exclusively looking at the contexts in which the pronouns occur (unsupervised experiments), b) training classifiers on the data enriched with the pronominal functions c) adding also PoS and lemma information to the training data.

The method of testing various types of linguistic information and algorithms in order to find the best appropriate classification methods to specific NLP tasks and corpora has been proposed by Daelemans et al. (2003).

All experiments were run in the WEKA system (Witten and Frank 2005). Only stress information for sn-pronouns were included in the data. The results of the classifiers were tested using 10-fold cross-validation. As baseline in the evaluation the results provided by the WEKA ZeroR class were used. This class predicts the most frequent attribute value for a nominal class (accuracy is the frequency of the most used category). The improvement with respect to the baseline using only n-grams and the function classes as training data was of 36.7% for the texts, 27.7% for the DanPASS monologues, 33.7% for the DanPASS dialogues and 19.1% for the LANCHART dialogues. The

results for the first three datasets can be compared with the results obtained by Hoste et al. (2007) and are quite positive given that the classification used in the Danish experiments is much more fine-grained than that used by Hoste et al. and that they only run their experiments on two text corpora (Navarretta 2009). The classification results for non-referential occurrences are also comparable with the results obtained by algorithms tested on English data. PoS and lemma information improved the performance of the classifiers, but the improvement was not significant.

In the present work we have repeated Navarretta's (2009) experiments on the two DanPASS corpora. The goal of these experiments was to investigate the influence of information on stress, pauses and hesitations on the recognition of the pronominal functions. The distribution of the pronominal function in the DanPASS data is in table 4 which is an extract of a table in (Navarretta 2009, p.20).

Corpus	Pronoun	Expl	IndAn	AbstAn	VagIndA	VagAbstA	Catap	Deict	Textdeic	Aband	Total
DanPASS monolog	unstressed	22	107	27	14	1	14	0	0	25	210
	stressed	1	74	10	8	13	11	1	0	12	130
	all	23	181	37	22	14	25	1	0	37	340
DanPASS dialogues	unstressed	34	177	100	25	5	17	0	4	72	434
	stressed	10	121	111	22	7	22	7	3	31	334
	all	44	298	211	47	12	39	7	7	103	768

Table 4: pronominal types and their functions in the DanPASS corpora

We run machine learning classifiers on the following data: a) simple n-grams without stress information; b) simple n-grams with stress information on all words; c) simple n-grams with stress information on all words and information about pauses and hesitations in speech. In the first experiment unstressed and stressed occurrences of the pronoun *det* are thus not distinguished. For the monologues the best results were obtained using the SMO class (Sequential Minimal Optimization) applied on a window of one word before and three words after the sn-pronouns. For the dialogues the best results were achieved with the same classifier using a window of two words preceding and three words following the sn-pronouns. The results of the three experiments are in table 5[5].

---

[5] . Precision, Recall and F-measure are calculated in WEKA as the average of Precision, Recall and F-measure achieved on each class.

Corpus	Data	Algorithm	Precision	Recall	F-measure
DanPASS monologues		Baseline	28.3	53.2	37
	no stress	SMO	65.2	67.6	64.6
	+stress	SMO	97.1	97.1	97
	+stress+pauses +hesitations	SMO	97.7	97.6	98.2
DanPASS dialogues	no stress	Baseline	15.1	38.8	21.7
		SMO	52.5	55.6	53.5
	+stress	Baseline	15.5	39.3	22.2
		SMO	51	53.4	51.2
	+stress+pauses +hesitations	SMO	51.9	54	51.8

Table 5: Results of classifiers on DanPASS data

The results obtained on the DanPASS monologues and dialogues are quite different. Stress information improves substantially classification on the monologues. Adding pause information to the data improves the performance of the SMO classifier even more, but the improvement is not significant. The recognition of the function of pronouns on the monologues is the highest achieved by any effort to recognize the function of sn-pronouns, but it must be taken into account the fact that the DanPASS monologues are a very homogeneous corpus and cannot represent spoken data in general.

The results obtained on the dialogues are not as good as for the monologues. Adding stress information to the data influences the baseline. The improvement of the classifier on the data enriched with stress information with respect to the baseline is lower than that obtained on the transcriptions without stress information. However, the decrease in performance is not as large as expected given that the occurrences of stressed and unstressed *det* are only distinguished in the second dataset. Adding information about pauses and hesitations to the data increases the performance of the classifier also on the dialogue dataset, but also this improvement is not significant.

6 Conclusions

In the paper we have described a study of the relation of pronominal types, pronominal functions and types of referents in spoken corpora which were annotated with information about sn-pronouns in the DAD project. The study accounts for the differences between unstressed and stressed occurrences of the pronoun *det* which is ambiguous with regard to its pronominal type in texts. These spoken data confirm Navarretta's (2002, 2007, 2008) observation that there are systematic differences in the way various types of pronoun are used in abstract reference in Danish and English.

In the second part of the paper we have described machine learning experiments run on the DanPASS corpora with the purpose of investigating the effect of information about stress, pauses and hesitations on the automatic recognition of the function of sn-pronouns. These experiments show that stress information improved classification on the DanPASS monologues significantly. Pause and hesitation information improved classification on both DanPASS corpora, but the improvement is not significant.

In future experiments we will add to the datasets the prosodic information has been marked in the DanPASS data after we had annotated the corpora with information about sn-pronouns.

References

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, 24(1):65–87.

Asher, N. (1993). *Reference to Abstract Objects in Discourse*, volume 50 of *Studies in Linguistics and Philosophy*. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Bergsma, S. Lin, D.Goebel, R. (2009) Distributional Identification of Non-Referential Pronouns. In *Proceedings of ACL 2008: HLT*, pp. 10-18. Columbus, Ohio, Jume 2008.

Byron, D.K.(2002).Resolving pronominal reference to abstract entities. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL '02)*, pages 80–87

Borthen, k., T., Fretheim, J.K., Gundel (1997). What brings a higher-order entity into focus of attention? Sentential pronouns in English and Norwegian. In Mitkov, R., & B., Boguraev (eds), *Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, pages 88–93.

Boyd, A., W., Gegg-Harrison, D., Byron (2005): Identifying non-referential *it*: a machine learning approach incorporating linguistically motivated patterns. In *Proceedings of the ACL Workshop on Feature Engineering for Machine Learning in NLP*, pages 40–47, Ann Arbor Michigan, June.

Carletta, J. (1996) Assessing agreement on classification tasks: the kappa statistics. *Computational Linguistics*, 22(2):249–254.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Education and Psychological Measurement*, 20:37–46.

Daelemans, W., Hoste, V., De Meulder, F., Naudts, B. (2003). Combined Optimization of Feature Selection and Algorithm Parameter Interaction in Machine Learning of Language. In *Proceedings of the 14th European Conference on Machine Learning (ECML-2003)*, pages 84–95, Cavtat-Dubrovnik, Croatia

Eckert, M., M. Strube (2001). Dialogue Acts, Synchronising Units and Anaphora Resolution. In: *Journal of Semantics 17*(1):51-89.

Evans. R. (2001). Applying machine learning toward an automatic classification of *it*. *Literary and Linguistic Computing*, 16(1):45.57.

Fraurud, K. (1992). *Processing Noun Phrases in Natural Discourse.* Department of Linguistics - Stockholm University.

Givón, T. (ed.) (1983). *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. John Benjamin, Amsterdam.

Gregersen, F. (2007). The LANCHART Corpus of Spoken Danish, Report from a corpus in progress. In *Current Trends in Research on Spoken Language in the Nordic Countries*, pages 130–143. Oulu University Press.

Grønnum, N. (2006) DanPASS - A Danish Phonetically Annotated Spontaneous Speech Corpus. In Calzolari, N., K., Choukri, A., Gangemi, B., Maegaard, J., Mariani, J., Odijk, D., Tapias (eds), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, Genova, Italy, May.

Gundel, J.K., N. Hedberg, R. Zacharski (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):274–307.

Gundel, J. K., M.,Hegarty, K., Borthen, (2003). Cognitive Status, Information Structure, and Pronominal Reference to Clausally Introduced Entities. *Journal of Logic, Language and Information*, 12:281-299, Kluwer Academic Publishers.

Gundel J. K., N., Hedberg, R., Zacharski (2004). Demonstrative Pronouns in Natural Discourse. In: Branco, A., T., McEnery, R., Mitkov (eds.): *Proceedings of DAARC 2004 -* 2004, Ediçoes Colibri, 81-86.

Gundel, J. K., n., Hedberg, R., Zacharski (2005). Pronouns without NP antecedents. In: Branco, A., McEnery, T.& R. Mitkov (eds.): *Anaphora Processing – Linguistic, cognitive and computational modelling.* Current Issues in Linguistic Theory. John Benjamins Publishing Company. 351-364.

Hajičová, E., P., Kuboň, V., Kuboň (1990) Hierarchy of Salience and Discourse Analysis and Production. In Karlgren, H. editor, *Proceedings of the 13th International Conference on Computational Linguistics (COLING'90)*, volume III, pages 144–148, Helsinki.

Hegarty. M. (2003) Semantic types of abstract entities. *Lingua*, 113:891–927.

Hoste, V., Hendrickx, I., Daelemans, W. (2007). Disambiguation of the Neuter Pronoun and Its Effect on Pronominal Coreference Resolution. In V. Matousek and P. Mautner (eds.)*Text, Speech and Dialogue, Proceedings of the 10th International Conference, (TSD 2007)*, volume 4629 of *Lecture Notes in Computer Science*, pages 48–55, Pilsen, Czech Republic, September. Springer.

Kaiser, E. (2000) Pronouns and demonstratives in finnish: Indicators of referent salience. In Baker, P., A., Hardie, T., McEnery, A., Siewierska, editors, *Proceedings of the Discourse Anaphora and Anaphor Resolution Conference*, volume 12 of *University Center for Computer Corpus Research on Language - Technical Series*, pages 20–27, Lancaster, UK.

Lyons, J. (1977) *Semantics*, volume I-II. Cambridge University Press.

Müller, C. (2006). Automatic detection of nonreferential It in spoken multi-party dialog. In Proceedings of EACL- 2006, pp. 49-56.

Müller, C. (2007) Resolving it, this and that in unrestricted multi-party dialog. In *Proceedings of ACL-2007*, pages 816–823, Prague.

Navarretta, C., S., Olsen (2008). Annotating abstract pronominal anaphora in the DAD project. In *Proceedings of LREC-2008*, Marrakesh, Morocco, May. ELRA.

C. Navarretta. (2000) Centering-Based Anaphora Resolution in Danish Dialogues. In: P.Sojka, I. Kopecek, K. Pala (eds.) *Text, Speech and Dialogue -* Proceedings of the Third International Workshop on Text, Speech, Dialogue - TSD 2000, Brno, Czech Republic, September 13-16, 2000, Springer-Verlag, 345-350.

Navarretta, C. (2002). *The Use and Resolution of Intersentential Pronominal Anaphora in Danish Discourse*. Ph.D. thesis - Copenhagen University.

C. Navarretta. (2004) The main reference mechanisms of danish demonstrative pronominal anaphors. In Branco, A., T., McEnery, R., Mitkov, editors, *Proceedings of DAARC-2004- 5th Discourse Anaphora and Anaphora Resolution Colloquium*, pages 115–120, Furnal, S.Miguel, Portugal. Ediçoes Colibri.

Navarretta, C. (2004a). Resolving Individual and Abstract Anaphora in Texts and Dialogues. In *Proceedings of the 20th International Conference of Computational Linguistics, COLING-2004*, Geneva, Switzerland, 233-239.

Navarretta, C. (2005) Combining information structure and centering-based models of salience for resolving danish intersentential pronominal anaphora. In A. Branco, T. McEnery, and R. Mitkov, editors, *Anaphora Processing. Linguistic, cognitive and computational modeling*, pages 329–350. John Benjamins Publishing Company.

Navarretta, C. (2007) A contrastive analysis of the use of abstract anaphora. In Branco, A., T., McEnery, R., Mitkov, and F., Silva, editors, *In Proceedings of DAARC-2007 - 6th Discourse Anaphora and Anaphora Resolution Colloquium*, pages 103–109, Lagos, Portugal, March. Centro de Linguistica da Universidade do Porto.

C. Navarretta. 2009. Automatic recognition of the function of third-person singular pronouns in texts and spoken data. In: S. Lalitha Devi, A. Branco and R. Mitkov (eds.) *DAARC 2009. LNAI 5847 proceedings*. pp. 15-28. Springer Verlag Berlin Heidelberg.

Orasan, C. (2003). PALinkA: a highly customizable tool for discourse annotation. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialog*, pages 39–43, Sapporo.

Poesio, M. (2004) "The MATE/GNOME scheme for anaphoric annotation, revisited", *Proc. of SIGDIAL*, Boston, April.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole, editor, *Radical Pragmatics*, Academic Press, pages 223-255.

Strube, M., Müller, C. 2003. A machine learning approach to pronoun resolution in spoken dialogue. In *Proceedings of the ACL'03*, pages 168–175

Webber, B.L. (1988). Discourse deixis and discourse processing. Technical report, University of Pennsylvania.

Witten, I.H., Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, second edition.