Project ref. no.	LE4-8370
Project title	MATE

Deliverable status	Public
Contractual date of delivery	31 July 1998
Actual date of delivery	31 July 1998
Deliverable number	D1.1
Deliverable title	Supported Coding Schemes
Type	RE
Status & version	Final
Number of pages	284
WP contributing to the deliverable	WP1
WP / Task responsible	Marion Klein, DFKI
Author(s)	Niels Ole Bernsen, Sarah Davies, Laila Dybkjær, Juanma Garrido, Henrik Kasch, Marion Klein, Andreas Mengel, Vito Pirrelli, Massimo Poesio, Silvia Quazza, Claudia Soria
EC Project Officer	Ray Hudson
Keywords	Coding schemes, communication problems, co-reference, cross-level issues, dialogue acts, multilevel annotation, morphosyntax, prosody, standardization, tools engineering
Abstract (for dissemination)	The first step of the MATE project is to define an overall mark-up formalism which is based on the TEI/CES standards. This formalism accommodates the needs of current and emerging coding schemes for the levels of prosody, (morpho-) syntax, co-reference, dialogue acts, communication problems, and for cross-level issues. In order to accomplish this, a survey of relevant existing coding schemes is required. These schemes should have proved their reliability by having been used by several novice and/or expert users for annotating at least one corpus of reasonable size. The report presents a survey of coding schemes which meet this criterion. The coding schemes are described in detail with regard to their coding book, number of annotators who have worked with it, number of annotated dialogues / segments / utterances, evaluation results, underlying task, list of annotated phenomena, and mark-up language used. Also annotation examples are provided.

MATE Deliverable D1.1
Supported Coding Schemes

31 July 1998

Authors

Marion Klein, Niels Ole Bernsen, Sarah Davies, Laila Dybkjær,
Juanma Garrido, Henrik Kasch, Andreas Mengel,
Vito Pirrelli, Massimo Poesio, Silvia Quazza, Claudia Soria

Executive Summary

This report gives an overview of the state of the art of coding schemes. Schemes for the levels of prosody, morpho-syntax, co-reference, dialogue acts, communication problems, and cross-level issues have been examined. In order to allow an appropriate comparison of schemes guidelines have been developed (s. section 1.3). These guidelines will guide the decision of which coding schemes will be supported by the MATE project and which ones might be of less interest as they lack reliability. The MATE annotation standards are going to be developed on the basis of the results of this report.

A brief overview of the chapters of this report is given below:

Chapter 1 gives a general introduction to the theme, summarizes the project's approach and discusses the guidelines which are used to standardize the retrieval of important information about schemes.

Chapters 2 – 7 present the state of the art of the five different annotation levels which MATE is going to investigate plus cross-level:

- **Chapter 2** - level of communication problems
- **Chapter 3** - level of coreference
- **Chapter 4** - level of dialogue acts
- **Chapter 5** - level of morpho-syntax
- **Chapter 6** - level of prosody
- **Chapter 7** - cross-level

Chapter 8 draws conclusions about the scheme comparisons on the different levels and outlines future work.

A detailed list of all schemes under consideration can be found in **Annexes.**

Related Deliverables

D1.2 Guidelines for specifying the formal representation of coding schemes. Report.

D2.1 DTDs and notes on schemes. Report.

Issue

Issue 1, Revision 0, July 1998

Acknowledgements

We would like to thank Masahiro Araki, Florence Bruneseaux, Sherri Condon, Mark Core, Barbara di Eugenio, Giovanni Flammia, Arne Jönsson, Staffan Larsson, Lori Levin, Christine Nakatani, Joakim Nivre, Laurent Romary, Jacques Terken, Ann Thyme-Gobbel, and Hans de Vreught for providing information on their schemes.

Contents

Glossary of Terms

Cross-level annotation	relation, that is established between any two or more linguistic units that are considered distinct in phenomenological classes
DRI	Discourse Resource Initiative
EAGLES	Expert Advisory Group on Language Standards
Explicit tags	tags are represented by character combinations and included in the representation of the target of description
Hierarchical annotation	theory dependent and consistent annotation of hierarchically structured levels of description
Layout tags	tags are represented by special positions of the representation of the targets of description
LE	Language Engineering
MUC	Message Understanding Conference
Multilevel annotation	annotation on more than one level of description
NLP	Natural Language Processing

1. Overview

1.1 Introduction

During the last years, corpus based approaches have gained significant importance in the field of natural language processing (NLP). Large corpora for many different languages are currently being collected all over the world. In order to reuse this amount of data for training and testing purposes of NLP systems, the corpora must be annotated in various ways [Carletta et al. 1997]. This annotation assumes an underlying coding scheme. The way such schemes are designed depends on the task and the linguistic phenomena on which developers focus. The author's own style also has its effects on the scheme. For these reasons reusability of annotated corpora is extremely complicated.

The Discourse Resource Initiative (DRI) was started as an effort to assemble discourse resources to support discourse research and application. The goal of this initiative is to develop a standard for semantic / pragmatic and discourse features of annotated corpora [Carletta et al. 1997]. Another project, LE-EAGLES, also has the goal to provide preliminary guidelines for the representation or annotation of dialogue resources for language engineering [Leech et al. 1998]. These guidelines cover the areas of orthographic transcription, morpho-syntactic, syntactic, prosodic, and pragmatic annotation. But instead of developing a standard they describe the most used schemes, mark-up languages and systems for annotation.

1.2 MATE's Aims

MATE aims to develop a preliminary form of standard concerning annotation schemes on the levels of prosody, morpho-syntax, co-reference, dialogue acts, and communication problems, as well as their interaction.

MATE's annotation standard is meant to be closely related to the standardisation efforts in the US, Europe and Japan and will thus build on the work of DRI and EAGLES, mentioned above.

The annotation standard will allow multi-linguality and the co-existence of a multitude of coding schemes. This report provides the basis for the decision on which existing coding schemes MATE should support. It represents a broad overview on current schemes and covers all levels under consideration.

The information collected in this report was collected from the web, from recent proceedings and through personal contact. In the future we will continue our search of schemes which, by accident, were not included in this report. A web version of this report will be available and regularly updated even after the deadline of deliverable D1.1.

The results of this report will feed into the work on implementation of the MATE workbench (WP3) which is a tool box in support of the MATE standard, and they will form the basis of the definition of level mark-up (WP2).

1.3 Scheme Evaluation Guidelines

Lots of research has been done in the field of annotation schemes. Therefore one has to carefully look at all schemes to make the right decision whether the scheme will be supported in the MATE project or if it doesn't seem to be reliable enough and, hence, has to be omitted. To ease this, decision guidelines have been used which are listed below:

- **Existence of coding book**

  The schemes have to be well-documented. Therefore a coding book has to be available that describes the purpose, the domain, and the application for which the scheme has been developed.

- **Number of annotators**

  The schemes must have been used by a decent number of different coders. This is because coding schemes that have only been used by their developers tend to be too subjective and difficult to use.

- **Number of dialogues / utterances / segments**

  The schemes must have been used for annotating a certain number of dialogues to prove their usability.

- **Evaluations of scheme**

  The evaluation of inter-coder agreement reflects the reliability of the coding scheme. As a common measurement the κ-value is used.

  The κ-coefficient is computed as

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)}$$

  where P(A) represents the probability that the annotators agree, while P(E) stands for the probability that the coders agree by chance. The per chance agreement is determined as:

  It appears to be settling on the interpretation that coding schemes with overall reliabilities at

$$P(E) = \sum_{i=1}^{n} p_i^2$$

  κ = 0.8 or higher are good enough that it is not necessary to try to improve on them, and values between 0.67 and 0.8 allow tentative conclusions to be drawn but indicate that the scheme should be improved [Carletta 96].

  Another measurement that should be mentioned is the α-value [Krippendorff 80]. It's calculated as:

$$\alpha = 1 - \frac{D_O}{D_E}$$

  Where

$$D_O = observed\_disagreements$$
$$D_E = \exp ected\_disagreements$$

- **Underlying task**

  Schemes for annotation are frequently linked to an underlying domain or task. This might reduce their general usability.

- **List of phenomena annotated**

  For comparison of schemes and the development of a standard a list of phenomena is essential.

- **Examples**

  For a better understanding of schemes examples are essential.

- **Mark-up language**

  The mark-up language of a scheme has to be known for writing mark-up translators, for instance. Also it is interesting to see which mark-up language is used most.

- **Existence of annotation tools**

  Annotation tools for schemes make annotation easier and are therefore more likely to be used. Also the tool might be integrated in the MATE workbench.

- **Usability**

  Schemes should be used in existing systems to show their usability.

- **Contact**

  In order to gain further information on schemes a contact address is given.

All schemes which have been observed are tested with these guidelines. A detailed listing of all schemes can be found in the Appendices.

1.4 Further Content of this Document

Chapters 2 to 7 provide insight into the level-stages on communication problems, co-reference, cross-level issues, dialogue acts, (morpho)-syntax and prosody. The observed schemes which can be found in Annexes are compared with regard to their levels. At the end of each chapter conclusions concerning the supported schemes are drawn from these comparisons.

In chapter 8, a summary of the results of the research on the different levels is presented and further work that could be done in this field is outlined.

2. Communication Problems

Responsibility: Laila Dybkjær, Niels Ole Bernsen, Henrik Kasch

2.1 Introduction

Cooperativity is a central issue both in human-computer dialogue and in human-human dialogue. Non-cooperative behaviour easily leads to miscommunication and an unnecessarily long, complicated and perhaps failed dialogue. In particular in human-computer interaction the consequences may often be a totally failed dialogue because of the system's limited abilities to detect, handle and recover from non-cooperative dialogue [Bernsen et al. 1996].

Research in aspects of communication, including various communication problems and cooperativity, is not new (see also below). However, there is no exhaustive theory on communication problems and cooperativity, and much of the research has focused on human-human dialogue. With an increasing number of more and more advanced spoken language dialogue systems becoming a commercial reality there is also an increasing need for methods and tools which can help developers in producing dialogue systems that offer the user as smooth and user-friendly interaction as possible.

Communication problems, if detected, typically lead to clarification or repair meta-communication. Human-human dialogue both allows for and is greatly assisted by clarification and repair meta-communication. If we are in doubt as to what our interlocutor said or meant, why a particular topic was raised, why it was raised at that particular point, or why it was raised in a particular way during dialogue, we initiate clarification and repair meta-communication to find out. Similarly, speakers often take advantage of the fact that their partners can demand elaboration at any point. This helps fine-tuning the speaker's contributions and indicates interest from the partner. The standard way of initiating clarification and repair meta-communication is by asking questions of the interlocutor [Schegloff et al. 1977].

However, for spoken language dialogue systems the situation is different. With current technologies, the possibilities of on-line handling of clarification and repair meta-communication are seriously limited. It is sometimes assumed that, as long as the system has powerful meta-communication abilities, it matters less how it behaves during domain communication. This is false already because the generation of bumpy interaction is always inefficient and induces user dissatisfaction. What is worse, however, is that really powerful meta-communication abilities are not feasible today. User needs for clarification meta-communication that arise from the way the system addresses the domain can easily surpass its meta-communication skills. For instance, if the system uses a patently ambiguous term it is unlikely that it will be able to respond sensibly to the user who asks what the system means by that term. And if the user unknowingly selects a non-intended meaning of an ambiguous term, the interaction may be well underway towards failure without the system being able to do much about it. Thus co-operative communication is important because it facilitates smooth interaction and prevents unnecessary user-initiated clarification and repair meta-communication, as well as other kinds of unexpected user behaviour with which the system cannot cope.

In spite of this the detection of communication problems in spoken language dialogue systems has so far usually been done on an ad hoc basis without much systematicity and, if at all, only at a fairly late stage as part of the evaluation. In order to support a more cost-efficient development process for interaction models for spoken dialogue systems we need a solid understanding of communication problems, their nature and why they occur. Not only an investigation of human-computer dialogue corpora is important. We also need to look at human-human dialogues because as spoken dialogue systems becomes increasingly advanced also the language used by the system as well as its users will change and become more like that of human-human dialogue. Annotation of communication problems in spoken dialogue corpora may not only help developers and researchers to extract information on deficiencies of a system and thereby perhaps also get clues to how to improve it. It may also help us to get the insight needed to develop methods or tools that might serve as an efficient and systematic development and evaluation tool during early analysis and design, thereby improving interaction model quality and at the same time reducing development costs.

The following sections briefly describe previous work in the area of communication problems, discuss communication problems and their annotation, provide a comparison of the four schemes presented in the Annexes under communication problems and summarise and conclude on existing and future work.

2.2 Previous work in the area

Several researchers have studied aspects of communication problems. These studies, however, have mainly focused on human-human dialogue, e.g. [Austin 1962] (locution, illocution and perlocution),[Searle 1969] (speech acts), and [Grice 1975] (the cooperative principle). Among the aspects studied are communication failure and cooperativity. For instance [Schegloff et al. 1977] have studied repair and clarification acts in human dialogue, [Clark and Shaefer 1987] have written about collaborative communication, and [Grice 1975] proposed the Cooperative Principle which can be explicated in terms of four groups of simple maxims of cooperative human-human conversation.

Communication theories from human-human dialogue are drawn upon in the area of human-computer dialogue. However, neither in the area of human-human dialogue nor in the area of human-computer dialogue are there any close-to-exhaustive theories with respect to communication problems. In addition of the work referenced in the preceding paragraph, relevant work on communication problems can be found in the context of evaluation of speech and natural language processing systems. There is a rapidly growing body of work on evaluation of natural language processing systems in general and of dialogue interaction in particular. Overviews are [Jones and Galliers 1996] and [Hirschmann and Thompson 1996]. Other recent references are [McRoy 1996], [Calzolari and McNaught 1996], [Gaizauskas 1997], [Hirschberg et al. 1997], [Gibbon et al. 1997], and [Bernsen et al. 1998].

2.3 Communication problems and their annotation

It thus appears that one has to look in several different places to collect information about communication problems. The reason for this is probably that communication problems are inherently cross-level issues to a large extent. Some problems are caused by speech phenomena such as overlaps. Other problems may be caused by an interlocutor expressing himself in an obscure way, e.g. through use of strange grammatical constructs. There are problems which are

mainly related to the opaque use of co-reference or indirect speech acts. A dialogue partner may fail to take into account the background knowledge of the other interlocutor(s), or he may not have made his own special communication situation clear to his communication partner, such as that of not being able to hear very well, or being on the phone in a noisy environment. The list of possible causes of miscommunication is long. To illustrate the variety of communication problems a few examples are provided in Figures 1 to 4. All four examples are from the corpora annotated using the Odense coding scheme.

S: Customer number 2.
Is that correct?

U: No id-numbers 1 and 4.

S: Customer number 4.

U: **Wrong, the customer number is 2.**

S: How many people will travel?

Figure 1: Boldface shows overlapping user speech. The user tries to make a change. However, the system does not listen during its own turns. Users were not informed on this as they should have been. This implied that the user thought that the system actually caught what s/he said. Only much later in the dialogue did s/he realise that this was not the case.

U: yes I was wondering (.3) whether flight number bee ay two three eight from orlando has arrived (.3) it was supposed to arrive at twenty five to eight (8.5)

S: please wait (37)

flight bee ay two three eight from orlando has been delayed (1.7)

it will be arriving at eleven fifteen (2)

U: is that eleven sixteen in the evening (5)

Figure 2: It is not clear whether the time provided by the system is a.m. or p.m. (ambiguity). Thus the user has to ask a new question in order to have it clarified.

S: When do you want to go from Ulm to Coburg?

U: No, no not to Coburg, to Stuttgart.

S: I'm sorry, I did not understand you. Please repeat when you want to go from Ulm to Coburg.

U: No, I did not mean Coburg, I mean Stuttgart.

S: When du you want to go from Neumuenster to Coburg.

U: No, you do not understand me. I want to go from Ulm to Stuttgart.

Figure 3: The system continuously misunderstands the user and produces a series of contextually irrelevant questions.

S: Id-number 89, Carl Frederiksen has two reservations. One on Friday September 11 from Aalborg to Copenhagen and one on Saturday September 10 from Aalborg to Copenhagen. Is it one of these?

U: Ahm ...something is wrong because Saturday September 10 must be a Thursday if September 11 is a Friday. Hmm ... Friday September 11 at 7:20.

Figure 4: The system's information on dates clearly cannot be true as pointed out by the user. The error was made by a wizard. Careful domain model design is needed to avoid such problems.

Coding schemes tailored to the description of communication problems are very few. However, as already indicated, aspects of communication problems may be reflected in many coding schemes for the other levels which MATE will be looking at, e.g. dialogue acts. Three of the four communication problems coding schemes presented in the Annexes do not have a direct focus on communication problems. Only the Odense coding scheme is exclusively on communication problems. The three others primarily address (morpho-)syntax and/or dialogue acts. There may be other schemes in the Annexes which, with the same right, could have been mentioned under communication problems because they include markup of some such problems.

The fact that only one of the coding schemes is entirely directed towards communication problems indicates that there is much work to be done in this area since there is little to build on. The Odense coding scheme will be used as the starting point in our further work on communication problems in MATE. This scheme was developed in support of the design of cooperative system utterances in spoken human-machine dialogue. The cooperativity guidelines [Bernsen et al. 1998] which form the basis of the scheme were developed from a set of simulated human-machine dialogues. They were then compared to Grice's maxims [Grice 1975] and shown to significantly extend these maxims, and they stood up to the test when we used the coding scheme on new Danish, English and German spoken human-machine corpora. Thus the guidelines are both theoretically founded and application-oriented.

To test and possibly extend the Odense scheme we propose the following two steps: (i) The coding scheme should be tested on new task-oriented spoken human-machine corpora. (ii) We should draw on the unique possibility which MATE provides by making available information on so many different coding schemes many of which include annotation of phenomena which potentially represent communication problems in spoken human-human dialogue. We would like to confront our cooperativity guidelines with these phenomena. This could provide an acid test of the theoretical closure of out approach, for the following reason: current spoken human-computer dialogues are limited in many respects, A discourse-level approach to communication problems aimed at preventing bad machine utterance design therefore risks not being confronted with the full set of phenomena it should be able to handle, if it is being tested only through the analysis of corpora from existing or simulated human-computer spoken language dialogue.

Moreover, we should look at i.a. the work on evaluation of natural language processing systems and dialogue interaction mentioned above. This may also provide input on phenomena of interest even if they have not directly been marked up in a corpus.

Work to be done moreover includes an investigation of when something should be marked as a communication problem. From one point of view only actually occurring problems should be marked. In some kinds of evaluation this may be the most fair viewpoint, e.g. if one compares final systems. On the other hand, if there is a potential risk for a spoken language dialogue system that a certain phenomenon may cause problems we may want to know. This is particularly interesting during development of a system and when carrying out diagnostic evaluation. In particular, we may want to know under which circumstances problems are likely to occur. This speaks in favour of marking up also potential communication problems. Thus we need to investigate which phenomena it may be relevant to mark up in the first place, the nature of these phenomena and, possibly, the circumstances under which they should be marked up.

2.4 Scheme comparison

Four coding schemes addressing, at least to some extent, communication problems are presented in the Annexes of this report. The table below provides an overview of the schemes and allows for comparison.

	Odense	**Davies**	**CHAT**	**Switchboard**
Coding book	Incomplete, in particular no operational coding decision support is available.	Incomplete, in particular no operational coding decision support is available.	Yes.	Yes.
No. of annotators	Three (2 experts, 1 novice).	Three (1 expert, 2 novices).	Used by around 60 groups all over the world.	9.
No. of dialogues	132.	32.	A huge number.	2400.
Language(s)	Danish, English, German.	English.	20 different languages.	English.
Degree of evaluation of scheme	Moderate.	Reliability evaluation on a very small data set: $0.69 < \kappa < 1.0$	Continuous evaluation and update. No statistical/quantitative evaluation of reliability.	Reliability evaluation: $0.8 < \kappa < 0.84$.
Underlying task	Flight ticket reservation, flight information and train timetable information.	HCRC Map Task dialogues are analysed in order to examine cooperation, effort and risk in task-oriented	Parent-to-child or child-to-child spontaneous conversation, task-oriented dialogues in play and story-telling	Spontaneous telephone conversations.

		dialogues.	situations.	
Task-oriented	Yes.	Yes.	Not only.	No.
Dialogue partners	Simulated human machine, human-machine	Human-human.	Human-human.	Machine-mediated human-human
Phenomena annotated	Communication problems (system errors and user errors).	Dialogue acts.	Speech management related phenomena (overlap, false starts, etc.), paralinguistic features (e.g. laughter), situational features (e.g. non-verbal actions), word-form related phenomena, syntax related phenomena, lexical and syntactic and other errors, and speech acts.	Speech management related phenomena (overlap, false starts, etc.), paralinguistic features (e.g. laughter), situational features (e.g. non-verbal actions), word-form related phenomena, syntax related phenomena, and speech acts.
Markup language	TEI extension.	Homegrown.	CHAT's own format.	DAMSL-variant
Supporting annotation tools	None.	None.	Editor (CED), analysis tools (CLAN).	No.
Theoretical basis	Cooperativity theory.	Discourse theory.	(Morpho-) syntax, dialogue acts.	(Morpho-) syntax, dialogue acts.
Coverage	Systematic approach to communication problems.	Communication problems are not in focus but a few phenomena are mentioned.	Communication problems are not in focus but several phenomena are included which may cause communication problems.	Communication problems are not in focus but several phenomena are included which may cause communication problems.
Scope of scheme	Not sufficiently analysed but assumed to cover at least task-oriented, shared-goal human-machine dialogues.	Not sufficiently analysed.	Meant to meet the needs of different languages and different users.	Machine-mediated switchboard operation.

2.5 Concluding Remarks

Annotation of communication problems is obviously in its infancy, and very few coding schemes exist. Moreover, communication problems are cross-level issues by their nature. Some communication problems are caused by flawed grammar or vocabulary design ((morpho-) syntactic level). Other problems may be due to misinterpretation or non-interpretation of co-references, etc. Our next steps will therefore be (i) to test the Odense coding scheme on new task-oriented spoken human-machine corpora, (ii) to look at coding schemes for other levels in which communications problems are being addressed, and (iii) to study literature concerning communication problems, in particular literature referring to empirical investigations of spoken human-machine dialogue. This work will allow us to consolidate and, if needed, extend the set of cooperativity guidelines listed under the Odense coding scheme (see the Annexes) and to use this list as a starting point for the definition of a first markup language for the level of communication problems.

3. Co-reference

Responsibility: Sarah Davies, Massimo Poesio

3.1 Introduction

In most texts the object under discussion are mentioned more than once, and what gets mentioned anew related in various ways to what has already been discussed. Subsequent mentions of an entity can have the same surface form - as when the expression *the Lord Provost* is encountered twice in a text - or different ones; indeed, there is a whole class of expressions called ANAPHORIC EXPRESSIONS used to indicate that elements of a text are correlated. The simplest forms of anaphoric expression are used to indicate a subsequent mention of an object already introduced: typical examples of this type of anaphoric expression are pronouns such as *he* in the text *John arrived. He looked tired.*. In the preferred reading of this text, the pronoun *he* is interpreted as an `abbreviated reference,' to the individual John which is denoted by the expression *John*.

Annotating corpora with information about such relations between elements of a text is useful both from a linguistic point of view and for applications such as information extraction; and unlike with other types of semantic information, there is sufficient agreement among researchers on the basic facts about anaphora that it may make sense to try this kind of annotation. Yet, work in this area is only starting; we are aware of only one large scale effort on annotating anaphoric information (at the University of Lancaster), and only a few annotation manuals have been produced. This document contains a brief overview of some of the issues that have to be considered when developing schemes for doing this type of annotation, as well as a review of some of the existing annotation schemes.

Following the terminology introduced by Sidner (1979), we will say that in the example just discussed the pronoun *he* CO-SPECIFIES with the proper name *John*, and we will call John the ANTECEDENT of the pronoun. We will also say that two strings CO-REFER when they point to the same entity in the world. In the example above, the pronoun *he* and the proper name *John* both co-specify and co-refer; more in general, two expressions may co-specify.

3.2 Expressions which may enter into co-specification relations

The typical use of (different kinds of) anaphoric expressions in a text is illustrated in anaphora-types: ex (a recipe posted to the newsgroup rec.food.cooking, reproduced verbatim including typos).

Ingredients:

1/8 cup warm water

1 package dry yeast

1 ¼ cup hot water

1/3 packed brown sugar

4-5 cups white flour

lots of kosher salt

Tbs of backing soda

Add THE DRY YEAST to THE WATER and let sit _ for a few minutes. Add THE REST OF THE WATER and sugar. Stir _ and let sit _ . Slowly add THE FLOUR and stir. Add enough flour so you can knead THE FLOUR in, but isn't sticky any more. Take golf ball size pieces of THE DOUGH and roll _ into pretzel shapes or any shape you want. In a frying pan or pot – boil water with backing soda (about 2 cups of water to a tablespoon of backing soda).

Take THE PRETZELS and place THEM in THE WATER for 30 seconds or until THEY float. Put THEM on a generously greased and salted cookie sheet. Salt THEM generously on top as well. Back _ at 375 for 8-15 minutes.

Figure 1: A recipe for pretzels.

The types of anaphoric expressions found in this text, include the pronouns *them* and *they*, DEFINITE DESCRIPTIONS such as *the dry yeast* and *the water*, and examples of NOMINAL ELLIPSIS, as *in Stir_and let sit _ ,* where the arguments of *stir* and let *sit* have been omitted. Other common types of anaphoric expressions not occurring in this example include POSSESSIVE DESCRIPTIONS such as *her computer*; and *one* ANAPHORA, as in *He saw that the plugs were worn, and advised John to fit some new ones*. Because of the high frequency of anaphoric expressions, resolving them - i.e., identifying their antecedent so that, for example, we can understand what exactly should be put in the water in the example recipe - is one of the main problems confronting the designer of a natural language processing system.

The classification of anaphoric expressions just presented is based on their form, i.e., their surface realization. These differences in syntactic form of anaphoric expressions correlate with differences in the way their antecedents are identified; in general, the more information is provided by the anaphoric expression, the less constrained is its use.

Anaphoric expressions can also be classified according to their semantic properties; this classification cuts across the syntactic classification just presented. For one thing, anaphoric expressions differ in the type of objects they denote: singular objects such as *the water* as well as plural objects such as *the pretzels* or *them*; uncountable entities such as *water* again, or countable objects such as pretzels; concrete objects such as those discussed in the recipe in Fig. 1, as well as more abstract objects such as events, like the one denoted by *it* in *John had a car accident yesterday. It happened at the corner of 34th Street with 2nd Avenue.*. Indeed,

anaphoric expressions do not only denote objects, but also properties; this is typically the case in VERBAL ELLIPSIS. In *Bill went home. John did too*, for example, the antecedent of the expression *did too* is the property denoted by the verb phrase *went home*; this property is attributed to *John* as well.

Anaphoric expressions also differ in the semantic relation that holds between the expression and its antecedent. Most of the anaphoric expressions in Fig. 1, as well the other examples just discussed, have the same denotation as their antecedent: for example, *the dry yeast* in Fig. 1 denotes the same object as the element of the ingredient list *1 package dry yeast*. In other cases, the semantic relation between the denotation of an anaphoric expression and the denotation of its antecedent is less direct. One example is the expression one in *Wendy prefers the red T-shirt to the yellow one*. In this case, we are talking about two distinct T-shirts, of different colors. The expression *one* thus denotes something like an object type rather than an object token. Pronouns can enter in the same type of semantic relation with their antecedents, albeit more rarely: the classical example of this are sentences such as *The man who gave his paycheck to his wife was wiser than the man who gave it to his mistress*, which give this kind of pronouns the name PAYCHECK PRONOUNS.

BOUND pronouns [Partee 1972] are another example of indirect relation between an anaphoric expression and its antecedent. In *Nobody likes to lose his job*, the pronoun *his* does not `refer' to the same object as its antecedent, the quantifier *nobody* (which does not refer to anything); this anaphoric expression is best seen as playing the role of a variable in first order logic. The sentence can be paraphrased as follows:

(1) $\neg\exists x$**likes**(x,**lose**(x,**jop-of**(x)))

where **job-of**(x) denotes a function from people to their (unique) job.

Finally, so-called BRIDGING REFERENCES [Clark 1977] are anaphoric expressions that denote objects only related to the denotation of their antecedent by (shared) generic knowledge. An example is *the indicators* in *John has bought a new car*. *The indicators use the latest laser technology*. We are able to interpret the description *the indicators* because we know that indicators are a part of cars, and a car was mentioned in the first sentence. Some of the relations that may hold between a bridging reference and its antecedent include part-whole as in the example just seen, and element-set (as in *The Italian team didn't play well yesterday until the centre-forward was replaced in the 30th minute*). A bridging reference may also refer to the object filling a role in an event, whether implicitly or explicitly introduced: an example is *the victim* in (2) (from the LOB corpus), which refers back to Dr. Slayton's wife by means of her role in the killing event. (A detailed survey of alternative classifications of bridging descriptions proposed in the literature can be found in [Vieira 1998].

(2) After Dr. Allen stepped down, Dr. Slayton took the stand and the questioning was brief. He told how he had gone to the board meeting, stopped for a couple of drinks at Phaedo's and come home to find his wife had been killed. Slayton was obviously suffering on the stand and Herring was gentle with him. Lt. Willis was next and he explained what had been done. The victim's robe, dust from the scene, and fingerprints from all over the house had been sent to the laboratory in Hartford. No clues had been found in the robe or the dust and the fingerprints were still being sorted.

So far, we have seen examples of anaphoric expressions which refer back to an object introduced in the text - Prince called these forms of anaphoric expressions DISCOURSE-OLD

[Prince 1992a] - or are somehow related to it (as in the case of bridging references). However, some of the linguistic expressions that we have seen--definite descriptions, in particular - can also be used to refer to objects which have not been mentioned before: Prince called these expressions DISCOURSE-NEW. An example are expressions such as *the pope or the Queen of England* that refer to entities whose existence can be assumed to be known to the reader. A second example are expressions that are meant to be interpreted with respect to the situation of utterance - called DEICTIC or also INDEXICALS: an example of indexical expression is *the salt* in an utterance of the sentence *pass me the salt, please* in a context in which nobody has mentioned the salt before. The important point about these expressions, as far as our discussion is concerned, is that a system processing definite descriptions must be able to recognize which expressions refer back to the text and which ones are discourse-new.

The anaphoric forms we have seen so far are only one instance of the relations that may exist between expressions in a text. Anaphoricity in the more general sense of dependence on context features is a property of a very large percentage of the tokens in a text. For example, the temporal location of an event often depends on the location of other events described in the text: thus, the preferred interpretation of (3) is one in which the event of turning on the TV occurs after the event of arriving home, whenever that happened [Partee 1973; Kamp 1979; Hinrichs 1981; Dowty 1986; Webber 1987; Lascarides and Asher 1993; Kameyama et al. 1993].

(3) John arrived home, and turned the TV on.

The interpretation of nouns, as well, often depends on the interpretation of other expressions in context: in (2), for example, the interpretation of *clues* in the last sentence (i.e., the set of clues quantified over by *no*) depends on the killing event--we are only talking about clues concerning the specific killing under discussion. The sentence would still be considered true even if the investigators had found clues indicating, for example, that the victim had been reading at the moment of the killing [Enc 1981; von Fintel 1994; Poesio 1994].

Indeed, very few sentences can characterized as providing only new information: most sentences relate whatever new information is provided to information that is presumed to be already known to the reader / listener [Clark and Haviland 1977; Prince 1981; Vallduvi 1990]. And it is not just noun phrases of verb phrases that are given or new; this separation between given and new information can be made also within such constituents --for example, the noun phrase *no clues* in the example above could be characterized as including a relation to an entity already introduced in the discourse, if only implicitly (the clues) as well as a quantification over that set. The term INFORMATION PACKAGING is often used to indicate the way information in a sentence is structured in such a way as to allow the reader to identify given and new information.

3.3 Discourse Models

Most theories of anaphora and anaphora resolution are based on the assumption that the interpretation of anaphoric expressions takes place with respect to a DISCOURSE MODEL that is updated as the text gets processed [Karttunen 1976; Webber 1979; Sidner 1979; Reichman 1981; Kamp 1981; Heim 1982]. A discourse model records information about the entities and events evoked by a discourse (typically called DISCOURSE ENTITIES or DISCOURSE PEGS) and the relationships between them. A reader processing the first sentence in (4), for example, would update her discourse model with several discourse entities, including a discourse entity *y* for *Cliff Thompson* and discourse entity *x* for *a cigarette*; these discourse

entities would be subsequently available when looking for suitable antecedents for the anaphoric expressions *he*, *the cigarette* and *it* in the second sentence. (In other words, the discourse model acts like the `search space' of anaphoric resolution) .

(4) Cliff Thompson lingered in the shade of the Oasis Saloon's overhang, smoking a cigarette, and watching. When the wagon was lost in the shimmering, heat-scourged distance, he flung the cigarette down and stamped on it.

The crucial property of discourse models - indeed, the reason for hypothesizing their existence - is that they `mediate' anaphoric relations. Earlier theories of anaphora assumed that resolving anaphoric expressions involved either `copying' bits of syntactic structure, or took place directly through direct reference to entities in the world. The so-called Bach-Peters sentences, an example of which is (5a), show that copying can't be the whole story, as in this example copying would lead to an infinite recursion (try replacing it with *the Mig who chased him*, then *him* with the resulting *the pilot who shot at the Mig who chased him*, and so forth). And the fact that anaphoric reference to quantified expressions is possible indicates that one may have anaphora without direct reference, since quantified expressions need not refer to anything, as shown in (5b) (from the LOB).

(5) a. The pilot who shot at it hit the Mig who chased him.

b. If there was a law, it should be impartial.

This simple picture is made more complicated by the fact, already discussed in the previous section, that discourse entities are not only introduced as the result of encountering noun phrases: events are added to the discourse model, as well. Indeed, it appears that under certain conditions, the search itself may result in the introduction of new discourse entities. A typical example is anaphoric reference to plural entities: examples such as (6a) show that the antecedent of a plural anaphoric expression may be the combination of discourse entities introduced by singular noun phrases; and examples such as (6b) (similar examples are discussed in) show that in general it's not a goo idea to simply add all possible combinations of discourse entities to the discourse model, as there may be a lot of them.

(6) a. John met Mary at a cafe. After dinner, they went to the movies.

b. John and Mary met Sue and Bill at a cafe. After dinner, they went to the movies.

3.4 Encoding Anaphoric Relations

Given that almost every word in a text may be anaphoric to some extent, hand-annotating all anaphoric expressions and all anaphoric relations is clearly impossible, except for small amounts of text. When designing a scheme for annotating anaphoric relations it is therefore necessary to identify the anaphoric expressions and relations more relevant for one's needs. Narrowing the scope of the scheme may also be necessary in order to achieve good agreement among subjects, as we will see shortly.

Some of the decisions that have to be taken when deciding which elements of the text to tag include: whether to tag all text elements that may enter into anaphoric relations, or only noun phrases, or just a subset of these (e.g., in many schemes first and second person pronouns are not marked); how to identify the boundaries of noun phrases (e.g., whether in *The Admiral's Head, that famous Portsmouth hostelry* the apposition *that famous Portsmouth hostelry* is to be

considered a separate noun phrase); whether to mark empty elements such as those seen in the recipe in Fig. 1; and whether to annotate anaphoric reference to events and other abstract objects and, if so, how. The range of relations of interest may also be limited: some annotators may only be interested in annotating true co-specification, others may be interested in the wider range of relations expressed by bridging references.

And indeed, the schemes for annotating anaphoric expressions proposed in the literature range from very basic schemes only concerned with the simplest cases of anaphora, to very ambitious schemes attempting to capture a large subset of the phenomena discussed in the previous sections. The scheme proposed for MUC-7 (Chinchor and Hirschman, 1997) is designed to identify the noun phrases in a text that may enter in co-specification relations, and only mark identity relations between them; the scheme therefore does not encode relations between abstract objects such as propositions, actions or clauses and a following pronoun referring to them. Also, the MUC annotation scheme only marks up the `identity' relation (where a referring string has the same denotation as its antecedent); all bridging relations are excluded. All markable noun phrases are assigned a COREF tag and an index (the value of the ID attribute); when a noun phrase is related to another discourse entity, that discourse entity is also specified as the value of the REF attribute, and the type of relation is specified as the value of the TYPE attribute. An example is shown in (7).

(7) utt2 : u: yes <COREF ID=”0” TYPE=”IDENT” REF=”1”>I
</COREF>’d like <sil> to
<sil> take <COREF ID=”3”>a tanker</COREF>
from <COREF ID=”12”>Corning</COREF> and
bring <COREF ID=”2” TYPE=”IDENT” REF=”3”>it</COREF>
to <COREF ID=”5”>Elmira</COREF>

The schemes proposed for linguistic purposes [Fligelstone 1992; Bruneseaux and Romary 1997; Passonneau 1997] are more general both in that a greater number of elements are marked and more relations among them are encoded. The scheme proposed by the University of Lancaster group, for example (Fligelstone, 1992), includes both ellipsis and verbal forms among the markables, and includes many types of 'bridging' relationships, rather than just co-specification.

Finally, the classification schemes proposed by Prince (1981; 1992) are designed to encode information packaging in general, not just basic anaphoric expressions and their relation to their antecedent. Prince proposes that each element of a sentence could be classified along two dimensions: whether it's DISCOURSE NEW or DISCOURSE OLD, and whether it's HEARER NEW or HEARER OLD. She also introduces the term INFERRABLES for those entities like bridging references that, properly speaking, are neither new nor old, but whose existence is somehow implicated by what has been explicitly asserted.

3.5 Reliability

Every large-scale annotation involving more than one annotator must address the problem of inter-coder reliability--indeed, even in the case of a single annotator one may raise the issue of whether the annotation scheme is such that we can expect consistency (for example, we may ask whether an annotator would annotate things the same way the second time). This is true with

coreference, as well. Informal evaluations of the reliability of the proposed annotation scheme have been done for MUC, but only systematic study we are aware of is the one presented in .

Poesio and Vieira ran two annotation experiments, using texts from the Wall Street Journal and concentrating on definite descriptions. In the first experiment, subjects were only asked to assign the (pre-determined) markables to one of three classes--anaphoric same head (as in a rig—the rig), associative (roughly corresponding to bridges), and discourse-new. In the second experiment, the subjects were also asked to indicate the antecedent in the text of those definite descriptions that were classified as either anaphoric or bridging, whereas for discourse-new definite descriptions, they had to specify whether they referred to some object whose existence could be expected to be known to the average reader, or to an object that could be expected to be unknown. The agreement on classes was evaluated using the K statistic ; the figures obtained were K=0.72 in the first experiment (involving 1040 definite descriptions), K=0.63 in the second experiment with four classes. When considering only three classes (i.e., when counting all discourse-new definite descriptions as being part of the same class), the measure of agreement in the second experiment improved to K=0.68. 164 descriptions were classified as co-referent by all three coders in the second experiment; of these, 155 (95) were taken by all coders to refer to the same entity. Again in the second experiment, only 7 definite descriptions were classified by all three annotators as associative; in 5 of these cases (71) the three annotators also agreed on a textual anchor.

The results of Poesio and Vieira confirm Fraurud's (1990) impression that the only distinction that can be marked reliably is that between first mentions and subsequent mentions; bridging references proved remarkably difficult to classify reliably. Annotators had trouble distinguishing bridging references both from discourse-new cases (as in (8a), where *the government* can either be interpreted as discourse-new or as related to *Koreans*) and from coreference (as in (8b), where *the reward* can either be interpreted as co-referring with *the payoff* or as bridging *on the bidding*).

(8) a. For the Parks and millions of other Koreans, the long-cherished

dream of home ownership has become a cruel illusion.

The government...

b. New England Electric System bowed out the bidding saying the

potential payoff was too far in the future to justify...

the reward seemed a long way off...

Poesio and Vieira did not consider the problem of agreement on markables, which has caused several problems within MUC. An already mentioned problem with identifying markables is whether to include appositions in noun phrases or not--particularly in cases such *as one of engines at Elmira, say engine E2*. Another issue is whether noun phrases in post-copular position in copular clauses should be considered markable or not. For example, it can be argued that in (9a), *a policeman* is clearly expressing a predicate, and therefore need not be marked, whereas in (9b) (to be imagined being said while looking at the sky at night), *both the planet on the left* and *Venus* are clearly referring expressions; it's not so clear how to handle *the president of the board* in (9c).

(9) a. John is a policeman.

b. The planet on the left is the Venus.

c. John is the president of the board.

3.6 Annotating Coreference in Dialogue

When annotating dialogues, new problems arise, both to identify markables and to express the semantic relation between them. One problem is what to do with hesitations and disfluencies such as repetitions or repairs. In (10a), for example, the noun phrase *one of engines at Elmira, say engine E2* is divided between several utterances, broken by pauses and other hesitations. In (10b), the definite description *the other kids* is repaired into *the kid*.

(10) a. 9.6 : I think that we should do

9.7 : is

9.8 : hook up

9.9 : uh one of the

[2sec]

9.10 : engine

9.11 : uh

9.12 : at Elmira

9.13 : say engine E2

b. and the g guy on the bike gives the other kids

gives the kid that returns his hat...

A second problem with dialogues is that the participants are not always very careful in establishing joint references, and therefore misunderstandings can ensue; this is very common in the MapTask, for example. In this case one has to decide whether to annotate only the interpretation of a given noun phrase intended by the speaker or the interpretations entertained by both participants (in which case, different coreference relations may hold).

3.7 Comparison of Existing Schemes

3.7.1 Brief Description

MUC Coreference Task [Chinchor and Hirschman, 1997]:

The MUC (Message Understanding Conference) (Chinchor et al. 1997) coreference annotation scheme is designed to identify referring strings in the text and mark the relations between these (rather than, say, classifying each referring string by syntactic or semantic type). Potential markables are nouns, noun phrases and pronouns, but these are only marked up if they enter into a coreference relationship with another string; not all referring strings are therefore marked up. Both a referring string and its antecedent have to fall into one of the above syntactic categories in order to be marked; the scheme therefore does not encode relations between a proposition, action or clause and a following pronoun referring to it. The MUC annotation scheme only marks up the 'identity' relation (where a referring string has the same denotation as its antecedent); all bridging relations are excluded.

DRAMA [Passonneau, 1997]:

Discourse Reference Annotation for Multiple Applications (DRAMA) (Passonneau 1997), like MUC, is based on identifying referential expressions in the text, and then marking the relationships between these. The main difference between this scheme and MUC is that DRAMA includes many types of 'bridging' relationships, rather than just dealing with coreference. These different types of bridging relationships are all classified in the mark-up. The set of markables is also greater in DRAMA than it is in MUC, and includes any clauses, verb phrases or adjectives which are deemed the antecedents of expressions like 'it' and 'that'. Unlike MUC, in DRAMA all strings which show the syntagmatic behaviour of noun phrases and introduce a discourse referent are identified, whether or not they enter into coreference relationships.

Poesio and Vieira (1):

The first scheme used by Poesio and Vieira (1998) was used to classify definite noun phrases based on their relationship to other NPs in the text or world. This contrasts with DRAMA and MUC, which link referential expressions in the text. As it was used only to classify definite NPs, the range of markables of this scheme is obviously more limited than that of others. The classification of each phrase included a broad class of bridging relations, though the kind of fine distinctions between types of bridging made in DRAMA is not implemented here. As this scheme was not orientated towards linking phrases, it included the classification of phrases which are not interpreted by reference to other items in the text (i.e. discourse-new), which is not a feature of other schemes.

Poesio and Vieira (2):

Like their previous scheme, Poesio and Vieira's (1998) second scheme was based on classification of definite NPs. However, in this scheme, where the NPs were classified as entering into a relationship with another string in the text, that link was also marked. Again, bridging relations and references external to the text were also included. The range of markables was wider than the previous scheme, as the annotators could link a definite NP with any other string in the text.

Bruneseaux and Romary:

Like MUC, DRAMA and UCREL, Bruneseaux and Romary's scheme (Bruneseaux 1997) identifies the referring expressions in the text and then marks up the relationships between them. Like MUC, only those strings which enter into some kind of relationships with other strings or with items in the visual context are marked up. With respect to the conventions used, this scheme is different to the others in that the relationships between strings are represented in a separate tag from the strings and their IDs - as a 'link' statement between 2 IDs. An important point about this scheme is that it encodes references to the visual context (many of the conversations in the target corpus are taken from human-computer interactions using a geological simulation program). Pointing and mouse-click gestures are also marked. This scheme has a large set of markables, including verbal structures.

*Map Task *U-line annotation:*

This scheme, developed for use with the HCRC Maptask corpus, has a somewhat different orientation to the others, in that it is designed to track the occurrence and form of references to landmarks on the maps which the speakers have in front of them. All references to these landmarks are marked up, linked to the relevant landmark, and classified as to whether they were the first or subsequent mention. Features are also assigned indicating the linguistic form of the reference (in/definite, pronoun, possessive phrase, etc.) Rather than being marked explicitly, coreference is indicated by two references pointing to the same landmark ID.

UCREL [Fligelstone, 1992]:

The coreference and annotation scheme developed at UCREL (Lancaster) (Fligstone, 1992), like MUC and DRAMA, is based on identifying referential expressions in the text, and then marking the relationships between these. Like DRAMA, this scheme includes many types of 'bridging' relationships, rather than just coreference. These different types of bridging relationships are all classified in the mark-up. Coreferring pronouns are also marked for directionality (anaphor or cataphor). The set of markables is even less restricted for UCREL than it is for DRAMA, including ellipsis and verbal forms.

3.7.2 Comparison of Markables

Structure	**Scheme**						
	MUC-7	**DRAMA**	**Poesio & Vieira 1**	**Poesio & Vieira 2**	**Bruneseaux & Romary**	***U-line**	**UCREL**
Nouns	Yes (unless only used as modifier throughout text)	Yes	No	No	Yes	No	Yes
Noun Phrases	Yes	Yes	Yes (definite only)	Yes (definite only)	yes	Yes (referring to landmarks only)	yes
Possessive Pronouns	Yes	Only for inference relation	No	No	Yes	Yes	Yes
Subj /Obj pronouns	Yes	Yes	No	No	Yes	Yes	Yes
Relative Pronouns	No	Only non-restrictiv e	No	No	?	No	No
Wh-Phrass	No	Yes	No	No	?	Yes	?
Clauses	No	Yes	No	No	Yes	No	Yes
Gerunds	Only if grammati-cal ly marked as nominal	Yes, unless clearly non-nominal	?	?	Yes	N/A	Yes
Empty String (i.e. zero or implicit pronouns)	No	Yes	No	No	?	Yes, if deemed elliptical ref. to landmark	Yes, if deemed recoverable ellipsis
Disfluencies which are 'verbally deleted'	No	Yes	N/A	N/A	?	Yes	?
Phrases with adj. as head and implicit pronoun	No	Yes	?	?	?	Yes	Yes
Verbs / VPs	No	No	No	No	Yes	No	Yes

Gestures	No	No	No	No	Yes	No	No

3.7.3 Comparison of Relationships

Example of Phenome-non	**Classification (if any) under each scheme**						
	MUC-7	**DRAMA**	**Poesio & Vieira 1**	**Poesio & Vieira 2**	**Bruneseaux & Romary**	***U-line**	**UCREL**
a/the tower ... the tower	identity (co-reference)	identity of reference	anaphoric same head (co-reference)	coreference	coreference	identity of reference, 'not as label'	coreference
a tower ... the giant steel construction	Identity (co-reference) – may be 'optional'	Identity of reference	associative	coreference	coreference	Identity of reference, 'not as label'	coreference
a house ... the kitchen	none	Bridging: part/whole	associative	Bridging reference	associative	Reference to part of landmark	Indirect anaphor
... the government ...	none	none	Larger situation / unfamiliar	Larger situation	none	none	none
... the door of the Bastille ...	none	none	Larger situation / unfamiliar	Unfamiliar	none	none	None
... in the soup/to the wall	none	none	idiom	none	none	none	none
<mouse click> 'this one'	none	none	none	none	designation	none	None
the pears ... three pears	none	Conceptual bridging: set/subset	none	none	?	Identity of reference, numerical	miscellane-ous
an explosion ... the noise	none	Conceptual bridging: cause	associative	Bridging reference	associative	none	Inferrable of com-plement
It's sunny ... that's a relief	none	Linguistic bridging: pro-positional	none	none	coreference	none	coreference

the boy ... his hair	coreference	Linguistic bridging: possessive	none	none	possessive	Identity of reference: possessive	coreference
Look at the car! The wheel just fell off ∅	none	Linguistic bridging: implicit argument	none	none	?	ellipsis	ellipsis
Here's one screw. Where's the other one?	none	Linguistic bridging: plural NP	?	?	?	Numericalised pronoun	Substitution form
The boy does one thing and the girl does another ... then they both start crying	coreference	Linguistic bridging: plural NP	none	none	?	Identity of reference with two landmarks	miscellane-ous
The Prime Minister, Tony Blair, ...	coreference	none	none	none	?	none	NP predication

4. Dialogue Acts

Responsibility: Marion Klein, Claudia Soria

4.1 Introduction

Dialogue acts, also called dialogue moves or illocutionary acts, are the basic elements of human communication rather than words or sentences. A dialogue is divided into units called turns which refer to speaker changes. A turn again consists of several utterances which are also called segments.

Dialogue act annotation schemes are used to mark important characteristics of utterances. These annotations indicate the role of an utterance in a specific dialogue and make the relationship between utterances more obvious.

Most of the dialogue schemes nowadays are task-oriented as we will see later on in this report. This is due to reduce the amount of annotation tags to a capable size for annotation and to increase the analyzing rate of the NLP system in which the scheme is used. The information content (or the semantics) of task-oriented dialogues can be basically split into task / domain related information and information that addresses the communication process. To guarantee generality and therefore more flexibility both information levels should be kept separately in the notion-choice of tags. Schemes which cover only those two fields are said to be schemes for rather shallow analysis.

As an example of a scheme that allows deep analysis DAMSL can be mentioned. With its forward and backward looking functions it keeps track of how an utterance constrains the future beliefs and actions of the participants, and affects the discourse and how an utterance relates to the previous discourse, respectively.

4.2 Scheme Comparison

First of all the schemes which have been observed are listed below together with information about their developer and the domain in which they are used. Further details about the schemes are given in the Annexes 133ff.

- **Alparon**

  Developer: Delft University of Technology, Netherlands

  Domain: information retrieving about the services offered by the Dutch public transport

- **Chat**

  Developer: Carnegie Mellon University, Department of Psychology, USA

  Domain: analysis of child language.

- Coding scheme of the University of **Chiba**

  Developer: University of Chiba, Japan

  Domain: route direction, scheduling, telephone shopping and tourist information.

- **COCONUT**

  Developer: The University of Pittsburgh, USA

  Domain: buying furniture for the living and dining rooms of a house
- **Condon & Cech**'s coding scheme

  Developer: University of Southwestern Lousiana, USA

  Domain: decision-making
- **C-STAR**

  Developer: C-STAR Consortium

  Domain: travel-planning
- **DAMSL**

  Developer: Discourse Representation Initiative

  Domain: general
- **Janus**

  Developer: Carnegie-Mellon University, USA

  Domain: appointment-scheduling
- **Flammia**'s coding scheme

  Developer: Massachusetts Institute of Technology, USA

  Domain: information-seeking
- **LinLin**

  Developer: Linköping University, Sweden

  Domain: information-retrieving
- **Maptask**

  Developer: The University of Edinburgh, UK

  Domain: route direction
- **Nakatani et al.**'s coding scheme

  Developer: Harvard University and AT&T Bell Laboratories, USA

  Domain: direction-giving
- **SLSA**

  Developer: Göteborg University, Sweden

  Domain: courtroom interaction
- **SWBD-DAMSL**

  Developer: University of Colorado, USA

  Domain: general/telephone conversation between strangers

- **Traum**'s coding scheme

  Developer: Université de Genève, Switzerland

  Domain: general/task-oriented

- **Verbmobil**

  Developer: DFKI, Germany

  Domain: appointment scheduling

  Not all of the above mentioned schemes are suitable to take under consideration for the MATE project. This might be because they are designed for a special task / domain and hence too much restricted.

  Another reason might be that they are not used very much which could, for example, lead to the assumption that they are too complicated.

  However, in the following some criteria for schemes are detailed which can be used to scale the observed schemes.

- **Task-Orientation:** almost all schemes refer to dialogues that are task-driven, i.e. performed to accomplish a specific task. LE research is obviously oriented towards task-driven dialogues, since it is not particularly useful for this area to analyse ordinary conversations. However, some schemes are nevertheless devised for non-task-driven dialogues. Values for this category are **TD** (task-driven) and **NTD** (non-task-driven)
- **Applications Orientation**: again, this is a relevant parameter for LE research. The majority of task-driven dialogues are also application-oriented, i.e. geared towards industrial or commercial applications. On the other hand other dialogues, and thus their schemes, are not applications-oriented. Values for this category are **positive** (applications-oriented) and **negative** (non-application-oriented).
- **Domain Restriction**: most dialogues in LE research are restricted to a relatively highly specific domain of subject matter. This represents a relevant parameter since schemes are often influenced by the type of domain of the dialogues annotated. Values for this category are **positive** (restricted domain) and **negative** (unrestricted domain). The positive value can be further categorized according to the following typology of domains:

  TR: travel

  TS: transport

  COS: computer operating systems

  COU: courtroom interaction

  BA: business appointments

  DES: directory enquiry services

  FUR: furnishing rooms interactively

  DIR: giving directions

  INST: giving instructions (e.g. about cooking)

- **Activity Type**: this category refers to the type of activity to which dialogues may belong, and is another dimension along which schemes can be classified. It seems that in current dialogue research, there is a major division between two leading paradigms: cooperative tasks between human participants (such as negotiating appointments) and information extraction tasks, in which a human agent interrogates a computer system (or a human surrogate for a computer system). A typology of activity types is the following (together with abbreviation used):

  CN: cooperative negotiation

  IE: information extraction

  PS: problem solving

  TI: teaching/instruction

  CO: counselling

  CH: chatting

- **Human / Machine Participation:**

  HH: human-human dialogues. This category is further divided into:

  MM: machine-mediated(computer, telephone)

  NMM: non-machine-mediated

  HM: human-machine dialogues. This category is further divided into:

  S: simulated

  NS: non-simulated

The results of the guidelines together with the criteria outlined above are applied to the observed schemes in the tables below:

Schemes		**ALPARON**	**CHAT**	**CHIBA**	**COCONUT**
Coding Book		yes	yes	yes	yes
Annotators	Number	3	huge	10	2
	Expertise	experts	experts	experts	experts
Information about annotated dialogues	Size	500 dialogues	160MB	22 dialogues	16 dialogues
	Languages	Dutch	many	Japanese	English
	Participants	2	2	2	2
	Task Orientation	TD	(NTD)	TD	TD
	Application Orientation	yes	no	no	yes
	Domain Restriction	DES	no	DIR, BA,TR	FUR
	Activity Type	IE	CH	CN, PS	CN
	Human / Machine Part.	HH, MM	HH, NMM	HH, NMM(?)	HH, MM (computer)
Evaluation		yes (77% agreement)	no	yes (0.57 < alpha < 0.68)	yes
Mark-up Language		yes, own	yes, own	yes, SGML-like	yes, DAMSL variant
Annotation Tools		yes, OVR coder	yes	yes, modification of dat	yes, Nb
Usability		yes	no	?	yes

Schemes		CONDON & CECH	C-STAR	DAMSL	FLAMMIA
Coding Book		yes	yes	yes	yes
Annotators	Number	5	5	4	7
	Expertise	fairly experts	experts	experts	trained
Information about annotated dialogues	Size	88 dialogues	230 dialogues	18 dialogues	25 dialogues
	Languages	English	Engl., Jap., Kor., It.	English	English
	Participants	2	2	2	2
	Task Orientation	TD	TD	NTD	TD
	Application Orientation	yes	yes	no	yes
	Domain Restriction	TS	TR	no	DES
	Activity Type	CN	CN	several	IE
	Human / Machine Part.	HH, MM, NMM	HH	HH	HH, MM
Evaluation		yes (91% agreement)	no	yes, κ=0.56	yes, κ=0,6+
Mark-up Language		yes, Nb's	yes	yes, DAMSL	yes
Annotation Tools		yes, Nb	no	yes, dat	yes
Usability		yes	yes	yes	?

Schemes		JANUS	LINLIN	MAPTASK	NAKATANI
Coding Book		yes	yes	yes	yes
Annotators	Number	4	4	4	6
	Expertise	experts	experts	experts	naive
Information about annotated dialogues	Size	many	140 dialogues	128 dialogues	72 dialogues
	Languages	English	Swedish	English	English
	Participants	2	2	2	1
	Task Orientation	TD	TD	TD	TD
	Application Orientation	yes	yes	yes	no
	Domain Restriction	BA	TR/TS	DIR	INSTR
	Activity Type	CN	IE	PS	TI
	Humpan / Machine Part.	HH	HM, NS	HH, NMM	HH, NMM
Evaluation		yes (89% agreement)	yes (97% agreement)	Yes, κ=0.83	no
Mark-up Language		yes, own	yes, Nb's	Yes, own SGML based	yes, Nb's
Annotation Tools		no	yes, Nb	Yes, own	yes, Nb
Usability		yes	yes	yes	yes

Schemes		SLSA	SWBD-DAMSL	TRAUM	VERBMOBIL
Coding Book		yes	yes	yes	yes
Annotators	Number	7	9	3	3
	Expertise	experts	experts	experts	naive
Information about annotated dialogues	Size	100 dialogues	1155 dialogues	36 dialogues	1172 dialogues
	Languages	Swedish	English	English	Eng., Jap., Ger.
	Participants	2 (?)	2	2	2
	Task Orientation	TD	NTD	NTD	TD
	Application Orientation	yes	no	yes	yes
	Domain Restriction	COU	no	no	BA
	Activity Type	several	several	CN	CN
	Human / Machine Part.	HH, NMM	HH, MM	HH, NM	HH, NMM
Evaluation		yes (not published)	yes, $0.8 < \kappa < 0.84$	yes (not published)	yes, $\kappa=0.84$
Mark-up Language		yes, own	yes, variant of DAMSL	yes, Nb's	yes, own
Annotation Tools		yes, TRACTOR	no	yes, Nb	yes, AnnoTag
Usability		yes	yes	yes	yes

In order to develop a standard it is necessary to compare schemes with regard to their underlying task and dialogue acts. The following tables present domain-grouped schemes and show the equivalence between their dialogue acts. Entries with italic font represent higher order expressions which can't be annotated.

Domain: information retrieval

Alparon	Flammia's	LinLin
Moves (Dialogue Acts)	Speech Acts	Initiative
Statement	-	Update
Question Check Alignment	Question-Confirm	Question
-	Response	Response
Clarification	-	Answer
-	Confirm Accept Reject	-

Acknowledgement Reconfirmation	Acknowledge Repeat	-
Greeting Bye	-	*Discourse Management* Opening Ending Continuation
Pause	-	-
Other	-	-

Domain: route direction

Chiba	**Maptask**
Initiation	*Initiating moves*
Inform Other assertion	Explain
Yes-no question Wh-question	Query-yn Query-w Check Align
Request Suggest Persuasion Propose Demand	Instruct
Promise	-
Response	*Response moves*
Positive Negative Answer Other response	Reply-y Reply-n Reply-w Clarify
Hold Confirm	-
-	Acknowledge
Follow-up Understanding	-
Conventional Opening Closing	-
Other initiation	-

Domain: appointment scheduling

Chiba	Verbmobil
Initiation	*Dialogue_Act*
Inform Other assertion	Inform Init Give-reason Digress Deviate_Szenario Refer_to_setting
Yes-no-question Wh-question	-
Request Suggest Persuasion Propose Demand	Suggest Request Request_Suggest Request_Clarify Request_Comment
Promise	-
Response	-
Positive Negative Answer Other response	Feedback Feedback_Positive Feedback_Negative Feedback_Backchanneling
Hold Confirm	Accept Confirm Reject Explained_Reject Clarify Clarify_Answer
Follow-up Understanding	-
Conventional Opening Closing	Convention Thank Deliberate Introduce Politeness_Formula Greeting Greeting_Begin Greeting_End
Other initiation	Not_Classifiable

Domain: general

DAMSL	SWBD-DAMSL	Traum's	Chat
Forward looking function	*Forward Communicative –Function*	*Illocutionary Function*	*Categories of Illocutionary Force*
Statement Assert Reassert Other	*Statement* Statement-no-opinion Statement-opinion	Inform Supp-Inf Supp-Sug	*Statement:* AC, CN, DW, ST, WS *Declarations:* DC, DP
Info-Request	*Influencing-Addressee-Future-Action (1)* Yes-No-Question Wh-Question Or-Clause Declarative-Yes-No-Question Declarative-Wh-Question Tag-Question Backchannel-in-Question Rhetorical-Question	YNO WHQ	*Questions:* AQ, AA, AN, EQ, NA, QA, QN, RA, SA, TA, TQ, YQ, RQ
Influencing-Addressee-Future-Action Action-directive Open-Option	*Influencing-Addressee-Future-Action (2)* Open-Question Action-Directive	Request Suggest	*Directives (1):* RP, RQ
Committing-Speaker-Future-Action Offer Commit Explicit-performative Exclamation	*Committing-Speaker-Future-Action* Offers Options Commits	Offer	*Commitments:* FP, PF, SI, TD *Directives (2):* CL, SS
-	-	Promise	PD
Backward looking function	*Backwards-Communicative-Function*	-	-
Answer	*Answer* Yes Answer No Answer Affirmative non-yes-answer Negative non-no answer Other answer Dispreferred	Eval	*Evaluations:* AB, CR, DS, ED, ET, PM *Directives (3):* AC

	answers		
Agreement Accept Accept-part Maybe Reject Reject-part Hold	*Agreement* Agree/Accept Maybe / Accept-part Reject Hold before answer/agreement	Accept Reject Check	*Directives (4):* AD, AL, CS, RD, GI, GR, DR *Declarations (2):* ND, YD
Understanding	*Understanding*	*Grounding*	-
-	-	RequestAck	-
Signal-understanding Acknowledge Repeat-rephrase Completion	Response-Acknowledgement Repeat-phrase Collaborative-Completion Acknowledge Summarize/Re-formulate Appreciation Downplayer	Acknowledge	Speech Elicitations: CX, EA, EI, EC, EX, RT, SC
Signal-Non-Understanding	Signal-non-understanding	Request-Repair	*Demands for clarification:* RR
Correct-Misspeaking		Repair	*Text editing:* CT
-	*Other-forward-function* Conventional-opening Conventional-closing Thanking Apology	Greet Apologise	-
-	-	-	-
-	*Other* Quotation Hedge	-	*Vocalisation:* YY, OO
-	-	-	*Markings:* CM, EM, EN, ES, MK, TO, XA
-	-	-	*Performances:* PR, TX

- There is no dialogue act information available on the Janus scheme.
- No specific dialogue acts are mentioned for SLSA.

- Nakatani et al.'s coding scheme is developed for discourse annotating only, therefore no dialogue acts are specified, too.
- All the other schemes, like COCONUT, Condon and Cech's and C-STAR are not stated in one of the tables above as they are designed for different tasks (e.g. problem solving, decision making or travel-planning, respectively). However, they are taken under consideration in the conclusion below.

4.3 Conclusion

The huge amount of coding schemes detailed in the Annexes 133ff shows the current research interests in dialogue act annotation schemes. There also seems to be a trend to shallow, task-oriented annotation as these schemes predominate those which focus on a general approach. The comparison of dialogue acts of schemes with equivalent domains reflect the similarities expected. But more surprisingly even a dialogue act comparison among all schemes regardless to their orientation shows quite a lot of parallelism although the general schemes are, of course, more comprehensive.

In order to decide, which schemes should be taken under consideration in MATE, schemes should have a coding book, they should be heavily used and should have good evaluation results. Also a scheme that is not related to a special task seems to be more appropriate as a task related and therefor possibly restricted one. If we look at the general comparison of schemes above, one can observe that all listed schemes provide a coding book. Amongst the schemes which are mostly used we can see Alparon, Chat, SWBD-DAMSL and Verbmobil. Unfortunately Chat hasn't been evaluated, but Alparon, SWBD-DAMSL and Verbmobil are judged to be good. As SWBD-DAMSL is the only one of these schemes which is not task related, it should be definitely supported in MATE. With regard to the MATE standard, of course, the dialogue acts of all schemes should be taken into account and analysed.

5. Morpho-Syntax

Responsibility: Vito Pirrelli, Claudia Soria

5.1 Preliminary Issues: Differences from Writing

It could be argued that morphosyntactic and syntactic annotation of dialogue does not represent a specific problem area as such, there being a reasonable expectation that schemes, techniques and software tools developed for robust annotation of unconstrained written texts should easily be amenable to annotation of dialogue as well.

In fact, such a principled expectation is not borne out entirely in practice. First, there is an issue of frequency and nature of the linguistic phenomena which are likely to appear in dialogue texts. Although it is true that even hesitators like 'ums' and 'ers' occur in fictional dialogue, it can hardly be ignored that interjections and hesitators are vastly more frequent in speech than in writing. As will be argued more extensively in the remainder of this report, frequency and variety make the difference for annotation: if an accidental feature of language becomes systematic, then an annotation scheme has to make provision for it.

Moreover, there is a wide range of phenomena (ranging from anacolutha to disfluencies) which are in fact specific to spoken language only. Again, they will be considered in this context insofar as they raise problems of annotation at the levels of morphosyntax and syntax proper.

Following Leech et al. 1998 (LE-EAGLES-WP4-3.1), it is useful to part the list of linguistic phenomena of specific interest in dealing with dialogue texts into two large and still closely interconnected classes:

1) disfluency phenomena,
2) linguistic phenomena which are characteristic of speech and not of writing.

In turn 1) includes:

a) hesitation fillers ('um', 'er' etc.),
b) word partials (e.g. when a speaker is interrupted in mid-word),
c) syntactic incompleteness or phrase partials (e.g. when the speaker fails to complete an utterance, owing to self-correction, interruption, trailing off[1] or other reasons),
d) retrace-and-repair sequences (when the speaker interrupts the production process, returns to an earlier point of the same utterance and restarts from there),

1 'trailing off' is said to occur when speakers shift attention away from what they are saying, sometimes even forgetting what they were going to say. Usually the trailing off is followed by a pause in the conversation. After this lull, the speaker may continue with another utterance or a new speaker may produce the next utterance.

e) dysfunctional repetitions,

f) syntactic blends due to the speaker failing to complete a sentence and changing tack (trailing off).

In turn 2) includes:

a) adverbs, interjections, dialogue markers,

b) some sort of semi-grammaticalised syntactic anacolutha hardly classifiable under the heading of disfluency in a strict sense (e.g., 'there is an accident by the Flying Fox, is it?', or 'Io speriamo che me la cavo/ I let's hope I can get away with it'),

c) the principled issue of both word-level and syntactic segmentation, namely: What is to be marked as a distinct morphosyntactic unit in spoken texts? What is an utterance in dialogue and how can it formally be identified?

It is clear that such a wealth of phenomena would inevitably lead to a considerable revision/integration of the schemata developed for annotation of written texts at the levels of morphosyntax and syntax.

5.1.1 Consequences in Software Development

Another related issue here bears on the availability of reliable pieces of software for fully automatic annotation of texts at the levels of morphology and syntax. To what extent should software tools be subjected to revision owing to dialogue-specific phenomena? The answer is not trivial. Basically, the problem has been dealt with according to two substantially different strategies:

- **normalization** of spoken text,
- **stretching annotation** so as to include the parsing of disfluencies and related phenomena.

In the case of normalization, the idea is to treat disfluency and related phenomena as fundamentally extraneous to the grammatical annotation of speech. The Penn Treebank and the International Corpus of English, for example, have adopted schemes for explicitly annotating disfluencies. These are eventually excluded from the syntactically annotated material, by applying annotation only to a normalized version of the data.

The alternative approach (CHRISTINE and UCREL) is to include the disfluent material in the syntactically annotated material.

As a matter of principle as well as of practice, both approaches require preliminary manual annotation of the critical (disfluent) material, but it is clear that while normalization calls for no major adaptation/augmentation of existing annotation software, inclusion of disfluencies during parsing does indeed call for a considerable stretching of common-or-garden notions of phrase structure, or any other suitable syntactic notion for that matter. It remains to be seen whether it

is possible to strike a compromise between the requirements of (automatic) parsing on the one hand, and the concern of departing from the attested linguistic evidence available as little as possible. As we will see in some detail later in this report, definition of more than one level of syntactic annotation, ranging from shallow parsing to annotation of grammatical functions such as subject and object, goes some way in providing room for such a compromise.

5.1.2 Segmentation / Parsing Issues: What is a word? What is a sentence?

In the compilation of a corpus of dialogues, orthographic transcription of speech recordings is usually the first level of linguistic interpretation of the actual acoustic signal. This involves a considerable amount of problems concerning, among other things, consistent use of punctuation marks, identification of word units, identification of sentences as basic syntactic units. Some of these issues are closely related to one another: e.g. marking of full stops highly correlates with identification of basic syntactic units; transcription of partially intelligible words or even word partials is often a matter of interpretation.

Although the notion of orthographic word and its associated notions of morphosyntactic word (or syntactic atom) and phonological word (or word stress domain) is in most cases well-established enough not to raise doubts about the orthographic rendering of a sequence of spoken words in an utterance, still there are cases where problems of segmentation may arise. Is a compound such as 'railway station' an orthographic unit as well as a single morphosyntactic and arguably phonological unit? Are expressions such as 'I mean, mind you, good morning, sort of, kind of' to be interpreted and accordingly tagged as one morphosyntactic unit? Although these problems are not confined to dialogue annotation, they get particularly thorny in this context due to i) their interaction with both orthographic transcription and interpretation of word partials and partially intelligible words, and ii) greater density of their occurrence in spoken language. This is particularly clear when things are looked at from the perspective of automatic annotation. For example, if the phenomenon of multi-words is ignored (as indeed it is ignored by some well-known taggers), their high frequency in dialogue is likely to introduce a considerable amount of syntactically non canonical (i.e. non compositional) sequences of tags (as in 'I kind of like it'), with the practical consequence of either introducing noise in the training of a probabilistic tagger, or repeatedly tripping up a probabilistic tagger trained on written texts. In fact, in many cases, the tagging of single constituents of a multi-word unit makes comparatively little sense from a syntactic point of view.

It is clear that, from the perspective of developing an annotation scheme, these issues should be addressed explicitly, with explicit guidelines for manual annotation. Eventually, the extent to which software tools for automatic tagging will be augmented/modified will greatly depend on decisions taken at the level of the annotation scheme.

The canonical sentence of written language, as a maximal parsable unit containing at least one finite verb, represents only a kind of ideal syntactic structure when it comes to segmenting real dialogues. In fact, in many cases, utterances typically consist of one word only, often not a verb. If we also consider anacolutha, syntactic incompleteness and other related disfluency phenomena, it becomes immediately apparent that any attempt to construe maximal syntactic projections on the basis of the concatenation of intermediate phrase structures is doomed to failure in many cases. In practice, some tree banks use maximal parse brackets to enclose the whole parsable unit, and make no assumption about its internal structure. This is what the guidelines of the British National Corpus call '**structure minimization principle**'. This is certainly preferable to the alternative strategy of editing out major disfluencies, so that parsing

is eventually only applied to relatively well-behaved utterances. In the first place, disfluencies, as we saw, are not the only source of difficulties in parsing dialogue sentences. Secondly, it is indeed useful to sketch syntactic (sub)structures even for incompleteness and repair phenomena, as this can represent a valuable source of information for both annotation and software development (as in this case of cross serial correspondences between the **reparandum** and the alternation in speech repairs, see Heeman and Allen 1997) and psycholinguistic studies. Thirdly, it is difficult to see how the normalization strategy can be applied to some markedly disfluent utterances without imposing an artifact interpretation on the dialogue, unless this sort of utterances are completely expunged from the corpus.

Another viable parsing strategy can be offered as an alternative to both structure minimization and normalization: so-called '**partial parsing**'.

In this context, partial or incomplete parsing will be understood as a form of underspecified parsing whereby a syntactic sketch is carried out by segmenting a sequence of word forms into nonrecursive constituents called chunks, as illustrated by the intermediate level of embedding in the following diagram:

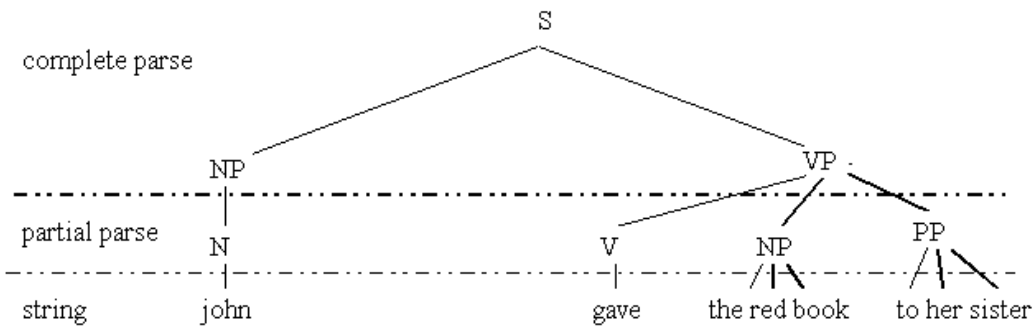


Figure 1

In fact, it is important to bear in mind that the diagram in figure 1 is only illustrative, and should not be interpreted as suggesting that text chunking is simply an intermediate stage in the process of generating a full constituency-based parse of a sentence. Later in this report, we will consider in more detail a particular chunking scheme which appears to be amenable to a dependency-based syntactic representation. Be that as it may, what all chunking schemes seem to have in common is the local character of their proposed analyses, with particular emphasis on the fact that the syntactic relations holding among chunks are not necessarily spelled out.

This aspect is particularly relevant to dialogue annotation, for example in connection with the analysis of phrase partials or syntactic blendings. Since partial parsing does not enforce any overall consistency checking of levels of structural embedding higher than chunks (that is, as to the way chunks are eventually related to the topmost nodes in a sentential tree), a shallow parser does not balk at the occurrence of phrase partials, nor does it attempt to keep an anacoluthon in line with the overall syntactic construction. In this respect, shallow parsing yields an output which is the mirror image of the representation required by the 'structure minimization principle': instead of providing the most comprehensive structure compatible with the data (with no indication of its internal substructures), shallow parsing outputs a list of unrelated minimum syntactic structures which are compatible with input data. This strategy can provide useful information also about portions of the original dialogue which are eventually expunged from a

normalized transcription of the dialogue in question (for example, in the case of retraced-and-repaired sentences).

As we will see in more detail later in this report, chunking is usually taken to be only the first stage in the process of augmenting input data with syntactic annotation. In the specific case of dialogue annotation, the transition from shallow to complete parse presupposes prior identification of disfluencies (either manual or automatic), for the latter to be conveniently pruned out from the final representation of the syntactic structure of an utterance, be that expressed in terms of a sentential phrase marker (a tree), or a dependency chain, or otherwise. It is our contention that this process of step-wise abstraction from concrete input data is inevitable and inherent in the idea of augmenting input data with richer and richer levels of annotation. The step-wise approach to syntactic analysis suggested here has the non-negligible advantage of providing graded levels of abstraction, from fairly local analyses to overall ones, so that even extra-grammatical phenomena such as disfluencies are annotated syntactically at some (low) level, to eventually be ignored at higher levels.

For a detailed description of schemes see Annexes 178ff.

5.2 Summary: preliminary recommendations

5.2.1 Morphology

5.2.1.1 Inflection

At the level of inflectional information, preference should be given to Eagles standards, as opposed to other alternative practices, such as Childes' "morphemization" or morpheme splitting in the main line (see relevant section for examples). This can be motivated on grounds of the principled difficulty, particularly in languages other than English, of providing a coherent morpheme-based segmentation of suppletive or, generally speaking, fusional word forms such as English *went* or *were*. It should be borne in mind that EAGLES recommendations have been implemented for a variety of different languages (e.g. in collateral projects such as MULTEXT), so that provision is made for different levels of granularity of encoding and underspecification.

5.2.1.2 Derivation and Compounding

Annotation of derivatives is not as widespread as annotation of inflectional information, owing to the practical difficulty of providing an exhaustive description of derivational phenomena as opposed to inflectional ones. Still, encoding of this level of information is a desirable objective, which would be of considerable practical utility for example for purposes of information and document retrieval, where information about the root of a derivative (commonly, but inaccurately referred to as "stemming") is heavily resorted to.

Unlike inflection, derivation lends itself more naturally to being dealt with in terms of morpheme splitting. "Morpheme segmentation", either immediate (e.g. signalling the most external affix only, as in "derivation-al"), or complete (as in "deriv-ation-al") or hierarchical (as in "(((deriv) ation) al)") is provided, for example, in the CELEX electronic lexica (Burnage 90). Yet, this type of representation is, in general, not able to account for, e.g., stem allomorphy, although this is admittedly far less frequent than in inflectional morphology. For lack of better

encoding practices, immediate morpheme segmentation (flat) could be proposed as a reasonable minimal annotation strategy for encoding derivational morphemes.

Although compounding represents another critical area for both theoretical and computational Morphology, annotation of compounds (as opposed to their identification or their interpretation) can be a relatively trivial issue if it is limited to signalling the membership of a sequence of word forms (such as *copy* and *editor* in "copy editor") to a morphosyntactically unique word. This problem is common to annotation of other types of multi word units.

5.2.2 Morphosyntax

In illustrating the SPARKLE level-wise standard for syntactic annotation (see section D of the Annex) we comment on some of the advantages of encoding immediate constituency information (chunking) and grammatical function information at separate and relatively independent levels of annotation. This practice was argued for on grounds of robustness of the required annotation software, as chunk-parsing is local in character, and naturally geared towards treatment of phrase partials, while functional annotation is by definition more brittle and is expected to carry out a consistency checking of the syntactic coherence of the complete sentence.

We contend that these two levels of syntactic annnotation, augmented with their linking to a common level of (edited) orthographic transcription of the acoustic signal, can be instrumental in getting around the stricture of the two radically different approaches to syntactic annotation of dialogue material proposed so far: namely i) normalization of orthographic transcription on the one hand, and ii) stretching of the annotation scheme to deal with unrestricted text on the other hand.

The problem with the alternative between i) and ii) lies in the fact that normalization gets rid of precious information, by disregarding material such as repaired speech which it would be useful to annotate at the syntactic level anyway (as illustrated by work of Core and Schubert, 1997). On the other hand, unedited spoken material is, in some extreme cases (e.g. child language), so difficult to deal with through any set of syntactic rules, than any notion of stretching the syntactic annotation here would inevitably lead to an uninformative output.

The compromise that we intend to suggest here is based on the idea of gradual abstraction from the raw, unedited orthographic transcription of the acoustic signal. Different levels of syntactic annotation can be developed which convey syntactic information at progressively higher levels of abstraction. Accordingly, different phenomena specific of speech are dealt differently depending on the level of syntactic annotation one is considering. For example, repaired speech should, in our view of things, annotated at the chunk level (see relevant examples provided in the overview). Among other things, this is also important for their identification, owing to the parallel structure usually exhibited by a *reparandum* and its *alteration*. On the other hand, it is generally meaningless to include the *reparandum* in a functional annotation, where only target or intended units (as opposed to actually uttered ones) are eventually taken into account. A possible exception to this general principle is represented by the case when a pronoun in the *alteration* refers back to a noun phrase in the *reparandum*, as in:

Take the oranges to Elmira, uh, I mean, take them *to Corning*

Here, it can be argued that functional annotation of the *alteration* requires intended information provided in the *reparandum* ("the oranges"). In fact, it is dubious that functional annotation should include disambiguation of the referential content of pronouns. Be that as it may, this case calls for exceptional reference to edited (repaired) material, and gives further support to an annotation practice whereby edited material is simply conveniently marked, but not expunged, so that it can be recovered if the need arises.

It remains to be seen how the two (or possibly more) levels of syntactic annotation should mutually be related. In short, two solutions can be envisaged: a) direct linking of the required levels, b) indirect linking through reference to a common level of edited orthographic transcription. Both solutions have pros and cons. On the one hand, it seems useful that functional annotation be built on a chunked text. On the other hand, the level of chunking is still too raw to provide an appropriate anchor for annotation at the functional level (for example, given a pronoun uttered thrice, which one of these tokens should be actually linked to the level of functional annotation? See discussion in the relevant section of the overview). Editing is felt useful in this context in order to i) allow a functional parser to disregard irrelevant phenomena (such as repetitions and repairs), ii) provide the target anchor to be referred to at the level of functional annotation.

Finally, it should be noted that the standard developed within SPARKLE offers the additional advantage of leaving room for underspecification depending on the specific requirements of the language being annotated. This is made possible thanks to the specification of a hierarchical typology of grammatical functions (see relevant section in the overview part), which has been designed so as to meet the grammatical requirements of English, French, German and Italian. We are aware of no other comparable effort along the same lines.

As to the treatment of phrase partials, CHILDES seems to provide a useful set of markers signalling the point where the expected phrase appears to be interrupted, and, possibly, the point where it is resumed, either by the same speaker or through completion by another interlocutor. The analogous scheme provided in Switchboard is, for what can be judged from the annotation manual, needlessly overspecific and of difficult application. Moreover, the strict assumption, made in Switchboard, that a phrase partial can only be completed by the same speaker who uttered the partial in the first place, strikes us as too abstract and not sufficiently motivated.

6. Prosody

Responsibility: Juanma Garrido, Silvia Quazza

6.1 Introduction

In the following a review of some existing coding schemes for prosody is presented. Section 6.2 gives a brief overview of prosodic phenomena, while section 6.3 discusses purposes and problems of prosodic transcription. Finally, section 6.4 summarizes and compares a number of current schemes for prosodic annotation, described in more detail in Annex E.

6.2 Prosodic Phenomena

The term 'prosody' covers a wide variety of facts, concepts and phenomena, defined by researchers working with different theories and frameworks. One of the first problems that arise when attempting the study of prosody (and of course, its representation) is the definition of the concept itself and its scope.

The description of prosody in any language can be approached from two opposite (and complementary) starting points:

1) From a linguistic point of view, the description of prosody can be viewed as the description of a series of suprasegmental units (syllables, stress groups, intonational units) and phenomena (stress, intonation, rhythm).

2) From a phonetic point of view, the description of prosody is mainly approached as the description of the different phonetic correlates (length, loudness, F0 variations) of these linguistically relevant prosodic events.

Considering this distinction, the prosodic phenomena can be classified in two main groups: a first group of 'linguistic' prosodic events, and a second group of phonetic prosodic events. They are closely related to each other but can be described separately.

These two subsets are reviewed in the following subsections.

6.2.1 Linguistic Prosodic Events

In the linguistic descriptions of prosody (mainly from a phonological point of view), usually two types of prosodic items are handled: a set of prosodic units (phonological units with a scope wider than a segment), and a set of prosodic phenomena which are 'superimposed' on these units.

6.2.1.1 Prosodic Units

Several types of prosodic units (differing mainly in their scope) have been proposed in the prosodic studies:

1) Paragraphs

2) Sentences

3) Intonation groups

4) Intermediate groups

5) Stress groups

6) Feet

7) Syllables

8) Mora

It is not the aim of this review to present a detailed description of each unit. Although some of these units have been proposed after experimental research (as in the case of the paragraphs), that is, using phonetic evidences, most of them are used in phonological analysis. Apart from sharing the feature that their scope is in all cases wider than a single segment, all of them have in common the fact that they have been proposed as the natural domain of specific suprasegmental or segmental processes (see, for example, [Nespor&Vogel 86]).

6.2.1.2 Prosodic Phenomena

We consider here as 'prosodic phenomena' the suprasegmental features of intonation, stress, rhythm and speech rate. They are not units, but take place usually at a specific domain. They are also holders of some kind of linguistic (or paralinguistic) meaning.

1) Intonation

As stated in [Roach 83, p. 112], "no definition [of prosody] is completely satisfactory, but any attempt at a definition must recognize that the pitch of the voice plays the most important part". No precise and universal definition of intonation has been given yet, but there is a general agreement about some facts: first, that intonation is clearly related to F0, although it determines changes in other phonetic parameters (for example, the length of prepausal syllables); there is also a general agreement in relating intonation to phenomena which occur at sentence level, leaving the word 'tone' for those F0 phenomena which are relevant at word level ([Lehiste 70]).

From a phonological point of view, intonation phenomena are usually described in the following terms ([Pierrehumbert 80], for example):

a) Pitch accents

b) Boundary tones

c) Phrase accents

d) Downstep

e) Upstep

In other cases, however, the phonological components of intonation are described using different concepts. This is the case, for example, of the British school, that uses the terms 'head', 'body' and 'tail' ([Palmer 22], [Crystal,69]).

2) Stress

In the case of stress, there is a wider agreement about its nature and phonetic correlates: it is usually associated with the presence of a special degree of prominence on specific syllables of the discourse. Several types of stress have been defined in the literature, some of them language-specific:

a) lexical (primary)

b) secondary

c) stød (accents I and II in Swedish and other Scandinavian languages).

d) emphatic (focus, contrast)

3) Rhythm

Rhythm can be defined as the perceptive effect produced by the periodical repetition of some phonetic phenomenon along the discourse. The nature of the rhythm may be different depending on the language: it can be based on the isochrony of syllables (syllable-timing), or in the placement of stressed syllables at regular intervals (stress-timing). It is then related to other prosodic phenomena (stress) and units (syllables), and produces variations in several phonetic parameters (duration of sounds or syllables, F0, intensity).

4) Tempo, speech rate

Tempo and speech rate depend on the speed at which the speaker produces their utterances. Speech rate is often measured as the number of sounds uttered per second. It produces then mainly changes in the length of the sounds, although differences in the shape of pitch movements due changes of speech rate have also been reported.

6.2.2 Phonetic Correlates of Prosody

Prosodic units and phenomena are physically realized in the speech chain by modifying a set of phonetic parameters. These phonetic cues (F0, length changes, pauses, loudness) are called here 'phonetic correlates of prosody'.

6.2.2.1 F0 Events

F0 changes are typically related to intonation phenomena, but stress and rhythm – as well as many other non-linguistic-factors – play also a role in the definition of the final F0 contour of a utterance.

These F0 changes (or 'events') seem to occur at different levels of description. At the first level (called here 'local'), some of them seem to affect syllables or groups of syllables. However, other F0 events seem to affect wider units, such as intonation phrases or even sentences or paragraphs. These type of events are called here 'global'.

1) Local F0 events

From a phonetic point of view, local F0 events can be described either as series of **F0 levels** or **F0 contours (movements)**. They are two different approaches to the description of the same phenomenon, the evolution of the F0 frequency along utterances.

2) Global F0 events

Several F0 variations seem to be related to more global phenomena, having a scope larger than the syllable or the stress group. They are concepts used mainly in phonetic descriptions of intonation:

a) global falling (declination) / rising

b) F0 reset

c) pitch range

However, these concepts still need to be integrated in phonological theories of intonation, which have been focused mainly on the description of 'local' phenomena.

6.2.2.2 Length

The length of a sound is the result of the interaction of several linguistic (stress, intonation, rhythm, speech rate) and non-linguistic factors (position in the utterance, phonetic context). Each sound of a given language seems to have also some kind of 'intrinsic duration', which is varied in the discourse by this set of factors. The length of a sound is then only partially related to prosody, because it depends also on segmental factors (the nature of each sound, the context where it appears).

6.2.2.3 Intensity - Loudness

As in the case of length, the intensity of a sound depends on several factors, being possibly stress and intonation those which affect mostly the final intensity of a sound. Each sound of a language seems to have also its 'intrinsic intensity', which can be estimated by 'removing' from the amplitude of a sound the influence of these affecting factors.

6.2.2.4 Pauses

The insertion of pauses in the discourse is one of the ways of marking prosodic phrasing; it is then closely related to intonational phenomena. Speech rate may also determine the location of pauses. And there are other non-linguistic factors that can determine the insertion of a pause: physiological, as the need of breathing; or psycholinguistic, as hesitations.

6.2.2.5 Voice Quality

Voice quality is a phonetic cue that is usually related to the idiosyncratic characteristics of the speaker's vocal tract. However, some changes in the voice quality may have a linguistic function, or may be determined by linguistic phenomena. This is the case, for example, of the changes in the spectrum of a sound affected by stress.

6.3 Prosodic Transcription

It is clear from the above review of prosodic concepts that prosody is a complex phenomenon that can be approached at different levels of description and can be studied for different purposes. From a linguistic point of view, it can be an object of analysis in itself, to be modelled

seeking its patterns and functions, or it can be analyzed as a correlate of discourse structure. In speech technology research, prosody has been studied mainly to achieve natural sounding synthetic speech, trying to associate the proper prosodic events with the input text and realize them with the proper manipulations of acoustic parameters. Also speech recognition is finding some interest in acoustic correlates of prosody as cues to text structure.

Each experimental study has adopted some kind of prosodic representation suited to its purposes, from abstract labels to acoustic measures. But due to the different perspectives of prosodic research a unique *coding scheme* for prosody is hard to conceive. Recently, the need for a *standard* coding scheme has been felt, in order to allow for easy data exchange in the era of large speech corpora.

But although several formal systems of representation of prosody have been used in the description of the prosodic events of the different languages, at this moment, it does not seem to exist a unique and complete system to represent all the prosodic phenomena listed in the previous section.

Some attempts to propose a standard coding scheme have been made, perhaps the most successful in terms of diffusion being ToBI (see Annex E.4). But the discussion on advantages and drawbacks of different schemes should take into account not only the complexity of the object - the different aspects of prosody - but also the various possible *objectives* of prosodic research.

If the purpose is an analysis of discourse, some diacritics marking prosodic boundaries or accents could be enough. For a study of the relations between prosody and discourse structure in a language for which accurate prosodic modelling is already available, symbolic labels concisely representing the prosodic patterns of that language are the proper choice. On the other hand, if one wants to gather experimental data to investigate prosodic patterns and build up a prosodic model, a more detailed phonetic transcription is necessary. For linguistic studies such transcription could be based on auditory analysis, but for speech technology implementations it should be assigned precise acoustic meaning.

Given that, it is rather a difficult task to review a number of different *coding schemes*, and compare them on the basis of quantitative categories such as the number of transcribers and transcriptions or the results of some evaluation test. In Annex E, we are not attempting a complete review, rather we give examples of transcription systems very different in nature and objectives. Some of them are *general approaches* to the study of prosody, which have been followed more or less thoroughly by many researchers in their experiments and studies. Others have been defined in the scope of some specific Project as a convention for *prosodic labelling of corpora*. In such cases, the purpose of the Project and the intended use of the corpus determine the kind of prosodic representation: corpora acquired for dialogue research, for example, often are not focused on prosody and need only abstract labels to mark some macroscopic prosodic features related with discourse events. Finally, some of the reviewed coding schemes have been defined with the explicit aim of providing a *standard*.

A final remark about the phenomena annotated in the different coding schemes: while it is admitted that prosody is a complex matter where intonation, rhythm and loudness are intertwined, the discussion on prosodic notation generally focuses on *intonation*, at least when coming to phonetic descriptions. Some phonological representations make explicit reference to *speech rate*, *lengthening* or more sophisticated rhythmical categories, and most coding schemes mark *phrase boundaries* and *accents*, which globally refer also to duration/intensity phenomena.

But in phonetic-level prosodic transcriptions the main point - perhaps because it is the most problematic aspect - is intonation. Generally, for annotated speech corpora a phonetic *segmentation* is available, so that duration is implicitly marked and intensity can be computed from the signal. The peculiarities of a coding scheme often concern its representation of fundamental frequency, so that a relevant feature of a notation system is its underlying *theory of intonation* or its reference methodology for intonation analysis.

Fully acoustic approaches such as the classical one by Fujisaki [Fujisaki 71], where the intonation profile is seen as a superposition of mathematically defined curves, can't be said to have developed into notational systems, although they provide descriptions of data. On the opposite side, linguistic approaches such as the traditional British School ([Crystal,69], [O'Connor 73]), based on *auditory analysis* and strong theoretical assumptions, have been largely used in phonological research and have also been recently adopted in corpora labelling (see Annex E.10). In this view, (English) intonation is subdivided into tone units, where the main intonation event, the *nuclear tone*, occurring on the last accented syllable, is described in its *height* and *shape*, for example as a *high fall* or a *low fall-rise*. Another family of phonological approaches, whose reference is [Pierrehumbert 80] (and whose first object is again English), describes intonation in terms of *levels* rather shapes: what seems relevant is the tone level reached by the different points in the pitch contour, which is described in terms of the contrast between *high* and *low* (H, L) and with the association with accents (*) and boundaries (%). The use of this notation (more than the underlying principles) is widespread, at least in scientific communication, and this theory has inspired the proposed standard ToBI (see Annex E.4). Experimental phonetic research and speech technology generally are more inclined to follow data-oriented *bottom-up* methodologies. For these approaches, an intonation model for a given language should keep a precise - implementable - phonetic/acoustic content. The starting point is the f0 curve, which is first *stylized* and then phonetically described by means of generalizations from the acoustic/perceptual data. The curve may be seen as a sequence of *pitch movements* or contours - as in the IPO view (see Annex E.9) - or as a series of interpolated target points or *pitch levels* connected by a continuous curve - as in the INTSINT approach (see Annex E.7).

Examples of coding schemes more or less explicitly inspired by such different intonation theories are included in the review presented in the Annexes 214ff The review, by no means exhaustive, gives brief descriptions of the following schemes:

1. PROSPA
2. IPA
3. TEI
4. ToBI
5. SAMPA
6. SAMPROSA
7. INTSINT
8. SAMSINT
9. IPO
10. TSM

11. TILT

12. VERBMOBIL

13. KIM

14. PROZODIAG (Lund)

15. Göteborg

For detailed surveys of prosodic transcription and encoding systems the reader is referred to [Llisterri 94, 96b], [Léon & Martin 70] -which contains a chapter devoted to classical approaches to prosodic transcription -and to [Gibbon 90], reviewing most of the work in this area carried out within the SAM (Speech Assessment Methodologies) project. A discussion of this topic is also found in the text representation chapter of the EAGLES Handbook on Spoken Language Systems [Gibbon *et al.* 1997].

6.4 Conclusion

As discussed above, prosodic research is too complex in contents and points of view to be codified in a standard coding scheme. The description of the different transcription systems reviewed in the Annexes 214ff should give an idea of the variety of theories and purposes underlying the attempts to give a representation of prosodic phenomena.

Comparing the reviewed schemes is not a trivial task. An attempt has been made to describe them according to a general pattern, but this has not always been possible, due to the different nature of the schemes: some of them are well defined, used in a single project to label a single corpus, others can be considered methodologies or theories. Even a quasi-standard like ToBI has indeed lots of variants, imitations or adaptations, some of which may loosen its basic assumptions (e.g. by admitting 'movements' beside 'pitch levels', [Mayo et al. 97]). There is no agreement about the prosodic phenomena which have to be represented. Some systems are intended only for f0 representation (e.g. INTSINT, IPO, TILT, TSM, ToBI, PROSPA), while others provide labels to mark rhythm, loudness, voice quality (e.g. TEI). Most systems delimit prosodic units (some implicitly, as breaks in the f0 curve), but units types range from the single 'tone unit' (e.g. PROSPA) to complex hierarchies (e.g. SAMPA). Approaches to the transcription of intonation can be acoustic-phonetic or phonological or allowing for different abstraction levels (e.g. IPO, INTSINT, TILT, PROZODIAG) and conceive the pitch profile in terms of 'levels' or 'movements'. Some systems are developed in the framework of specific prosodic theories or methodologies (e.g. IPO, INTSINT, ToBI, TSM, PROZODIAG). In some cases labels are strictly linked to language-dependent models (e.g. ToBI, PROZODIAG), while in other cases they are more general or ‘phonetic’, although abstract (e.g. PROSPA, IPA, SAMPA). Manual labelling is for many schemes based both on auditory analysis and on visual inspection of the f0 curve and waveform, but for some schemes labels are not aligned with the signal but merely associated with linguistic units (e.g. TEI, IPA, Göteborg). Only a few systems have a ‘real’ coding book, in most cases the scheme is described in the literature. Formal evaluations of the performance have been carried out in very few cases and only one coding scheme (TEI) has been developed within a standard markup language (SGML). Some systems insert labels directly in the orthographic or phonetic transcription, while others have different tiers for prosodic annotation. Few systems have specific annotation tools, while many are compatible with standard signal analysis environments such as ESPS/Waves+.

In order to help a possible comparison, the following Table provides a synopsis of the different schemes, just a summary of their relevant features and underlying principles. For each scheme, the Table specifies which is its underlying intonation theory, which are the labelled prosodic units and phenomena (see section 6.2) and how labels are aligned with speech. Some abbreviations are used, 'p.' stands for 'phrase', '>' stands for 'labels are symbolically associated with' and '|' means 'labels are time-aligned with'. Schemes are roughly ordered according to their level of abstraction. The first schemes listed in the Table are those conceived in bottom-up approaches, where the analysis of intonation starts from the f0 curve, 'stylized' and represented with labels or parameters keeping a precise acoustic content, and reaches, as a second step, a more abstract phonological representation (e.g. contour labels for manual labelling in TILT, 'pitch configurations' in IPO, tonal labels for accent, focus, juncture in PROZODIAG). Such systems align their prosodic labels with the speech signal, in some cases at the phonetic boundary of relevant units (stressed vowel, syllable) or at turning points in the f0 curve (peaks, valleys). The link with the f0 curve is less strict for the schemes listed at the bottom of the Table. Labels are often inserted in the phonetic or orthographic representation or refer to linguistically defined units. Phonological assumptions may be more or less strong, but labels tend to have a qualitative interpretation (e.g. ToBI labelling rules are strict and rely on a predefined language-dependent phonological model, but labels may be aligned with the f0 curve; systems like TEI or Göteborg are model-independent but are more qualitative and their links with the signal are looser).

Scheme	**Prosodic Units**	**Prosodic Phenomena**	**Alignment**	**Intonation Theory**
TILT	(implicit)	*intonation:* intonational events described with starting f0, duration, amplitude, shape (numerical values) and classified as accents and boundaries, rises, falls, connections	> accents, boundaries | signal, vowel onset	Taylor: sequence of intonational events (movements)
IPO	(implicit)	*intonation:* pitch movements described with direction, timing, rate of change, size (categorical values)	> accents, boundaries | syllable | signal	IPO: f0 stylization with straight pitch movements; search for recurring f0 patterns (language dependent models)
INTSINT	(implicit)	*intonation:* transcription of the f0 curve by means of target points (classified according to pitch level)	| signal	Hirst: pitch levels, absolute tones and relative tones

SAMSINT	Intonation unit	*intonation:* local events+global trend	?	Hirst: pitch levels, global trends
PROZODIAG	Minor p. Major p.	*intonation:* global: register and range (numerical values) local: tonal labels for accent, focus, juncture	>accents, boundaries | signal	Bruce: pitch levels
KIM	Syntactic clause Phrase Parenthesis Interruption	*speech rate* *lexical stress* *sentence accent* *intonation:* downstep, intonation cont. (peaks, valleys)	> phrase, word, stressed syll. | 3 positions in accented vowel	Kohler: pitch movements
TSM	Tone units: Minor Major	*intonation:* 'tones', described with starting level and shape of the contour	> syllables, accents, tone units | accented syllables	British School: nuclear tones: pitch movements on accented syllables
ToBI	Clitic Word Intermediate p. Intonational p.	*intonation:* pitch and phrase accents, boundary tones, downstep	> accents, boundaries | signal	Pierrehumbert: pitch levels
VERBMOBIL	Word Intermediate p. Full p. Interruption (Syntactic-prosodic units)	*accents:* phrase acc., secondary acc., emphasis (*intonation:* see ToBI)	> syllables | signal	(levels)
SAMProsa	Syllable Word Tone group	*intonation:* global tones, local tones, nuclear tones *duration:* phoneme lengthening *pauses*	?	both levels and movements
IPA	Syllable Minor p. Major p.	*stress:* primary, secondary *duration:* 3 classes *intonation:* local f0 variations, global (downstep, etc.)	> phonetic transcription	symbols both for levels and movements
SAMPA	Syllable Morpheme Word	*stress:* (primary, second., .scandinav.) *duration:* phoneme	> phonetic transcription	pitch movements

	Tone group Intonational p. Rhythm group Phonological p.	lengthening *intonation:* contours *pauses*		
PROSPA	Intonation unit	*intonation:* global contour, local accents	> phrases, accents	movements; for each phrase: global slope, local pitch accents, slope of the 'tail' (after last accent)
TEI	Tone unit	*stress rhythm:* global, syllable lengthening, speech rate *pauses loudness intonation:* contours, pitch range, trend *voice quality*	>orthograph. transcription	movements (global trend, global range, local contours)
Goeteborg	/	*stress duration:* lengthening *pauses speech properties*	>orthograph. transcription	/

Choosing one of the existing schemes as a possible standard for prosodic annotation, requires a clear picture of which phenomena we want to represent and which use we intend to make of the annotated corpora. The simple search for a *de facto* standard may not be the best strategy. A widespread system like ToBI, which is intended only for *intonation* transcription, is indeed open to criticism and can't be said to be an unquestionable standard. Its extension to languages other than English is not trivial and often requires to define in advance their intonation model, rather than deriving it from the annotated corpora. The separation between phonetic and phonological representations is not clear in ToBI, which may be considered as an "uneasy compromise" between the two [Nolan et al. 97].

In conclusion, the choice (or definition) of a standard for prosodic transcription of discourse should take into account the following points.

a) It can be inferred from the great variety of phenomena underlying the term ‘prosody’ that a first step towards the choice of a standard should be the selection of the prosodic phenomena to be covered by the scheme. Also, it has to be decided if it is more adequate to define a ‘general purpose’ notation scheme, which could be used to annotate the prosody of any kind of text, or to restrict the scope of the scheme to those phenomena which play a relevant role in discourse.

b) It can be concluded from the present review that prosodic analysis can be approached from several points of view, using different theoretical models and for very different purposes. One way of handling this variety without loosing ‘standardisation’ is to allow different levels of transcription, which should include at least a phonetic representation of prosody (not limited to intonation, but including also other information, such as length), a phonological representation (in terms of pitch accents or boundary tones, for example, but including also other information, such as stress or the location of prosodic boundaries) and a functional representation (indicating the uses of prosodic phenomena to express different linguistic or pragmatic functions, and even cross-level references).

c) In order to ensure the usability of the annotated corpora both in language and speech applications, it seems also important to choose a scheme which allows the alignment of the notation symbols both with the speech signal and the orthographic (or phonetic) transcription of the annotated utterance.

7. Cross-level

Responsibility: Andreas Mengel

7.1 Introduction

This chapter gives an overview of existing standards and methods of cross level annotation. Before describing existing schemes, a set of properties and a definition of the scope of *cross-level annotation* is given. A description of the state of the art cannot be done from scratch and go into every detail, rather it has to be conducted with a set of categories. The development of categories implies a description in terms of functionality. Thus, it is necessary to have criteria which do not only allow the description, but do also evaluate existing schemes.

7.2 Cross-level Annotation

The term *cross-level annotation* implies a set of properties which have to be made clear.

- For the time being, the term of *unit of description* will be used to denote the linguistic objects under analysis; traditionally, phonemes, syllables, morphemes, words, chunks, phrases, sentences, utterances, turns, text, etc. are all examples of such units.
- *Level* refers to every unit of a communication situation that is described. By this definition, a level can be an orthographic representation of a word, a character or a sentence; it can be the value of a sample of a speech signal or the nodding of a dialogue participant. Note, that in this sense, every unit is seen as a different level. Although this might look counter-intuitive at first glance, because one would rather only regard all of the orthographic, phonetic, intonation etc. representation as one level, one has to concede that aligning consecutive units in time is a conventionalized technique common to users of language. Usually, levels of linguistic description, such as phonology, prosody, morphology and morpho-syntax, syntax, semantics and pragmatics are distinguished, for example when the 'packaging' of information in tagsets or in lexicons is at stake. Of course, each such aspect of the description of linguistic object is internally structured. For example, valency phenomena (complementation of predicates) can be described with respect to phrasal constructs (NP, PP, V, ...) appearing as complements, but also with respect to the grammatical function of the complement (subject, object, ...), or the relation governing the predicate-argument-structure. Analogously, most 'levels' can themselves be broken up into layers of description, and such a more fine grained analysis is necessary for cross-level annotation. Thus, one can also regard sequences of sounds, words, phrases, sentences as sequences of different levels.
- C*ross-level* refers to every relation, that can be established between any two or more units. These units may belong to different representation areas, e.g.., f0 and body movements, but may also be found within one area of representation, e.g.., word order phenomena. The very advantage of this view of *level* and *cross-level* is that it allows to link units of every kind with one another.
- *Annotation* refers to every (re-)presentation of a cross-level relation. These representations may be lines on a piece of paper, pointers to other locations in the memory of a computer or

the display of a query result. In fact, as will be shown below, in the existing schemes used, there are different methods of representing cross-level information. Also, this definition abstracts away from the actual encoding, it does not distinguish between the SGML representation of an annotation of cross-level phenomena by a human coder and the representation of cross-level information as a complex query expression.

In this sense, every cross-level annotation of linguistic data is the result of a query process applied to a number of units.

7.3 Cross-level Annotation Requirements

In order to come up with useful categories for the description and evaluation of existing coding schemes, some requirements have to be identified. These requirements fall into two categories, namely basic and scientific ones.

Basic requirements:

- *Multiple levels*: It is obvious that the first basic requirement is that there is more than one level and more than one unit. Again, the number of units, one or more, does not depend on the number of units physically present, but on the number of different descriptions that are attached to it. Thus, even a single word can be described with respect to many different aspects: Its orthographic representation, the sequence of sounds, the spectrogram, the duration, the part of speech, the meaning, the pitch etc.
- *Common ground*: Having more than one level of description is not sufficient for the query across and linking of different units; an infrastructure for symbolic reference from one to another is mandatory. This common infrastructure is present if it is made sure that the representations of the units to be linked can be related to each other.

Scientific requirements:

- *Theories*: Everything can be linked, but then, there is no information in a link. Thus, there must be theories or questions about the interrelation of units and its communicative relevance. These questions can then be reformulated into queries applied to the data.
- *Results*: Depending on the results of queries (no result, no theory conformant result) there will be a need to represent the relation of specific units. This representation may be called *cross-level annotation* or be transformed to a *new level of description.* In either case will it be necessary to add this new information to the existing data. Again, there will be a need for an infrastructure (*common ground*) of representation that does not only allow to relate existing descriptions to each other but does also allows to add new levels of description to the set of existing descriptions attached to the data.

Since only the fulfilment or the basic requirements can be operationalized, the remainder of this document will only deal with those.

7.4 Linking Infrastructure

Annotated corpora offer different linking infrastructures. Here, four types of linking infrastructure are described: no infrastructure, theory-inherent infrastructure, explicit link infrastructure, and linking potential infrastructure. The distinction made between these four categories is guided by the intentions of the producers of databases.

- *No infrastructure*: *No infrastructure* refers to all those corpora that possibly allow retrieval and identification of cross-level phenomena, but do not provide explicit infrastructure for this purpose. Thus, there might be a possibility of investigating cross-level phenomena because the data are described on different layers although no special effort has been made to facilitate this. However, in this case, cross-level queries are neither intended nor especially supported.
- *Theory-inherent infrastructure*: *Theory-inherent infrastructure* refers to those data represented in a structure where - according to a *given theory* - it is intended to include units of different abstraction levels in one uniform way of representation. A typical example for this kind is would be the annotation on different levels of syntactical complexity: words, phrases, sentences: All of the left boundaries of any given higher level are shared by the equivalent lower level, e.g., all sentence boundaries are phrase boundaries which are always word boundaries. In this case, cross-level annotation and query are possible simply because of theory-inherent considerations that consequently call for theory-conformant annotation. Within these theory aligned levels of descriptions, query might be possible, cross-level annotation between this area and other representations will rather not be supported.
- *Explicit link infrastructure*: *Explicit link infrastructure* refers to those data represented in a structure where - according to a *research question* rather than according to an existing theory - the connection between two or more areas or levels is provided by explicit references or tags between units of these levels. A good example of this kind is building links between f0-peaks and words or syllables. Mostly - although not always - explicit linking is provided in situations where no theory about the interaction of units across levels exists, but knowledge and theories about the structure of the areas involved. Explicit links may be used when hypotheses about interactions of two or more levels are being developed and verified: an explicit link may then have a sort of 'note-book' function. At a later stage of the research, the theory may cover the interaction facts. In the example of f0-peaks and words, another level would be introduced, encoding co-occurrences of f0-patterns and syntactic structures as theme/rheme structures. At that point of time, this way of encoding could again be a form of implicit linking, and corpora might be consistently structured. Thus, explicit linking can be considered as a research oriented and pre-theoretical way of encoding data what later on leads to new levels of description.
- *Linking potential infrastructure*: *Linking potential infrastructure* applies to those data represented in a structure with the potential for finding co-occurrences between units of different levels. This view is theory-independent as long as the positions of any units can be compared to each other. The difference between *explicit link infrastructure* and *linking potential infrastructure* is that in the case of the former, corpora are designed to meet the needs of a given set of research questions whereas in the case of the latter, data are structured with the aim of allowing to search for and represent correspondences of any two or more levels of description. This approach of linking potential infrastructure seems to be most powerful. It does not limit the application of the data to special questions and is open for the application of any new investigation.

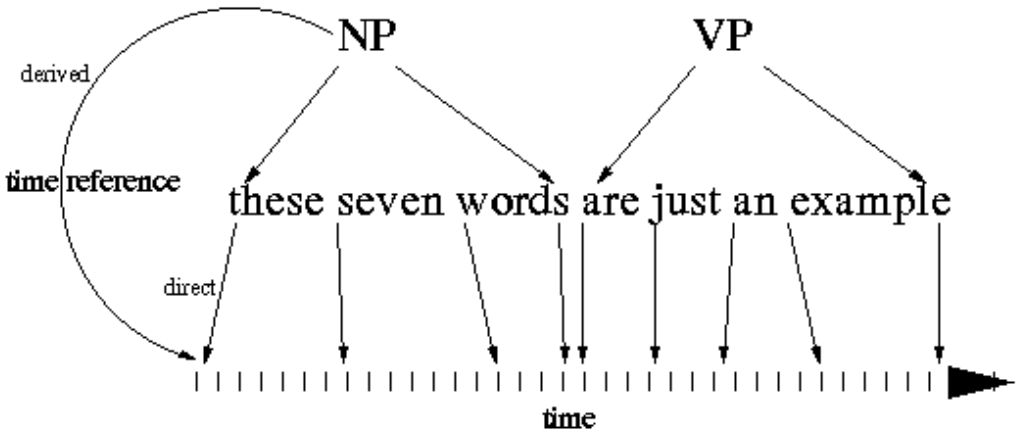
7.5 Directions of Cross-level Links

Different directions of links can be identified. Again, they can be subcategorized by their reference to time related phenomena (synchronous/asynchronous links) and the representation area of the levels (form vs. function).

7.5.1 Tim-Related Distinction

Distinctions in terms of temporal aspects serve to distinguish whether the units of levels to be linked co-occur synchronously or not. A central issue for this topic is the identification of the location of units. The location of units may be expressed in a position number and with reference to a structured list, an enumeration or time. If it is expressed in time relatively to the beginning of a (speech) file or another unit, this information may be encoded directly (e.g., in msec from the start). For hierarchically structured units it might be better to infer the time position of higher level units from the location of lower level units, e.g.. the starting point of a sentence should be derived from the starting point of its first word.

- *Synchronous links*: Synchronous links are links between different units of different levels of description that co-occur at some point or region of time. A subset of synchronous links could be named *vertical links*, i.e. those links that can be found between different hierarchically structured levels of abstraction within one description area, e.g. syntax (which may internally comprise grammatical categories, chunks, grammatical functions, etc.). In other cases, although the term *level* is used, it is more appropriate to use the word *parallel* or simply *synchronous*.



- *Asynchronous links*: Applied to one level, links of this sort connect units that appear at different points of time, e.g. repetitions. An example, where units of different levels and different points of time are linked, is the case of co-reference. Also, all phrases within a dialogue that relate to the same topic might be linked to one unit on the semantically level used for the representation of topics introduced.

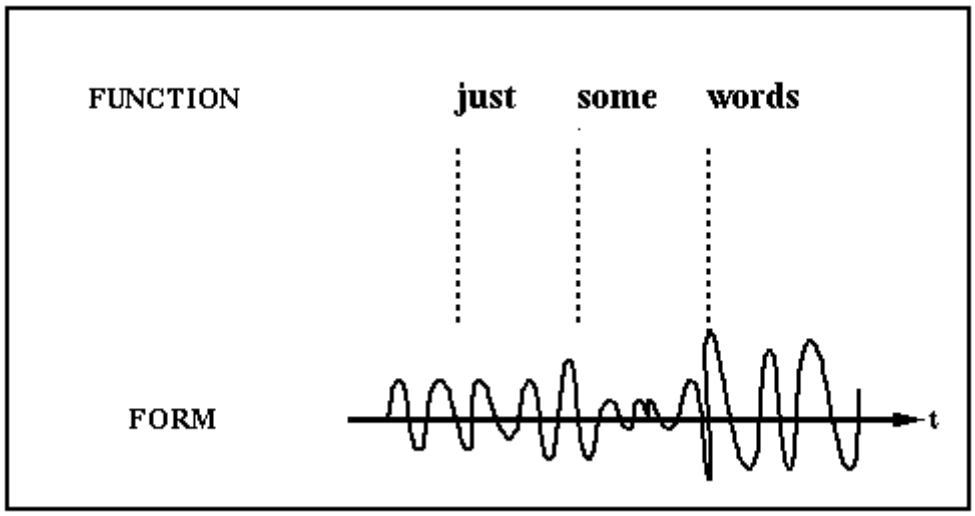
7.5.2 Representation Area Distinction

Representation area links are links within or across representational areas of spoken dialogue. The two representational areas are the form and the function of human behavior: All physical and physiological manifestations of utterances, movements, and the utterance situation belong to the form, all categories that either refer to classes of physiological or physical configurations (phones, f0-patterns, rhythm, gestures, signs) or to classes of configurations of functional units (morphemes, phrases, utterance types, speech acts) belong to the function.

- *Form links*: Form links are links that connect physical or physiological aspects of utterances. A typical example might be the correlation of pitch and energy, another example

would be the duration of pauses compared to the signal to noise ratio in the communication situation.

- *Function links*: Function links are links that connect functional units of language. Examples for this kind of links are the occurrence of personal pronouns in commands or the relation of passive constructions and tense.
- *Form-function links*: Form-function links are links that connect form aspects and function units. Two examples of this kind of relation are the pitch of parts of speech or the duration of pauses before questions.



7.6 Contact-Types of Cross-level Links

If the phenomena under investigation are called units then it is probable that most of these units also have an extension, mainly in time. If this holds for most units to be linked to each other, one will have to define and distinguish the kind of contact between the units compared.

7.6.1 Contact Points

Contact points are those parts of a unit that are used as reference for the alignment or the localization. Thus, for most units it is mainly the beginning that can be located and is taken as a reference. In the majority of cases, this might be sufficient. Thus, when simultaneous speech is marked by an alignment of orthographic information, the beginning of the utterances that are simultaneous will be marked. Unfortunately, in this case, nothing can be said about the synchronicity of the following words, syllables or sounds, although this might be relevant for some investigations. Interesting for cross-level investigations might not only be the starting points of units but also their end, some point in between or larger regions.

7.6.2 Contact Overlap

In cases where the contact refers to a region of units rather than just to one point of time, it might also be useful to define the kind of contact. *Contact overlap* defines in what way the units

under investigation can be described according to their relative position. In cases, where there is no time overlap, the *no contact* situation can also be defined more precisely.

- *No contact*: The no-contact relations seem trivial, but it might be useful to distinguish distance situations from contact situations, and it might help to provide information on the relative distance of the units.
- *Contact*: On the time scale, contact can be defined by the fact that there are points of time where parts of all units that are investigated are present. These contacts can further be divided into partial overlap, total overlap, and inclusive overlap.

7.7 Encoding

Three different aspects of the representation of data representing text can be named: The distribution of information to files, the machine representation of the data and the notation style. These aspects are important for the question of *common ground* addressed above.

- *Distribution to files*: The information for different levels of speech can be put into one file or into as many files as there are levels of description. If the latter method is chosen, core reference units have to be selected, which all information can be related to, e.g., words. In the first case (one file), the common technique for the identification of units, is 'physical neighbourhood', i.e., putting all tags into the same place. In the other case (multiple files), a common reference to symbolic labels of the units, e.g., names, numbers or time stamps, are used.
- *Machine representation:* In most cases, standard ASCII is used to represent text, tags and comments. This way, only one kind of software is needed for the display and access to the data. Yet, some data might in the first place not be accessible with normal text editors, i.e. speech files and video tapes. In order to edit these data, different software is needed. In some cases, however, a decision, which way of representation to use, might be difficult, e.g., when representing f0 values.
- *Notation style*: There may be different ways of representing tags: Depending on different theoretical backgrounds tags might be put into the text using brackets, hash marks etc.

7.7.1 Description of Existing Corpora

In this section, existing corpora representing more than one level of description will be described using the background of classification described above.

Schemes described in this report are the following:

BAS	Bavarian Archive for Speech Signals
BNC	British National Corpus
CHI	Child Language Data Exchange System
DAM	Dialogue Act Markup in Several Layers
KCS	Kiel Corpus of Read and Spontaneous Speech
SAB	Synthesis Markup Language
SAM	SAM Standards

TEI	Text Encoding Initiative
TRA	Natural Spoken Dialogue and Interactive Planning
VER	Verbmobil II Conventions for Spontaneous Speech
BFR	Bonn Focus Research

For each of the schemes, further background information can be found in the Annexes 260ff.

The aim of this text is not to describe in depth what phenomena were coded but rather if and how cross-level phenomena have been and can be addressed.

7.8 Description Items

- *Multilevel vs. hierarchy*: As described above, consistent labelling of phenomena that - according to a given theory - can be described in one hierarchical framework, will be more likely than consistent labelling of phenomena that are considered belonging to different areas. Thus a distinction will be made between multi-level and hierarchical description of data, although in some cases, it might be difficult to tell the difference between the two.
- *Corpus vs. coding scheme*: Within the data to be compared here, there are examples of real coding schemes that were explicitly designed to meet special coding and retrieval interests whereas some other examples are only corpora were there is only little documentation on the coding scheme used. Some examples can be grouped in between. Thus, part of the coding schemes can only be inferred from existing data coded.
- *Explicit tags vs. layout tags*: In some cases, the tagging syntax is very much formalized (cf., TEI) putting a tag around every item that can be identified, whereas in others, the encoding of these concepts is rather implicit and not uniform. In the latter case (cf. esps label files) the tagging syntax could be referred to as a formal layout. This difference has implications for parsing: In the formalized version, one parser can read and extract all information, in the second case, for each corpus and all concepts encoded, a special treatment is inevitable.
- *Uniform tagging vs differential tagging*: In the case of explicit tagging, i.e. providing special character sequences for marking the relevant phenomena, the characters used for the tags can be uniform (e.g. all beginning with '<') or different (e.g., '#', '$', '@', etc.). Again, in the case of SGML, the syntax and semantics of the tags is very much structured. Other tag systems may use different ways of tagging different phenomena. Layout tags are never uniform.
- *Open tagging vs. closed tagging*: Open tagging is that kind of tagging of a corpus that allows further iterations and addition of annotation. If the description of a data-base offers only orthographic reference to a dialogue, but no time or speech signal information, i.e., reference to the original data, then certain aspects of the data cannot be added. Thus, for an open tagging structure of speech, it is crucial that the accessibility to the raw data or at least some time aligned phonetic transcription is at hand.
- *Direct vs. indirect cross-level tags*: Tagging cross-level phenomena or allowing cross-level queries are two extremes of bringing different phenomenological layers together. Consider the following example where some (virtual) piece of dialogue is transcribed:

  A: I was saying something when suddenly you also started to talk.
  B: I did not want to interrupt you.

In this case, cross-level annotation is directly coded (by layout). This is done by aligning the beginning of the first word of the utterance of the second speaker (B) with a word from the utterance of the first speaker (A). Only one single event has been marked, namely the moment where B starts to talk during A's turn. This cross-level coding can not be used for any other purpose. No information is coded or can be retrieved about what other onsets of words coincide or when exactly the utterance of B stops. A structured encoding of different levels need not be direct in marking the co-occurrence of events but offers the potential in retrieving any simultaneous phenomena. This structured coding does not require direct links but a common reference level, e.g., time.

- *Area of description*: The areas covered by schemes and corpora may vary a lot, thus, the following areas are distinguished.

SIG	signal, e.g., a speech file, time reference, video signal etc.
PHO	phonetics, e.g., sound duration, f0 values, ToBI annotation
ORT	orthography, e.g., a transcription of utterances
MOR	morphology, e.g., a morphological analysis of the words
POS	part-of-speech, e.g., noun, adjective etc
SYN	syntax, e.g., phrase and sentence structure
SEM	semantics, e.g., coreference
CNT	context, e.g., where the recording was made, about other events, paralinguistic elements
CMT	comments, e.g. on reliability, problems etc.

The following table lists the abbreviations used in the evaluation table.

Code	**Description**
ML	more than one level is annotated
HC	hierarchical concepts are annotated
CO	scheme is implied in or has been applied to a corpus
CS	scheme has been defined
ET	there are explicit tags
LT	there are layout tags
UT	the syntax of the tags is uniform
OT	open tagging
CT	cross-level tags are coded

PT	automatic cross-level tagging is possible

	BAS	BNC	CHI	DAM	KCS	SAB	SAM	TEI	TRA	VER
ML	+	+	+	+	+	+	+	+	+	+
HC		+	+	+				+	+	
CO	+	+	+		+	+	+	+	+	+
CS	+	+	+	+	+	+	+	+	+	+
ET	+	+	+	+	+	+	+	+	+	+
LT	+		+	+	+		+		+	
UT		+		+		+		+		
OT	+	+	+	+	+			+		+
CT				+					+	
PT	+	+	+	+				+	+	+
DESCRIPTION AREAS										
SIG	+	+			+		+	+		
PHO	+	+			+			+	+	
ORT	+	+	+	+	+	+	+	+	+	+
MOR			+		+			+		
POS		+	+		+			+		+
SYN			+					+		+
SEM				+		+		+		+
CNT	+		+	+				+	+	+
CMT	+	+	+		+	+	+	+	+	+

7.9 Concluding Remarks

In general, most of the coding schemes and corpora have been designed for special purposes (applications, theoretical frameworks). To nearly none of the corpora, cross-level annotation has been applied, and if, then only for one purpose and in a direct way. For most of the corpora, cross-level annotation could be added, although there is a wide variety: Depending on the number of different areas of description, the number of cross-level phenomena that can be investigated may be very small for some schemes/corpora.

One final remark: From the considerations and examples above it seems that the diversity of levels of description that can be applied to dialogue corpora grows by the uniformity of the description styles used.

8. Summary

8.1 Achieved work

This report gives an overview of the state of the art of coding schemes for the levels of prosody, morpho-syntax, co-reference, dialogue acts, communication problems and cross-level issues.

For the level of communication problems the results of the research show that little work has been done in this field. In fact, only one scheme was found which has its focus point on communication problems. Three other schemes were included which encompass phenomena of relevance to communication problems. However, probably many other coding schemes in this report could have been mentioned under communication problems with the same right since they also include phenomena of relevance to communication problems coding. It is quite obvious that MATE has to do some basic work in this area.

In the case of co-reference annotation, more work has been done. One-person annotations of substantial pieces of text have been done by Prince and by Fraurud [1990], and substantial multi-person annotations by UCREL at the University of Lancaster and by Poesio and Vieira, who also ran a reliability test, if only for definite descriptions; more informal tests have been performed in the context of MUC. Annotation manuals have been written by Passonneau, for MUC, and by UCREL. So, even if a standard hasn't been proposed yet, we can build upon a substantial body of work. In MATE we propose to build on the closest thing that there is to a standard now – the MUC proposal – by adding ways for annotating more types of coreference information, more dialogue-related phenomena, and some issues that arise when languages other than English are considered. However, because of the reliability results discussed above, we will organize the annotation scheme in layers – a core scheme of instructions to annotate those aspects of coreference that can be marked fairly reliably, plus instructions for marking other aspects for those researchers interested in more complete annotations. We also propose to make the scheme more TEI conformant by adopting some suggestions by Brunesaux and Romary.

Compared with the level of communication problems and cross-level issues quite a lot of schemes on dialogue acts are available. Most of them are pretty well described and evaluated with (very) good results. A clear distinction can be drawn between the schemes for shallow analysis and schemes for deep analysis. So in contrast to the level of cross-level issues, for example, nothing new has to be invented. Instead the work of MATE should involve developing a scheme which is created out of all existing schemes – as far as this is possible – or one of the existing schemes is selected as "the" standard scheme.

Dialogue annotation at the morphosyntactic level is fraught with a considerable amount of new germane problems, basically foreign to the practice of annotating free written texts. Adaptation of existing schemata/practices developed for written texts is indeed possible, with the proviso that, in some cases, it does not possibly boil down to a straightforward extension of available sets of tags. For example, we argued that, while annotation of phenomena such as hesitators and pauses simply requires addition of further tags specific of annotation of spoken material, syntactic annotation of dialogues calls for a preliminary process of editing raw transcriptions, which are eventually overlaid with progressively more abstract levels of syntactic information.

In this context, editing is conceptually different from the practice of text normalization introduced, e.g., by Switchboard among others. In fact, edited material can, and in fact should, be annotated at the level of shallow parsing, while being eventually ignored for the purposes of functional annotation. Accordingly, edited material is not simply expunged (as is done in the normalization approach), but considered for annotation at some level only. This way of approaching syntax requires specification of more than one level of syntactic analysis, so as to mark word order phenomena at a distinct level from functional dependencies. The approach tends to disfavour (although it is not incompatible with) syntactic representations whereby these two levels are actually compounded into one (as with complete constituency-based representations). Finally, it appears to strike a reasonable compromise between the requirements of (automatic) parsing on the one hand, and the concern of departing from the attested linguistic evidence available as little as possible.

The level of prosody differs from the above mentioned levels in terms of that there exists a great variety of schemes, defined with different purposes and covering different prosodic phenomena. A clear understanding of which phenomena we want to represent and which use we intend to make of the annotated corpora is preliminary to the choice (or definition) of a standard for prosodic transcription. What seems reasonable is that the selected scheme(s) should allow for multiple level annotation (phonetic, phonological, functional) and for synchronization with the signal.

The amount of research that has been done on cross-level issues is as low as the one for communication problems. Only two real annotation schemes are stated which are DAMSL and TEI. And even those don't cover all levels considered in the MATE project. Therefore, again, MATE can lead to new standards in this field.

To sum up, one can say that this report fulfils its task in the way that it presents a broad view on existing schemes. A lack of schemes on the level of communication problems and cross-level issues have been enlightened. Insights of the structure of dialogue act schemes have been shown and shortcomings of existing schemes on the level of prosody have been discussed. With all this information the report represents a well prepared starting point for the MATE project.

8.2 Future Work

This report represents the background needed for deliverables D1.2 and D2.1.

Deliverable D1.2 will present guidelines for specifying the formal representation of coding schemes. One result of D1.1 that influences D1.2 is the point that the majority of the observed schemes take as their format an SGML variant. This confirms the decision to use SGML or XML as mark-up language in the MATE project.

Deliverable D2.1 will describe document type definitions (DTDs) and will give notes on schemes. D2.1, or rather WP2, will be heavily based on the annotated phenomena we can observe in the collected coding schemes. The level mark-up will take these phenomena as its starting point and then eventually, if needed, add other phenomena as a result of testing the level mark-up on different corpora. WP2 is the most direct consumer of D1.1 results.

Since the workbench (WP3) must be able to handle the level mark-up, the D1.1 results influence WP3, too.

References

[Alexandersson et al. 97] Jan Alexandersson, Bianka Buschbeck-Wolf, Tsutomu Fujinami, Elisabeth Maier, Norbert Reithinger, Birte Schmitz, Melanie Siegel. Dialogue Acts in VERBMOBIL-2. Verbmobil Report 204. DFKI Saarbrücken, Universität Stuttgart, TU Berlin, Universität des Saarlandes, 1997.

[Allen et al. 97] James Allen and Mark Core. Draft of DAMSL: Dialog Act Markup in Several Layers., October 1997.

[Austin 62] J. Austin. How to Do Things with Words. Oxford University Press, 1962.

[Bard et al. 95] E. G. Bard, C. Sotillo, A. H. Anderson, M. M. Taylor. The DCIEM map task corpus: Spontaneous dialogue under sleep deprivation and drug treatment. Proc. ESCA-NATO Tutorial and Workshop on Speech under Stress. Lisbon, 1995.

[Batliner et al. 96] A. Batliner et al. Syntactic-Prosodic Labeling of large Spontaneous Speech Database. ICSLP96, 1996.

[Batliner et al. 97] A. Batliner et al. Can we tell apart intonation from prosody (if we look at accents and boundaries)?. ESCA Workshop on Intonation. Athens, 1997.

[Beaugendre et al. 92] F. Beaugendre et al. A perceptual study of French intonation. ICSLP 92, Banff, 1992.

[Beaugendre 94] F. Beaugendre. Une étude perceptive de l'intonation du francais. PhD thesis. Université d'Orsay, 1994.

[Beckman et al. 94] M. E. Beckman, G. M. Ayers. Guidelines for ToBI Labelling. Version 2.0, February 1994.

[Beckman et al. 94b] M. E. Beckman, J. Hirschberg (1994).- The ToBI Annotation Conventions. Appendix A of [Beckman et al. 94a]

[Bernsen et al. 96] N.O. Bernsen, H. Dybkjær, L. Dybkjær. Cooperativity in Human-Machine and Human-Human Spoken Dialogue. Discourse Processes. Vol. 21, No. 2, 1996, 213-236.

[Bernsen et al. 98] N.O. Bernsen, H. Dybkjær, L. Dybkjær. Designing Interactive Speech Systems. From First Ideas to User Testing. Springer Verlag 1998.

[Bernsen et al. 97] N.O. Bernsen, H. Dybkjær, L. Dybkjær, and V. Zinkevicius. Generality and transferability. Two issues in putting a dialogue evaluation tool into practical use. Proceedings of Eurospeech'97, September 1997, 1911-1914.

[Bertenstam et al. 95] J. Bertenstam, M. Blomberg, R. Carlson, K. Elenius, B. Granström, J. Gustafson, S. Hunnicutt, J. Högberg, R. Lindell, L. Neovius, A. de Serpa-Leitao, L. Nord, N. Ström. The Waxholm Application Data-Base. Proc. of Eurospeech '95 vol1 pp 833-836, Madrid, 1995.

[Bod et al. 97] R. Bod, R. Scha. Data-Oriented Language Parsong. An Overview. 1997.

[Bosch 93a] L. ten Bosch. On the automatic classification of pitch movements. ESCA Workhop on Prosody, Lund,1993.

[Bosch 93b] L. ten Bosch. Algrithmic classification of pitch movements. Eurospeech 93.

[Brindopke et al. 97] C. Brindopke, A. Phade, F. Kummert, G. Sager (Bielefeld Univ.). An environment for the labeling and testing of melodic aspects of speech. Eurospeech 97

[Bruce 77] G. Bruce. Swedish word accents in sentence perspective. Lund: Gleerup, 1977.

[Bruce 88] G. Bruce. 2.3. Suprasegmental categories and 2.4. The symbolization of temporal events", Journal of the International Phonetic Association 1, 8,2: 75-76, 1988.

[Bruce 89] G. Bruce. Report from the IPA working group on suprasegmental categories. Lund University Department of Linguistics, General Linguistics, Phonetics, Working Papers 35: 25 – 40, 1989.

[Bruce et al. 95a] G. Bruce et al.. Towards an enhanced prosodic model adapted to dialogue applications. ESCA Workshop on Dialogue Systems, 1995a.

[Bruce et al. 95b] G. Bruce et al.. Text-to-Speech synthesis in spoken dialogue research. Eurospeech 95, 1995.

[Bruce et al. 96] G. Bruce et al.. Developing the modelling of Swedish Prosody in spontaneous dialogue. ICSLP 96, 1996.

[Bruce et al. 97a] G. Bruce et al.. Text-to-intonation in spontaneous Swedish. Eurospeech 97, 1997.

[Bruce et al. 97b] G. Bruce et al.. Global features in the modelling of intonation in spontaneous Swedish. ESCA Workshop on Intonation, 1997.

[Bruce et al. 97c] G. Bruce et al.. Models of Intonation – from a Lund-horizon. ESCA Workshop on Intonation, 1997.

[Bruce et al. 97d] G. Bruce et al.. On the Analysis of Prosody in Interaction. In: Y. Sagisaka, N. Campbell & N. Higuchi (eds.). Computing Prosody. Springer, New York, 1997.

[Bruneseaux & Romary 97] F. Bruneseaux and L. Romary. Codage des references et coreferences dans le dialogues homme-machine. In: Proc. of ACH-ALLC, Kingston, 1997.

[Bunt 98] Harry Bunt. Dynamic Interpretation and Dialogue Theory (Draft). To appear in: M. M. Taylor, D. G. Bouwhuis & F. Neel (eds.). The Structure of Multimodal Dialogue. Vol. 2. Amsterdam: John Benjamins (30 pp), 1998.

[Burnage 90] G. Burnage. CELEX - A Guide for Users. Nijmegen: Centre for Lexical Information, University of Nijmegen, 1990.

[Calzolari & McNaught 96] N. Calzolari, and J. McNaught. *EAGLES—Expert Advisory Group on Language Engineering Standards*. http://www.ilc.pi.cnr.it/EAGLES/home.html, 1996., N. and McNaught, J.: *EAGLES—Expert Advisory Group on Language Engineering Standards*. http://www.ilc.pi.cnr.it/EAGLES/home.html, 1996.

[Campione 97] E. Campione, E. Flachaire, D. J. Hirst, J. Véronis. Stylisation and symbolic coding of F0, a quantitative approach. In: A. Botinis, G. Kouroupetroglou & G. Carayiannis (eds*) Intonation: Theory, Models and Applications.* (Proceedings of an ESCA Workshop. 18.-20.September. Athens, Greece), 71-74, 1997.

[Carletta 96] J. Carletta. Assessing agreement on classification tasks: the kappa statistic. Computational Linguistics, 22(2), 249—254, 1996.

[Carletta et al. 97] Jean Carletta, Nils Dahlbäck, Norbert Reithinger, Marilyn A. Walker (editors). Standards for Dialogue Coding in Natural Language Processing. Dagstuhl-Seminar-Report 167, 1997.

[Carlson & Granstrom 85] R. Carlson, B. Granstrom. Rule controlled data base search. Speech Transmission Laboratory – Quarterly Status Report 4, 1985.

[Carroll et al. 96] J. Carroll, T. Briscoe, N. Calzolari, S. Federici, S. Montemagni, V. Pirrelli, G. Grefenstette, A. Sanfilippo, G. Carroll, M. Rooth. Specification of Phrasal Parsing. Deliverable 1, Work Package 1, Progetto CEE SPARKLE "Shallow Parsing and Knowledge Extraction for Language Engineering" (LE-2111), 1996.

[Chinchor & Hirschman 97] N. Chinchor and L. Hirschman. MUC-7 coreference task definition, verison 3.0. Available from the authors, chinchor@gso.saic.com, 1997.

[Clark 77] H. H. Clark. Bridging. In P. N. Johnson-Laird and P. Wason, editors, Thinking: Readings in Cognitive Science. Cambridge University Press, London and New York, 1977.

[Clark & Haviland 77] H. H. Clark and S. E. Haviland. Comprehension and the given-new contract. In R. O. Freedle, editor, Discourse Production and Comprehension, chapter 1, pages 1- 40. Ablex, Norwood, NJ, 1977.

[Clark & Schaefer 87] H. H. Clark and E.F. Schaefer. Collaborating on contributions to conversations. Language and Cognitive Processes, 2, 19-41, 1987.

[Coile et al. 94] B. van Coile et al. ProTran: a Prosody Transplantation Tool for Text-to-Speech Application. ICSLP 94, Yokohama, 1994.

[Cole et al. 96] R.A. Cole, J. Mariani, H. Uszkoreit, A. Zaenen, and V.W. Zue (Editorial Board), G. Varile and A. Zampolli (Managing Editors). Survey of the State of the Art in Human Language Technology. Sponsors: National Science Foundation, Directorate XIII-E of the Commission of the European Communities, Center for Spoken Language Understanding, Oregon Graduate Institute, 1996. URL:

http://www.cse.ogi.edu/CSLU/HLTsurvey/.

[Core et al. 97] M. G. Core, L. K. Schubert. Handling Speech Repairs and Other Disruptions Through Parser Metarules. In: Working Notes AAAI Spring Symposium on Computational Models for Mixed Initiative Interaction. Stanford CA, March 1997. Available as AAAI TR SS-97-04.

[Crystal 69] D. Crystal. Prosodic Systems and Intonation in English. Cambridge, Cambridge University Press, 1969.

[Davies 97] Bethan L. Davies. An Empirical Examination of Cooperation, Effort and Risk in Task-Oriented Dialogue. University of Edinburgh, 1997.

[Dowty 86] Dowty, D. R. The effects of aspectual class on the temporal structure of discourse: Semantics or pragmatics? Linguistics and Philosophy, 9(1). 1986.

[Dusterhoff et al. 97] K. Dusterhoff, A. W. Black. Generating f0 contours for speech synthesis using the tilt intonation theory. ESCA Workshop on Intonation, Athens, 1997.

[Dybkjær 97] L. Dybkjær, N.O. Bernsen, and H. Dybkjær. Generality and Objectivity. Central Issues in Putting a Dialogue Evaluation Tool into Practical Use. Proceedings of the ACL/EACL Workshop on Spoken Dialog Systems, Madrid, July 1997, 17-24.

[Enç 81] M. Enç. Tense Without Scope: An Analysis of Nouns as Indexicals. Ph.D. thesis, University of Wisconsin, Madison, WI. Distributed by IULC, 1981.

[Faurud 90] K. Fraurud. Definiteness and the processing of NPs in natural discourse. Journal of Semantics, 7, 395 - 433, 1990.

[Federici et al. 96] S. Federici, S. Montemagni, V. Pirrelli. Shallow Parsing and Text Chunking: a View on Underspecification in Syntax. In: Atti del Workshop On Robust Parsing, 12 –16 agosto 1996, nell'ambito della Scuola Estiva Europea in Logic, Language and Information, Prague, Czech Republic.

[Federici et al. 98] S. Federici, S. Montemagni, V. Pirrelli. Chunking Italian: Linguistic and Task-oriented Evaluation. In: J. Carroll (ed.) Proceedings of the Workshop on the Evaluation of Parsing Systems. Granada, Spain, May 26, 1998.

[Fintel 94] K. von Fintel. Restrictions on Quantifier Domains. Ph.D. thesis, University of Massachusetts, Amherst, 1994.

[Fligelstone 92] S. Fligelstone. Developing a scheme for annotating text to show anaphoric relations. In: G. Leitner, editor, New directions in English language corpora, pages 153 - 170. Mouton de Gruyter, Berlin, 1992.

[Fraurud 90] K. Fraurud. Definiteness and the processing of NPs in natural discourse. Journal of Semantics, v. 7, pages 395-433, 1990.

[Fujisaki & Sudo 71] H. Fujisaki and H. Sudo. Synthesis by rule of prosodic features of connected Japanese. Proc. VII Int. Congr. Acoustics Budapest 1971.

[Gaizauskas 97] R. Gaizauskas (Ed.). Proceedings of the SALT Club Workshop on Evaluation in Speech and Language Technology, Sheffield, 1997.

[Gibbon 90] D. Gibbon. Survey of Prosodic Labelling for EC Languages. SAM-UBI-1/90, 12 February 1989; Report e.6, in ESPRIT 2589 (SAM) Interim Report, Year 1., Ref. SAM-UCL G002. University College London, February 1990.

[Gibbon et al. 97] D. Gibbon, R. Moore, R. Winski (Eds.). Handbook on Standards and Resources for Spoken Language Systems. Berlin: Mouton De Gruyter, 1997.

[Godfrey et al. 92] J. Godfrey, E. Holliman, J. McDaniel. SWITCHBOARD: Telephone Speech Corpus for research and development. ICASSP92, 1992.

[Grice 75] P. Grice. Logic and conversation. In P. Cole and J. L. Morgan (Eds.), Syntax and Semantics Vol. 3: Speech Acts. New York: Academic Press 1975, 41-58. Reprinted in Paul Grice: Studies in the Way of Words. Cambridge, MA, Harvard University Press, 1989.

[Grice et al 95a] M. Grice, R. Benzmueller. Transcription of German intonation using ToBI tones. The Saarbruecken system", Phonus 1, University of the Saarland, pp. 33-51, 1995.

[Grice et al. 95b] M. Grice, M. Savino. Low tone versus ësagí. In: Bari Italian intonation; a perceptual experiment. *Proc. XIII International Congress of Phonetic Sciences, Stockholm*, 1995.

[Grice et al. 96] M. Grice, M. Reyelt, R. Benzmuller, J. Mayer, A. Batliner. Consistency in Transcription and Labelling of German Intonation with GtoBI. Proc. ICSLP 96.

[Grønnum 87] N. Grønnum Thorsen. Suprasegmental transcription. Annual Report of the Institute of Phonetics, University of Copenhagen (ARIPUC), 21, pp. 1-28, 1987.

[Hart et al. 90] J. 't Hart, R. Collier, A. Cohen. A perceptual study of intonation. Cambridge University Press, 1990.

[Heeman & Allen 97] P. A. Heeman, J. Allen, Intonational Boundaries, Speech Repairs and Discourse Markers: Modeling Spoken Dialogue, in Proceedings of ACL, 1997, Madrid, 1997.

[Heim 82] I. Heim. The Semantics of Definite and Indefinite Noun Phrases. Ph.D. thesis, University of Massachusetts at Amherst, 1982.

[Heuven & Pols 93] V. van Heuven, L. Pols, Analysis and Synthesis of Speech, Berlin, 1993.

[Hinrichs 81] E. Hinrichs. Temporale Anaphora in englischen Staatsexamen, Universität Tübingen, 1981.

[Hirschberg et al. 97] j. Hirschberg, C. Kamm and M. Walker (Eds.). Proceedings of the ACL Workshop on Interactive Spoken Dialog Systems: Bringing Speech and NLP Together in Real Applications. Madrid, 1997.

[Hirschmann & Thompson 96] L. Hirschmann and H.S. Thompson. Overview of evaluation in speech and natural language processing. In Cole et al. 1996, Section 13.1.

[Hirst 91] D. J. Hirst. Intonation models: Towards a third generation. In: Actes du XIIème Congrès International des Sciences Phonétiques. 19-24 août 1991, Aix-en-Provence, France. Aix-en-Provence: Université de Provence, Service des Publications. Vol. 1 pp. 305-310, 1991.

[Hirst 94] D. J. Hirst. The symbolic coding of fundamental frequency curves: from acoustics to phonology. In FUJISAKI, H. (Ed.) *Proceedings of International Symposium on Prosody.* Satellite Workshop of ICSLP 94, Yokohama, September 1994.

[Hirst et al. 93] D. J. Hirst, A. di Cristo, M. le Besnerais, Z. Najim, P. Nicolas, P. Roméas. Multilingual modelling of intonation patterns. In: HOUSE, D. TOUATI, P. (Eds.) *Proceedings of an ESCA Workshop on Prosody. September 27-29, 1993, Lund, Sweden*. Lund University Department of Linguistics and Phonetics, Working Papers 41. pp. 204-207, 1993.

[Hirst et al. 94] D. J. Hirst, N. Ide, J. Veronis. Coding fundamental frequency patterns for multilingual synthesis with INTSINT in the MULTEXT project. In: Proceedings of the ESCA/IEEE Workshop on Speech Synthesis, New York, September 1994.

[Hirst et al. ?] D. J. Hirst, A. di Cristo (Eds.) (forthcoming).- Intonation Systems. A Survey of 20 Languages. Cambridge: Cambridge University Press.

[Johansson 95a] S. Johansson. The approach of the Text Encoding Initiative to the encoding of spoken discourse. In LEECH, G.- MYERS, G.- THOMAS, J. (Eds).- Spoken English on Computer: Transcription, Markup and Applications. Harlow: Longman. pp. 82-98.2, 1995.

[Johannson 95b] S. Johansson. The Encoding of Spoken Texts. Computers and the Humanities 29,1: 149-158; in IDE, N.- V...RONIS, J. (Eds) (1995) The Text Encoding Initiative. Background and Context. Dordrecht: Kluwer Academic Publishers. pp. 149-158, 1995.

[Jones and Galliers 96] K.S. Jones and J.R. Galliers. Evaluating Natural Language Processing Systems. An Analysis and Review. Springer Verlag, Berlin, 1996.

[Kameyama et al. 93] M. Kameyama, R. Passonneau, and M. Poesio. Temporal centering. In: Proc. of the Meeting of the Association for Computational Linguistics, pages 70 -77, Columbus, OH, 1993.

[Kamp 79] H. Kamp. Events, instant and temporal reference. In: R. Bauerle, U. Egli, and A. von Stechow, editors, Semantics from Different Points of View, pages 376 -417. Springer-Verlag, 1979.

[Kamp 81] H. Kamp. A theory of truth and semantic representation. In: J. Groenendijk, T. Janssen, and M. Stokhof, editors, Formal Methods in the Study of Language. Mathematical Centre, Amsterdam, 1981.

[Kamp et al. 93] H. Kamp and U. Reyle. From Discourse to Logic. D. Reidel, Dordrecht, 1993.

[Karttunen 76] L. Karttunen. Discourse referents. In: J. McCawley, editor, Syntax and Semantics 7 – Notes from the Linguistic Underground. Academic Press, New York, 1976.

[Knowles et al. 66] G. Knowles, A. Wichmann, P. Anderson. Working with Speech: Perspective on research into the Lancaster/IBM Spoken English Corpus. London and New York. Longman, 1966.

[Kohler 95] K. J. Kohler. PROLAB: The Kiel system of prosodic labeling. Proc. XIIIth ICPhS, Stockholm, 1995.

[Kohler et al. 95] K. J. Kohler, M. Patzold, A. P. Simpson. From Scenario to Segment: the Controlled Elicitation, Transcription, Segmentation and Labelling for Spontaneous Speech. AIPUK 29, Kiel University, 1995.

[Kohler 96] K. J. Kohler. Labelled Data Bank of Spoken Standard German. The Kiel Corpus of Read/Spontaneous Speech. ICSLP 96, 1996.

[Kohler 97] K. J. Kohler. Parametric Control of Prosodic Variables by Symbolic Input in TTS. In: van Santen et al.. Progress in Speech Synthesis. New York, 1997.

[Krippendorf 80] K. Krippendorf. Content Analysis. Sage Publications, 1980.

[Larsson 98] Staffan Larsson. Coding Schemas for Dialogue Moves. http://www.ling.gu.se/~sl. 1998.

[Lascarides et al. 93] A. Lascarides and N. Asher. Temporal Interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, **16**(5), 437 – 493, 1993.

[Leech et al.98] Geoffrey Leech, Martin Weisser and Andrew Wilson. Draft chapter: Survey and guidelines for the representation and annotation of dialogue. LE-EAGLES-WP4-3.1. Integrated Resources Working Group. 1998.

[Leeuwen et al. 93] H. van Leeuwen, E. te Lindert. Speech Marker: A flexible framework for constructing text-to-speech systems. In: V. van Heuven, L. Pols. Analysis and Synthesis of Speech. Berlin, 1993.

[Lehiste 70] Lehiste. Suprasegmentals. Cambridge, Mass, MIT Press, 1970.

[Léon & Martin 70] P. Léon, P. Martin. Prolegomènes á l'étude des structures intonatives. Montréal. Didier (Studia Phonetica 2), 1970.

[Llisterri 94] J. Llisterri. Prosodic encoding survey. Multext (LRE 62-050) report. WP1, 1994.

[Llisterri 96a] J. Llisterri. Preliminary recommendations on Spoken Texts. EAGLES Document EAG-TCWG-SPT/P, May 1996.

[Llisterri 96b] J. Llisterri (ed.). Prosody Tools Efficiency and Failures. Deliverable 4.5.2, WP4 Corpus T4.6 Speech Markup and Validation, 1996.

[MacWhinney 91] B. MacWhinney. The CHILDES project: Tools for Analyzing Talk. Erlbaum. Hillsdale, NJ, 1991.

[MacWhinney 94] B. MacWhinney. The CHILDES project: Tools for Analyzing Talk. 2nd edition. http://sunger2.uia.ac.be/childes/manual.pdf

[Mayer 95] Jörg Mayer. Transcription of German intonation - the Stuttgart System. Ms., University of Stuttgart, 1995.

[Mayo et al. 97] C.Mayo, M.Aylett, R. Ladd. Prosodic transcription of Glasgow English: an evaluation study of GlaToBI. Proc.ESCA Workshop on Intonation, Athens, 1997.

[McRoy 96] S. McRoy (Ed.). Detecting, Repairing and Preventing Miscommunication, Portland Oregon, 1996.

[Meeter et al. 95] M. Meeter et al. Disfluency annotation stylebook for the Switchboard Corpus. Available via: ftp.cis.upenn.edu/pub/treebank/swbd/doc/DFL-book.ps

[Mertens et al. 97] P. Mertens et al. Comparing Approaches to Pitch Contour Stylization for Speech Synthesis. In: van Santen et al. Progress in Speech Synthesis. New York, 1997.

[Nespor & Vogel 86] M. Nespor, I. Vogel. Prosodic Phonology. Dordrecht, Foris. Studies in Generative Grammar, 28, 1986.

[Nivre 98] J. Nivre. Transcription Standard. Semantics and Spoken Language. Department of Linguistics, Göteborg University. Forthcoming.

[Nolan & Grabe 97] F. Nolan, E. Grabe. Can ToBI transcribe intonational variation in British English? Proc.ESCA Workshop on Intonation, Athens, 1997.

[O'Connor et al. 73] J. D. O'Connor, G. F. Arnold. Intonation of Colloquial English. Longman, 1973.

[Odé 97] C. Odé. The realization of types of tones in different positions of utterances in Mpur (West Papuam Phylym): an experimental phonetic analysis. ESCA Workshop on Intonation, Athens, 1997.

[Ostendorf et al. 95] M. Ostendorf, P. J. Price, S. Shattuc-Hufnagel. The Boston University Radio News Corpus. Technical Report ECS- 95-001. Boston University, 1995.

[Palmer 22] H. Palmer. English Intonation, with systematic exercises. Cambridge, Heffer, 1922.

[Partee 72] B. H. Partee. Opacity, coreference, and pronouns. In D. Davidson and G. Harman (editors). Semantics for Natural Language. Pages 415 – 441. D. Reidel, Dordrecht, Holland, 1972.

[Partee 73] B. H. Partee. Some structural analogies between tenses and pronouns in English. Journal of Philosophy, 70, 601—609, 1973.

[Passonneau 97] R. Passonneau. Instructions for applying discourse reference annotation for multiple applications (DRAMA). Unpublished manuscript, 1997.

[Payne 92] J. Payne. Report on the compatibility of J P Frenchís spoken corpus transcription conventions with the TEI guidelines for transcription of spoken texts. *Working Paper, COBUILD Birmingham and IDS Mannheim, December 1992*, NERC WP8/WP4 –122, 1992.

[Pierrehumbert 80] J. B. Pierrehumbert. The Phonology and Phonetics of English Intonation. Bloomington, Indiana University Linguistics Club, 1980.

[Pitrelli et al. 94] J. Pitrelli, M. Beckman, J. Hirschberg. Evaluation of prosodic transcription labelling reliability in the tobi framework. In: *Proceedings of the third International Conference on Spoken Language Processing*, Yokohama, ICSLP, Vol. 2, pp. 123-126, 1994.

[Poesio 94] M. Poesio. Discourse Interpretation and the Scope of Operators. Ph.D. thesis. University of Rochester, Department of Computer Science, Rochester, NY, 1994.

[Poesio et al. 98] M. Poesio and R. Vieira. A corpus-based investigation of definite description use. Computational Linguistics, 24(2), 183 - 216. To appear. Also available as Research Paper CCS-RP-71, Centre for Cognitive Science, 1998.

[Price 92] P. Price. Summary of the Second Prosodic Transcription Workshop: the TOBI (TOnes and Break Indices) Labeling System. Nynex Science and Technology, Inc. 5 - 6 April, 1992. Linguist List vol. 3-761, 9 October 1992.

[Prince 81] E. F. Prince. Toward a taxonomy of given-new information. In: P. Cole, editor, Radical Pragmatics, pages 223 - 256. Academic Press, New York, 1981.

[Prince 92] E. F. Prince. The ZPG letter: subjects, definiteness, and information status. In S. Thompson and W. Mann, editors, Discourse description: diverse analyses of a fund-raising text. Pages 295--325. John Benjamins, 1992.

[Quazza 91] Silvia Quazza. Modeling Italian Intonation on a Text-to-Speech System. Eurospeech 91. Genova, 1991.

[Reichman 81] R. Reichman. Plain-speaking: A theory and grammar of spontaneous discourse. Ph.D. thesis, Harvard University. Also appeared as BBN report 4681, 1981.

[Reyelt 93] M. Reyelt (Braunschweig University). Experimental investigation on the perceptual consistency and the automatic recognition on prosodic units in spoken German. ESCA Workshop on Prosody. Lund, 1993.

[Reyelt & Batliner 94] M. Reyelt and A. Batliner. Ein Inventar prosodischer Etiketten für VERBMOBIL", Verbmobil Memo 33, 1994.

[Roach 83] P. Roach. English Phonetics and Phonology. A practical course. Cambridge, Cambridge University Press, 1983.

[Roach 94] P. Roach. Conversion between prosodic transcription systems: ‘Standard British’ and ToBI. Speech communication 15, 91-99, 1994.

[Sampson 95] G. Sampson. English for the Computer. Oxford: Clarendon Press, 1995.

[Sanfilippo et al. 96] A. Sanfilippo, R. Barnett, N. Calzolari, S. Flores, P. Hellwig, P. Kahrel, G. Leech, M. Melero, S. Montemagni, J. Odijk, V. Pirrelli, S. Teufel, M. Villegas, L. Zaysser. Subcategorization Standards. Report of the EAGLES Lexicon/Syntax Group. SHARP Laboratories of Europe, Oxford (UK), 1996.

[Schegloff et al. 77] E. A. Schegloff, G. Jefferson and H. Sacks. The preference for self-correction in the organizations of repair in conversation. Language 53, 696-735, 1977.

[Searle 69] J. R. Searle. Speech Acts. Cambridge University Press, 1969.

[Selting 87] M. Selting. Descriptive categories for the auditive analysis of intonation in conversation. Journal of Pragmatics, 11, pp. 777-791, 1987.

[Selting 88] M. Selting. The role of intonation in the organization of repair and problem handling sequences. In: Conversation. *Journal of Pragmatics*, 12, pp. 293-322, 1988.

[Sidner 79] C. L. Sidner. Towards a computational theory of definite anaphora comprehension in English discourse. Ph.D. thesis, MIT, 1979.

[Silverman et al. 92] K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, J. Hirschberg. TOBI: A standard for labeling English prosody. In: OHALA, J.J. et al. (eds.).- *Proceedings ICSLP 92*, pp. 867-870, 1992.

[Sokolov et al. 94] J. Sokolov, C. Snow (Eds.). Handbook of research in language development using CHILDES. Erlbaum. Hillsdale, NJ, 1994.

[Sperberg 94] C. M. Sperberg-McQueen, L. Burnard. *Guidelines for Electronic Text Encoding and Interchange*. TEI P3. Chapter 11: Transcriptions of Speech. Association for Computational Linguistics / Association for Computers and the Humanities / Association for Literary and Linguistic Computing: Chicago and Oxford1994.

[Taylor 94] P. Taylor. The Rise/Fall/Connection Model of Intonation. Speech communication, 15, 1994.

[Taylor 97] P. Taylor. Analysis and Synthesis of Intonation using the Tilt Model. Draft Journal paper on Tilt model, 1997.

[Taylor & Knowles 88] L. J. Taylor, G. Knowles. Manual of Information to Accompany the SEC Corpus: the Machine Readable Corpus of Spoken English. Unit for Computer Research on the English Language, University of Lancaster.

[Taylor et al. 94] P. Taylor, A. Black. Synthesizing conversational intonation from a linguistically rich input. Proc. ESCA Workshop on Speech Synthesis. New Paltz, 1994.

[Ten Bosch 93] L. Ten Bosch: Algorithmic classification of pitch movement. ESCA Workhop on Prosody, Lund ,1993.

[Terken 93a] J. Terken. Human and synthetic intonation: A case study. In: Analysis and Synthesis of Speech. Editor: V. van Heuven. L. Pols, Berlin, 1993.

[Terken 93b] J. Terken. Synthesizing natural-sounding intonation for Dutch: rules and perceptual evaluation. Computer Speech and Language, 7,1, 1993.

[Vallduvi 90] E. Vallduvi. The Informational Component. Ph.D. thesis, University of Pennsylvania, Philadelphia, 1990.

[Vieira 98] R. Vieira. Definite Description Resolution in Unrestricted Texts. Ph.D. thesis, University of Edinburgh, Centre for Cognitive Science, 1998.

[Webber 79] B. L. Webber. A Formal Approach to Discourse Anaphora. Garland, New York, 1979.

[Webber 87] B. L. Webber. The interpretation of tense in discourse. In Proc. ACL-87, pages 147—154, 1987.

[Wells 92] J. Wells, W. Barry, M. Grice, A. Fourcin, D. Gibbon. Standard Computer-Compatible Transcription. SAM Stage Report Sen.3 SAM UCL-037, 28 February 1992. In SAM (1992) ESPRIT PROJECT 2589 (SAM) Multilingual Speech Input/Output Assessment, Methodology and Standardisation. *Final Report. Year Three*: 1.III.91-28.II.1992. Ref. SAM-UCL-G004. London: University College London, 1992.

[Willems et al. 88] N. J. Willems, R. Collier, J. t'Hart. A synthesis scheme for British English Intonation. Journal of the Acoustical Society of America 84, 1988.

Annexes

Contents of the Annexes

Communication Problems Schemes

A.1 Bethan L. Davies' Coding Scheme

Coding book:

Information is available in [Davies 1997: chapters 5 and 6 and appendices A and B].

Number of annotators:

Three annotators have used the scheme, 1 expert and 2 novices.

Number of dialogues:

32 HCRC Map-Task dialogues.

Evaluations of scheme:

Inter-coder and intra-coder reliability tests on a very small data set.

Intra-coder reliability test: Davies has tested her own coding after an interval of 18 months with a kappa value equal to one. Two naive coders were used to kappa-test the scheme on 32 HCRC Map-Task Corpus dialogues [Davies 1997, 207ff] with kappa values ranging from 0.69 to 1.0 The data set is however, as Davies remarks herself, insufficient for the testing of coder agreement for some of the markups.

Underlying task:

To empirically examine co-operation, effort and risk in task-oriented dialogues (ad-hoc annotation scheme [Davies 1997, 1]). The scheme was not meant to become a standard or to build on annotation standardisation initiatives.

List of phenomena annotated:

Dialogue acts graded according to effort and negative coding for the absence of felicitous or required dialogue acts [Davies 1997, 208ff] drawing on discourse theory such as especially the Birmingham School of Discourse Analysis (Davies 1997, 104ff).

INITIATE moves

+NEW-QUESTION, +NEW-SUGGESTION, +RELEVANT-INFO, +CHECK, + QUERY, +OBJECTION

RESPONSE moves

+ACK-SHORT, + REPLY-YN, + REPLY-FULL

FOLLOW-UP moves

+ACK-SHORT, +ACK-REP, +ACK-FULL

4. +INFO-INTEG (only in conjunction with either +ACK-FULL or +REPLY-FULL),

5. +KNOWLEDGE-MISMATCH (only in conjunction with either +QUERY or +OBJECTION

6. -REPLY-FULL (inadequate response move)

7. -ACK-FULL (inadequate follow-up move)

Examples:

[Davies 1997, 120]:

T= Turn, A = dialogue partner A, B = dialogue partner B

#TA 33

Can you move down, eh, three inches.

Could you move to your left, ah, approximately six inches.

#TB 34

Am I allowed to say if I'm going to go into one of these obstacles?

It's got right in the middle of the water.

Six inches.

#TA 35

Mmhmm.

#G +REPLY-MIN -REPLY-FULL

Markup Language:

Homegrown.

Existence of annotation tools (manual/automatic):

No annotation tools supporting this coding scheme have been developed.

Usability:

The coding scheme was only designed for the narrowly defined underlying task and analysed dialogues, cf. above.

Contact person:

Bethan L. Davies, HCRC Edinburgh University, UK.

A.2 CHAT

Coding book:

Information about the purpose and domain of the CHAT system as well as instructions for use are described in [MacWhinney 1994].

Number of annotators:

The CHAT system is a widespread standard system for the transcription and coding of child language in many European and non-European languages. Approximately 60 groups of researchers around the world are currently actively involved in new data collection and transcription using the CHAT system. As a consequence of its widespread use, it is impossible to calculate the exact number of annotators.

Number of annotated dialogues:

A huge number of dialogues has been/is being annotated with the CHAT coding scheme. This number exceeds the amount of dialogues in the database, as many projects concerning child language make use of CHAT without contributing to the overall CHILDES database. The internationally recognized CHILDES database (http://sunger2.uia.ac.be/childes/database.html) include transcripts from over forty major projects in English and additional data from 19 other languages. The additional languages are Brazilian Portuguese, Chinese (Mandarin), Chinese (Cantonese), Danish, Dutch, French, German, Greek, Hebrew, Hungarian, Italian, Japanese, Mambila, Polish, Russian, Spanish, Swedish, Tamil, Turkish, and Ukrainian. The total size of the database is now approximately 160 million characters (160 MB). Full documentation about the database can be found at http://sunger2.uia.ac.be/childes/database.pdf.

Evaluation of scheme:

As a result of its worldwide use, CHAT is continuously evaluated and updated to meet the needs of different languages and different users. We are not aware of statistical/quantitative evaluations of its reliability.

Underlying task:

Being first created as a tool for the study of language acquisition, the data collected mainly refer to parent-to-child or child-to-child spontaneous conversations, task-oriented dialogues in play and story-telling situations.

Some of the data coded by CHAT also include second language learners and adults recovering from aphasic disorders.

List of phenomena annotated (with examples):

1. speech-management-related phenomena

1.1 Speaker Overlap

Overlapping is annotated in CHAT using two complementary symbols, as the following example illustrates:

*MOT: no # Sarah # you have to <stop doing that [] !

*SAR: <Mommy I don't like this [<].

*SAR: it is nasty.

When several overlaps occur in a single sentence they are identified by numbers, as in the following example:

*SAR: and the <doggy was [1] really cute and it <had to go [2] into bed.

*MOT: <why don't you [<1] ?

*MOT: <maybe we could [<2] .

1.2 False starts

False Start without retracing

The symbols [/] and [//] are used when a false start is followed by a complete repetition or by a partial repetition with correction. If the speaker terminates an incomplete utterance and starts off on a totally new tangent, this can be coded with the [/-] symbol:

*BET: <I wanted [/-] uh when is Margie coming?

[Note that if this coding is not in contrast with the coding of incomplete utterances, this uniquely depends on the decisions about what a coder wants to count as an utterance] Retraced false starts can also be coded on the %err line.

1.3 Incomplete utterances and interruptions

Interruption

The symbol +/. is used for an utterance which is incomplete because one speaker is interrupted by another speaker.

Example:

*MOT: what did you +/.

*SAR: Mommy.

If the utterance that is interrupted is a question, then the symbol +/? is used.

Quick uptake

The symbol +^ is used when an utterance of one speaker follows quickly on the heels of the last utterance of the preceding speaker without the customary short pause between utterances. Example:

*MOT: why did you go?

*SAR: +^ I really didn't.

Other-completion

The symbol ++ can be used at the beginning to mark latching, or the completion of another speaker's. It is complementary to the trailing off symbol. Example:
*HEL: if Bill had known +...

*WIN: ++ he would have come.

1.4 Unintelligible/Unidentifiable material

A group of tags in CHAT refers to unidentifiable material for the purposes of transcription. These tags are put on the main line and allow to annotate:

A: unintelligible speech, either with or without reference to the possible phonological form. Example:

*SAR: xxx.

*MOT: what?

*SAR: I want xx.

B: material left untranscribed, either because the transcriber does not know how to transcribe it or because does not want to transcribe it.

For example, it could be that the material is in a language that the transcriber does not know, or that the speaker says something that has no relevance to the interaction taking place and the experimenter would rather ignore them. The symbol is used in conjunction with an %exp tier (explanatory).

Example:

*MOT: www.

%exp: talks to neighbor on the telephone

CHAT also allows to indicate (with an [?] on the main line) when a word or group of words are simply the transcriber's best guess at what was being said and there is some doubt in the transcriber's mind whether this guess is correct.

1.5 Pauses, hesitations, stuttering

Pauses are treated in CHAT under the prosodic annotation. The following phenomena are annotated:

Unfilled pauses

Pauses that are marked only by silence are coded on the main line with the symbol #. The number of # symbols represents the length of pauses. Alternatively, a word after the symbol # is added to estimate the pause length, as in #long. Example:

*SAR: I don't # know -.

*SAR: #long what do you ### think -?

CHAT also allows coding of exact length of the pauses, with minutes, seconds, and parts of seconds following the #. Example:

*SAR: I don't #0_5 know -.
*SAR: #1:13_41 what do you #2 think -?

Disfluent pauses

Fluent pauses occur at grammatical junctures where commas are generally used. They also occur at other sites that are determined by discourse rules. Pauses that occur elsewhere are typically considered disfluent, and are marked by the symbol #d. Example:

*CHI: well -.

*CHI: # how I felt about that -?

*CHI: I had to //put #d in my arms -.

*CHI: because I had to //put on a special coat -.

*MOT: we'll see -.

*MOT: #long maybe //tomorrow -.

*CHI: my brother does-'nt //sleep #dlong so much now -.

Hesitations and stuttering

CHAT uses a special symbol (&) to annotate phonological fragments, i.e. word partials and nonce/nonsense words frequently uttered in stuttering and false starts.

Use of this symbol signals that the following material is being transcribed in a correct phonological form in UNIBET and is not to be treated as a word by the analysis programs. For example, the utterance 't- t- c- can't you go? is transcribed as follows:

*MAR: &t &t &k can't you go?

The ampersand can be used for nonce and nonsense forms, as in:

*DAN: &glNk &glNk

%com: weird noises

Filled and unfilled pauses and word fragments due to hesitation or stuttering can also be coded on the %err line.

1.6, 1.7 Self-Interruptions, Restarts, Repetitions, Repairs

Retracing without correction (simple repetition)

When speakers repeat words or whole phrases without change. The retraced material is put in angle brackets.

Example: *BET: <I wanted [/] I wanted to invite Margie.

Several repetitions of the same word can be indicated in the following way:

*HAR: It's(/4) like # a um # dog.

Retracing with correction

When a speaker starts to say something, stops, repeats the basic phrase, changes the syntax but maintains the same idea. Usually, the correction moves closer to the standard form, but sometimes it moves away form it. . The retraced material is put in angle brackets. Example:

*BET: <I wanted [//] uh I thought I wanted to invite Margie.

Retracing with Reformulation

When retracings involve full and complete reformulations of the message without any specific corrections. Example: *BET: all of my friends had [///] uh we had decided to go home for lunch.

Unclear Retracing Type [/?]

Trailing Off

An incomplete, but not interrupted, utterance, is marked with the 'trailing off' symbol. This phenomenon is said to occur when speakers shift attention away from what they are saying, sometimes even forgetting what they were going to say. Usually the trailing off is followed by a pause in the conversation. After this lull, the speaker may continue with another utterance or a new speaker may produce the next utterance. Example:

*SAR: smells good enough for +Ö *SAR: what is that?

If the utterance that is being trailed off is a question, then the symbol +..? is used.

Self-Interruption

The symbol +//. is used to distinguish an actual self-interruption from an incompletion that involves a trailing off. Example:

*SAR: smells good enough for +//. *SAR: what is that?

This category may actually overlap with the trailing off; transcribers are advised to use the trailing off symbol +... whenever in doubt. If the utterance that is self-interrupted is a question, then the symbol +//? can be used.

Self-completion

The symbol +, can be used at the beginning of a main tier line to mark the completion of an utterance after an interruption. It is complementary to the trailing off symbol. Example:

*CHI: so after the tower +... *EXP: yeah. *CHI: +, I go straight ahead.

1.8 Errors

CHAT provides a system for the detailed coding of errors. It is worth noting that CHAT does not always force the user to comply with any particular theoretical assumption. The same phenomena are sometimes provided with two different codings, depending on whether the coder wants to consider them as errors or not (cf. infra). Errors are coded by CHAT on a dependent tier (the %err line); however, all errors that are to be coded on the %err line are also coded with the [*] symbol on the main line, right after the error. This symbol can be given multiword scope.

The CHAT system for error coding has the following features:

1. it indicates what the speaker actually said, or the erroneous form
2. it indicates that what the speaker actually said was an error
3. it allows the transcriber to indicate the target form
4. it facilitates retrieval, both toward target forms and actually produced forms
5. it allows the analyst to indicate theoretically interesting aspects of the error by delineating the source of the error, the processes involved, and the type of the error in theoretical terms.

An error, thus, can be described along the following dimensions:

a) the actual form produced and its target form;

b) the possible source that may have originated the error (introduced by the symbol $=)

c) the type of the error, in theoretical terms.

Error codes involve two levels of analysis, hierarchically organized. The first level distinguishes between errors occurring at the phonological, morphological, lexical, etc. level. The second level further specifies the error type, whether the error involves addition, loss, substitution, anticipation, etc. A finer taxonomy is given for the phonological, morphological and syntactic error types.

General Codes:

$PHO error involving specific phonological units (gutter = butter)

$MOR omissions, additions, and substitutions of closed- class items (jump = jumped)

$LEX choice of the wrong word on a semantic basis (coat = sweater)

$SYN syntactic error, accommodation, stranding, etc.

$ALL morphophonological errors in allosegments or allomorphs (guve = gave)

$CWFA complex word finding attempts as in Wernicke's aphasia (a binny, a figgy, a fig, no an eagey)

$MAL malapropism (mix of phonological and lexical sources) (croutons = coupons)

$INT intonational error usually detected during a retraced false start

$NW nonword with an unknown or unclear basis griff

$CIR circumlocution, as in Wernicke's aphasia (do with car = drive)

Type Codes :

$ADD addition (blunch = bunch)

$LOS loss (garet = garnet)

$SUB substitution (batter = tatter)

$SYN syntactic error, accommodation, stranding, etc.

$HAP haplology (Sancisco = San Francisco)

$BLE a blend (flaste = flavor + taste)

$EX1 first part of an exchange error (broudy klight = cloudy bright)

$EX2 second part of an exchange error

$SH1 first part of a shift error

$SH2 second part of a shift error people different = different people

$ANT anticipation: an item is produced early and also where it belongs (bould be = would be)

$PER perseveration: an item is produced late and also where it belongs (would we = would be)

$A/P both anticipation and perseveration are possible (thingle = single thing)

$ACH a chain in which A anticipates B which anticipates C

$PCH a chain in which C perseverate B which perseverates A

$INC the production is incomplete

$UNC unclear error type

Errors of $PHO type are further characterizable as follows:

$VOW error involving a vowel or diphthong (bonny = bunny)

$CON error involving consonants (munny = bunny)

$CC error involving consonant clusters (tickle = trickle)

$SYL error where the target or source are syllables (perfacial performance = spatial performance)

$FEA error involving particular features (munny = bunny)

$STS error involving stress (capiTULate = caPITulate)

$MRA moraic error in Japanese and other mora languages

$TON error involving tone

Errors of $MOR type are further characterizable as follows:

$PRE error involving prefix (misforgiving = unforgiving)

$SUF error involving suffix (taked = taken)

$NFX error involving infix

$NFL error involving inflection (taked = taken)

$DER error involving derivational processes (taked = taken)

$RED error involving morphological reduplication (sevenses = sevens)

$AGA error of agreement where agreer is wrong (el palma = la palma)

$AGC error of agreement where controller is wrong (la palmo = la palma)

$AGB error with both agreer and controller wrong (el palmo = la palma)

$REG regularization (eated = ate)

$FUL full regularization (throwed = threw)

$PAR partial regularization (threwed = threw)

$HAR vowel harmony error (ablakek = ablakok)

Errors of $SYN type are further characterizable as follows:

$ACC error where accommodation was present (a apper = an ëAí)

$STR error where affixes are stranded the flood (was roaded = the road was floaded)

$SBL syntactic blend (thingle = single thing)

$POS position error (gave him it = gave it him)

Note that this implies, theoretically, that such phenomena are considered as errors.

2. Paralinguistic features

paralinguistic events, such as coughing, laughing, crying, etc. are marked in CHAT by special symbols on the main line. Example:

*CHI: that's mine [=! cries]

if the child cries only while uttering the word 'mine' or

*CHI: < that's mine [=! cries]

if the entire utterance is made while crying.

Paralinguistic events can also be marked on a separate line, by using the %par line.

Turn-inclusion: CHAT uses a special symbol [+trn] to indicate that an utterance must be treated as a turn, even if it would normally not be treated as such. For example, utterances containing no verbal actions are usually not treated as turns. However, if the annotator believes that the accompanying nonverbal gesture constitutes a turn, he/she has the possibility to note this using [+ trn], as in the example:

*MOT: where is it?

*CHI: 0. [+trn]

%act: points at wall.

3. Kinesic features

not annotated

4. Situational Features

CHAT annotates the following phenomena:

4.1 Back Channel Utterances

In the following example, the mother makes a remark to the investigator. Use of this symbol indicates that the utterance is to be excluded from analysis.

*CHI: here one.

*MOT: no -, here.

%sit: the doorbell rings.

*MOT: just a moment. [+bch]

*MOT: I'll get it. [+bch]

4.2 Actions without speech

A symbol on the main line is used when the speaker performs some action that is not accompanied by speech. Example:

*FAT: wher's your doll?

*DAV: 0 [=! cries].

Situational features are also annotated on a separate special dependent tier %sit.

5. Word-form related phenomena

5.1 Assimilations

For assimilated forms, such as 'gonna' for 'going to' and 'whynt cha' for 'why don't you', CHAT allows the transcriber to place the assimilated form on the main line followed by a fuller form in square brackets, as for example:

gonna [: going to]

5.2 Non-standard Forms

Annotation is provided in CHAT for many categories of non-standard forms. These include task-specific forms such as babbling or child-invented forms, but also more general categories such as dialectal forms or neologisms. Categories for which a special coding is provided are the following: babbling, child-invented forms, dialect forms, family-specific forms, filled pauses, interjections, neologisms, phrasal repetitions, or other general special forms. Other non-standard forms can be coded placing the all-purpose symbol @ after the form. In addition, standardized transcription conventions are provided.

Note that recording of these phenomena is not made at the coding level, but at the transcription level.

Examples:

Coding	Description	Example	Meaning	Coded Example
@b	babbling	abame	-	abame@b
@c	child-invented form	gumma	sticky	gumma@c
@d	dialect form	younz	you	younz@d
@f	family-specific form	bunko	broken	bunko@f
@fp	filled pause	huh	-	huh@fp
@i	interjection, interactional	uhhuh	-	uhhuh@i
@l	letter	b	letter b	b@l
@n	neologism	breaked	broke	breaked@n
@o	onomatopoeia	woof woof	dog barking	woof@o
@p	phonol. consistent forms	aga	-	aga@p
@pr	phrasal repetition	its a, it's a	-	its+a@pr
@s	second-language form	istenem	my God	istenem@s
@sl	sign language	apple sign	apple	apple@sl

5.3. Word-partials

Non-completion of a word

A symbol is used to signal the missing parts of words that are incomplete, but whose meaning still seems clear. Example:

*RAL: I been sit(ting) all day.

Shortenings

The same symbol is used to annotate shortened forms, such as 'bout' for 'about'. CHAT also provides a list of the most common shortened words in English.

6. Syntax related phenomena

CHAT annotates the following phenomena:

6.1. Discourse markers, pragmatic particles

Interactional Markers

This category of interjection has the specific function of signaling agreement, disagreement, and pauses. CHAT gives a list of standardized spelling for the most common ones.

6.2. Syntactic incompleteness

Omission of a word

To code an omission, a zero symbol is placed before a word on the text tier. If what is important is not the actual word omitted, but its part of speech, then a code for the part of speech can follow the zero. The best guess is placed on the main line. Example:

*EVE: I want 0to go.

In addition, CHAT annotates the following phenomena:

Incorrect Omission

A symbol 0*word is used when the omission is clearly ungrammatical and the transcriber wishes to code that fact.

Ellipsis

A special symbol is used when the omission of a word is licensed by the standard grammatical and discourse pattern of the language. Example:

*FAT: where did you go?

*ABE: 00sub 00verb 0*prep the store

Alternatively, omissions are marked by using the forms [0 text], [0* text] and [00 text].

7. Communication Problems at the Speech Act Level

see the description of speech act categories in the dialogue act level report.

Mark-up language:

CHAT's own format.

Existence of annotation tools:

The CHILDES system contains several separate, yet integrate, programs which are clustered around two major tools. The first tool is a full-fledged and ASCII-oriented editor (CED, Childes EDitor), specifically designed to facilitate the editing of CHAT files and to check for accuracy of transcriptions. CED also allows the user to link a full digitized audio recording of the interaction directly to the transcript. This is the system called "sonic CHAT". The CED editor is currently being extended to facilitate its use with videotapes. The plan is to make available a floating window in the shape of a VCR controller that can be used to rewind the videotape and to enter time stamps from the videotape into the CHAT file. An alternative way

of analyzing video is to record from tape onto QuickTime movies and to link these digitized movies to the transcript.

The second tool, actually a bunch of several smaller tools, is a set of computer programs called CLAN (Child Language ANalysis) which serves different analysis purposes. The full system is presented in detail in MacWhinney (1991) and illustrated through practical examples in Sokolov and Snow (1994).

Usability:

CHAT-encoded databases have been set up as a result of nearly a hundred major research projects in 20 languages. New databases are continuously being set up worldwide.

Contact person:

Brian MacWhinney (macw@cmu.edu)

A.3 Odense University Scheme

Coding book:

Information is available in [Bernsen et al. 1998 Chapters 4 and 8, Bernsen et al. 1997, Dybkjær et al. 1997, and http://www.mip.ou.dk/~laila/dialogue/index.html/]

Number of annotators:

Three annotators have used the scheme, two experts and one novice.

Number of dialogues:

57 Danish Dialogue System user test dialogues (Danish, 2 expert annotators) + 48 Sundial WOZ dialogues (English, 2 expert annotators and one novice annotator) + 27 Philips field test dialogues (German, 1 expert annotator so far).

Evaluations of scheme:

The scheme is currently being evaluated in the DISC project through use on various task-oriented shared-goal human-machine dialogues. For results see e.g. [Bernsen et al. 1997, Dybkjær et al. 1997, Bernsen et al. 1998]

Underlying task:

The Danish Dialogue System dialogues are about flight ticket reservation. The Sundial dialogues are about flight information. The Philips dialogues are about train timetable information. The task domain for which the cooperativity guidelines are assumed to work is task-oriented, shared-goal human-machine dialogues.

List of phenomena annotated:

Any dialogue interaction problem due to dialogue design errors or user errors. User errors were mainly looked at in the corpus from the Danish Dialogue System. A taxonomy of cooperativity problems due to flawed dialogue design, and user errors have been established, cf. [Bernsen et al. 1998]:

COOPERATIVITY PROBLEM	
Aspect	**Generic or specific guideline**
Group 1:	GG1: System provides less information than required.
Informativeness	SG1: System is not fully explicit in communicating to users the commitments they have made.
	SG2: Missing system feedback on user information.
	GG2: System provides more information than required.
Group 2:	GG3: System provides false information.
Truth and evidence	GG4: System provides information for which it lacks evidence.
Group 3: **Relevance**	GG5: System provides irrelevant information.
Group 4:	GG6: Obscure system utterance.
Manner	GG7: Ambiguous system utterance.
	SG3: System does not provide same formulation of the same question to users everywhere in its dialogue turns.
	GG8: Too lengthy expressions provided by system.
	GG9: System provides disorderly discourse.
Group 5: **Partner asymmetry**	GG10: System does not inform users of important non-normal characteristics which they should, and are able to, take into account to behave cooperatively in dialogue.
	SG4: Missing or unclear information on what the system can and cannot do.
	SG5: Missing or unclear instructions on how to interact with the system.
Group 6: **Background**	GG11: System does not take user relevant background knowledge into account.
knowledge	SG6: Lacking anticipation of domain misunderstanding by analogy.
	SG7: System does not separate when possible between the needs of novice and expert users.
	GG12: System does not consider legitimate user expectations as to its own background knowledge.
	SG8: Missing system domain knowledge and inference.
Group 7: **Repair and**	GG13: System does not initiate repair or clarification meta-communication in case of communication failure.
clarification	SG9: System does not initiate clarification if it has failed to understand the user.
	SG10: Missing clarification of inconsistent user input.
	SG11: Missing clarification of ambiguous user input.

User Error Types	Error Sub-Types
Misunderstanding of scenario.	Careless reading or processing.
Ignoring clear system feedback.	Straight ignorance.
Responding to a question different from the clear system question.	Straight wrong response.
Answering several questions at a time.	Slip.
Thinking aloud.	Natural thinking aloud.
Non-cooperativity.	Unnecessary complexity.

Examples:

<u id="U1:9-1”>

#h em I'm enquiring about bee ay nine oh three the flight from frankfurt to heathrow coming in tomorrow morning could you tell me what time it arrives, at heathrow (4)

<u id="S1:9-2”>

please wait (9) That flight is scheduled for twelve noon (1.2)

<violation ref=”S1:9-2” guideline=”SG2”>

Missing feedback. It is impossible to see if the system has understood the user correctly and is talking about the flight mentioned by the user. The system does not allow the user to verify that the query has been correctly understood (”that flight”).

<violation ref=”S1:9-2” guideline=”GG6”>

The answer is obscure. Is ”scheduled” only used about flights already departed? In that case the system does not provide the desired information.

For an explanation and walk-through of a full dialogue, see e.g.
http://www.mip.ou.dk/~laila/dialogue/index.html/walk.htm#Sundial

Markup language:

For the Danish Dialogue System dialogues a very home-grown way of marking up communication problems was used. For the Sundial and the Philips dialogues TEI was used to markup system and user utterances and was extended with the two new tags ”violation” and ”usererror”.

Existence of annotation tools (manual/automatic):

There are no annotation tools available for the markup of dialogue interaction problems. All markup of communication problems was inserted manually.

Usability:

The set of cooperativity problems listed above has been sufficient to account for the dialogue design problems we have observed in the corpora analysed, i.e. Danish Dialogue System dialogues, Sundial dialogues and Philips dialogues.

Contact:

Laila Dybkjær,

The Maersk Institute, Odense University, Forskerparken 10, 5230 Odense M., Denmark.

Phone: +45 65 57 35 53, Fax: +45 63 15 72 24, Email: laila@mip.ou.dk

A.4 SWBD-DAMSL

Coding book:

information can be found in [Meeter et al. 1995]

Number of annotators:

not specified, but more than one/two.

Number of dialogues/material annotated:

SWITCHBOARD is a collection of about 2400 two-sided telephone conversations among 543 speakers (302 male, 241 female) from all areas of the United States. A computer-driven "robot operator" system handled the calls, giving the caller appropriate recorded prompts, selecting and dialing another person (the callee) to take part in a conversation, introducing a topic for discussion, and recording the speech from the two subjects into separate channels until the conversation was finished. About 70 topics were provided, of which about 50 were used frequently. Selection of topics and callees was constrained so that: (1) no two speakers would converse together more than once, and (2) no one spoke more than once on a given topic.

Evaluation of scheme:

missing information

Underlying task:

spontaneous telephone conversations surrounding prompted topics

List of phenomena annotated (with examples):

1. speech-management related phenomena

1.1 Speaker Overlap

Speaker overlap is annotated in the transcript using the #...# symbol

1.2 False starts

see retracing

1.3 Incomplete utterances and interruptions

When an utterance is left incomplete, without continuation even in the following turns, the symbol -/ is used. Completions by other speakers are ignored, and an utterance completed this way is still marked as incomplete. Non-speech sound that stands alone as turns are ignored. A dangling filler taking up a turn on its own counts as a slash-unit, but dangling coordinating conjunctions and discourse markers for which there is no continuation in the next turn are counted as incomplete slash-units, except for you know which is a special case.

1.4 Unintelligible/Unidentifiable material

?

1.5 Pauses, hesitations, stuttering

Filled pauses {F...}

Switchboard allows for coding of filled pauses. Each filler is enclosed separately in its own brackets and preceded by the symbol F. Specific criteria are provided for deciding when an item counts as a filler in a filled pause as opposed to when it does not.

The Switchboard coding contemplates the following phenomena:

Simple restarts:

- restarts without repair or simple repetition. Example:

but [I, + I] do know some of these places were doing that

- restarts with repair. These category is further broken down into:

a) repetition and insertion (at least an item is repeated on the right side, plus something new is inserted or added). Example:

people would [dr-, + have a drink] of soda

b) repetition and deletion (at least one item is repeated on the right side, plus something is deleted). Example:

[I just, + I] noticed it

c) repetition and substitution (at least one item is repeated on the right side, plus there is a substitution of another element on the left, by a different one on the right). Example:

[why didn't he, + why didn't she] keep him home with her?

Substitutions:

Substitution plus repetition was described in the previous section. This category concerns substitution without repetition, which is only possible if the left side contains a single word. something, however, must carry over from the left to the right. In the case of single words this will most often be the part of speech, as illustrated by the following examples:

Article: I gave her [a + the] dog

Pronoun: [he + she] reported it

Preposition: i put it [in + on] the table

Adjective: a [blue + green] block

Noun: a blue [triangle + square]

Verb: I [liked + loved] the opera

Complex restarts: these are those with more than one interruption point (+). there are two main types of complex restarts: chaining and nesting.

Chaining: in chaining, something from each part carries over into the next one. This category is further subdivided as the preceding Repetition category:

- simple repetition: [[I + I] + I] don't want it
- - simple repetition with insertion: [[the + the boy] + the big boy]
- - simple repetition with deletion: [[I really + I] + I] like it
- - single words can be repeated or substituted (following the rules under 'Simple
- restarts'): [[he + she] + she]
- - multi-words complex restarts: [[the flights + Delta flights] + Delta fares]

Nesting: when you have a restart within a restart. Example:

[I liked, + {F uh, } [I, + I] liked] it a lot

1.7 Repairs, self corrections

These phenomena are partially treated in Switchboard under the preceding category. In addition, the following phenomena are considered:

Asides: asides are non-sentence elements that interrupt the flow of the sentence and are most commonly found in repairs. They are coded with the symbol {A...}. Example:

-- I think those satellites, {D you know, } {A or not satellites, but the spaceflights } could really spy /

Explicit editing terms: these are elements that usually occur between the restart and the repair. They are coded with the {E...} symbol. Example:

[We've had kid, + {E or } we've had dogs] when I was a kid,

1.8 Errors

not annotated

2. Paralinguistic features

Paralinguistic events, such as laughter, noise, breathing, lipsmack etc. are coded in angle brackets.

3. Kinesic Features

not annotated

4. Situational Features

Switchboard annotates situational features such as TV or non-verbal actions by using the same symbol as for paralinguistic events.

5. Word-form related phenomena

Switchboard only allows for coding of fragmented or incomplete words (word partials) by use of the symbol -, as in Ut- or you kn-.

6. Syntax related phenomena

Switchboard annotates the following word classes: fillers {F...}, discourse markers {D...}, and coordinating conjunctions {C...}.

7. Communication Problems at the Speech Act Level

see the description of the coding scheme for dialogue acts.

Mark-up language:

missing information

Existence of annotation tools:

None, annotation was done manually

Usability:

Switchboard is designed to support telephone-based speech technology development as well as basic research on spontaneous conversational speech and language.

Contact person:

Linguistic Data Consortium at UPenn (ldc@ldc.upenn.edu).

Coreference Schemes

A.5 Bruneseaux and Romary

Contact:

Florence Bruneseaux (brunesea@loria.fr)

Coding book:

No - just papers describing scheme, rather than anything which could serve as instructions to annotators.

Number of annotators:

10 non-naive annotators (Linguistics students)

Number of dialogues:

(It is not clear whether all of these have been fully annotated)

32 dialogues from the GOCAD Corpus (about 100 lines each)

3 trialogues + 1 quad from the Microfusées corpus (about 400

turns each)

34 dialogues (phone calls) from the CIO Corpus (about 190 pages of

transcription on paper)

169 dialogues (phone calls) from the SNCF Corpus

Magnetoz : A wizard of Oz experiment.

Evaluations of scheme:

None

Underlying task:

Designed to represent the range of referring expressions found in task-based human-machine dialogue, including the types of deictics accompanied by mouse clicks or other gestures.

Description of the scheme:

Like MUC, DRAMA and UCREL, this scheme identifies the referring expressions in the text and then marks up the relationships between them. Like MUC, only those strings which enter into some kind of relationships with other strings or with items in the visual context are marked up. With respect to the conventions used, this scheme is different to the others in that the relationships between strings are represented in a separate tag from the strings and their IDs - as a 'link' statement between 2 IDs. An important point about this scheme is that it encodes references to the visual context (many of the conversations are taken from Human-Computer interactions using a geological simulation program). This is done by listing all the items in the 'universe', giving them IDs, and then linking references to these items to their ID. Pointing and mouse-click gestures are also marked. This scheme has a large set of markables, including verbal structures.

Relationships:

Coreference

Partitive relations

Possessive relations

Bridging reference

'Designation' (link between a mouse-click and vocal designation e.g. 'this one')

Markables:

NPs

'actions' (e.g. VPs, clauses, or a whole sentence like 'zoom in')

Pointing gestures

Objects, people and actions in the current universe

Examples:

<u id="u11" who="S2">Create <rs type="object" id="O1">the surface.</rs> </u>

<u id="u12" who="C">Done. </u>

<u id="u17" who="S2">Model <rs type="object" id="O2">the surface.</rs> </u>

<link type="coref" targets="O2 O1">

<u id="u22" who="C">Which name do you want to give to <rs type="object" id="O3">the surface</rs> ? </u>

<link type="coref" targets="O3 O2">

<s> Paul likes <rs type="object" id="O1">his car</rs> : <rs type="object" id="O2">the seats</rs> are confortable. </s>

<link type="partitif" relation="internal" target="O2 O1" targetOrder="y">

<rs type="person" id="O1">Mary</rs> 's <rs type="object" id="O2">car</rs>.

<link type="possession" relation="internal" target="O2 O1" targetOrder="y">

<u id="u145" who="C"><rs type="object" id="O1">Surfaces S1 and S2</rs> are not visualized on the screen.</u>

<u id="u146" who="U7">Visualize <rs type="object" id="O2">them</rs>.</u>

<link type="coref" relation="internal" target="O2 O1" targetOrder="y">

We came into <rs type="object" id="O1">a village</rs>.

<rs type="object" id="O2">The church</rs> was on a hill.

<link type="association" relation="internal" target="O2 O1" targetOrder="y">

<u id="u41" who="S4">Remove <rs type="object" id="O1>this part</rs> <anchor id="a1" synchro="K1">

<kinesic id="K1" desc="click on the dome"></kinesic>

</u>

<link type="designation" relation="external" target="O1 K1">

<u id="u42" who="C">Message received.</u>

<universe type="object" content="O1 O2 O3" id="U1"> </universe>

<u id="u33" who="S5">Densify the grid</u>

<u id="u34" who="C">Okay. <rs type="object" id="Ox">Which surface</rs> do you want to densify ?

</u>

<u id="u35" who="S5"><rs type="object" id="Oy">The 3</rs></u>

<link type="TargetUniverse" targets="Oy, U1 ">

Markup language:

SGML

Existence of annotation tools (manual/automatic):

No

Usability:

Not implemented in any other system.

A.6 DRAMA - Discourse Reference Annotation for Multiple Applications

Contact:

No contact available: Passonneau no longer works in NLP.

Coding book:

Passonneau, RJ (1996) Instructions for Applying Discourse Reference Annotation for Multiple Applications (DRAMA)

Number of annotators:

6+ (all linguists)

Number of dialogues:

3+ (all in English - total 943 words)

Evaluations of scheme:

DRI subgroup at Dagstuhl (1997): 5-6 people annotated 1 TRAINS dialogue (221 words), 1 Coconut dialogue (311 words), and a spoken narrative from Chafe's 'Pear Stories' (411 words); the first two were annotated solely for identity of coreference, and the last for other relationships among referents (e.g. set/subset, set/member, part/whole). Scores were calculated for the precision (avoidance of spurious relations) and recall (inclusion of real relations) of each annotation, compared to an answer key. For the TRAINS dialogue, recall ranged from .55 to > .80, and precision clustered at > .80. For the Coconut dialogue, recall ranged from .2 to > .7, and precision from .6 to > .7. For the Pear Stories, a score was calculated for each of the relationships (see below for relationships):

genitive/possessive: recall .47, precision 1.0

member: recall .57, precision .65

subset: recall .48, precision .84

Underlying task:

To feed into NLP work on language understanding

Description of the scheme:

This scheme, like MUC, is based on identifying referential expressions in the text, and then marking the relationships between these. The main difference between this scheme and MUC is that DRAMA includes many types of 'bridging' relationships, rather than just coreference. These different types of bridging relationships are all classified in the markup. The set of markables is also greater in DRAMA than it is in MUC, and includes any clauses, verb phrases or adjectives which are deemed the antecedents of expressions like 'it' and 'that'. Unlike MUC, in DRAMA all strings which show the syntagmatic behaviour of noun phrases and introduce a discourse referent are identified, whether or not they enter into coreference relationships.

Relationships:

Identity of reference AND the following bridging inferences are annotated:

set/subset/member relationships (the pears... three pears)

part/whole and physical connection (a house... the kitchen)

causal inference (an explosion... the noise)

propositional inference (It's sunny... that's a relief)

genitive/possessive pronouns (the boy... his hair)

implicit arguments (Look at the car! The wheel just fell off_)

implicit partitives and pseudo-partitives (Here's one screw. Where's the other one?)

set relationships in plural NPs (The boy does one thing and the girl does another... then they both start crying)

(apposition relations are NOT marked)

Markables:

nouns

noun phrases

pronouns (possessive pronouns are only marked as inference, not identity)

phrases with adj. as head (with noun implied) - 'the largest', 'poorest of the seven'

phrases with quantifiers as head (with noun implied) - 'a great many'

wh-phrases

gerundive clauses (if they occur in the same structural position as could be occupied by an NP)

non-restrictive relative pronouns

phonologically reduced phrases (if they otherwise fit the criteria)

zero pronouns

subjects of imperatives

repairs, repetitions, phrase fragments

physical or vocal gestures and deictic adverbs

Non-markables:

negated noun phrases

non-referential/pleonastic NPs (e.g. _It's_ easy to see what happened here)

Examples:

<REFEXP ID="100" REFINDEX="49">IBM</REFEXP> announced

<REFEXP ID="101" REFINDEX="50"> a dividend.

<REFEXP ID="102" REFINDEX="49"> It </REFEXP> showed

<REFEXP ID="103" REFINDEX="51"> a hefty fourth quarter profit. </REFEXP>

Markup language:

SGML (?)

Existence of annotation tools manual/automatic):

A Discourse Tagging Tool (DTT) is available for this scheme, which annotators would use with a text-based crib sheet.

Usability:

Information not available.

A.7 MUC-7 Coreference Task

Contact:

Lynette Hirschman (lynette@mitre.org) or Patricia Robinson (parann@mitre.org)

Coding book:

Hirschman (1997) MUC-7 Coreference Task Definition, Version 3.0. in Proc. MUC-7

Number of annotators:

15+ Human annotators (linguists) and many automatic annotations done by text understanding systems.

Number of dialogues:

100 texts/dialogues for automatic annotation, plus ~10 dialogues for evaluation.

Evaluations of scheme:

MUC-7 task: text understanding systems have to annotate the text without pre-processing or human intervention, and their output is compared with that of human annotators

Evaluations of older versions of the scheme:

DRI subgroup at U.Penn (1996): 10 people annotated 4 texts:

1 from MUC corpus (newswire)

2 from human-human dialogue (map, trains)

1 from narrative monologue (directions)

Interannotator agreement = 70-80%. Agreement between annotators and key: more variation.

DRI subgroup at Dagstuhl (1997): 5 people annotated a Wall Street Journal article (176 words) and an ABC news item (403 words). Scores were calculated for the precision (avoidance of spurious relations) and recall (inclusion of real relations) of each annotation. Recall ranged from .54 to .85, clustering at > .80, and precision ranged from .80 to .95.

Underlying task:

The annotation scheme's primary purpose is to support the MUC information extraction tasks - i.e. to provide a tractable scheme for machine annotation and to support the comparison of machine- and human-annotated materials. The coreference scheme also directly provides input to one of the other MUC tasks - the template element task. The task definition outlines its criteria in order of importance as follows:

1) Support for the MUC information extraction tasks

2) Ability to achieve good (ca. 95%) interannotator agreement

3) Ability to mark text up quickly (and therefore cheaply)

4) Desire to create a corpus for research on coreference and discourse phenomena, independent of the MUC extraction task.

They state that it is more important to preserve high inter-annotator agreement than to capture every possible phenomenon that could fall under the heading of 'coreference'.

Description of the scheme:

The MUC coreference annotation scheme basically marks up all the strings in the text (see 'markables' below) which enter into coreference relationships, assigning each a unique ID, and indicating the relationship between two strings by marking the later string with a REF pointer to its antecedent. This could therefore be considered a scheme which links items in the text to each

other. (Note that only the strings which enter into coreference relationships are marked up, NOT all referring strings.)

Relationships:

The annotation scheme covers only the 'IDENTITY' (or IDENT) relationship for noun phrases, (including apposition). It does NOT include coreference among clauses, nor does it cover other kinds of coreference relations (set/subset, part/whole, etc.)

So, the markables are nouns, noun phrases, and pronouns (see below for fuller list). Each markable is assigned an ID, and, if it corefers with another markable, a REF pointer to the antecedent. Each REF may be designated optional (OPT) if the reader may not be certain that the two expressions corefer. If applicable, a substring of each markable is designated as MIN - the minimum string to which the coreference could be considered to refer (usually the head of the phrase).

Markables:

nouns

noun phrases

pronouns (inc. possessives)

Non-markables:

wh-phrases

clauses

substrings of named entities (e.g. 'Iowa' in 'Equitable of Iowa Co.')

gerunds (unless modified by other nouns or adjs., preceded by an article, or followed by an 'of' phrase

nouns only used as modifiers throughout a coreference chain

relative pronouns

empty string (i.e. zero or implicit pronouns)

disfluencies which are 'verbally deleted'

Examples:

Transcript of a TRAINS Dialogue:

utt1 : s: hello can I help <COREF ID="1">you</COREF>

utt2 : u: yes <COREF ID="0" TYPE="IDENT" REF="1">I</COREF>'d like <sil> to <sil> take <COREF ID="3">a tanker</COREF> from <COREF ID="12">Corning</COREF> and bring <COREF ID="2" TYPE="IDENT" REF="3">it</COREF> to <COREF ID="5">Elmira</COREF>

utt3 : s: alright

utt4 : u: and from <COREF ID="4" TYPE="IDENT" REF="5">Elmira</COREF> <COREF ID="6" TYPE="IDENT" REF="0">I</COREF>'d like to load <COREF ID="17" MIN="juice">orange juice</COREF> <sil> into <COREF ID="7" TYPE="IDENT" REF="2">the tanker</COREF>

utt5 : s: mm-hm

utt6 : u: <COREF ID="8" TYPE="IDENT" REF="6">I</COREF>'d like then to take the el- <sil> <COREF ID="9" TYPE="IDENT" REF="7">the tanker</COREF> back from <COREF ID="10" TYPE="IDENT" REF="4">Elmira</ COREF> to <COREF ID="11" TYPE="IDENT" REF="12">Corning</COREF>

utt7 : s: alright

utt8 : u: now from <COREF ID="13" TYPE="IDENT" REF="11">Corning</COREF> what would be the quickest route to <COREF ID="23">Avon</COREF>

utt9 : s: uh through <COREF ID="21">Dansville</COREF>

utt10 : u: okay then <COREF ID="14" TYPE="IDENT" REF="8">I</COREF>'d like to take <sil> <COREF ID="15" TYPE="IDENT" REF="9" MIN="tanker">the <sil> tanker of <COREF ID="16" TYPE="IDENT" REF="17" MIN="juice">orange juice</ COREF></COREF> through <COREF ID="20" TYPE="IDENT" REF="21">Dansville</ COREF> and then on to <COREF ID="22" TYPE="IDENT" REF="23">Avon</COREF>

utt11 : s: alright <sil> um which engine would <COREF ID="24" TYPE="IDENT" REF="14">you</COREF> like to use

utt12 : u: <COREF ID="26" MIN="E two">engine <sil> E two</COREF>

utt13 : s: alright

utt14 : u: well is is <COREF ID="25" TYPE="IDENT" REF="26">E two</COREF> which is a <sil> will either one take <COREF ID="27" TYPE="IDENT" REF="24">me</COREF> there any quicker

utt15 : s: uh no <COREF ID="101">they</COREF>'re a <COREF ID="100" TYPE="IDENT" REF="101">they</COREF>'re both the same

utt16 : u: okay then <COREF ID="28" TYPE="IDENT" REF="27">I</COREF>'ll take <COREF ID="29" TYPE="IDENT" REF="25" MIN="E two">engine E two</COREF>

utt17 : s: alright

utt18 : u: and the tankers move at the same speed

utt19 : s: right

utt20 : u: okay

utt21 : s: so uh the plan is to take <brth> <sil> <COREF ID="30" TYPE="IDENT" REF="29" MIN="E two">engine E two</COREF> <sil> to <COREF ID="31" TYPE="IDENT" REF="13">Corning</COREF> pick up <COREF ID="32" TYPE="IDENT" REF="7">a tanker</COREF> <sil> back to <COREF ID="33" TYPE="IDENT" REF="10">Elmira</COREF>

utt22 : um <sil> load <COREF ID="34" TYPE="IDENT" REF="32">it</COREF> with <COREF ID="35" TYPE="IDENT" REF="16" MIN="juice">orange juice</COREF> <sil> and then <sil> to <sil> <COREF ID="37" TYPE="IDENT" REF="22">Avon</COREF>

utt23 : u: by way of <sil> + <COREF ID="38" TYPE="IDENT" REF="31">Corning</COREF> + <sil> and <COREF ID="39" TYPE="IDENT" REF="20">Dansville</COREF>

utt24 : s: + by way of +

utt25 : right

utt26 : u: how long will that take

utt27 : s: eleven hours

utt28 : u: okay <sil> <COREF ID="40" TYPE="IDENT" REF="28">I</COREF> am now finished

Markup language:

SGML

Existence of annotation tools manual/automatic):

Tools to aid human annotators: yes, developed by SRA. Also, many systems have been developed which can use the scheme to produce a fully automatic annotation. (Which systems?)

Usability:

Systems using the annotation scheme (apart from those which just perform the annotation): not known

A.8 Poesio and Vieira (1)

Contact:

Massimo Poesio (poesio@cogsci.ed.ac.uk)

Coding book:

Described, with summary sheets to help decisions, in Poesio and Vieira (forthcoming) A Corpus-based Investigation of Definite Description Use, Computational Linguistics

Number of annotators:

4 (inc. 2 authors - all 4 are linguists)

Number of dialogues:

Dialogues: none. Texts: 20 articles from Wall St Journal

Evaluations of scheme:

Agreement between 2 annotators: K = 0.68 - 0.73

Agreement between annotators and authors: K = 0.68 - 0.72

(these figures allow the tentative conclusion to be drawn that there is broad agreement between the annotators)

Underlying task:

Scheme was based on Hawkins' and Prince's classification schemes, and was developed at least in part to provide input to the evaluation task. The wider aim was to study definite description use in written text, with a view to developing automatic interpretation algorithms.

Description of the scheme:

This scheme was used to CLASSIFY definite noun phrases based on their relationship to other NPs in the text or world. This contrasts with DRAMA and MUC, which LINK referential expressions in the text.

The annotators had to assign a definite NP to one of the following classes:

Anaphoric same head (a car... the car)

Associative (anaphoric descriptions in which assoc. is based on more complex lexical or common-sense knowledge: a car... the steering wheel, a car... the vehicle)

Larger situation/unfamiliar (NPs relying on real-world knowledge, or which fully identify their referent within the NP: the Querecho Plains of New Mexico, the third quarter, the government, the Iran-Iraq war)

Idiom (in the soup, to the wall etc)

(Doubt)

Markables:

Definite NPs only

Examples:

none

Markup language:

Annotation was stored as prolog assertions

Existence of annotation tools (manual/automatic):

Yes - presentation of NPs to annotator in discourse context and offered choices of classification

Usability:

Implemented in automatic system for classifying definite NPs.

A.9 Poesio and Vieira (2)

Contact:

Massimo Poesio (poesio@cogsci.ed.ac.uk)

Coding book:

Described, with summary sheets to help decisions, in Poesio and Vieira (forthcoming) A Corpus-based Investigation of Definite Description Use, Computational Linguistics

Number of annotators:

3 non-linguists

Number of dialogues:

Dialogues: none. Texts: 14 articles from Wall St Journal

Evaluations of scheme:

Agreement between 3 annotators: K = 0.58 - 0.63

(below the level at which we can tentatively assume agreements between the annotators)

Underlying task:

Scheme was based on Hawkins' and Prince's classification schemes, and was developed at least in part to provide input to the evaluation task. The wider aim was to study definite description use in written text, with a view to developing automatic interpretation algorithms. It was also hoped that this revised scheme would throw some light on the causes of inter-annotator disagreement under their first scheme.

Description of the scheme:

Like the previous scheme, this one was based on CLASSIFICATION of definite NPs. However, in this scheme, where the NPs were classifying as entering into a relationship with another string in the text, that link was also marked.

Relationships/classes:

The annotators had to assign a definite NP to one of the following classes:

Co-referent (a car... the car, a car... the vehicle)

Bridging references (anaphoric descriptions in which assoc. is based on more complex common-sense knowledge: a car... the steering wheel)

Larger situation (NPs relying on real-world knowledge: the Querecho Plains of New Mexico, the third quarter, the government)

Unfamiliar (NPs which fully identify their referent within the NP: the Iran-Iraq war, the door of the Bastille)

(Doubt)

Where the NP was deemed coreferential or bridging reference, the annotators had to identify the antecedent in the text.

Markables:

The annotators had to classify the only definite NPs in the text. However, where there was an antecedent for one of these definite NP, this could be of any syntactic class.

Examples:

none

Markup language:

Annotation was stored as prolog assertions

Existence of annotation tools (manual/automatic):

none

Usability:

The data collected through this annotation scheme was converted into Poesio and Vieira's other scheme (1), before being used to evaluate an automatic annotator.

A.10 UCREL anaphoric annotation

Contact:

University Centre for Computer Corpus Research on Language (UCREL), Lancaster University.
Email: ucrel@lancaster.ac.uk

Coding book:

100+ pages of internal guidelines, plus:

Description of scheme in:

Fligelstone, S (1990) A description of the conventions used in the Lancaster Anaphoric Treebank Scheme. Lancaster: Department of Linguistics and Modern English Language, Lancaster University.

Summary provided in:

Fligelstone, S. (1992). Developing a Scheme for Annotating Text to Show Anaphoric Relations. In: G. Leitner (ed.), New Directions in Corpus Linguistics. Berlin: Mouton de Gruyter. pp153-170.

Number of annotators:

Information not available.

Number of dialogues:

Not used on dialogues (?). Texts: used on approximately 100,000 words from the Associated Press Treebank Corpus, and some work done on the Spoken English Corpus (radio broadcasts).

Evaluations of scheme:

Information not available.

Underlying task:

Corpus-based NLP research.

Description of the scheme:

This scheme, like MUC and DRAMA, is based on identifying referential expressions in the text, and then marking the relationships between these. Like DRAMA, this scheme includes many types of 'bridging' relationships, rather than just coreference. These different types of bridging relationships are all classified in the markup. Coreferring pronouns are also marked for directionality (anaphor or cataphor). The set of markables is even less restricted for UCREL than it is for DRAMA.

Markables:

Noun Phrases

Pronouns (inc. possessives)

part sentences (where this is needed to mark up recoverable ellipsis eg '(jim plays) tennis better than (ellip) football.

clauses

'pro-verbs' (eg do, do so)

Relationships:

Relationships are marked for direction as well as type.

Directions:

< anaphoric

> cataphoric

?<> nondirectional or ambiguous directionality

>< not endophoric

Relationship types:

The following relationships were marked:

REF= coreference: pronoun

(1 1), (1 1) coreference: NPs

SUBS= substitution form (eg {more changes} than {those} recommended by the company; He was looking for {a taxpayer group}. Finding {none}, he...

ELLIP= ellipsis (The UK isn't {a republic}. If it were {},...)

definite indirect anaphor - 'implied' antecedent (The 'love triangle' case... the defendant)

OF= NP with inferrable 'of'-complement (Edinburgh High School... the headmaster)

MISC= miscellaneous cohesion (inc. set/subset, part/whole) (the killings... a motive; local residents... some local citizens; Liverpool... the area)

{{... ...} NP predication (apposition) (The Prime Minister, Mr Blair...)

MTR metatextual reference

Each non-pronomial reference was enclosed in brackets and given an ID. (1 The woman 1). Coference is marked by giving a later referring string the same ID as its antecedent: (1 Judge Clairmont Treefell 1) handed down a sentence of 4 years. The defendant's lawyers said they thought (1 the judge 1) had been very harsh.

Examples:

(from Spoken English Corpus)

(1 The deputy leader of (2 the Labour-controlled Liverpool Council 2) 1),

{{1 Mr Derek Hatton 1}, has said any of (3 <REF=1 his town hall staff 3) who (4 cooperate with (5 the district auditor 5) 4) faced the threat of suspension. Yesterday, <MISC=3(6 two

senior officials who'd <SUBS=4:3done just that 6) were sent home, pending an investigation. (5 The auditors 5) were called in immediately after (5 the council 5) agreed to a nine per cent rates increase to fund spending beyond Government limits. (1 Hr. Hatton 1) said last night <REF=1 he didn't know?(7 whether <MISC=3,6(8 officers 8) were obliged by law to cooperate with (5 the district auditor 5) 7)?. He added: ' <REF=1[S]I'm not bothered about <REF=7 that'.

The following text was annotated under a slightly revised version of the scheme - note the following changes to the conventions:

In the following list, i represents an arbitrary index:

(i i) OR

[i...] enclose a constituent (normally a noun phrase) entering into an equivalence `chain'

{{i i} enclose a noun phrase entering into a copular relationship with a preceding noun phrase

{i i}} enclose a noun phrase entering into a copular relationship with a following noun phrase

(0) represents an anaphoric barrier, in effect, the beginning of a new text.

ANAPHORIC ANNOTATION OF AP NEWSWIRE

S.1 (0) The state Supreme Court has refused to release
{1 [2 Rahway State Prison 2] inmate 1}} (1 James Scott 1) on
bail .

S.2 (1 The fighter 1) is serving 30-40 years for a 1975 armed
robbery conviction .

S.3 (1 Scott 1) had asked for freedom while <1 he waits for an
appeal decision .

S.4 Meanwhile , [3 <1 his promoter 3] , {{3 Murad Muhammed 3} ,
said Wednesday <3 he netted only $15,250 for (4 [1 Scott 1] 's
nationally televised light heavyweight fight against {5 ranking
contender 5}} (5 Yaqui Lopez 5) last Saturday 4) .

S.5 (4 The fight , in which [1 Scott 1] won a unanimous
decision over (5 Lopez 5) 4) , grossed $135,000 for [6
[3 Muhammed 3] 's firm 6], {{6 Triangle Productions of
Newark 6} , <3 he said .

Key: The use of the same index 1, 2, ... n binds one syntactic constituent to another to which it is coreferential or semantically equivalent.

Markup language:

Custom developed for scheme.

Existence of annotation tools (manual/automatic):

Yes, XANADU, X-Windows editor

Usability:

Information not available.

A.11 *-U line

("star-you-line")

Contact:

Matthew Aylett, HCRC (matthewa@cogsci.ed.ac.uk)

Coding book:

Short internal document describing the possible codes and examples of where they might be applied.

Number of annotators:

7 linguists

Number of dialogues:

150+, all English, all map task (128 English, 18 Canadian, ? with children)

Evaluations of scheme:

None

Underlying task:

To investigate whether the form of referring expression used changed between first (introductory) and subsequent mention. It only covers named items on the maps which the speakers are holding, to make labelling more concrete and simple.

Description of the scheme:

This scheme basically marks up all references to the landmarks on the speakers' maps, links the reference to the relevant landmark, and classifies these references as to whether they were the first or subsequent mention. Features are also assigned indicating the linguistic form of the reference (in/definite, pronoun, possessive phrase, etc.) All landmarks mentioned in the conversation are given a unique ID at the start of the transcript of that conversation, and each mention of a landmark is given a pointer to that ID. Coreference is therefore encoded by two references pointing to the same landmark ID.

Relationships:

coreference (references pointing to the same landmark)

partitive relations (references pointing to part of a landmark)

Classification:

first/subsequent mention

def - definite 'THE BRIDGE'

indef - indefinite 'A BRIDGE'

null - no article 'go to BRIDGE'

pro - pronoun 'IT'

el - elliptical reference. 'go round <IT> to the left'

relpro - relative pronoun 'WHICH is on the left'

dctc - deictic 'I don't have THAT'

poss - possessive definite 'MY BRIDGE'

posspro - possessive pronoun 'MINE'

possnum - possessive numerical 'MY ONE'

dem - demonstrative 'THAT BRIDGE', 'THIS BRIDGE'

demnum - demonstrative numerical 'THOSE TWO', 'THAT ONE'

num - numerical 'TWO BRIDGES'

numpro - numericalised as a pronoun 'I've got ONE to the left'

l- literal mention - as the label or close enough. i.e. 'antelope' == 'antelopes', 'site of plane crash' == 'site of the plane crash'

nl - not as the label - 'pile of stones' for 'tor'

rl - reduced literal - 'bridge' for 'ropebridge'

Each code has a disfluent variant

Markables:

Phrases referring to the landmarks marked on the speakers' maps.

Elliptical references to landmarks

Examples:

(The whole conversation is followed by the SGML coding of just references to relevant landmarks, first for the giver, then the follower)

GIVER: Are you above a rocket warehouse.

FOLLOWER: Yes.

GIVER: Okay. Right. You're gonnae slants... go to the end of the rocket warehouse if you know what I mean. See

how the top.

FOLLOWER: Uh-huh.

GIVER: Of the rocket warehouse. Go to the end of that.

FOLLOWER: Parallel to the... the rocket warehouse.

GIVER: Uh-huh.

FOLLOWER: Right.... Okay.

GIVER: Right. Slope s--..... slope... down again..... to the left... directly.

FOLLOWER: To... directly to the left? So to.

GIVER: Directly to.

FOLLOWER: To the left.

GIVER: Like you're sloping down like a seven o'clock line.

FOLLOWER: Uh-huh.

GIVER: Okay.

FOLLOWER: To the left of the rocket warehouse.

GIVER: Uh-huh.

FOLLOWER: How far down.

GIVER: Ehm,..... you should be about,..... ehm..... six centimetres

above the bottom of the page.

FOLLOWER: So... below the rocket warehouse.

GIVER: No. You're at the left-hand side of the ro--.

FOLLOWER: S--... stop at the left.

GIVER: You should be just an. Inch away from the edge of the page now.

FOLLOWER: Yeah.

GIVER: Right. Okay.

FOLLOWER: Uh-huh.

GIVER: Go straight down again for three centimetres.

FOLLOWER: Uh-huh.

GIVER: And then curve towards the old lighthouse.

FOLLOWER: Round which side of the cave.

GIVER: Hmm.... I don't have a cave. Ehm,... go towards the lighthouse.

FOLLOWER: Uh-huh. Ehm,..... just slope towards it.

Giver:

<U MEN=intro_q START=268.2955 DUR=1.0120 OMO=1 DEF=indef NAME='rocket warehouse' SIM=same ID=rocketwhouse LIT=l ASSIM=glott STAT=0>a rocket warehouse</U>

<U MEN=rep0 START=273.4618 DUR=0.8618 OMO=1 DEF=def NAME='rocket warehouse' SIM=same ID=rocketwhouse LIT=l ASSIM=glott STAT=0>the rocket warehouse</U>

<U MEN=rep0 START=276.6470 DUR=0.6840 OMO=1 DEF=def NAME='rocket warehouse' SIM=same ID=rocketwhouse LIT=l ASSIM=glott STAT=0>the rocket warehouse</U>

<U MEN=rep0 START=278.1811 DUR=0.2720 OMO=1 DEF=dctc NAME='rocket warehouse' SIM=same ID=rocketwhouse ASSIM=glott STAT=0>that</U>

<U MEN=rep0 START=323.3644 DUR=0.2 OMO=1 DEF=def NAME='rocket warehouse' SIM=same ID=rocketwhouse LIT=ld ASSIM=glott STAT=0>the ro--</U>

<U MEN=intro START=334.1719 DUR=0.7360 DEF=def NAME='old lighthouse' SIM=same ID=oldjjlighthouse LIT=l ASSIM=ddel STAT=0>the old lighthouse</U>

<D DUR=0.8123 LANDMARK=cavejj START=341.7959 SIM=dif_f10 STAT=0>i don't have a cave</D>

<U MEN=rep0 START=342.1722 DUR=0.4360 DEF=indef NAME=cave SIM=dif_f10 ID=cavejj LIT=l STAT=0>a cave</U>

<U MEN=rep0 START=344.0529 DUR=0.6300 DEF=def NAME='old lighthouse' SIM=same ID=oldjjlighthouse LIT=rl ASSIM=ddel STAT=0>the lighthouse</U>

Follower:

<U MEN=rep0 START=281.2534 DUR=0.956 OMO=1 DEF=def NAME='rocket warehouse' SIM=same ID=rocketwhouse LIT=l ASSIM=glott STAT=0>the rocket warehouse</U>

<U MEN=rep0 START=298.9796 DUR=0.9760 OMO=1 DEF=def NAME='rocket warehouse' SIM=same ID=rocketwhouse LIT=l ASSIM=glott STAT=0>the rocket warehouse</U>

<U MEN=rep0 START=318.8779 DUR=0.9160 OMO=1 DEF=def NAME='rocket warehouse' SIM=same ID=rocketwhouse LIT=l ASSIM=glott STAT=0>the rocket warehouse</U>

<U MEN=intro START=338.5999 DUR=0.4700 DEF=def NAME=cave SIM=dif_f10 ID=cavejj LIT=l STAT=0>the cave</U>

<U MEN=rep0 START=348.4719 DUR=0.0900 DEF=pro NAME='old lighthouse' SIM=same ID=oldjjlighthouse ASSIM=ddel STAT=0>it</U>

Markup language:

SGML

Existence of annotation tools (manual/automatic):

Yes; Python-based tool to aid annotation.

Usability:

None.

Dialogue Act Schemes

A.12 Alparon

(Delft University of Technology)

Coding book:

ftp://ftp.twi.tudelft.nl/TWI/publications/tech-reports/1996/DUT-TWI-96-137.ps

Authors: R.J. van Vark, J.P.M. de Vreught, L.J.M. Rothkrantz

Title: Analysing OVR dialogue coding scheme 1.0; Report 96-137

Number of annotators:

3 (all computer scientists)

Number of annotated dialogues:

about 500 (in Dutch), with 12000 turns and 16000 segments

Evaluations of scheme:

ftp://ftp.kbs.twi.tudelft.nl/pub/alparon/publications/1997/ L.J.M.Rothkrantz-SALT-97.ps.gz

No κ-evaluations are published. But the scheme has been used in the VIOS system for dialogues concerning public transport services in the Netlands.Miscommunication

between the system and its client occurred in 23% of the dialogues.

Underlying task:

The scheme was designed within the domain of information retrieving about the services offered by the Dutch public transport.

List of phenomena annotated:

- Phases
  - Greeting (G)
  - Query (Q)
  - Pause (P)
  - Information (I)

  - Subquery (S)
  - Goodbye (B)
- Moves (Dialogue Acts)
  - Greeting (Gre)
  - Acknowledgement (ack)
  - Alignment (Ali)
  - Bye (Bye)
  - Statement (Sta)
  - Check (Che)
  - Clarification (Cla)
  - Question (Que)
  - Pause (Pau)
  - Reconfirmation (Rec)
  - Other (Oth)
- Coding of Information
  - Personal (Per)
  - Price (Pr)
  - Timetable (Tt)
  - Time (Tm)
  - Rout_Infortmation (RI)
  - Location (Loc)
  - Acknowledge (Ack)
  - Announce (Ann)
  - Break (Bre)
  - Ready (Rea)
  - Bye (Bye)
  - Garbage (Gar)
  - Other (Oth)

Examples:

2: goedemorgen reisinformatie (good morning travel information)	Gre(G,[],[])
1: goedemorgen [achternaam] kunt u mij zeggen hoe laat de[uh] bus van Lochem naar Deventer toe gaat (good morning [last name] can you tell me what time the[uh] bus departs from Lochem to Deventer)	Gre(G,[Per(Nam])],[]), Que(Q,[Tt([DaS([Unspecified])]),RI([TrT([Bus_Traum(0)])]),Loc([DeP([City(1)]), ArP([City(2)])])],[])
2: hoe laat ongeveer zou u mee willen (how late approximately would you like to go)	Que(S,[Tm([DeT([Unspecified])])],[])
1: [uh] rond een uur of negen ([uh] about nine o'clock)	Sta(Q,[Tm([DeT([About(3)])])],[])
2: oke ok)	Ack(Q,[Ack([Pos])],[])
2: negen uur twee mevrouw buslijn zesenvijftig wo past nine madam bus line fifty- six)	Sta(I,[Tm([DeT([Exact(4)])]), RI([TrT([Bus_Tram(5)])])],[])
1: oke dank u wel (ok thank you)	Ali(B,[Ann],[]),Bye(B,[Bye([Tha])],[])
2: tot uw dienst (at your service)	Bye(B,[Bye([ReG])],[])
1: ja hoor dag (yes fine bye)	Bye(B,[Bye([ReG,Goo])],[])
2: goodemorgen (good morning)	Bye(B,[Bye([Goo])],[])

Mark-up language:

The dialogue acts are Prolog terms with some extra punctuation marks for making list of dialogue acts and to mark which agent was talking.

Existence of annotation tool:

OVR coder:

- manual coding tool
- implementation language: tcl/tk

Usability:

Used in VIOS, the ASP (Automated Speech Processing) prototype of OVR (Openbaar Vervoer Reiseinformatie, Public Transport Travel Information).

Contact person:

L.J.M. Rothkrantz (L.J.M.Rothkrantz@cs.tudelft.nl)

A.13 Chiba Coding Scheme

(Chiba University)

Coding book:

Instead of having one scheme the Chiba scheme consists of three different coding schemes according to

- the unit of tagging: discourse marker tagging for word or phrase **(A)**
- utterance unit tagging for one utterance (sentence) **(B)**, and
- discourse unit tagging for so called discourse segment **(C)**.

These schemes are applied for any tasks.

The coding book will be available by WWW in the near future. But it is written in Japanese. The outline of the work is reported at the First International Conference on Language Resources and Evaluation, Spain, May 1998.

"Standardising Annotation Schemes for Japanese Discourse", A. Ichikawa, et. al.

Number of annotators:

10 coders (WG members)

Number of annotated dialogues:

Task	Dialogues	Utterances
Schedule management	14	509
Route direction	3	131
Telephone shopping	4	277
Tourist information	1	68

Evaluations of scheme:

	A	B	C
alpha	0.577	0.680	0.612

Underlying task:

route direction, scheduling, telephone shopping, tourist information

List of phenomena annotated

- conventional: *opening, closing*
- initiation: *request, suggest, persuasion, propose, confirm, yes-no question, wh-question, promise, demand, inform, other assertion, other initiation*
- response: *positive, negative, answer, hold, other response*
- follow-up: *understanding*
- response with initiation: The element of this category can be represented as response/initiation

Examples:

66 U: *hai, etto, shinkanseN waNji hatsu desu ka.*

67 **(I)** (What's the departure time of the bullet train?)

68 S: *e, jyuu nana ji haN ni natte orimasu.*

69 **(R)** (It's 17:30.)

70 U: *hai.*

71 **(F)** (I see.)

Mark-up language:

The markup language is as follows:

```
<Dialog>
<Utt Id=0000 Utterance_unit=open_dialogue Speker="S"
   Topic=scheduling Depth_of_segment=2 >
[Well] <then> please start.
```

Discourse markers are tagged in the transcriptions. Utterance unit and discourse unit are described in SGML manner.

Existence of annotation tool:

A modification of DAT (DRI) is used. It includes prediction of utterance unit tag. (Accuracy of prediction is about 70 % in open test.)

Usability:

Information not available.

Contact person:

Masato Ishizaki (masato@jaist.ac.jp)

A.14 Chat

(Carnegie Mellon University, Department of Psychology)

Coding book:

http://poppy.psy.cmu.edu/childes/index.html

http://atila-www.uia.ac.be/childes

Author: Brian MacWhinney

Title: The CHILDES Project: Tools for Analysing Talk

Number of annotators:

The CHAT system is a real standard for transcription and coding of child language in a number of European and non-European languages. This means that a great number of annotators has been using CHAT for different purposes, so that it is difficult to state an exact number of annotators. Most of the annotators were linguists.

Number of annotated dialogues:

A huge number of dialogues has been/is being annotated with the CHAT coding scheme. This number exceeds the amount of dialogues in the database, as many projects concerning child language make use of CHAT without contributing to the overall CHILDES database. The internationally recognized CHILDES database include transcripts from over forty major projects in English and additional data from 19 other languages. The additional languages are Brazilian Portuguese, Chinese (Mandarin), Chinese (Cantonese), Danish, Dutch, French, German, Greek, Hebrew, Hungarian, Italian, Japanese, Mambila, Polish, Russian, Spanish, Swedish, Tamil,

Turkish, and Ukrainian. The total size of the database is now approximately 160 million characters (160 MB).

Evaluation of scheme:

As a result of its worldwide use, CHAT is continuously evaluated and updated to meet the needs of different languages and different users. Anyway, we are not aware of statistical/quantitative evaluations of its reliability.

Underlying task:

Analysis of child language.

List of phenomena annotated:

Speech act codes:

- Interchange type categories ("x"):
  - CMO comforting to comfort and express sympathy for misfortune
  - DCA discussing clarification of action to discuss clarification of hearer's non-verbal communicative acts
  - DCC discussing clarification of communication to discuss clarification of hearer's ambiguous verbal communication or a confirmation of the speaker's understanding of it
  - DFW discussing the fantasy world to hold a conversation within fantasy play
  - DHA directing hearer's attention to achieve joint focus of attention by directing hearer's attention to objects, persons and events in the environment
  - DHS discussing hearer's sentiments to hold a conversation about hearer's non observable thoughts and feelings
  - DJF discussing a joint focus of attention to hold a conversation about something in the environment that both participants are attending to, e.g., objects; persons; ongoing actions of hearer and speaker; ongoing events
  - DNP discussing the non present to hold a conversation about topics which are not observable in the environment, e.g., past and future events and actions, distant objects and persons, abstract matters (excluding conversations about inner states)
  - DRE discussing a recent event to hold a conversation about immediately past actions and events
  - DRP discussing the related-to- present to discuss non observable attributes of objects or per-sons present in the environment or to discuss past or future events related to those referents
  - DSS discussing speaker's sentiments to hold a conversation about speaker's non-observable thoughts and feelings

- MRK marking to express socially expected sentiments on specific occasions such as thanking, apologizing, or to mark some event
- NCS negotiate copresence and separation to manage the transition
- NFA negotiating an activity in the future to negotiate actions and activities in the far future
- NIA negotiating the immediate activity to negotiate the initiation, continuation, ending and stopping of activities and acts; to direct hearer's and speaker's acts; to allocate roles, moves, and turns in joint activities
- NIN non interactive speech speaker engages in private speech or produces utterances which are clearly not addressed to present hearer
- NMA negotiate mutual attention to establish mutual attentiveness and proximity or withdrawal
- PRO performing verbal moves to perform moves in a game or other activity by uttering the appropriate verbal forms
- PSS negotiating possession of objects to determine or discuss who is the possessor of an object
- SAT showing attentiveness to demonstrate that speaker is paying attention to hearer.
- TXT read written text to read or recite written text aloud.
- OOO unintelligible utterances unknown function
- YYY uninterpretable utterances unknown function

- Categories of Illocutionary Force ("i"):
  - Directives
    - AC Answer calls; show attentiveness to communications.
    - AD Agree to carry out act requested or proposed by other.
    - AL Agree to do for the last time.
    - CL Call attention to hearer by name or by substitute exclamations.
    - CS Counter suggestion; an indirect refusal.
    - DR Dare or challenge hearer to perform action.
    - GI Give in; accept other's insistence or refusal.
    - GR Give reason; justify a request for action, refusal or prohibition.
    - RD Refuse to carry out act requested or proposed by other.
    - RP Request, propose, or suggest an action for hearer, or for hearer and speaker.
    - RQ Yes/no question/suggestion about hearer's wishes and intentions
    - SS Signal to start performing an act, such as running or rolling a ball.
    - WD Warn of danger.

- Speech Elicitations
  - CX Complete text, if so demanded.
  - EA Elicit onomatopoeic or animal sounds.
  - EI Elicit imitation of word or sentence by modelling or by explicit command.
  - EC Elicit completion of word or sentence.
  - EX Elicit completion of rote learned text.
  - RT Repeat or imitate other's utterance.
  - SC Complete statement or other utterance in compliance with request.
- Commitments
  - FP Ask for permission to carry out act.
  - PD Promise.
  - PF Prohibit/forbid/protest hearer's performance of an act.
  - SI State intent to carry out act by speaker; description of one's own on-going activity.
  - TD Threaten to do.
- Declarations
  - DC Create a new state of affairs by declaration.
  - DP Declare make-believe reality.
  - ND Disagree with a declaration.
  - YD Agree to a declaration.
- Markings
  - CM Commiserate, express sympathy for hearer's distress.
  - EM Exclaim in distress, pain.
  - EN Express positive emotion.
  - ES Express surprise.
  - MK Mark occurrence of event (thank, greet, apologize, congratulate, etc.).
  - TO Mark transfer of object to hearer.
  - XA Exhibit attentiveness to hearer.
- Statements
  - AP Agree with proposition or proposal expressed by previous speaker.
  - CN Count.
  - DW Disagree with proposition expressed by previous speaker.
  - ST State or make a declarative statement.
  - WS Express a wish.

- Questions
  - AQ Aggravated question, expression of disapproval by restating a question.
  - AA Answer in the affirmative to yes/no question.
  - AN Answer in the negative to yes/no question.
  - EQ Eliciting question (e.g., hmm?).
  - NA Intentionally non satisfying answer to question.
  - QA Answer a question with a wh-question.
  - QN Ask a product-question (wh-question).
  - RA Refuse to answer.
  - SA Answer a wh-question by a statement.
  - TA Answer a limited-alternative question.
  - TQ Ask a limited-alternative yes/no question.
  - YQ Ask a yes/no question.
- Performances
  - PR Perform verbal move in game.
  - TX Read or recite written text aloud.
- Evaluations
  - AB Approve of appropriate behaviour. Express positive evaluation of hearer's or speaker's acts.
  - CR Criticize or point out error in non-verbal act.
  - DS Disapprove, scold, protest disruptive behaviour. Express negative evaluation of hearer's or speaker's behaviour as inappropriate.
  - ED Exclaim in disapproval.
  - ET Exclaim in surprise or enthusiasm, express enthusiasm for hearer's performance.
  - PM Praise for motor acts, i.e. for non-verbal behaviour.
- Demands for clarification
  - RR Request to repeat utterance.
  - Text editing
  - CT Correct, provide correct verbal form in place of erroneous one.
- Vocalizations
  - YY Utter a word-like utterance without clear function.
  - 00 Unintelligible vocalization.

Example:

*****MOT: are you okay?

%spa: $x:dhs $i:yq

Mark-up language:

CHAT's own format.

Existence of annotation tools:

The CHILDES system contains several separate, yet integrate, tools, subdivided in two major tools. The first tool is a full-fledged and ASCII-oriented editor (CED, Childes EDitor), specifically designed to facilitate the editing of CHAT files and to check for accuracy of transcriptions. The second tool, actually a bunch of several smaller tools, is a set of computer programs called CLAN (Child Language ANalysis) which serve different analysis purposes. The full system is presented in detail in MacWhinney (1991) and illustrated through practical examples in Sokolov and Snow (1994).

MacWhinney, B. (1991). The CHILDES project: Tools for analyzing talk. Hillsdale, NJ: Erlbaum. Sokolov, J. and C. Snow (Eds.). (1994). Handbook of research in language development using CHILDES. Hillsdale, NJ: Erlbaum.

Usability:

Used in the CHILDES project.

Contact person:

Brian MacWhinney (macw@cmu.edu)

A.15 COCONUT

(The University of Pittsburgh Intelligent Systems Program; The Natural Language Group at SRI International)

Coding book:

http://www.isp.pitt.edu/~intgen/research-papers.html

Author: Barbara de Eugenio, Pamela W. Jordan, Liina Pylkkänen

Title: The COCONUT project: dialogue annotation manual (draft)

Number of annotators:

3 for development, but only 2 out of these 3 for the real annotation effort.

One annotator is linguist, the others are comutational linguists.

Number of annotated dialogues:

16 dialogues (about 800 utterances) have been annotated, of which 9 have been doubly annotated. All dialogue are in English.

Evaluation of scheme:

http://www.isp.pitt.edu/~intgen/research-papers.html

Underlying task:

Linguistic Motivation: Exploration how conversation correlates with the problem solving process where two participants are involved in.

Task: Buying furniture for the living and dining rooms of a house.

List of phenomena annotated

- Information level
  - Task
    - EvaluatePlan
    - GameProcedure
  - Task Management
    - Strategize Action
  - Communication Management
  - Other Level
- Forward-Communication Function
  - Statement
    - Assert
    - Reassert
    - Other-Statement
  - Influence-on-Listener
    - Open-Option
    - Directive
      - Info-Request
      - Action-Direction
  - Influence-on-Speaker
    - Offer
    - Commit

- Other-forward-function
  - ConventionalOpening
  - ConventionalClosing
  - ExplicitPerformative
  - Exclamation
- Topic
  - Topic proper (furniture items)
    - needItem
    - haveItem
    - getItem
    - elaborateItem
    - otherItem
  - Topic proper (money)
    - budgetAmount
    - budgetRemains
    - costAccum
  - Topic proper (points)
    - pointAmount
    - pointAccum
  - Attitude
    - Eval
    - Relate
- SurfaceFeatures
  - Word-Surface-Features
    - Matrix
    - Modal
    - Subject
  - Syn-Surface-Features
    - Tense
    - Mood
    - Neg-Polarity
- Backward Communication Function
  - Initiate

- Agreement
  - Accept
  - Accept-Part
  - Maybe
  - Reject-Part
  - Reject
  - Hold
  - ClarificationRequest
- Understanding
  - Signal-non-understanding
  - Signal-understanding
    - Acknowledge
    - RepeatRephrase
    - (Completion)
  - CorrectMisspeaking
  - CorrectAssumption
- Answer
- Information Relations
- Coreference / Set Relations
- Segment Tag
- Fragment

Example:

S1: (a) so we should move to the engine at Avon engine E to

S2: (b) engine E one

CorrMisspeak(a)

S1: (c) E one to Bath

Accept(b)

Mark-up language:

Variant of DAMSL

Existence of annotation tools:

Nota Bene (Nb) by Flammia

- manual
- implementation language: Tcl/Tk (Tcl Version 7.4 and Tk Version 4.0 or higher)

Usability:

COCONUT system

Contact person:

Barbara di Eugenio (dieugenio@cs.pitt.edu)

A.16 Condon and Cech's Coding Scheme

(Discourse Intervention Project, University of Southwestern Louisiana)

Coding Book:

ftp://sls-ftp.lcs.mit.edu/pub/multiparty/coding/condon
Author: Sherri Condon, Claude Cech

Title: Manual for Coding Decision-Making Interactions

Number of Annotators:

Five students, two undergraduate and three graduate students, were trained on the original scheme. One undergraduate did not achieve acceptable levels of agreement and did not code any dialogues. The other undergrad coded only a couple. Three graduate students have been trained on the new scheme and another is in progress. All students are non-linguists, although they may have had course work in linguistics.

Number of Annotated Dialogues:

First corpus (original coding scheme) contains 4141 utterances from the 16 face-to-face interactions and 918 utterances from the 16 computer-mediated interactions. The utterance is defined as a main clause together with all complements and adjuncts (including subordinate clauses). The new scheme has been used for 8 face-to-face and 60 synchronous computer-mediated interactions. In addition, we are working on about 20 asynchronous (e-mailed) computer-mediated interactions. The task for these interactions was to plan the MTV video awards ceremony, and participants were again dyads, but with all combinations of male and female.

Evaluation of scheme:

Evalution hasn't been performed yet but there has been a test at the discourse annotation workshop at Penn in which computational linguists whose only training was reading the manual were given Verbmobil data (as opposed to the planning tasks that the scheme was designed for) and achieved perfect agreement on 36 of 33 utterances.

Underlying task:

Decision-Making tasks

List of phenomena annotated: Top-Level Functions:

- **Move** (MOVE)
- **Response** (RESP)
- **Other** (OTHR)

Each utterance must be associated with exactly one function in each of the three broad categories.

- **Move Functions :**
  - *SA* Suggests Action (for the decision task) locations, activities and orders for them
  - *RA* Requests Action (requires immediate action)
  - *RV* Request Validation/Verification/Acknowledgement (of some statement)
  - *RI* Requests Information (Information Questions)
  - *ER* Elaborates, Explains, Supports, Repeats previous utterance (Relevant comment)
  - *NC* No Clear MOVE function
- **Response Functions:**
  - *AS* Agrees with Suggestion
  - *DS* Disagrees with Suggestion, Refuses to Comply with Request
  - *CR* Complies with Request
  - *AO* Acknowledges Only
  - *NC* No Clear RESPONSE Function
- **Other Functions:**
  - *DM* Discourse Marker
  - *ML* Metalanguage
  - *OS* Orientation of Suggestion
  - *PI* Requests, Offers, Refers to, Evaluates Personal Information
  - *JE* Jokes, Exaggerates
  - *NC* No Clear OTHER Function

Examples:

- Move Functions:

  SA: Let's go to New Orleans

  RA: Write that down

  RV: right?, you know?, agreed?, To New Orleans? (checking questions)

  RI: Where do you want to go?, How long does it take to drive to New Orleans?

  ER: This is fun, I love New Orleans

  NC: Fillers

- Response Functions:

  AS: ok, good idea, we should have a great time there

  DS: no, sounds boring, that is too much in one day

  CR: ok, it takes about an hour to drive to Baton Rouge

  AO: me, too, really, I know

- Other Functions:

  DM: so, well, let's see

  ML: Let's decide where the party will be first, We're finished

  OS: To go to New Orleans, let's hire a jet, In New Orleans we can go on a riverboat

  PI: Were you in the service?, Have you ever been there?, I go there all the time

  JE: yeah/ mall warriors, party on!

Mark-up Language:

N.b's mark-up language. This is not fully compliant with SGML, but a program is distributed with Nb that converts Nb-annotated files into standard SGML files.

Existence of Annotation Tools:
N.b. Tcl/Tk interface by G. Flammia.

Usability:
Used in the Discourse Processing Project.

Contact person:

Sherri Condon (slc6859@usl.edu)

Department of English

University of Southwestern Louisiana

Lafayette, La 70504-4691

USA

Phone/Voice Mail (318) 482-5476 Fax (318) 482-5071

A.17 C-STAR

(C-STAR Consortium)

Coding book:

available via ftp.cs.cmu.edu in project/enthusiast/cstar/current/manual.ps

Author: not mentioned

Title: Template translation and Dialogue Act Annotation

Number of annotators:

5 (linguists, computational linguists, computer scientists)

Number of annotated dialogues:

	Number of Dialogues	**Number of Turns**	**Number of Segments**
English, CMU	31	1605	2523
Korean with English translation, ETRI (Korea)	70	453	1140
Italian with English translation, IRST (Italy)	5	132	233
Japanese with English translation, ATR (Japan)	124	4424	5887
Total	230	6614	9783

Evaluation of scheme:

No inter-coder agreement tested.

Underlying task:

The scheme is developed for two-agent travel planning domain dialogues in which a travel agent and a customer are involved in various travel scenarios like hotel/flight reservation, ticket purchase, transportation inquiry, tour and sight seeing information seeking etc.. The dialogue act annotation scheme is designed for shallow representation of spoken utterances. The current inventory of dialogue acts is mainly based on hotel reservation dialogues, although the scheme is general enough to expand into other domains (i.e. scheduling).

List of phenomena annotated

about 25 speech acts

about 500 domain-specific domain actions

Dialogue acts are compositional. A dialogue act consists of three representation levels indication different aspects of the utterance: the speech act (e.g. whether the speaker performs the act of accepting, giving a requesting information etc.), the concept which denote the informational focus of the utterance in question (i.e. whether the speaker is giving information about the availability of rooms, about a trip, a flight, etc.) and the arguments denoting the specific contents of the utterance (e.g. whether the speaker is giving information about single or double rooms, about one or two flights, etc.). Arguments are inherited either by the speech act or by the concepts.

Speech acts: accept, acknowledge, affirm, apologize, closing, give-information, greeting, introduce-self, introduce-topic, negate, offer, please-wait, reject, request-action, request-affirmation, request-information, request-suggestion, request-verification, suggest, suggest-action, thank, verify;

Some concepts: Actions (change, reservation, confirmation, cancellation, preference, help), Attributes (availability, size, temporal, price, location, features, etc.), Objects (room, hotel, flight, payment, etc.), Other (arrival, departure, numeral, expiration date);

Example:

The week of the twelfth we have both singles and doubles available.

a:give-information+availability+room (room-type=(single & double), time=(week, md12))

Mark-up language:

Interchange format

Existence of annotation tools:

No annotation tool.

Usability:

Used in the systems of the consortium members.

Contact person:

Lori Levin (Lori.Levin@alexis.boltz.cs.cmu.edu)

A.18 DAMSL

(Discourse Representation Initiative)

Coding book:

http://www.cs.rochester.edu:80/research/trains/annotation

Authors: James Allen, Mark Core

Title: Draft of DAMSL: Dialog Act Markup in Several Layers

Number of annotators:

At Rochester, only 2 of the 18 DAMSL annotated dialogues were coded by linguistics undergraduates. The other 16 were coded by a computer science grad student and an undergraduate in engineering.

Number of annotated dialogues:

18 (1037 turns, 1524 utterances) - all in English

Evaluation of scheme:

- Forward Looking Function
  - Statement: $\kappa = 0.66$
  - Influencing Addressee Future Action: $\kappa = 0.70$
  - Committing Speaker Future Action: $\kappa = 0.15$
  - Other Forward Functions: $\kappa = 0.48$
- Backward Looking Function
  - Agreement: $\kappa = 0.42$
  - Understanding: $\kappa = 0.57$
  - Answer: $\kappa = 0.76$
  - Response to: $\kappa = 0.77$

Underlying task:

The annotation scheme has been defined in order to provide a top-level structure for annotating a range of dialogues for many different purposes.

List of phenomena annotated:

- Communicative-Status (Records whether the utterance is intelligeble and wheter it was succesfully completed)
  - Uninterpretable
  - Abandoned
  - Self-talk
- Information Level (A characterization of the semantic content of the utterance)
  - Task
  - Task-management
  - Communication-management
  - Other-level
- Forward Looking Function (How the current utterance constrains the future beliefs and actions of the participants, and affects the discourse)
  - Statement
    - Assert
    - Reassert
    - Other-statement
  - Influencing-addressee-future-action
    - Open-option
    - Action-directive
  - Info-request
  - Committing-speaker-future-action
    - Offer
    - Commit
  - Conventional
    - Opening
    - Closing
  - Explicit-performative
  - Exclamation
  - Other-forward-function
- Backward Looking Function (How the current utterance relates to the previous discourse)
  - Agreement
    - Accept
    - Accept-part

    - Maybe
    - Reject-part
    - Reject
    - Hold
- Understanding
    - Signal-non-understanding
    - Signal-understanding
        - Acknowledge
        - Repeat-rephrase
        - Completion
    - Correct-misspeaking
- Answer
- Information-relation

Example:

utt1: u: mm <click okay *Reassert*

utt2: four hours from Avon to Bath *Action-directive*

utt3: and then I guess attach that to the boxcar to Corning

utt4: it's four hours and

utt5: how long Info-request *Abandoned*

utt6: it is two hours from Bath to Corning *Info-request*

Mark-up language:

DAMSL (a variant of SGML)

Existence of annotation tools:

dat

- manual
- implementation language: Perl 5.004_04, Perl Tk 402.003

Usability:

COCONUT, SWBD-DAMSL, CLARIFY

Contact person:

Mark G. Core (mcore@cs.rochester.edu)

A.19 Janus

(Carnegie Mellon University, Language Technology Institute; Universität Karlsruhe)

Coding book:

Coding book not public available.

Number of annotators:

Between 4 and 10. (linguists, computational linguists, computer scientists)

Number of annotated dialogues:

Lots of test sets. Each test set consists of roughly 100 utterances.

Evaluations of scheme:

The correct word to state classification rate is 89%.

("Statistical Analysis of Dialogue Structure"; Ye-Yi Wang, Alex Waibel)

"scores from different judges may vary by as much as 10 percentage points" in Donna Gates et al. : *End-to-End Evaluation in JANUS: A Speech-to-Speech translation system* in: E. Maierr, M. Mast, S. LuperFoy (Eds.):Lecture Notes in Artificial Intelligence 1236 *Dialogue Processing in Spoken Language Systems*, Springer-Verlag Berlin Heidelberg, 1997

Underlying task:

Appointment scheduling

List of phenomena annotated:

Related to C-STAR phenomena.

Examples:

[nicety] (Hello Dr. Noah)

[nicety] (Hi Tor)

[suggest-meeting] (let's set up a meeting for a couple of hours)

[temporal] (in the next two weeks)

[your-availability] (when's good for you)

[interject] (let's see)

[suggest-time] (how about Friday the second in the morning)

[my-unavailability] (I'm busy that morning)

Markup language:

Own format.

Existence of annotation tools:

Annotation by hand.

Usability:

Janus system

Contact person:

Lori Levin (Lori_Levin@alexis.boltz.cs.cmu.edu)

A.20 Flammia's Coding Scheme

(Spoken Language Systems Group, Laboratory for Computer Science, Massachusetts Institute of Technology)

Coding book:

ftp://sls-ftp.lcs.mit.edu/pub/multiparty/coding_schemes/flammia
Author: Giovanni Flammia

Title: Instructions for Annotating Segments in Dialogues

Number of annotators:

16 graduate students with some knowledge of computer science and linguistics

Number of annotated dialogues:

25, with an average number of dialogue turns of 40 and with 29 to 120 utterances per dialogue. The language of the dialogues is American-English.

Evaluations of scheme:

http://www.sls.mit.edu/~flammia/publications.html

- ”Empirical evaluation of human performance and agreement in parsing discourse constituents in spoken dialogue”
- ”Learning the structure of mixed initiative dialogues using a corpus of annotated conversations”

The pairwise agreement for segment boundaries placing is 84.3%.

The average pairwise symbolic accuracy for segment purposes is 80.1%.

$\kappa = 0.6$ (in a previous trial - probably better now!)

Underlying Task:

Information-seeking dialogues; telephone conversations between customers and operators of the BellSouth *Movies Now* service - a telephone number that people can call to get information about current movie schedules in Atlanta.

List of phenomena annotated:

Structural/functional phenomena, such as the division of dialogues into segments, each one concerning a given topic. A segment is thus defined as a sequence of two or more dialogue turns (including at least one utterance by each one of the speakers), where one relevant piece of information is exchanged between conversation participants. Relevance is defined in terms of necessity to the continuation of the task defined in the dialogue. Flammia's coding scheme does not provide categories with which segments should be annotated; instead, annotators are free to choose what they consider to be the most appropriate description for a given segment. However, some speech act tags that are exemplified in Flammia's approach are the following: Request, Response, Acknowledge, Accept, Reject, Repeat, Confirm, and Question Confirm. A decision procedure concerning how to carve segments out of dialogues is specified, together with 'rules of the thumb' regarding possible correspondences between surface forms and segments boundaries. Discourse phenomena such as greetings, introductions, offers to help, back-channel phenomena, prompts for continuation, thanks and closings are not recognized as having a relevant status for segmentation. Only segments directly dealing with task-relevant information are signaled and annotated.

Examples:

http://sls-www.lcs.mit.edu/~flammia/Nb/example_output.gif

Mark-up language:

N.b.'s mark-up language. This is not fully compliant with SGML, but a program is distributed with Nb that converts Nb-annotated files into standard SGML files.

Existence of annotation tools:

N.b. Tcl/Tk interface by G. Flammia.

Usability:

Information not available.

Contact person

Giovanni Flammia (flammia@sls.lcs.mit.edu)

A.21 LinLin

(Linköping University)

Coding book:

http://www.cs.umd.edu/users/traum/DSD/arne2.ps

Authors: Nils Dahlbäck and Arne Jönsson

Title: A coding manual for the Linköping dialogue model

Further information can be found in

- Staffan Larsson (1998): Coding Schemas for Dialogue Moves. Göteborg University, January 1998
- Lars Ahrenberg, Nils Dahlbäck, and Arne Jönsson (1995): Coding Schemes for Studies of Natural Language Dialogue. in *Working Notes from AAAI Spring Symposium*, Stanford, 1995
- Arne Jönsson (1995): Dialogue Actions for Natural Language Interfaces. in *Proceedings of IJCAI-95*, Motreal, Canada, 1995
- Arne Jönsson (1995): A Dialogue Manager for Natural Language Interfaces. in *Proceedings of the Pacific Associaton for Computational Linguistics, Second Conference*, The University of Queensland, Brisbane, Australia, 1995
- Arne Jönsson (1993): A Method for Development of Dialogue Managers for Natural Language Interfaces. in *Proceedings of AAAI-93*, pp. 190-195, Washington DC

Number of annotators:

None at the moment. Previously, four people used the scheme.

The dialogues has been analyzed by linguists, a psychologist and computer scientsits. However, the main work on tagging the dialogues was done by two students, one cognitve science student and one computer science student.

Number of annotated corpora:

The corpus used when developing Linlin consists of 30 dialogues with 1749 utterances. The dialogue model was also applied to 100 SUNDIAL dialogues with around 700 utterances and to 10 Waxholm dialogues with around 400 utterances. (all Swedish)

Evaluations of scheme:

No κ-statistic is available for the LinLin scheme, but in a pairwise agreement for LinLin1 97% was achieved.

Underlying task:

LinLin was designed to a written human-to-(simulated)computer information retrieval dialogue. It is now also applied to the AIRPLANE corpus with a human-to-human instructional dialogue.

List of phenomena annotated:

- Initiative
  - Update (*U*): User provides information to the system
  - Question (*Q*): User obtains information from the system
- Response
  - Answer (*A*): System database answer, answer to clarification request
- Discourse management
  - Opening (*DO*)
  - Ending (*DE*)
  - Discourse Continuation (*DC*)

Example:

S: [Welcome to Cardata] *DO*

U: [show mercedes] *Q*

S: [Wait...] *DC*

[Cardata can answer questions about a number of car models, concerning manufacturer, model, year, country of manufacturing, disposition to rust and size class as well as question about price, security, space and technical data.] *A*

[Any particular wishes?] *Q*

U: [cost and space] *A*

...

Mark-up language:

Nb's mark-up language (pseudo-SGML - not fully compliant)

Existence of annotation tools:

Nb (NotaBene) for coding, perlscript for analysis.

Usability:

Used in the S-DIME (Swedish dialogue move engine) project.

Contact person:

Arne Jönsson (arnjo@ida.liu.se)

Dept. of Computer and Information Science

Linköping University

S-581 83 LINKÖPING

SWEDEN

A.22 Maptask

(HCRC)

Coding book:

http://www.hcrc.ed.ac.uk/~jeanc/

Authors: Carletta, J. C., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G. and

Anderson, A.

Title: HCRC Dialogue Structure Coding Manual

Human Communication Research Centre HCRC TR-82, University of Edinburgh, Edinburgh, Scotland; 1996

A slightly shortened version of the coding instructions can be found in

Authors: Carletta, J. C., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G. and Anderson, A.

Title: The Reliability of a Dialogue Structure Coding Scheme. Computational Linguistics, 23, 13-31. 1997

Number of annotators:

Main Map Task corpus was annotated using four different coders. In all, at least 50 people have tried the scheme, with around a dozen research projects employing it. Most of the annotators were PhD students in linguistics or psychology; one was just someone with a degree in modern languages.

Number of dialogues:

128 in the original Map Task corpus (English) , plus at least as many again coded using the same scheme or minor variants, comprised of Map Task in other languages, in other conditions (audio-only, video-mediated, children), and dialogues for other tasks (e.g., travel planning, financial services simulations, simpler children's tasks).

Evaluations of scheme:

Full results published in Carletta, J. C., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G. and Anderson, A. (1997) The Reliability of a Dialogue Structure Coding Scheme. Computational Linguistics, 23, 13-31.

For act segmentation, κ=.92 (N=4079,k=4).

For act classification on main corpus, κ=.83 (N=563,k=4).

For act classification using naive coders and written instructions, κ=.67

(N=139,k=3; agreement κ =.69 when coding developer added to pool).

For main distinction between initiation, response, or ready, using naive coders, κ=.84.

Disagreements were between CHECK and QUERY-YN, INSTRUCT and CLARIFY, and ACKNOWLEDGE, READY, and REPLY-Y.

Underlying task:

Linguistically motivated, but developed on map task (and therefore likely to be missing categories for goal negotiation).

List of phenomena annotated:

Primarily dialogue acts, but the papers also describe coding and reliability for higher level discourse structure built from the acts, in terms of goal-oriented dialogue games and transactions relating to dialogue planning divisions. These levels are not part of the MATE specification.

- Initiating Moves
- *Instruct*
- *Explain*
- *Check*
- *Align*
- *Query-YN*
- *Query-W*
- Response Moves
- *Acknowledge*

- *Reply-Y*
- *Reply-N*
- *Reply-W*
- *Clarify*
- *Ready* Move

Examples:

*TA15
*A 3 3,4

*E 7 IG instruct

And go up to about the middle of the map.

*M instruct

*TB16

B 7,

The middle of the map.

*M acknowledge

*TA 17

And stop.

*M instruct

Markup language:

Current mark-up specified in an HCRC internal document; SGML-conformant and based on the TEI.

Existence of annotation tools:

No tools publicly available; in-house tools for move coding operate in python using LT-XML and Tk, and in Microsoft Word. Nb can be configured to perform the annotation (implemented in TCL/Tk).

No semi-automatic annotation available.

Usability:

Has been used to structure the dialogue planning element of an SDS, to learn how to mark dialogue moves based on topic spotting, and to train the relationship between prosody and move type.

Contact address:

Maptask@cogsci.ed.ac.uk

A.23 Nakatani et al.'s Coding Scheme

(Harvard University, USA and AT&T Bell Laboratories, USA)

Coding book:

ftp://sls-ftp.lcs.mit.edu/pub/multiparty/coding_schemes/nakatani
Author: Christine H. Nakatani, Barbara J. Grosz, David D. Ahn and Julia Hirschberg (1995)

Title: "Instructions for Annotating Discourses". Technical Report Number TR-21-95. Center for Research in Computing Technology, Harvard University: Cambridge, MA.

Number of annotators:

A team of six annotators was trained to use the manual for the project on the Boston Directions Corpus at Harvard University, involving the authors of the manual. The annotators did not have any linguistic backgrounds, intentionally. "Naive" users have been desired to provide "unbiased" codings (compared to codings done by the researchers themselves for example).

Number of annotated dialogues:

Approx. 72 direction-giving monologues have been coded, from four different speakers. The coding was done while listening to the speech. The language is American English. The monologues have been broken down into intermediate prosodic phrases for discourse coding.

Evaluation of scheme:

The scheme is the result of augmenting and refining instructions given to students in discourse classes; in this sense, there has been evaluation of the scheme. Statistical/quantitative evaluation is done but not published yet.

Underlying task:

The scheme is not meant to be limited to any particular task or purpose. However,

it is mainly applied to direction-giving. The scheme is not meant for conversational speech without clear communicative intentions.

List of phenomena annotated:

The scheme aims at annotating *discourse segment purposes*, that is, the reasons why the speaker utters a given discourse segment. The purpose of each segment is described at the start of the segment, on a line that begins with a simple WHY? tag. Purposes are figured out by making reference to annotators own background knowledge and general intelligence. Annotators are advised to use the most possible specific expression that suitably describes the speaker's purpose, and thus to prefer a description like "Give tip on removing vein under faucet" instead of an expression like "Explain rinsing/washing of vein". In general, one segment is associated with one purpose, but a segment can be related to many purposes, and vice versa.

Purposes corresponding to different discourse segments are hierarchically organized, from the WHY? for the discourse overall to the smaller subsidiary purposes of smaller segments. Segments range from the whole dialogue/discourse to sentences; adverbial/prepositional phrases (called *mini-segments*) that supply additional information are not labeled with a WHY? tag. There are no rules about the number of subsegment allowed within one segment. Segments/purposes at the same level do not need to be at the same level of detail or about the same kind of information. Segments/purposes at the same level may not be directly related to each other, but must be related to their immediately higher segment/purpose. Two consecutive phrases may or may not share the same purpose: if they do, their purposes belong to the same level; if they don't, this means that one of the two purposes is subsidiary to the other and thus one of the two phrases starts an embedded subsegment. Discontinuous segments (as for digressions, asides, elaborations etc., which suspend the current topic flow) appear as a subsegment within a bigger segment "wrapped" around it.

Examples:

WHY? Teach new cook how to make stuffed sole

We're going to be making sole, stuffed with shrimp mousse.

WHY? Explain steps of initial preparation of ingredients and equipment

WHY? Identify ingredients

In the small bag is the sole and the shrimp.

And there are ten small sole fillets and there's half a pound of medium shrimp

WHY? Instruct new cook to get equipment ready.

Okay, and you're going to need a blender to make the mousse. So you should get your blender out.

WHY? Explain how to make shrimp mousse

Okay, the first thing we want to do, we should do is we should make the shrimp mousse.

WHY? Tell how to prepare shrimp

And, what you want to do is you want to take the shrimp, okay and

you want to peel and devein them.

WHY? Describe peeling

Okay, what you do is you peel the outer shell off.

WHY? Describe deveining processes

WHY? Tell how to find vein by cutting

Okay, and then you hold the shrimp and
you run a knife down the outside,
it's like the back of the shrimp, okay,
just cut in about a sixteenth of a inch.
What you'll see, is there'll be a vein, there.

WHY? Tell how to remove vein

Okay, it, it'll either be a pinkish vein or a black vein.

WHY? Explain removal of pink vein

Okay, if there's a pink vein you can just pull it out,

WHY? Explain removal of dark vein

Okay, if there's a dark colored vein, you can, you
wash that out. Run your thumb down one of your
fingers down the back to get that out.

WHY? Give tip on removing vein under faucet

And you know, what I usually do is, to rinse or
wash out the vein, I just hold the shrimp under
the sink, under the uh, the faucet. I cut it and then
I put it under the faucet.

WHY? Explain how to blend shrimp and other ingredients to make mousse

Okay now um, let's see, take the shrimp and place the shrimp in the blender.

...

WHY? Describe how to prepare sole for "stuffing"

Now, get out a large casserole, like a nine by twelve.

...

Now you want to place five of the um, the sole fillets side by side in the baking dish.

WHY? Explain how to "stuff" sole with shrimp mousse

Okay, and now you take the shrimp mousse and you uh,

you place a fifth of the mousse on each of the fillets.

...

Use all the mousse. Spread it evenly over each fillet.

Mark-up language:

N.b.'s mark-up language. This is not fully compliant with SGML, but a program is distributed with Nb that converts Nb-annotated files into standard SGML files.

Existence of annotation tools:

N.b. Tcl/Tk interface by G. Flammia.

Usability:

Boston Directions Project, also in the work on the intonational correlates of discourse structure (Barbara Grosz, Julia Hirschberg, Christine Nakatani).

Contact person:

Christine Nakatani (chn@research.att.com)

A.24 SLSA

(Göteborg University)

Coding books:

a) feedbacks and turn management:

Authors: Joakim Nivre, Jens Allwood, Elisabeth Ahlsén

Title: Interaction Management

b) own communication management:

Authors: Jens Allwood, Elisabeth Ahlsén, Joakim Nivre, Staffan Larsson

Title: Own Communication Management (Swedish)

Number of annotators:

7, linguists (but often undergraduate students)

Number of annotated dialogues:

around 100 (Swedish)

Evaluations of scheme:

Reliability tests have been done but the results are not public available.

Underlying task:

- domain dependent (courtroom interaction)
- linguistically motivated - distinguishing between
- social activity types
- speaker types
- subsequences
- medium
- written vs. spoken
- human partner vs. non-human partner
- face-to-face vs. mediated

List of phenomena annotated:

- for feedbacks and turn management:
- grammatical categories
- structural operation (phonological, morphological, contextual)
- position (of structural units within the FBU/within a large utterance)
- function of FBUs
- for own communication management
- structure (pause, word, phrase, repetition, etc.)
- function (mainly choice vs. change)
- communicative acts in a courtroom interaction
- holistic communication act
- main expressive function
- main evocative function
- obligations
- relations on preceding contributions

Examples:

A: kommer du

(Are you coming?)

B: ja

(Yes.)

A: kan du [1 ta me en] 1 penna

(Can you take a pencil with you?)

B: [1 va sa du] 1

(What did you say?)

B: okey // vill du ha en egen

(All right, would you like to have your own one?)

A: ja de vill ja

(Yes, please!)

Mark-up language:

The mark-up language is not based on one of the standard mark-up languages as it was felt that they are lacking in expressive power. Instead it is based on logic as an annotation language.

Existence of annotation tools:

Tractor (developed at Göteborg University): works on iX-Window UNIX and Power Macs

- possibility of identifying structural and descriptive levels
- possibility of distinguishing between multi-valued and single valued scheme
- possibility of distinguishing between properties and relations
- fair reliability support
- support of automatic analysis
- allowance of hierarchical schemes
- allowance of easy inclusion of different, alternative schemes

Usability:

Used in the SLSA system.

Contact person:

Joakim Nivre (nivre@ling.gu.se)

A.25 Switchboard SWBD-DAMSL

(University of Colorado)

Coding book:

http://stripe.Colorado.EDU/~jurafsky/manual.august1.html

Authors: Dan Jurafsky, Liz Shriberg, Debra Biasca

Title: Switchboard SWBD-DAMSL, Ahallow-Discourse-Function Annotation; Coders Manual, Draft 13

Number of annotators:

9 (linguists)

Number of dialogues annotated:

1155 conversations / over 250 000 utterances / 1.4 million words

Evaluation of scheme:

$0.80 < \kappa < 0.84$

Underlying task:

telephone conversation between strangers

List of phenomena annotated:

- Communicative-Status
- Uninterpretable (%): *But, uh, yeah.*
- Non-verbal (x): *[Laughter]*
- Abandoned or Turn-Exit (% -): *So,-*
- Self-talk (t1): *What's the world I'm looking for...*
- 3rd-party-talk (t3): *My goodness, Diane, get down from there.*
- Forward-Communicative-Function
- Statement
- Statement-non-opinion (sd): *Me, I'm in the legal department.*
- Statement-opinion (sv): *I think it's great.*
- Influencing-addressee-future-action
- Yes-No-Question (qy): *Do you have to have any special training?*
- Wh-Question (qw): *Well, how old are you?*

- Open-Question (qo): *How about you?*
- Or-Clause (qrr): O*r is it more of a company?*
- Declarative Yes-No-Question (qy^d): *So you can afford to get a house?*
- Declarative Wh-Question (qw^d): *You are what kind of buff?*
- Tag-Question (^g): *Right?*
- Action-directive (ad): *Why don't you go first?*
- Backchannel in question form (bh): *Is that right?*
- Rhetorical-Questions (qh): *Who would steal a newspaper?*
- Committing-speaker-future-action
- Offers, Options Commits (oo,cc,co): *I'll have to check that out.*
- Other-forward-function
- Conventional-opening (fp): *How are you?*
- Conventional-closing (fc): *Well, it's been nice talking to you.*
- Thanking (ft): *Hey thanks a lot.*
- Apology (fa): *I'm sorry.*
- Backwards-Communicative-Function
- Agreement
- Agree/Accept (aa): *That's exactly it.*
- Maybe/Accept-part (aap/am): *Something like that.*
- Reject (ar): *Well, no.*
- Hold before answer/agreement (^h): *I'm drawing a blank.*
- Understanding
- Signal-non-understanding (br): *Excuse me?*
- Response Acknowledgement (bk): *Oh, okay.*
- Repeat-phase (b^m): *Oh, fajitas.*
- Collaborative Completion (^2): *Who aren't contributing?*
- Acknowledge (b): *Uh-huh.*
- Summarize/reformulate (bf): *Oh, you mean you switched schools for the kids.*
- Appreciation (ba): *I can imagine.*
- Downplayer (bd): *That's all right.*
- Answer
- Yes answers (ny): *Yes.*
- No answers (nn): *No.*

- Affirmative non-yes answers (na,ny^e): *It is.*
- Negative non-no answers (ng,nn^e): *Uh, not a whole lot.*
- Other answers (no): *I don't know.*
- Dispreferred answers (arp,nd): *Well, not so much that.*
- Other
- Quotation (^q): *You can't be pregnant and have cats.*
- Hedge (h): *I don't know if I'm making any sense or not.*

Examples:

ad	A63 utt2:	{C and } think [what, + what's] is going to be like for [youngest,] + [an +] my oldest] son, when he goes to school.
qh	A.63 utt3:	What's going to happen? /
sd	A.63 utt4:	{E I mean } [I, + I'm] afraid for him to go. /

Mark-up language:

Variant of DAMSL

Existence of annotation tool:

None. Utterances are hand-labeled.

Usability:

Used for training stochastic discourse grammars so as to build better Language Models (LM) for Automatic Speech Recognition (ASR) of Switchboard. A variant of SWBD-DAMSL is CLARIFY currently developed at the Carnegie Mellon University (Lori_Levin@alexis.boltz.cs.cmu.edu).

Contact person:

Daniel Jurafsky (jurafski@colorado.edu)

A.26 Traum's Coding Scheme

(Université de Genève, Switzerland)

Coding book:

ftp://sls-ftp.Lcs.mit.edu/pub/multiparty/coding_schemes/traum
Author: David Traum

Title: Coding Schemes for Spoken Dialogue Structure

Number of annotators:

2 for the inter-turn coherence coding

1 (the author) for the grounding coding

Number of annotated dialogues:

26 from the TRAINS-93 corpus for the inter-turn coherence coding. (English)

10 from the TRAINS-91 corpus for the grounding coding. (English)

Evaluation of scheme:

Done but no results published.

Underlying task:

The scheme is designed for application to any kind of dialogue; actual application is for task-oriented dialogues.

List of phenomena annotated:

- **Relatedness**
- **e** Explicit Acknowledgment
- **Related**
- **0** related to the most recent utterance by the previous speaker
- **1** related to the utterance one previous to the most recent but not related to the most recent
- **2** related to utterance two previous to the last one (and not to anything more recent)
- **etc**. higher numbers for related utterances further back
- **,** related to previous material by the other speaker, but it is unclear to the coder whether they are related to the immediately previous utterance unit or to an utterance further back.
- **u** Unrelated to previous speech by the old speaker
- **?** Uncertain whether these utterances relate to previous speech by the other speaker

- **u-e** Unrelated following an explicit acknowledgement
- **1-e** Related to the penultimate utterance unit by the other speaker, when last utterance contained just an explicit acknowledgement
- **Grounding Acts**
- **Initiate:** begins a new DU (Discourse Unit), content separate from previous uncompleted DUs
- **Continue:** continuation of previous material by the same speaker
- **Acknowledge**: demonstrates or claims understanding of previous material by other speaker
- **Repair:** correct (potential) misunderstanding of DU content
- **Request Repair**: signal of lack of understanding
- **Request Ack**: signal for other to acknowledge
- **Cancel**: top work on DU, leaving it ungrounded and ungroundable
- **Surface Form**
- **Declarative**
- **Interrogative**
- **Imperative**
- **Fragment**
- **Cue Word**
- **Non-linguistic**
- **Illocutionary Function**
- **INFORM:** The speaker provides new information (including providing requested information when answering a question)
- **YNQ:** The speaker asks a yes-no question, trying to determine the polarity of a proposition
- **CHECK:** The speaker is attempting to verify that a certain (suspected) proposition is true
- **WHQ:** The speaker asks a wh-question, trying to determine the value of some term in a proposition
- **SUGGEST:** The speaker proposes a new item (action, proposition, plan constraint)
- **REQUEST:** The speaker aims to get the hearer to perform some action
- **ACCEPT:** The speaker agrees to a prior proposal by the hearer
- **REJECT**: The speaker rejects a prior proposal by the hearer
- **SUPP-INF:** The speaker provides additional information (perhaps already known) that augments, or help the hearer interpret some other accompanying speech act. this is usually performed by a subordinate clause or appositive phrase.
- **SUPP-SUG**: The speaker makes a supplementary suggestion of content, which is presupposed to be part of the plan by other accompanying suggestion or request. This is often performed by a purpose clause.

- **EVAL**: The speaker provides an evaluation of some item. This includes both factual evaluations, (e.g. of the likelihood of success of an action), and personal evaluations, describing how the speaker feels about something
- **GREET**: The speaker seeks to establish connection, e.g. by saying 'hello', or naming the hearer at the beginning of a dialogue.
- **APOLOGIZE:** The speaker apologizes for some action (e.g., speaking out of turn) or mistaken interpretation.
- **Argumentation Structure**
- (: so *act1 act2*): act1 is relevant to the interpretation of act2. If act2 is an informational act, then the truth of its content should be partially supported by act1. If act2 is a suggestion, then the suggestion should be about (a part of) the plan dominated by act1.
- (: and *act1 act2*): the interpretation of act2 is connected to act1 in some way to form a coherent whole. If act2 is a suggestion, then it should be part of the same plan as act1.
- (: and-then *act1 act2*): the interpretation of act2 is connected to act1 in some way to form a coherent whole. If act2 is a suggestion, then it should be part of the same plan as act1. In addition, act2 should temporally follow act1.
- (: purpose *act form*): act is to be done for the purpose of achieving form
- (: background *act1 act2*): act1 is performed for the purpose of making act2 more clear to the hearer.

Examples:

u: so we have to start in Avon

s: okay

u: how long does it take to bring engine one to Dansville

S: three hours

u: okay <sil> and then <sil> back to Avon to get the bananas

S: three more hours si(x) - six in all

u: how long does it take to load the bananas

UU#	Speaker	Utterance	grounding act label
31.9	M	it would get there at 3,	
31.10		is that what you're saying?	repair
32.1	S	it would get there at 4.	
33.1	M	it would get there at 4.	

Mark-up language:

N.B.'s mark-up language. This is not fully compliant with SGML, but a program is distributed with Nb that converts Nb-annotated files into standard SGML files.

Existence of annotation tools:

N.b. Tcl/Tk interface by G. Flammia.

Usability:

TRAINS-93 system

Contact person:

David Traum (traum@cs.umd.edu)

A.27 Verbmobil

(DFKI)

Coding book:

http://www.dfki.de/cgi-bin/verbmobil/htbin/doc-access.cgi

Authors: Jan Alexandersson, Bialnka Buschbeck-Wolf, Tsutomu Fujinami, Elisabeth Maier, Norbert Reithinger, Birte Schmitz, Melanie Siegel;

Title: Dialogue Acts in VERBMOBIL-2

REPORT 204, Mai 1997

Number of annotators:

3 (all non-linguists)

Number of annotated dialogues:

English: 252 (1869 turns, 4923 segments), Japanese: 400 (7057 turns, 8747 segments), German: 520 (5530 turns, 13932 segments)

Evalutaions of schemes:

For the dialogue scheme used in VERBMOBIL-1 we achieved κ=.83 for 10 presegmented dialogues labelled by two coders with equal expertise. The value for stability over time was κ=.84 for the same coder labelling identical dialogues with a time of about one year being between the two experiments.

Underlying task:

Appointment scheduling

List of phenomena annotated:

ACCEPT, CLARIFY, CLARIFY_ANSWER, CONFIRM, CONVENTION, DELIBERATE, DEVIATE_SCENARIO, DIGRESS, EXPLAINED_REJECT, GREETING, GREETING_BEGIN, GREETING_END, FEEDBACK, FEEDBACK_BACKCHANNELING, FEEDBACK_NEGATIVE, FEEDBACK_POSITIVE, GIVE_REASON, INFORM, INIT, INTRODUCE, NOT_CLASSIFIABLE, POLITENESS_FORMULA, REFER_TO_SETTING, REJECT, REQUEST, REQUEST_CLARIFY, REQUEST_COMMENT, REQUEST_SUGGEST, SUGGEST, THANK

Examples:

RMW002: <:<# <#Klicken <A <#Klicken how 'bout <;comma <#Klicken <#Klicken at <;comma three on <;comma <A February third <;quest <A <;seos @SUGGEST

#Rascheln would that be okay <;quest <# <#Klicken <# <# <;seos @REQUEST_COMMENT

Mark-up language:

Verbmobil

Existence of annotation tools:

AnnoTag

- manual
- implementation language: Tcl/Tk

Usability:

Verbmobil system

Contact person:

Norbert Reithinger (Norbert.Reithinger@dfki.de)

Morpho-syntactical Schemes

In this section we will overview some existing corpus annotation schemes for both morphosyntax and syntax. We will consider them insofar as they have something of interest to say about typical problems encountered in dialogue annotation in connection with the following typology of phenomena:

- word-level classification issues
- adverbs, interjections, interactional markers
- word partials, non standard forms
- segmentation issues
- pauses, hesitators
- multi-words
- phrase partials
- trailing off, interruption, completion
- retrace-and-repair sequences
- anacolutha (syntactic blending)

This typology says nothing about whether the phenomena considered are classifiable as disfluent material or should rather be taken as germane linguistic phenomena characteristic of speech and not of writing. The classificatory perspective entertained here lays emphasis on the impact that the listed phenomena are likely to have on issues of annotation: e.g. if they would simply require introduction of an extra part of speech category, or if they are rather bound to have repercussions on syntactic parsing and segmentation issues in general.

Note that, in some cases, the same phenomenon can be treated under two different headings: interactional markers, for example, pose both a problem of categorial classification (how should they be labelled?) and an issue of segmentation, when they happen to be multi-word units (e.g., is 'I see' in its interactional usage to be treated as a single morphosyntactic unit, or should it rather be treated as a complex syntactic constituent?). Clearly, the two perspectives interact to a large extent.

Not all the annotation schemes overviewed here have actually explicitly addressed all problems in our list. Most of them simply came up with interesting practices which can easily/usefully be extended to dialogue annotation proper with a view to the treatment of such phenomena. For example, we will mention here Eagles 1996 recommendations on both morphosyntax and syntax annotation, although they were initially intended to deal with written material only. As pointed out in Leech et al. 1998, they can in fact be taken as a useful starting point for dialogue annotation too, with the proviso that a certain amount of customization be carried out. Hopefully, this should pave the way to the ultimate integration of practices in the scientific communities of NLP and speech.

A.28 Childes

Coding book:

Information about the purpose and domain of the CHAT system as well as instructions for use are described in MacWhinney (1994).

Number of annotators:

The CHAT system is a widespread standard system for the transcription and coding of child language in many European and non European languages. Approximately 60 groups of researchers around the world are currently actively involved in new data collection and transcription using the CHAT system. As a consequence of its widespread use, it is impossible to calculate the exact number of annotators.

Number of annotated dialogues:

A huge number of dialogues has been/is being annotated with the CHAT coding scheme. This number exceeds the amount of dialogues in the database, as many projects concerning child language make use of CHAT without contributing to the overall CHILDES database. The internationally recognized CHILDES database (http://sunger2.uia.ac.be/childes/database.html) includes transcripts from over forty major projects in English and additional data from 19 other languages. The additional languages are Brazilian Portuguese, Chinese (Mandarin), Chinese (Cantonese), Danish, Dutch, French, German, Greek, Hebrew, Hungarian, Italian, Japanese, Mambila, Polish, Russian, Spanish, Swedish, Tamil, Turkish, and Ukrainian. The total size of the database is now approximately 160 million characters (160 MB). Full documentation about the database can be found at http://sunger2.uia.ac.be/childes/database.pdf.

Evaluations of scheme:

As a result of its worldwide use, CHAT is continuously evaluated and updated to meet the needs of different languages and different users. We are not aware of statistical/quantitative evaluations of its reliability.

Underlying task:

Being first created as a tool for the study of language acquisition, the data collected mainly refer to parent-to-child or child-to-child spontaneous conversations, task-oriented dialogues in play andstory-telling situations.

Some of the data coded by CHAT also include second language learners andadults recovering from aphasic disorders.

List of phenomena annotated:

See below.

Examples:

See below.

Mark-up language:

CHAT's own format.

Existence of annotation tools:

The CHILDES system contains several separate, yet integrate, programs which are clustered around two major tools. The first tool is a full-fledged and ASCII-oriented editor (CED, Childes EDitor), specifically designed to facilitate the editing of CHAT files and to check for accuracy of transcriptions. CED also allows the user to link a full digitized audio recording of the interaction directly to the transcript. This is the system called "sonic CHAT". The CED editor is currently being extended to facilitate its use with videotapes. The plan is to make available a floating window in the shape of a VCR controller that can be used to rewind the videotape and to enter time stamps from the videotape into the CHAT file. An alternative way of analyzing video is to record from tape onto QuickTime movies and to link these digitized movies to the transcript.

The second tool, actually a bunch of several smaller tools, is a set of computer programs called CLAN (Child Language ANalysis) which serves different analysis purposes. The full system is presented in detail in MacWhinney (1991) and illustrated through practical examples in Sokolov and Snow (1994).

Usability:

CHAT-encoded databases have been set up as a result of nearly a hundred major research projects in 20 languages. New databases are continuously being set up worldwide.

Contact person:

Brian MacWhinney (macw@cmu.edu)

1 Word-Level Classification Issues

CHAT makes provision for two physically and in part also conceptually distinct ways of encoding morphological information in a corpus: i) morpheme splitting on the 'main line', that is the line of orthographic transcription, ii) morphological categorization on the 'morphology line', that is a separate tier of encoding specifically devised for containing morphological information.

In order to indicate the ways that words on the main line are composed from morphemes, CHAT uses the symbols -, +, #, ~, &, and 0: they are all used as concatenative operators and accordingly placed between two consecutive morphemes. These same six symbols are also used for parallel purposes on the morphology line, where these symbols form a part of a more extensive system.

Morphemization on the main line is intended mostly for initial morphemic analysis or general quantitative characterization of morphological development. For more thorough analyses the morphology line is strongly recommended, especially for languages other than English.

The basic scheme for coding of words on the morphology line is:

'part-of-speech' |

'pre-clitic' ~

'prefix' #

'stem'

= 'english translation'

& 'fusionalsuffix'

- 'suffix'

~ 'post-clitic'

where the gloss between quotes indicates the content and position of corresponding encoded information relative to the symbol/operator. For example, part-of-speech information precedes '|', while fusional suffix follows '&'. Furthermore the delimiter '**+**' is used between words in a compound (see infra).

The order of elements after the | symbol is intended to correspond to the linear order of morphemes within the word, as shown by the following example:

`'sing-s' v|sing-3s`

There are no spaces between any of these elements. The English translation of the stem is not a part of the morphology, but is included here for convenience in retrieval and data entry. The morphological status of the affix is identified by the type of delimiter.

In particular, '**&**' is used to signal that the affix is not realized in its usual phonological shape. For example, the form "men" cannot be broken down into a part corresponding to the stem "man" and a part corresponding to the plural marker "s", hence it is coded as **n|man&PL**. Similarly, the past forms of irregular verbs may undergo ablaut processes, e.g. "came", which is coded **v|come&PAST**, or they may undergo no phonological change at all, e.g. "hit", which is coded **v|hit&PAST** Sometimes there may be several codes indicated with the **&** after the stem. For example, the form "was" is coded **v|be&PAST&13s**.

Clitics are marked by a tilde, as in **`v|parl=speak&IMP:2S~pro|DAT:MASC:SG`** for Italian "parlagli" and **`pro|it~v|be&3s`** for English ”it’s.” Note that part of speech coding is repeated for clitics. Both clitics and contracted elements are coded with the tilde. The use of the tilde for contracted elements extends to forms lijke ”sul” in Italian, ”ins” in German, or ”rajta” in Hungarian in which prepositions are merged with articles or pronouns.

1.1 Adverbs, Interjections, Interaction Markers

The category 'communicator' is used in CHAT for interactive and communicative forms which fulfil a variety of functions in speech and conversation. Many of these are formulaic

expressions such as **hello**, **good+morning**, **good+bye**, **please**, **thank+you**. Also included in this category are words used to express emotion, as well as imitative and onomatopeic forms, such as **ah**, **aw**, **boom**, **boom-boom**, **icky**, **wow**, **yuck**, **yummy**.

1.2 Pauses, Hesitators

Pauses are treated in CHAT on the prosodic annotation tier. Pauses that are marked only by silence are coded on the main line with the symbol #. The number of # symbols represents the length of pauses. Alternatively, a word after the symbol # is added to estimate the pause length, as in #long.

Example:

*SAR: I don't # know -.

*SAR: #long what do you ### think -?

CHAT allows coding of exact length of the pauses, with minutes, seconds, and parts of seconds following the #.

Example:

*SAR: I don't #0_5 know -.

*SAR: #1:13_41 what do you #2 think -?

1.3 Word Partials, Non Standard Forms

When an item on the main line is incorrect in either phonological or semantic terms it is marked by a following '[*]'. The coding of that item on the morphology line should be based on its target, as given in the 'error line'. If there is no clear target, the form should be represented with 'xxx', as in the following example:

***PAT:** **the catty [*] was on a eaber [*].**

%mor: **det|the *n|kitty v|be&PAST prep|on**

det|a *n|xxx.

%err: **catty = kitty $BLE $=cat,kitty ; eaber = [?]**

In this example the symbol '*' on the morphology line indicates the presence of an incorrect usage, in this case due to blending two different words into one. The detailed analysis of this error should be conducted on the 'error line'. Errors involving segmentation issues (such as omission of a syntactically obligatory unit etc.) will be treated in the following section.

A non standard or incorrect usage can be encoded directly on the main line by trailing after it the replacing standard form in square brackets: example, **gonna [: going to]**. The material on the %mor line corresponds to the replacing material in the square brackets, not the material that is being replaced. For example, if the main line has **gonna [: going to],** the %mor line will code **going to.**

Some special characters are intended to give information about, for example, babbling, child-invented forms, dialect forms, family-specific forms, filled pauses, interjections, neologisms, phrasal repetitions, or other general special forms, according to the following conventions. Note that recording of these phenomena is not made at the coding level, but at the transcription level.

Letters	Categories	Example	Meaning	Coded Example
@b	babbling	Abame		abame@b
@c	child-invented form	Gumma	sticky	gumma@c
@d	dialect form	Younz	you	younz@d
@f	family-specific form	Bunko	broken	bunko@f
@fp	filled pause	Huh	-	huh@fp
@i	interjection, interactional	Uhhuh	-	uhhuh@i
@l	letter	B	letter b	b@l
@n	neologism	Breaked	broke	breaked@n
@o	onomatopeia	woof woof	dog barking	woof@o
@p	phonol. consistent forms1	Aga	-	aga@p
@pr	phrasal repetition	its a, it's a	-	its+a@pr
@s	second-languag e form	Istenem	my God	istenem@s
@sl	sign language	apple sign	apple	apple@sl
@	general special form	Gongga	-	gongga@

2 Segmentation Issues

2.1 Multi-Words

Those compounds that are usually written as one word, such as "birthday" or "rainbow," should not be segmented. Those compounds that are generally separated by a hyphen in English orthography are separated by a + symbol in CHAT transcription (e.g., "jack-in-the-box" should

be transcribed as ”jack+in+the+box”). Rote forms to be counted as a single morpheme may also be joined with a + symbol (e.g., all+right).

Multi-word expressions which are concatenated through a '+' are assigned a unique part-of-speech tag at the level of mrophosyntax. For example, the following idiomatic phrases can be coded: **qn|a+lot+of**, **adv|all+of+a+sudden**, **adv|at+last**, **co|for+sure**, **adv:int|kind+of**, **adv|once+and+for+all**, **adv|once+upon+a+time**, **adv|so+far**, and **qn|lots+of**.

2.2 Error Coding

The symbol *0 is used in CHAT to indicate omission (recall that the symbol * is used to indicate incorrect usage), as in the following examples:

***CHI:** **dog is eat.**

%mor: ***0det|the n|dog v:aux|be&PRES v|eat-*0PROG.**

***PAT:** **the dog was eaten [*] the bone.**

%mor: **det|the n|dog v:aux|be&PAST&3S v|eat-*PERF det|the n|bone.**

%err: **eaten = eating $MOR $SUB**

Here is an example of coding on the morphology line that indicates how the omission of an auxiliary is coded:

***BIL:** **he going.**

%mor: **pro|he *0v|be&3S v|go-prog.**

Note that the missing auxiliary is not coded on the main line, because this information is available on the morphology line. If a noun is omitted, there is no need to also code a missing article. Similarly, if a verb is omitted, there is no need to also code a missing auxiliary.

The CHAT system for error coding has the following features:

1. it indicates what the speaker actually said, or the erroneous form

2. it indicates that what the speaker actually said was an error

3. it allows the transcriber to indicate the target form

4. it facilitates retrieval, both toward target forms and actually produced forms

5. it allows the analyst to indicate theoretically interesting aspects of the error by delineating the source of the error, the processes involved, and the type of the error in theoretical terms (on the error line)

2.3 Phrase Partials

In CHAT, the syntactic role of each word can be notated before its part-of-speech on the morphology line. To capture syntactic groupings, provision is made for coding syntactic structure on the syntactic line. Clauses are enclosed in angle brackets and their type is indicated in square brackets, as in the following example:

***CHI: if I don't get all the cookies you promised to give**

%syn: <C S X V M M D < S V < R V I > [CP] > [RC] > [CC] <

me, I'll cry.

S V > [MC].

In this notation, each word plays some syntactic role. The rules for achieving one-to-one correspondence to words on the main line apply to the syntactic line also. Higher order syntactic groupings are indicated by the bracket notation. The particular syntactic codes used in this example come from the following list. This list is not complete, particularly for languages other than English.

A	Adverbial Adjunct		V	Verb
C	Conjunction		X	Auxiliary
D	Direct Object		AP	Appositive Phrase
I	Indirect Object		CC	Coordinate Clause
M	Modifier		CP	Complement
P	Preposition		MC	Main Clause
R	Relativizer/Inf		PP	Prepositional Phrase
S	Subject		RC	Relative Clause

2.3.1 Trailing off, Interruption, Completion

An incomplete, but not interrupted, utterance, is marked with the "trailing off" '+=8A' symbol on the main line.

Example:

*SAR: smells good enough for +=8A

*SAR: what is that?

If the speaker does not really get a chance to trail off before begin interrupted by another speaker, then the interruption marker '+/.' is used instead. If the utterance that is being trailed off is a question, then the symbol '+..?' is used.

The symbol '+' can be used at the beginning of a main tier line to mark the completion of an utterance after an interruption. It is complementary to the trailing off symbol.

Example:

*CHI: so after the tower +...

*EXP: yeah.

*CHI: +, I go straight ahead.

Others' completion is marked through '++'. This symbol can be used at the beginning to mark "latching", or the completion of another speaker's. It is complementary to the trailing off symbol.

Example:

*HEL: if Bill had known +...

*WIN: ++ he would have come.

2.3.2 Retrace-and-Repair Sequences

Retracing without correction (simple repetition) [/] takes place when speakers repeat words or whole phrases without change. The retraced material is put in angle brackets.

Example:

*BET: <I wanted> [/] I wanted to invite Margie.

Several repetitions of the same word can be indicated in the following way:

*HAR: It's(/4) like # a um # dog.

Retracing with correction [//] takes place when a speaker starts to say something, stops, repeats the basic phrase, changes the syntax but maintains the same idea. Usually, the correction

moves closer to the standard form, but sometimes it moves away form it. The retraced material is put in angle brackets.

Example:

*BET: <I wanted> [//] uh I thought I wanted to invite Margie.

Retracing with Reformulation [///] takes place when retracings involve full and complete reformulations of the message without any specific corrections.

Example:

*BET: all of my friends had [///] uh we had decided to go home for lunch.

Unclear Retracing Type is marked by [/?].

CHAT distinguishes a False Start without retracing [/-], from false starts with correction. False starts with no retracing are dealt with in the following section. The symbols [/] and [//] are used when a false start is followed by a complete repetition or by a partial repetition with correction.

2.3.3 Anacolutha (syntactic blending)

If the speaker terminates an incomplete utterance and starts off on a totally new tangent, this can be coded with the [/-] symbol:

*BET: <I wanted> [/-] uh when is Margie coming?

Note that if this coding is not in contrast with the coding of incomplete utterances (either trailed off or interrupted); this uniquely depends on the decisions about what a coder wants to count as an utterance.

A.29 CHRISTINE (SUSANNE)

The CHRISTINE corpus is, for spoken dialogues, what SUSANNE was for written corpora: a carefully annotated collection of real spoken material of British English only.

The CHRISTINE project is using the structural annotation scheme defined for the SUSANNE Corpus (which is probably the most detailed thing of its kind yet produced). The definition of the SUSANNE scheme can be found in G. Sampson's book, "English for the Computer" (see Sampson, 1995). The EAGLES group asked for a copy of this book when it was in proof and its contents (Chapter 6 in particular, which deals with extending annotation to spoken material) played a significant part in their decisions (see Section D.3 in this report for further details). In the CHRISTINE project, the annotation rules of Chapter 6 are being redefined on the basis of experience in actually applying them to sizeable quantities of spontaneous spoken English. G. Sampson (personal communication) reports that in most respects what is being done is only adding to already existing rules, not changing them. Additional annotation rules are not, at the present stage, into a form fit to circulate yet.

The CHRISTINE project is due to be completed at the end of 1999. there may be a few months' "polishing" after that, but then or soon afterwards the annotated corpus will be made available freely to all comers in the same way that the SUSANNE Corpus already is.

Some documentation available at:

http://www.cogs.susx.ac.uk/users/geoffs/RChristine.html

A.30 EAGLES 1996-98

Coding book:

documentation available at

http://www.ilc.pi.cnr.it/EAGLES96/annotate/annotate.html

Number of annotators:

not available

Number of annotated dialogues:

not available

Evaluations of scheme:

indirect evaluation through instantiation in many different projects (see usability)

Underlying task:

standard development

List of phenomena annotated:

list of relevant phenomena provided below

Examples:

list of relevant examples provided below

Mark-up language:

not available

Existence of annotation tools:

EAGLES-conformant annotation tools developed in other projects

Usability:

schemes adopted in Multext, Sparkle, Parole

EAGLES is the ancestor of a family of standardization efforts for corpus annotation. It is then worth looking into EAGLES' methodology in some detail, as this will also offer a key to an understanding of the design and development of other Eagles-related annotation schemes.

Word-level classification issues

EAGLES provides a list of (major) morpho-syntactic categories.

1.	N [noun]	2.	V [verb]	3.	AJ [adjective]
4.	PD [pronoun/ determiner]	5.	AT [article]	6.	AV [adverb]
7.	AP [adposition]	8.	C [conjunction]	9.	NU [numeral]
10.	I [interjection]	11.	U [unique/ unassigned]	12.	R [residual]
13.	PU [punctuation]				

They represent the most general and obligatory level of morphosyntactic annotation, in the sense that any set of morphosyntactic tags is expected to convey at least information about morphosyntactic categories.

The set of Eagles category tags is not formally consistent, in that it does not provide a minimal set of mutually exclusive morphosyntactic classes. See, for example, the umbrella-category PD, including both determiners and pronouns, and its coexistence with the overlapping category AT for articles. Accordingly there is no general expectation that the mapping between the EAGLES category tags and a language specific instantiation of it should be one-to-one.

Morphosyntactic categories can further be specified by means of appropriate morphosyntactic features (such as gender, number, case etc.),expressed as supplementary tags. The combination of a category tag with its morphosyntactic feature specification yields complex tags of considerable length and granularity. As an illustration, we provide below the feature matrix for the category verb as detailed .

Verbs (V)

(i)	Person:	1. First	2. Second	3. Third	
(ii)	Gender:	1. Masculine	2. Feminine	3. Neuter	
(iii)	Number:	1. Singular	2. Plural		
(iv)	Finiteness:	1. Finite	2. Non-finite		
(v)	Verbform/ Mood:	1. Indicative	2. Subjective	3. Imperative	4. Conditional
		5. Infinite	6. Participle	7. Gerund	8. Supine
(vi)	Tense:	1. Present	2. Imperfect	3. Future	4. Past
(vii)	Voice:	1. Active	2. Passive		
(viii)	Status:	1. Main	2. Auxiliary		

Examples of use of this matrix are provided for what is called "Itermediate Tag Set", a specific instantiation of a subset of the list of categories above:

A 3rd person, singular, finite, indicative, past tense, active, main verb,

non-phrasal, non-reflexive, verb is represented: **V3011141101200**

Wherever an attribute is inapplicable to a given word in a given tag-set, the value **0** fills that attribute's place in the string of digits. When the **0**s occur in final position, without any non-zero digits following, they can be dropped.

Eagles makes provision for disjunctive specification of morphosyntactic categories in cases of i) genuine systematic ambiguity in a given language (e.g. present indicative and present subjunctive forms in English, or some past participles and adjectives in Italian), ii) practical demands of fully automatic tagging.

1.1 Adverbs, Interjections, Interactional Markers

The interjection and adverb categories are much broader and variegated than usually assumed in traditional grammar. Eagles 98 provides two illustrative lists of the level of granularity at which both categories can be subclassified, taken from Sampson (1995) and the London Lund Tagset respectively. In both cases a fine-grained functional or semantic analysis of the role of each subclass in dialogue interaction is presupposed. This aspect makes both proposals prohibitive for the purposes of automatic annotation. A practical strategy could be to add interjection to the Eagles inventory of part-of-speech categories and provide a rich feature matrix for subclassification, under the assumption that only the topmost attribute (part-of-speech) be disambiguated in automatic tagging.

1.2 Pauses, Hesitators

Eagles 98 recommends to treat pauses and hesitators as punctuation marks, to eventually be attached as high in the syntactc tree as possible during parsing.

1.3 Word Partials, Non Standard Forms

No specific recommendations are provided for word partials, and the suggestion is tentatively put forward to use the peripheral part-of-speech category U ('unique' or 'unassigned', see list above) for their tagging. Non standard forms (e.g. 'gonna') are recommended to be transcribed with standard spelling. Deviations from this practice should be documented and justified.

2 Segmentation Issues

2.1 Multi-Words

Eagles 98 leaves the matter open of whether multi-word units should be assigned a single tag or rather a multi-tag. Representation issues are not addressed either in any detail.

2.2 Error Coding

Coding of mistakes is neither envisaged nor excluded by Eagles 98 recommendations.

2.3 Phrase Partials

2.3.1 Trailing Off, Interruption, Completion

Eagles 98 provides a couple of illustrative examples of how syntactic incompleteness could be annotated. In the first one (drawn from the British National Corpus) syntactic incompleteness is annotated by means of a special marker (a slash following the non- terminal constituent label) tagging the incomplete constituent as a whole. In the second example (from Sampson 1995), no new label is introduced to mark the incomplete constituent, but only a place holder, '#', which marks the position of the missing element within the incomplete constituent.

It is emphasized that the examples provided are only indicative and should not be taken as standards in any way.

2.3.2 Retrace-and-Repair Section

Only one example is provided by means of illustration. Once more, it is drawn from Sampson 1995 and recast into an Eagles-conformant style. Both the retrace and the repair are within the minimal superordinate constituent, with the marker '#' used to signal the interruption point:

and that [NPs any bonus [RELCL he] # money [RELCL he gets over that]] is a bonus

It is not immediately clear from the example what word stretch the repair is meant to replace.

2.3.3 Anacolutha (syntactic blending)

Cases of syntactic blending are illustrated by means of a drastically incoherent sentence, annotated through maximal parse brackets to enclose the whole parsable unit, and no information about its internal structure. This is what the guidelines of the British National Corpus call 'structure minimization principle':

[and this is what the # the <unclear>] # [what's name now # now] # <pause> [that when it's opened in nineteen ninety-two <pause> the communist bock will be able to come through Germany this way in]

A.31 LE Sparkle

The syntactic annotation schemes developed within SPARKLE are an example of instantiation of Eagles recommendations at the morphosyntactic and syntactic levels, specifically geared towards the completion of two different tasks: i) use of morphosyntactically and syntactically annotated corpora for (semi)automatic acquisition of lexical information from them, and ii) use of annotated material for multi-lingual information retrieval and speeh recognition. Both tasks are being carried out on four different languages (namely English, French, German and Italian).

In Sparkle, bootstrapping lexical information from a corpus is modelled as the process of extracting typical contexts of usage of a given lexical item in a shallow-parsed corpus. The acquired information is eventually put to use by either providing a lexicalized version of the shallow parser, or by augmenting the lexicon of another independent parser. In both cases, the ultimate goal of the lexicalized parser is to provide the analysis of a sentence in terms of

functional relations holding between head words. Usefulness of this level of analysis is eventually assessed through industrial demonstrators for multilingual information retrieval and monolingual speech recognition.

Accordingly, Sparkle defines the following three possible levels of syntactic annotation:

i) **chunking**

ii) **phrasal parsing**

iii) **functional parsing**

In the following we will review in detail levels i) and ii) only.

Coding book:

documentation available at

http://www.ilc.pi.cnr.it/sparkle.html

Number of annotators:

>5

Number of annotated material:

600 annotated sentences of English, German and Italian

Evaluation of scheme:

Evaluation of automatic annotation over all levels available at:
http//www.ilc.pi.cnr.it/sparkle.html

Underlying Task:

Language modelling for Speech Recognition, Multilingual Information Retrieval

List of phenomena annotated:

List of relevant phenomena provided below.

Examples:

Provided below.

Mark-up language:

SPARKLE's own format.

Existence of annotation tool:

Software available for English, German and Italian.

Usability:

Speech Recognition and Multilingual Information Retrieval.

Contact Person:

Vito Pirrelli (vito@ilc.pi.cnr.it)

1 Word-level Classification Issues

SPARKLE did not develop a specific set of word-level tags, but it simply built on pre-existing part-of-speech Eagles96-conformant encoding schemes. A straightforward extension of these schemes should make provision for the additional tags needed to cover phenomena which are specific of dialogues.

2 Segmentation Issues

In SPARKLE, segmentation problems are dealt with differently, depending on which level of syntactic annotation one is considering. For the specific purposes of the present overview, we will limit ourselves to consideration of chunking and functional annotation only. This is done for ease of exposition, as these two levels, unlike complete phrase-structure trees, are clearly complementary, and exemplify two profoundly different perspectives on syntactic annotation: one based on the linear arrangement of word forms in a sentence and on the internal cohesion of relatively small syntactic islands, the other on an abstract representation of grammatical functions relative to a verb head. Traditionally, complete phrase-structure trees are assumed to simultaneously convey both types of information. For reasons that will be clear in a moment, syntactic annotation of dialogue favours a view whereby linear adjacency of word forms on the one hand and encoding of functional annotation on the other hand are to be dealt with separately.

1.1 Chunking in Sparkle

In what follows, we first exemplify the SPARKLE approach to chunking through detailed illustration of the Italian chunking scheme.

The typology of phrase chunks in the Italian chunking annotation scheme is summarised in the table below.

NAME	TYPE	POTGOV	EXAMPLES
ADJ_C	adjectival	Adj	*bello* ‘nice’,

NAME	TYPE	POTGOV	EXAMPLES
	chunk		*molto bello* 'very nice'
BE_C	predicative chunk	Adj past part	*è bello* '(it/(s)he) is nice', *è caduto* '(it/he) fell'
ADV_C	adverbial chunk	Adv	*sempre* 'always'
SUBORD_C	subordinating chunk	Conj	*quando* 'when', *dove* 'where'
N_C	nominal chunk	noun pron verb adj	*la mia casa* 'my house', *io* 'I', *questo* 'this', *l'aver fatto* 'having done', *il bello* 'the nice (one)'
P_C	prepositional chunk	Noun pron verb adj	*di mio figlio* 'of my son', *di quello* 'of that (one)', *dell'aver fatto* 'of having done', *del bello* 'of the nice (one)'
FV_C	finite verbal chunk	Verb	*sono stati fatti* '(they) have been done', *rimangono* '(they) remain'
G_C	gerundival chunk	Verb	*Mangiando* 'eating'
I_C	infinitival chunk	Verb	*per andare* 'to go', *per aver fatto* 'to have done'
PART_C	participial chunk	Verb	*finito* 'finished'

Table 1: Typology of phrase chunks

The following informal definitions are intended to make the assumptions underlying this schema fully explicit. More on this can be found in SPARKLE WP1 final report (Carroll et al. 1996), and related papers (Federici et al. 1996 and 1998).

ADJ_C

ADJ_Cs are chunks beginning with any premodifying adverbs and intensifiers and ending with a head adjective. This definition provides a necessary but not sufficient condition for identification of ADJ_C. In fact, adjectival phrases occurring in pre-nominal position are not marked as distinct chunks since their relationship to the governing noun is unambiguously identified within the nominal chunk (see example sentence above). The same holds in the case of predicate adjectival phrases governed by the verb *essere* 'be', which are part of BE_C (see below).

BE_C

BE_Cs consist of a form of the verb *essere* 'be' and an ensuing adjective/past participle including any intervening adverbial phrase. E.g.:

[BE_C *è intelligente* BE_C] '(he) is intelligent'

[BE_C *è molto bravo* BE_C] '(he) is very good'

[BE_C *è appena arrivato* BE_C] '(he) just arrived'

ADV_C

ADV_Cs extend from any adverbial pre-modifier to the head adverb. Once more, this definition provides a necessary but not sufficient condition for ADV_C. In fact, adverbial phrases that occur between an auxiliary and a past participle form are not identified as distinct chunks due to their unambiguous dependency on the verb. By the same token, adverbs which happen to immediately premodify verbs or adjectives are respectively part of a verbal chunk and an adjectival chunk. Finally, noun phrases used adverbially (e.g. *questa mattina* 'this morning') are treated as nominal chunks (see below). E.g.:

[FV_C *ha sempre camminato* FV_C] [ADV_C *molto* ADV_C] '(he) has always walked a lot'

[FV_C *ha finito* FV_C] [ADV_C *molto rapidamente* ADV_C] '(he) has finished very quickly'

SUBORD_C

SUBORD_Cs are chunks which include a subordinating conjunction. Subordinating conjunctions are chunked as an independent chunk in its own right only when they are not immediately followed by a verbal group. Compare, for example, the chunk structure of the following sentence

[FV_C *non so* FV_C] [SUBORD_C *quando* SUBORD_C] [N_C *il direttore* N_C] [FV_C *mi riceverà* FV_C] '(I) do not know when the director will receive me'

with the chunk structure of the following sentence, which differs from the previous one in having the subject of the subordinate clause in postverbal position:

[FV_C *non so* FV_C] [FV_C *quando mi riceverà* FV_C] [N_C *il direttore* N_C].

N_C

N_Cs extend from the beginning of the noun phrase to its head. They include nominal chunks headed by nouns, pronouns, verbs in their infinitival form when preceded by an article (i.e. Italian nominalised infinitival constructions) and proper names. Noun phrases functioning adverbially (e.g. *questa mattina* 'this morning') are also treated as nominal chunks. All kinds of modifiers and/or specifiers occurring between the beginning of the noun phrase and the head are included in N_Cs. E.g.:

[N_C *un bravo bambino* N_C] 'a good boy'

[N_C *tutte le possibili soluzioni* N_C] 'all possible solutions'

[N_C *i sempre più frequenti contatti* N_C] 'the always more frequent contacts'

[N_C *questo* N_C] 'this'

[N_C *il camminare* N_C] 'walking'

[N_C *il bello* N_C] 'the nice (one)'

In the chunking scheme, nominal chunks cover only a portion of the range of linguistic phenomena normally taken care of by nominal phrases: namely only noun phrases with prenominal complementation.

P_C

P_Cs go from a preposition to the head of the ensuing nominal group. Most of the criteria given for N_Cs also apply to this case. Typical instances of P_Cs are:

[P_C *per i prossimi due anni* P_C] 'for the next two years'

[P_C *fino a un certo punto* P_C] 'up to a certain point'

FV_C

FV_Cs include all intervening modals, ordinary and causative auxiliaries as well as medial adverbs and clitic pronouns, up to the head verb. E.g.:

verbal chunk with auxiliary or modal verb and medial adverb:

[FV_C *può ancora camminare* FV_C] '(he) can still walk'

verbal chunk with pre-modifying adverb:

[FV_C *non ha mai fatto* FV_C] [ADV_C *così* ADV_C] '(he) has never done so'

the auxiliary *essere* 'be' in periphrastic verb forms (whether active or passive) such as *sono caduto* 'I fell', *sono stato colpito* 'I was hit', or *mi sono accorto* 'I realized', is dealt with as part of a finite verb chunk, unless the verb *essere* is followed by a past participle which the

dictionary also categorises as an adjective; in the latter case it is chunked as a BE_C (see above).

[FV_C *è* FV_C] [N_C *un simpatico ragazzo* N_C] '(he) is a nice guy'

fronted auxiliaries constitute separate FV_Cs:

[FV_C *può* FV_C] [N_C *la commissione* N_C] [I_C *deliberare* I_C] [P_C *su questa materia* P_C]? 'can the Commission deliberate on this topic?'

periphrastic causative constructions:

[FV_C *fece studiare* FV_C] [N_C *il bambino* N_C] '(he) let the child study'

clitic pronouns are part of the chunk headed by the immediately adjacent verb:

[FV_C *lo ha sempre fatto* FV_C] '(he) has always done it'

G_C

G_Cs contain a gerund form. When part of a tensed verb group (e.g. in progressive constructions), the gerundival verb form is not marked independently. G_C also includes gerund forms functioning as noun phrases.

[FV_C *sta studiando* FV_C] '(he) is studying'

[G_C *studiando* G_C] [FV_C *ho imparato* FV_C] [ADV_C *molto* ADV_C] 'by studying (I) have learned a lot'

I_C

Infinitival chunks (I_Cs) include both bare infinitives and infinitives introduced by a preposition.

[FV_C *ha promesso* FV_C] [I_C *di arrivare* I_C] [ADV_C *presto* ADV_C] '(he) has promised to arrive early'

[FV_C *desidera* FV_C] [I_C *partire* I_C] [ADV_C *domani* ADV_C] '(he) wishes to leave tomorrow'

PART_C

A past participle chunk (PART_C) includes participial constructions such as:

[PART_C *finito* PART_C] [N_C *il lavoro* N_C] , [N_C *Giovanni* N_C] [FV_C *andò* FV_C] [P_C *a casa* P_C] '(having) finished the job, John went home'

1.2 Examples of usage

In this section we illustrate, by way of exemplification, the chunking of linguistic phenomena which are typical of dialogues. Examples are only indicative and represent an adaptation to English material of the principles underlying the Italian chunking schema outlined above.

multi-words

Chunking presupposes prior identification and marking of multi-word units.

error coding

Chunking presupposes prior identification and marking of errors and non standard forms.

trailing off, interruption, completion

*SAR: [FV_C *smells* FV_C] [Adj_C good enough Adj_C] [P_C for P_C]

retrace-and-repair sequences

*BET: [FV_C *I wanted* FV_C] [filler_C *uh* filler_C][FV_C *I thought* FV_C] [FV_C *I wanted* FV_C] [I_C *to invite* I_C] [N_C *Margie* N_C].

anacolutha (syntactic blending)

*BET: [FV_C *I wanted* FV_C] [filler_C *uh* filler_C] [WH_C *when* WH_C] [FV_C *is Margie coming* FV_C] [Punct_C ? Punct_C]

1.3 Sparkle: Functional Annotation

In EAGLES, a three-layered approach to the specification of grammatical dependencies for verbal arguments was followed (Sanfilippo et al., 1996). The first layer identifies the subject/complement and predicative distinctions as the most general specifications; this layer is regarded as encoding mandatory information. The second layer provides a further partition of complements into direct and indirect as recommended specifications. Finally, a more fine-grained distinction qualified as useful is envisaged introducing further labels for clausal complements and second objects.

The first step in tailoring the EAGLES standards to the needs of SPARKLE, has been to make provisions for modifiers. These were not treated in EAGLES since only subcategorizable functions were taken into consideration. Secondly, the relationship among layers of grammatical dependency specifications has been interpreted in terms of hierarchical links.

In general, grammatical relations (GRs) are viewed as specifying the syntactic dependency which holds between a head and a dependent. In the event of morphosyntactic processes modifying head-dependent links (e.g. the passive, dative shift and causative-inchoative diatheses), two kinds of GRs can be expressed:

1. the initial GR, i.e.\ before the GR-changing process occurs
2. the final GR, i.e.\ after the GR-changing process occurs

For example, *Paul* in *Paul was employed by Microsoft* is the final subject and initial object of *employ*. The hierarchical organisation of GRs is shown graphically in Figure2 below.

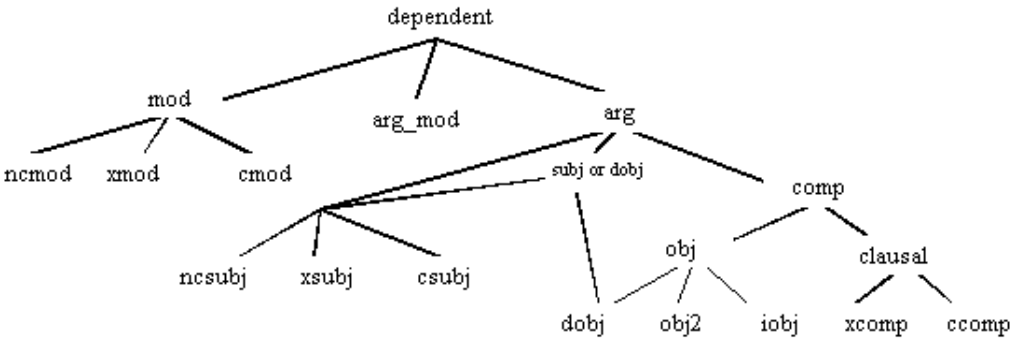


Figure 2: GR Hierarchy

Each GR in the current version of the scheme is described individually below.

mod(type,head,dependent)

The relation between a head and its modifier; where appropriate, type indicates the word introducing the dependent; e.g.

mod(_,flag,red)

a red flag

mod(_,walk,slowly)

walk slowly

mod(with,walk,John)

walk with John

mod(while,walk,talk)

walk while talking

mod(_,Picasso,painter)

Picasso the painter

mod is also used to encode the relation between an event noun (including deverbal nouns) and its participants; e.g.

mod(of,gift,book)

the gift of a book

mod(by,gift,Peter)

the gift of a book by Peter

mod(of,examination,patient)

the examination of the patient

mod('s,doctor,examination)

the doctor's examination of the patient

cmod,xmod,ncmod

Clausal and non-clausal modifiers may (optionally) be distinguished by the use of cmod / xmod, and ncmod respectively, each with slots the same as mod. The GR cmod is for when the adjunct is controlled from within, and xmod for control from without. E.g.

cmod(because,eat,be)

he ate the cake because he was hungry

xmod(without,eat,ask)

he ate the cake without asking

arg_mod(type,head,dependent,initial_gr)

The relation between a head and a semantic argument which is syntactically realised as a modifier; thus a by-phrase can be analysed as a `thematically bound adjunct'. The

type slot indicates the word introducing the dependent: e.g.

arg_mod(by,kill,Brutus,subj)

killed by Brutus

subj(head,dependent,initial_gr)

The relation between between a predicate and its subject; where appropriate, the initial_gr indicates the syntactic link between the predicate and subject before any

GR-changing process:

subj(arrive,John,_)

John arrived in Paris

subj(employ,Microsoft,_)

Microsoft employed 10 C programmers

subj(employ,Paul,obj)

Paul was employed by Microsoft

With pro-drop languages such as Italian, when the subject is not overtly realised the annotation is, for example, as follows:

subj(arrivare,Pro,_)

arrivai in ritardo '(I) arrived late'

where the dependent slot is filled by the abstract filler Pro, which indicates that person and number of the subject can be recovered from the inflection of the head verb form.

csubj,xsubj,ncsubj

The GRs csubj and xsubj may be used for clausal subjects, controlled from within, or without, respectively. ncsubj is a non-clausal subject. E.g.

csubj(leave,mean,_)

that Nellie left without saying good-bye meant she was still angry

xsubj(win,require,_)

to win the America's Cup requires heaps of cash

dobj(head,dependent,initial_gf)

The relation between a predicate and its direct object--the first non-clausal complement following the predicate which is not introduced by a preposition (for English and German); initial_gf is iobj after dative shift; e.g.

dobj(read,book,_)

read books

dobj(mail,Mary,iobj)

mail Mary the contract

iobj(type,head,dependent)

The relation between a predicate and a non-clausal complement introduced by a preposition; type indicates the preposition introducing the dependent; e.g.

iobj(in,arrive,Spain)

arrive in Spain

iobj(into,put,box)

put the tools into the box

iobj(to,give,poor)

give to the poor

obj2(head,dependent)

The relation between a predicate and the second non-clausal complement in ditransitive constructions; e.g.

obj2(give,present)

give Mary a present

obj2(mail,contract)

mail Paul the contract

ccomp(type,head,dependent)

The relation between a predicate and a clausal complement which does have an overt subject; type indicates the complementiser / preposition, if any, introducing the clausal XP. E.g.

ccomp(that,say,accept)

Paul said that he will accept Microsoft's offer

ccomp(that,say,leave)

I said that he left

xcomp(type,head,dependent)

The relation between a predicate and a clausal complement which has no overt subject (for example a VP or predicative XP). The type slot is the same as for ccomp above.

E.g.

xcomp(to,intend,leave)

Paul intends to leave IBM

xcomp(_,be,easy)

Swimming is easy

xcomp(in,be,Paris)

Mary is in Paris

xcomp(_,be,manager)

Paul is the manager

Control of VPs and predicative XPs is expressed in terms of GRs. For example, the unexpressed subject of the clausal complement of a subject-control predicate is specified by saying that the subject of the main and subordinate verbs is the same:

Paul intends to leave IBM

subj(intend,Paul,_)

xcomp(to,intend,leave)

subj(leave,Paul,_)

dobj(leave,IBM,_)

arg(head,dependent)

The hierarchical organisation of GRs makes it possible to use underspecified GRs where no reliable bias is available for disambiguation. For example, both Gianni and Mario

can be subject or object in the Italian sentence

Mario, non l'ha ancora visto, Gianni

'Mario has not seen Gianni yet' / 'Gianni has not seen Mario yet'

In this case, the parser could avoid having to try to resolve the ambiguity by using the underspecified GR arg, e.g.

arg(vedere,Mario)

arg(vedere,Gianni)

dependent(introducer,head,dependent)

The most generic relation between a head and a dependent (i.e. it does not specify whether the dependent is an argument or a modifier). E.g.

dependent(in,live,Rome)

Marisa lives in Rome

1.4 Examples of usage

It can be argued quite convincingly that the level of functional annotation (or any other syntactic representation which abstracts away dramatically from the surface ordering of syntactic units in a sentence) is relatively independent of the specific utterance through which grammatical functions happen to be concretely realized. For example, given the following orthographic transcription

i)

I I I go away

where the pronoun "I" is uttered thrice, it still makes sense to say that the subject of "go away" is one (namely the pronoun "I"), and that it just happens to be repeated more than once, owing to some extra-grammatical factors. The neat separation between chunked representations (where concretely realized syntactic units matter) on the one hand and the level of functional representation on the other hand, allows the annotator to get around somewhat puzzling issues such as "which one of the three overtly realized instances of 'I' is the subject of this utterance?". In fact it makes comparatively little sense to associate the label "subject" with any particular token of "I" in i) above. A level of annotation which abstracts away from the level of linear representation embodied in i) achieves this purpose:

subj(go, I,_)

Still linking the functionally annotated material with elements of i) can be useful. This could be achieved as follows: a) first, the three pronouns in a row are signalled as a repetition at some level of "edited" orthographic transcription; b) a target form ("I") is then added to the surface representation; c) finally, the target form is linked to the functionally annotated material.

A.32 OVIS

Coding book:

No coding book is publicly available. References can be found at http://grid.let.rug.nl:4321

See also Bod and Scha (1997).

Number of annotators:

missing information

Number of annotated dialogues:

21000 sentences, Dutch

Evalutation of scheme:

missing information

Underlying task:

Information-seeking, telephone-mediated human-machine dialogues for travel/transport domain.

Examples:

no examples available

Mark-up Language:

missing information

Existence of annotation tools:

Annotation was done semi-automatically, using a tool called SEMTAGS.

Usability:

Used in the OVIS interactive spoken language system for travel information to users using public transport in the Netherlands.

Contact person:

Rens Bod (Rens.Bod@let.uva.nl)

List of phenomena annotated:

The OVIS system aims at reaching large vocabulary, speaker-independent continuous speech recognition technology, combined with natural language processing using a probabilistic partial parsing approach. The NLP Ovis component is a statistically based language processing system, based on the 'Data-Oriented Parsing' System developed and implemented at the Department of Computational Linguistics of Amsterdam University.

Hesitations, false starts, and additional noises produced by speakers are annotated at the morpho-syntactic level. The following is a slightly more detailed description of information represented at the syntactic and semantic levels of analysis.

1. Syntactic annotation

Syntactic annotation starts from a minimum level consisting in bracketing of constituents. Sentences are annotated with labelled constituent trees, as in the ATIS corpus. The syntactic categories have been reconsidered to fit the needs of the application. The original linguistically inspired annotation convention has received considerable revision: in particular, certain rather broad categories were introduced that are non-standard in linguistic theories. For instance, a notion of 'modifier-phrase' which includes adverbs, PP's, and various kinds of conjunctions and other combination of such constituents. Other ad hoc categories have been introduced to deal with peculiarities of Dutch word order which do not fit well in a purely surface-based syntactic description without features.

The grammar covers most of the common verbal subcategorization types (intransitives, transitives, verbs selecting app, and modal and auxiliary verbs), np-syntax (including pre- and postnominal modification, with the exception of relative clauses), pp-syntax, the distribution of vp-modifiers, various clausal types (declaratives, yes/no and wh-questions, and subordinate clauses), all temporal expressions and locative phrases relevant to the domain, and various typical spoken language constructs.

2. Semantic/pragmatic annotation

Every meaningful node is annotated with a formula expressing that meaning; if the meaning of a node depends on its daughter nodes, this formula contains variables referring to those daughter node meanings. When a new tree is constructed out of subtrees with such annotations, it is obvious how to compute the meaning of this tree.

A.33 The Lancaster/IBM Spoken English Corpus (SEC)

Annotation for the Spoken English Corpus (SEC) is based on the LOB Corpus tag-set. Almost every SEC tag is identical to its LOB equivalent. The major difference between the tag-sets is that LOB differentiates between relative and interrogative WH-pronouns whereas SEC does not. For example, the LOB tag pair WP (WH-pronoun, interrogative, nominative or accusative) and WPR (WH-pronoun, relative, nominative or accusative) are covered by the same SEC tag. Confusingly, this tag is also called WP, but, unlike for LOB, does not imply that the WH-pronoun is interrogative. The following table details the major differences between LOB and SEC with regard to WH-pronouns:

Tag	Description in SEC	Description in LOB
WP	WH-pronoun, nominative or accusative	WH-pronoun, interrogative, nominative or accusative
WPR	Not used in SEC (use WP instead)	WH-pronoun, relative nominative or accusative
WP$	WH-pronoun, genitive	WH-pronoun, relative, genitive
WP$R	not used in SEC (use WP$ instead)	WH-pronoun, relative, genitive
WPO	WH-pronoun, accusative	WH-pronoun, interrogative, accusative
WPOR	Not used in SEC (use WPO instead)	WH-pronoun, relative, accusative

As its name implies, the Spoken English Corpus is composed of transcriptions of spoken English. This inherently means that there will be differences between it and the LOB corpus which is comprised of written texts only. Phenomena that are used primarily for English in its written form will not be found in SEC. A good example is written abbreviations. These were marked in LOB in a pre-automatic-tagging phase by adding the sequence '\0' to the start of the abbreviated token whereas this is not required in SEC.

Some of the LOB tags do not appear in SEC even though, in theory, they would have been allowable. This is because, at just over 52 thousand words, SEC is much smaller than LOB which has over a million words. Naturally, in such a small corpus the coverage of rare parts-of-speech was reduced. This can also explain why annotation of SEC did not call for a significant extension of the LOB tagset.

Further information on the SEC can be found in Taylor and Knowles (1988) and at the International Computer Archive of Modern English (ICAME) corpus collection (http://nora.hd.uib.no/corpora.html).

A.34 SWITCHBOARD

Coding book:

Marie Meeter et al. 1995. Disfluencyannotation stylebook for the Switchboard Corpus.

(ftp://ftp.cis.upenn.edu/pub/treebank/swbd/doc/DFL-book.ps)

Number of annotators:

missing information

Number of annotated dialogues:

2430 conversations, more than 240 hours, 3 million words

Evaluations of scheme:

missing information

Underlying task:

missing information

List of phenomena annotated:

list of relevant phenomena provided below

Examples:

list of relevant phenomena provided below

Mark-up language:

missing information

Existence of annotation tools:

missing information

Usability:

missing information

Contact person:

Linguistic Data Consortium (ldc@ldc.upenn.edu)

2 Word-Level Classification Issues

1.1 Adverbs, Interjections, Interactional Markers

Explicit editing terms (such as 'I mean') and discourse markers (such as 'Well') are annotated respectively as '{E...}' and '{D...}'. Use of curly brackets allows annotation of a sequence of words, by simply including it into brackets.

Example:

{E I would say}

1.2 Pauses, Hesitators

Only filled pauses are markes (hesitators) by '{F}'.

1.3 Word Partials, Non Standard Forms

Fragmented or incomplete words are marked in the transcription with '-'.

Example:

you kn-

2 Segmentation Issues

Transcribed texts are subdivided primarily into so-called "slash units". A slash unit is maximally a sentence but can be a smaller unit. Slash units below the sentence level correspond to those parts of the narrative which are not sentential but which the annotator interprets as complete.

2.1 Multi-Words

Annotation makes provision for marking sequences of more than one word with one label only by encopassing them between curly brackets.

2.2 Error Coding

No specific marker is envisaged for this purpose.

2.3 Phrase Partials

2.3.1 Trailing Off, Interruption, Completion

When a turn does not constitute a complete constituent, it is marked as incomplete with the symbol '-/'. It is possible for the speaker to continue over more than one turn. In this case, the annotation guidelines make provision for use of the symbol '- -'. Combination of the two symbols means the following:

'- - -/' interruption with constituent left incomplete and following completion

Example:

A: *I'll do it if - - -* /

B: *Yeah*/

A: *- - you wish*/

'- - /' interruption with complete slash unit and following completion

Example:

A: *I'll do it - -* /

B: *Yeah*/

A: *- - if you wish*/

'- -' interruption with neither incomplete constituent nor complete slash unit, and following completion

Example:

A: *If you wish - -*

B: *Yeah*/

A: *- - I'll do it*/

2.3.2 Retrace-and-Repair Sequences

The entire restart with its repair is contained in square brackets. The Interruption Point is marked by a '+'.

Example:

[*we're + at the same time we're*] *real scared*

2.3.3 Anacolutha (syntactic blending)

Syntactic blending is treated as a kind of incomplete slash unit, if the speaker continues speaking but has obviously begun a new slash-unit.

Example:

when it comes to being alone -/ now if you give him the freedom to walk around, he likes that/

A.35 TRAINS

The TRAINS project at the University of Rochester Department of Computer Science is a long-term effort to develop an intelligent planning assistant that is conversationally proficient in natural language. The goal is a fully integrated system involving on-line spoken and typed natural language together with graphical displays and GUI-based interaction. The primary

application has been a planning and scheduling domain involving a railroad freight system, where the human manager and the system must co-operate to develop and execute plans.

The current system prototype, named TRIPS (The Rochester Interactive Planning System), involves a more realistic domain and more complicated planning problems, while continuing the emphasis on dialogue-based, mixed-initiative interaction.

Coding book:

No coding book is available, but information can be found in Core and Schubert (1997).

Number of annotators:

missing information

Number of annotated material:

Altogether, the Trains-93 corpus includes 98 dialogs, collected using 20 different tasks and 34 different speakers. This amounts to six and a half hours of speech, about 5900 speaker turns, and 55,000 transcribed words. The collection and transcription of the dialogues is documented in the technical note "The Trains 93 Dialogues"

(ftp://ftp.cs.rochester.edu/pub/papers/ai/94.tn2.Trains_93_dialogues.ps.gz)

The transcriptions themselves are available at
http://www.cs.rochester.edu/research/speech/93dialogs

Evaluations of scheme:

missing information

Underlying task:

Task-driven, application-oriented problem solving dialogues. The dialogues involve two participants: one who plays the role of a user and has a certain task to accomplish, and another who plays the role of the system by acting as a planning assistant.

List of phenomena annotated and examples:

For some of the phenomena annotated at the morpho-syntactic level, see the general description below.

Mark-up language:

missing information

Existence of annotation tools

For collecting and annotating ``The Trains 93 Dialogues'', a set of tools has been developed for converting a DAT recording into a fully segmented and annotated dialogue. These tools allow the user to progress stepwise through this process, from creating the initial dialogue audio file, breaking up the dialogue into a sequence of single-speaker utterance files that preserve the sequentiality of the dialogue, annotating the utterance files, printing the contents of the dialogue, and updating the breakup of the dialogue. These tools are described in the Trains technical note, "Dialogue Transcription Tools"

(ftp://ftp.cs.rochester.edu/pub/papers/ai/94.tn1.Dialogue_transcription_tools.ps.Z)

and are available through ftp, as well as on the CD-ROM. The toolset itself is available in a tar file at ftp://ftp.cs.rochester.edu/pub/packages/dialog-tools/toolset.tar.gz.

Usability:

Used in the TRAINS system.

The collected dialogues have played an integral part in the Trains project. They have also been used to train a parser that uses statistical preferences, and to train a part-of-speech tagger that models speech repairs (cfr. Heeman and Allen, 1994)
(ftp://ftp.cs.rochester.edu/pub/papers/ai/94.heeman.ARPA_HLT.ps.Z)

Contact person:

James Allen (james@cs.rochester.edu)

A short description

The TRAINS project is to be mentioned as an example of how the exigencies of spoken language can be accommodated in software development. In particular, the TRAINS project is especially relevant for our purposes in that it adopts an integration vs. normalization strategy (see. the section 5.1.1 in the report).

The traditional approach consists in removing disfluencies before they reach the parser or in having the parser skip over such material. However reasonable, this approach not only abstracts from real data but also neglects the important roles such segments can play in the dialogue structure. Repairs, for example, can contain referents that are needed to interpret subsequent text (e.g., *Take the oranges to Elmira, uh, I mean, take* them *to Corning*).

In contrast to the above strategy, the alternative adopted in TRAINS is a parser-level approach that includes in phrase structure those disfluencies (such as repairs, hesitations and overlapping backchannel acknowledgments) that constitute a common problem for parsers for mixed-initiative dialogues.

To handle the disfluencies in mixed-initiative dialogues caused by repairs, hesitations and acknowledgments, the dialogue parser uses *metarules* that allow the chart of a dialogue parser to contain parallel syntactic structures (what was first said and its correction) in the case of repairs, and interleaved syntactic structures in the case of interruptions.

The *editing term metarule* allows constituents to skip over words signaling turn keeping (*um, ah*) and repairs (*I mean*).

In the structure allowed by the metarule a constituent may be interrupted between two subconstituents by one or more editing terms, and a constituent can be interrupted in more than one location.

In the case of overlapping acknowledgments and continuation prompts, such as 'okay', 'right' etc. uttered by the second speaker in overlap with the 'main' talk, the *continuation metarule* allows a constituent to overlap or be embedded inside another constituent to which it is unconnected. In this way, a constituent can be built across tracks.

An *interruption metarule* is used to deal with interjected corrections, questions, and comments separately from any repair that may follow. An example of interruption is the following:

```
u: then e1 will have

s: oh e1

u: right

two boxcars of oranges
```

In the case of repairs, a *repair metarule* operates on what is being corrected (or *reparandum*) and the correction (or the *alteration*), to build parallel phrase structure trees: one with the reparandum and one with the alteration. For example, for an utterance such as "*Take the ban-um the oranges*", the repair metarule would build two VPs: *take the ban-* and *take the oranges.*

This parsing framework has two relevant consequences. First, it allows the parser to accommodate disfluency phenomena, thus leaving important aspects of dialogue structure untouched. In addition, in this way the parser has information about the syntactic structure of the utterance and the range of allowed structures. These sources of information are absent from preprocessing, normalizing routines, and the dialogue parser can still use acoustic cues, pattern matching, and other sources of information used in preprocessing techniques.

Prosody Schemes

The review of coding schemes for prosodic annotation follows slightly modified evaluation guidelines, due to the fact that prosodic schemes are intended for speech in general rather than for dialogues only. "Number of annotators", "Number of dialogues" and "Usability" have been gathered into "Applications", while "Underlying task" has been substituted with "Purpose and underlying approach".

A.36 Prospa

PROSPA was developed by Margaret Selting and Dafydd Gibbon [Selting 87, 88], specially to meet the needs of discourse and conversation analysis, but it has also been discussed within the Prosody Group in the ESPRIT 2589 SAM (Multilingual Speech Input/Output Assessment, Methodology and Standardization) project.

Coding book:

[Selting 87, 88] provides a description of the scheme.

Applications:

Information not available.

Evaluation:

Information not available.

Purpose and underlying approach:

Specially oriented to discourse and conversation analysis.

List of phenomena annotated:

PROSPA focuses on the transcription of F0 variations.

It seems to be a rather ‘phonetic’ system of notation (only movements and levels are reported, without reference to any specific theoretical model).

Transcriptions consist of:

1) an overall inclination or declination specified over the domain of an intonation unit

2) peaks and troughs internal to the unit

3) a final dynamic tone

Global categories are defined according to rhythmical or pitch contour properties in a cohesive series of accents. Length of a global contour and the direction of pitch or tone level are indicated as follows:

( )	extent of a sequence of cohesive accents
F	globally falling intonation
R	globally rising intonation
H	level intonation on high tone level
M	level intonation on middle tone level
L	level intonation on low tone level
H/F	falling intonation on a globally high tone level
…	sequence of weakly accented or unaccented syllables

Local categories are defined as accent and accent types or ”short range pitch movements usually realized on lengthened vowels” [Selting 87]; they include:

+	upward pitch movement
-	downward pitch movement
=	level pitch accent

Since accents can be realized together with pitch changes, the following symbols are introduced:

?+	upward local pitch jump co-occurring with an upward accent
Ø+	downward local pitch jump co-occurring with an upward accent”

The intonation after the last accent of a global unit - or ‘tails’ - is noted after the parentheses in the following manner:

`	falling tails
/	rising tails
-	level tails
/`	combinations of tails (rising-falling here)”

Examples:

Information not available.

Markup language:

Symbols are inserted in the phonetic transcription.

Annotation tools:

Information not available.

A.37 IPA

The IPA (International Phonetic Alphabet) has a set of symbols for the representation of suprasegmental elements. On the occasion of the Kiel convention in 1989 a working group on Suprasegmental Categories coordinated by Gösta Bruce was set up [Bruce 88, 89]. It was concluded that additions were needed to represent suprasegmentals within the IPA framework.

The symbols of the IPA, including those referred to the transcription of suprasegmentals, are available at: http://www.arts.gla.ac.uk/IPA/ipachart.html

Coding book:

[Bruce 89] provides a description of the scheme.

Applications:

Information not available.

Evaluation:

Information not available.

Purpose and underlying approach:

The scheme was designed to provide an extension of the IPA phonetic alphabet to suprasegmental phenomena.

List of phenomena annotated:

The IPA alphabet includes symbols for transcription of prosodic boundaries, prosodic phenomena and phonetic cues of prosody

Prosodic boundaries:

.	Syllable break
|	Minor (foot) group
||	Major (intonation) group
	Linking (absence of a break)

Prosodic phenomena:

- Stress (primary and secondary):

>	Primary stress
,	Secondary stress

Phonetic cues:

- Duration (Long, extra-long, short).

:	Long
‘	Half-long
&	Extra-short

- Local F0 variations
  The IPA transcription system provides symbols to transcribe local F0 variations using both a level or a contour approach:

a) Levels

The IPA alphabet assumes the existence of 5 different tonal levels: extra high, high, mid, low and extra low

ä	Extra high
ë	High
ï	Mid
ö	Low
ü	Extra low

b) Contours

For the transcription of F0 variations using a contour approach, IPA suggests a series of symbols, although the set presented in the IPA chart does not seem to be exhaustive:

Ë	Rising
Ü	Falling
ÿ	High rising
	Low rising
Ö	High falling
Ÿ	Low falling
	Rising-falling

- c) Global F0 variations
- In the case of global F0 variations, the IPA phonetic alphabet provides some symbols to transcribe downstep and upstep, as well as global rises and falls:

õ	Downstep
ã	Upstep
Ã	Global rise
Õ	Global fall

Examples:

Information not available.

Markup language:

Symbols are inserted in the phonetic transcription.

Annotation tools:

Information not available.

A.38 TEI

Chapter 11 of the Text Encoding Initiative Guidelines [Sperberg 94] discusses the transcription of spoken language. Since the main aim of this standardization effort concerns written texts, the guidelines presented in this chapter are oriented towards the transcription of speech as a text enriched with a set of conventions for phenomena that can not be adequately described with

standard spelling. TEI Guidelines on spoken texts were mainly the result of work carried out within a subgroup composed of Stig Johansson -chair-, Jane Edwards and Andrew Rosta [Johansson 95a, b].

More information on the TEI can be found at:

http://etext.virginia.edu/TEI.html;

http://www-tei.uic.edu/orgs/tei ; http://info.ox.ac.uk/archive/teilite.

Coding book:

Chapter 11 of the Text Encoding Initiative Guidelines [Sperberg 94] is the basic reference manual to apply the TEI conventions to the transcription of prosody.

Information about the Text Encoding Initiative Guidelines can be found at http://www.uic.edu/orgs/tei/. There is also a ftp site where documents about TEI are available: ftp-tei.uic.edu (in the "pub/tei" directory).

Applications:

The TEI web page includes a list of 63 projects using TEI Guidelines for text annotation. This list is available at:

http://www-tei.uic.edu/orgs/tei/app/index.html.

The list contains references to some projects involving the annotation of dialogues by means of TEI. Some of these are:

- Danish Spoken Language Dialogue Systems Project

(http://www.cog.ruc.dk/projects/Dialogue/user-95)

- Chiba Corpus of Map Task Dialogues in Japanese

(http://cogsci.L.chiba-u.ac.jp/MapTask)

- Edinburgh Map Task Corpus

(http://www.cogsci.ed.ac.uk/elsnet/Resources/Map-Task/ mt_corpus.html)

Evaluation:

Information not available.

Purpose and underlying approach:

The scheme is intended to enhance TEI conventions concerning written text with labels for prosodic phenomena that can not be adequately described with standard spelling.

List of phenomena annotated:

Prosodic boundaries:

TEI conventions allow to indicate tone units or intonational phrase boundaries by means of the elements <seg> (beginning of a unit) and </seg> (end of a unit).

Prosodic phenomena:

1) Stress

The stressed syllable can be indicated using the label &stress, after the stressed syllable.

2) Rhythm

A set of labels is proposed to specify different types of rhythm:

rh	beatable rhythm
arrh	Arrhythmic
spr	spiky rising
spf	spiky falling
glr	glissando rising
glf	glissando falling

Phonetic cues of prosody:

1) Duration

TEI includes one symbol to indicate the extra lengthening of syllables:

:	lengthened syllable

2) Pauses

The presence of a pause is indicated with the element <pause>

3) Tempo (speech rate)

TEI proposes a set of symbols to transcribe tempo:

a	allegro (fast)
aa	very fast
acc	accelerando (getting faster)
l	lento (slow)
l l	very slow
rall	rallentando (getting slower)

4) Loudness

TEI also proposes a set of symbols to transcribe different degrees of loudness:

f	forte (loud)
ff	very loud
cresc	crescendo (getting louder)
p	piano (soft)
pp	very soft
dimin	diminuendo (getting softer)

5) F0 events

5.1. F0 contours

The following set of symbols is defined in the TEI conventions to transcribe pitch patterns (contours):

.	low fall intonation
,	fall rise intonation
?	low rise intonation
!	rise fall intonation

5.2. Global F0 events

Variations in pitch range can be transcribed using the TEI conventions using the following set of labels:

high	high pitch range
low	low pitch range
wide	wide pitch range
narrow	narrow pitch range

Global falling or rising intonation can be transcribed using the following labels:

Asc	Ascending
Desc	Descending
Monot	Monotonous
Scand	scandent (each succeeding syllable higher than the last, generally ending in a falling tone)

6) Voice quality

The following set of labels is proposed to indicate voice quality

Whisp	Whisper
Breath	Breathy
Husk	Husky
Creak	Creaky
Fals	Falsetto
Reson	Resonant
Giggle	unvoiced laugh or giggle
Laugh	voiced laugh
Trem	Tremulous
Sob	Sobbing
Yawn	Yawning
Sigh	Sighing

Some critique

[Llisterri 96a]:

”[Payne 92:51ff] mentions the lack of development of guidelines for encoding prosody in the TEI scheme and discusses some inconsistencies of the statements about prosody in the TEI Guidelines. The favoured solution would be to incorporate basic prosodic information in the orthographic transcription and to use a fundamental frequency tracing aligned with the text in cases where a detailed prosodic analysis is needed. Tone units: Although an easy conversion can be made between French’s boundary markers and TEI tags delimiting tone units, [Payne 92] notes the difficulties of transcribing melodic contours with TEI conventions. Tonic syllables: TEI Guidelines do not provide an indication of tonic syllables as straightforwardly as in French’s system. As [Payne 92:55] points out if the tonic syllable is going to be marked, it should be marked in the orthographic transcription, and the TEI Guidelines should be extended to provide a way of doing this in a straightforward manner. Tones: [Payne 92:56] suggests the extension of TEI Guidelines to allow distinguishing tones as in French’s conventions; such an extension could be based in different specifications for the tag <syllable>. Prominent non-tonic syllables: Prominent non-tonic syllables are marked in French’s system, but no provision for such feature is found in the TEI Guidelines. Speech management: TEI has no specific guidelines for the transcription of disfluency phenomena, recommending transcription using IPA or other systems of phonemic transcription. On the other hand, French’s conventions, adopted by NERC, are much more specific and deal with different phenomena not covered by TEI, such as guessed or unintelligible fragments.”

Examples:

Information not available.

Markup language:

TEI conventions have been defined using SGML as markup language. This is one of the advantages of this transcription scheme.

Annotation tools:

Information not available.

A.39 ToBI

ToBI (Tones and Break Indices) was proposed in 1992 [Silverman et al. 92] by "a group of researchers with expertise in a variety of approaches to prosodic analysis and speech technology" with the aim of defining a notational system, "analogous to IPA for phonetic segmentation", that could become a "standard for prosodic transcription of most varieties of American English".

A description of the ToBI system is available at:

http://julius.ling.ohio-state.edu:80/Phonetics/ToBI/

Coding book:

ToBI creators have developed two labelling guides (Beckman & Ayers, 1994; Beckman & Hirschberg, 1994).

They are available at:

via ftp at kiwi.nmt.edu.

via URL at http://ling.ohio-state.edu/Phonetics/ToBI/ToBI0.html

Applications:

Although primarily developed for English, it has been used also to transcribe intonation events of English dialects [Mayo et al. 97] or other languages such as Italian [Grice et al. 95 b] or German [Grice et al. 95a].

ToBI has also been integrated, with adjustments and ehancements, into other transcription systems, such as VERBMOBIL [Reyelt et al. 94] or the Stuttgart System [Mayer 95] (see http://www.ims.uni-stuttgart.de/phonetik/joerg/

labman/STGTsystem.html)

Evaluation:

An evaluation of the performance of ToBI is presented in [Pitrelli *et al.* 94].

The German version GToBI has been evaluated in [Grice et al. 96]

Purpose and underlying approach:

ToBI is an adaptation of Pierrehumbert's phonological model of English intonation [Pierrehumbert 80].

[Llisterri 96a]:

"In the domain of prosodic transcription systems to be used in speech research and in speech technology, ToBI (Tone and Break Index Tier was developed to fulfill the need of a prosodic notation system providing a common core to which different researchers can add additional detail within the format of the system; it focuses on the structure of American English, but transcribes word grouping and prominence, two aspects which are considered to be rather universal [Price 92].

As described by [Silverman et al. 92] the system shows the following features: (1) it captures categories of prosodic phenomena; (2) it allows transcribers to represent some uncertainties in the transcription; (3) it can be adapted to different transcription requirements by using subsets or supersets of the notation system; (4) it has demonstrated high inter-transcriber agreement; (5) it defines ASCII formats for machine-readable representations of the transcription; and (6) it is equipped with software to support transcription using Waves and UNIX programmes.

A ToBI transcription for an utterance consists of symbolic labels for events on four parallel tiers: (1) orthographic tier, (2) break-index tier, (3) tone tier and (4) miscellaneous tier. Each tier consists of symbols representing prosodic events, associated to the time in which they occur in the utterance. The conventions for annotation according to TOBI are defined for text-based transcriptions and for computer-based labeling systems such as Waves."

ToBI is based on a phonological model of English intonation, but several attempts have been made to extend it to other languages (and English dialects), by means of additions and adjustments. Criticism has been raised against it, see for example [Nolan et al. 97]

List of phenomena annotated:

ToBI system has been conceived for the transcription of intonation phenomena and prosodic boundaries. There are no existing symbols for the transcription of the phonetic cues of prosody. Boundaries and tones are represented in separate tiers, aligned with the text by means of temporal coordinates.

Prosodic boundaries

Prosodic boundaries are annotated in ToBI by means of the Break Indices:

0	clitic group boundary
1	word boundary
2	boundary with no tonal mark
3	Intermediate Phrase boundary
4	Intonative Phrase boundary

Prosodic phenomena

ToBI provides a set of symbols for the transcription of intonation phenomena: pitch accents, phrase accents and boundary tones. Such symbols are associated with the accented syllable and with phrases, respectively. They *can* be time-aligned with f0 peaks and valley.

1.1. Pitch accents

H*	peak accent (high pitch accent)
L*	low accent (low pitch accent)
L*+H	scooped accent
L+H*.	rising peak accent
H+!H*	downstepped accent

1.2. Boundary tones

L%	final low boundary tone
H%	final high boundary tone
%H	initial high boundary tone

1.3. Phrase accents

L-	low phase accent
H-	high phase accent

ToBI also provides one symbol for the transcription of downstep:

!	Downstep

Examples:

Using the transcriber tool and xwaves, a series of files are created during the transcription process which contain the information related to the different tiers. The following are examples of files containing the transcription of the utterance 'Show me the cheapest fare from Philadelphia to Dallas excluding restriction" (obtained from the TOBI-TRAINING material):

Orthographic tier:

signal cheapest2

type 1

color 123

font -*-times-medium-r-*-*-17-*-*-*-*-*-*-*

separator ;

nfields 1

#

2.105000 123 show

2.245000 123 me

2.355000 123 the

2.935000 123 cheapest

3.315000 123 fare

3.565000 123 from

3.836919 123 Da(llas)-

4.325000 123 from

5.015000 123 Philadelphia

5.225000 123 to

5.855000 123 Dallas

7.399125 123 excluding

8.585000 123 restriction

8.825000 123 V
9.115000 123 U
9.595000 123 slash
9.880000 123 one

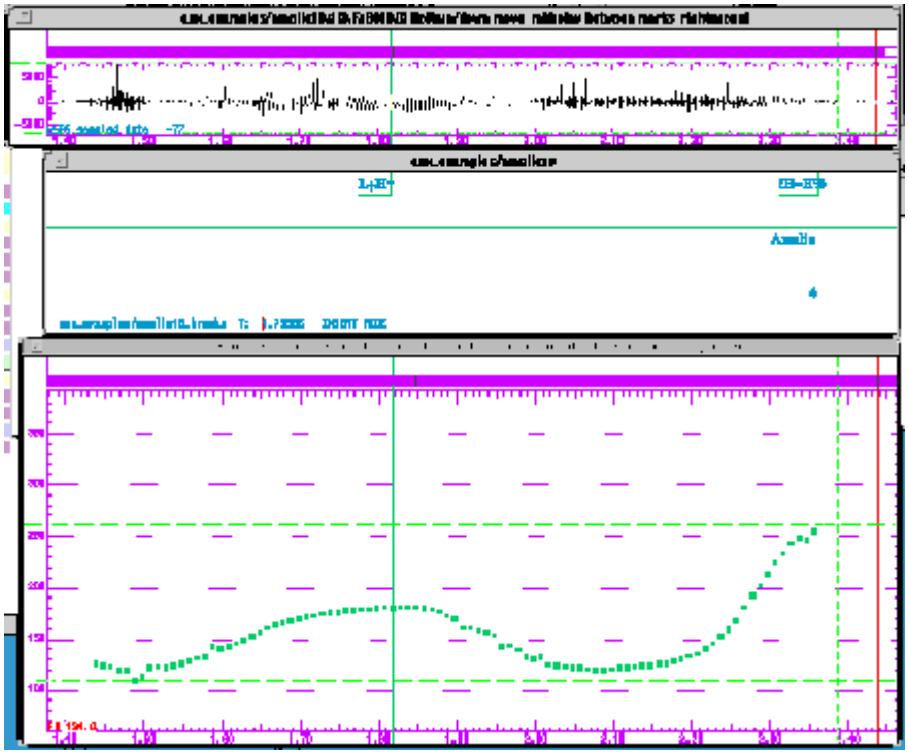
Break index-tier:

signal cheapest2
type 0
color 123
comment created using xlabel Fri Sep 3 17:24:47 1993
font -*-times-medium-r-*-*-17-*-*-*-*-*-*-*
separator ;
nfields 1
#
2.105000 123 1
2.245000 123 1
2.355000 123 1
2.935000 123 1
3.315000 123 4
3.565000 123 1
3.836919 123 1p
4.325000 123 1
5.015000 123 3
5.225000 123 1
5.855000 123 4
7.399125 123 4
8.585000 123 4
8.825000 123 1
9.115000 123 3
9.595000 123 1
9.880000 123 4

Tone tier:

```
signal cheapest2
type 0
color 115
comment created using xlabel Fri Sep  3 17:24:48 1993
font -*-times-medium-r-*-*-17-*-*-*-*-*-*-*
separator ;
nfields 1
#
    2.052696  115 H*
    2.579923  115 L+H*
    3.065052  115 !H*
    3.315635  115 L-L%
    4.149572  115 %r
    4.470318  115 L+H*
    4.771018  115 !H*
    5.015584  115 L-
    5.388451  115 H*
    5.855538  115 L-L%
    6.984159  115 L+H*
    7.399114  115 L-L%
    8.154402  115 H*
    8.585841  115 L-L%
    8.711954  115 H*
    8.928780  115 !H*
    9.114631  115 L-
    9.353582  115 H*
    9.694309  115 H*
    9.880160  115 L-L%
```

The following picture gives an example of x-waves visualization of a ToBI transcription, aligned with waveform and f0 curve.



Markup language:

Symbolic labels in separate tiers for each type of information (orthography, boundaries, tones, miscellaneous), time-aligned with the signal.

Annotation tools:

Two annotation tools have been developed using the xwaves environment, a transcriber and a checker. The transcriber is a UNIX script that simplifies the transcription task, but doesn't produce the transcription automatically. The checker is also a UNIX script that validates the coherence of the transcribed sequences of symbols. They are available via ftp at kiwi.nmt.edu.

A.40 SAMPA

SAMPA (SAM Phonetic Alphabet) is a multi-lingual computer-readable transcription system developed within the ESPRIT project 2589 SAM (Multilingual Speech Input/Output Assessment, Methodology and Standardization). The SAMPA final standard system is presented in Wells *et al.* (1992).

Information about SAMPA is available at:

http://www.phon.ucl.ac.uk/home/sampa/home.htm

Coding book:

Information not available.

Applications:

SAMPA has been applied not only by the SAM partners collaborating on EUROM 1, but also in other speech research projects (e.g. BABEL, Onomastica) and by Oxford University Press.

Evaluation:

Information not available.

Purpose and underlying approach:

SAMPA aims to provide ASCII encodings for the IPA symbols required for European languages. SAMPA includes a number of symbols for prosodic transcription, attempting to avoid any model-dependency. It is mainly intended to support signal-oriented labelling and provides a basis for cross-language comparisons.

List of phenomena annotated:

Prosodic boundaries:

$	syllable boundary
+	morpheme boundary
#	word boundary
|	tone group/intonation phrase boundary
§	phonological phrase/rhythm group boundary
##	sentence boundary

Prosodic phenomena

- Stress

"	primary stress and accent I in Norwegian and Swedish
%	secondary stress
""	accent II in Norwegian and Swedish

Phonetic cues of prosody

- Duration

:	length mark

- Intonation
- SAMPA includes a set of symbols for transcribing intonation contours:

’	rising tone
‘	falling tone
‘ ’	fall-rise
’ ‘	rise-fall

- Pauses

…	silent pause

Examples:

Information not available.

Markup language:

Diacritics inserted in the phonetic transcription.

Annotation tools:

Information not available.

A.41 SAMPROSA

SAMPROSA SAM Prosodic Alphabet has been initially proposed by [Gibbon 90], incorporating results from discussions within the SAM Prosody Working Group.

The relevant information can be found at URL :

http://www.phon.ucl.ac.uk/home/sampa/samprosa.htm

Coding Book:

It is documented in [Wells *et al.* 92]

Applications:

Information not available.

Evaluation:

Information not available.

Purpose and underlying approach:

[Llisterri 96a]:

"The system is intended both for prosodic transcription for linguistic purposes, and for prosodic labelling in speech technology and experimental phonetic research. The system allows the transcription of global, local, terminal and nuclear tones, length, stress, pauses and prosodic boundaries."

It is designed for multi-tier transcription, where independent parallel symbolic representations can be made using different segmental or prosodic criteria , related via association with phonetic segments or with temporal alignment with the signal (synchronization).

List of phenomena annotated:

Prosodic boundaries:

$	Syllable boundary
#	Word boundary
|	Tone group boundary (non-directional)
[	Tone group boundary (left)
]	Tone group boundary (right)

Prosodic phenomena:

1) Stress

"	Primary stress
%	Secondary stress

2) Intonation

Intonation contours are assumed to be made of a series of local, global, terminal and nuclear tones.

Local tone symbols:

H	High pitch
L	Low pitch
T	Top pitch (extreme H)
B	Bottom pitch (extreme L)
M	Mid pitch
+	Higher pitch
++	Much higher pitch
+-	Peak (upward-downward)
-	Lower pitch
--	Much lower pitch
-+	Trough (downward-upward)
= or > or S	Level or same tone
^	Upstep
^^	Wide upstep
!	Downstep
!!	Wide downstep

Global tone symbols:

Global tones are transcribed using symbols extracted from Local and Nuclear tone repertoire.

Terminal tone symbols:

As Global tones, terminal tones are transcribed using symbols extracted from Local and Nuclear tone repertoire.

Nuclear tone symbols:

-	Level tone (before tone group boundary)
' or / or R	Rising tone
` or \ or F	Falling tone
`' (etc.)	Fall-rise
'` (etc.)	Rise-fall

Phonetic cues of prosody:

Length

:	Segment length mark

Pause

...	Silence

Examples:

Information not available.

Markup language:

Symbolic labels, organized in separate parallel tiers related via synchronization with the signal (or, alternatively, via phonological rules for association with segments).

Annotation tools:

Information not available.

A.42 INTSINT

INTSINT is a coding system of intonation developed by Daniel Hirst and his colleagues at the CNRS centre of the Aix-en-Provence University. Descriptions of this method can be found in [Hirst 91,94]; [Hirst et al. ?].

More information about the INSTINT system can be found at
http://www.lpl.univ-aix.fr/~hirst/intsint.html.

Coding book:

There is no coding book to apply INTSINT, although the tutorial describing the use of the automatic annotation tool includes some references about the philosophy of INTSINT.

Applications:

[Llisterri 96a]

"The system has already been applied to several languages (see, for example, [Hirst *et al.* 93] and is being used in MULTEXT Multilingual Text Tools and Corpora project ([Hirst et al. 94]; more information on the project is available at URL
http://www.lpl.univ-aix.fr/projects/multext/index.html)
for the encoding of intonation in the paragraphs contained in the EUROM.1 corpus."

Within the MULTEXT project, three different groups have been working in the annotation of read paragraphs extracted from the EUROM.1 corpus:

a) Laboratoire Parole et Langage, CNRS, Aix-en-Provence, France

b) Grup de Fonètica, Departament de Filologia Espanyola, Universitat Autònoma de Barcelona, Spain

c) Department of Phonetics, Umeå University, Sweden

These three groups have been working in the annotation of English (40 passages), Spanish (40 passages), French (40 passages), German (20 passages) and Swedish (5 passages).

In all cases, the annotated material were read-aloud paragraphs. No real dialogues were annotated.

Evaluation:

There is no formal evaluation of INTSINT, but an evaluation of the performance of the annotation tool using INTSINT ('mes') has been carried out within the MULTEXT project. The results of this evaluation are presented in [Llisterri 96b].

Purpose and underlying approach:

INTSINT is based of the intonation model developed by Daniel Hirst and the group of the Aix-en-Provence University [Hirst 91, 94], [Hirst et al. ?].

It is conceived "to provide a purely formal encoding of the macroprosodic curve. Each target point of the stylized curve is coded by a symbol either as an absolute tone, defined globally with respect to the speakers pitch-range or as a relative tone, defined locally with respect to the immediately neighbouring target-points" ([Campione *et al.* 97], p. 72).

List of phenomena annotated:

INTSINT includes only symbols to transcribe F0 events:

1) Absolute Tones

INTSINT includes three symbols to label the Absolute Tones, which are defined according to the speaker's pitch range.

T	top of the speaker's pitch range
M	initial, mid value
B	bottom of the speaker's pitch range

2) Relative tones

Relative tones are coded in INTSINT considering the height of the preceding and following target points. Five different symbols exist to transcribe these Relative Tones:

H	target higher than both immediate neighbours
L	target lower than both immediate neighbours
S	target not different to preceding target
U	target in a rising sequence
D	target in a falling sequence

Examples:

The INTSINT coding is usually stored in a set of files generated with the annotation tool ‘mes’, which is described below. An example of the files containing the INTSINT coding of the speech utterance ‘Il faut que je sois a Grenoble Samedi vers quinze heures’ is provided here:

Orthographic representation:

il 1745

faut 3605

que 6669

je 8710

sois 10678

a 14461

Grenoble 15334

Samedi 26463

vers 37025

quinze 39730

heures 45645

50164

F0 target points:

-99.95 140.012

106.3 163.589

265.6 217.241

521.7 148.048

617.25 190.525

827.95 130.806

1249.35 223.515

1614 139.595

1822.45 172.134

1983.25 144.184

2078.75 185.903

2248.35 152.624

2505.85 99.1518

2730 152.606

INTSINT transcription:

M -1999

L 2126

T 5312

M 10434

H 12345

L 16559

T 24987

M 32280

H 36449

L 39665

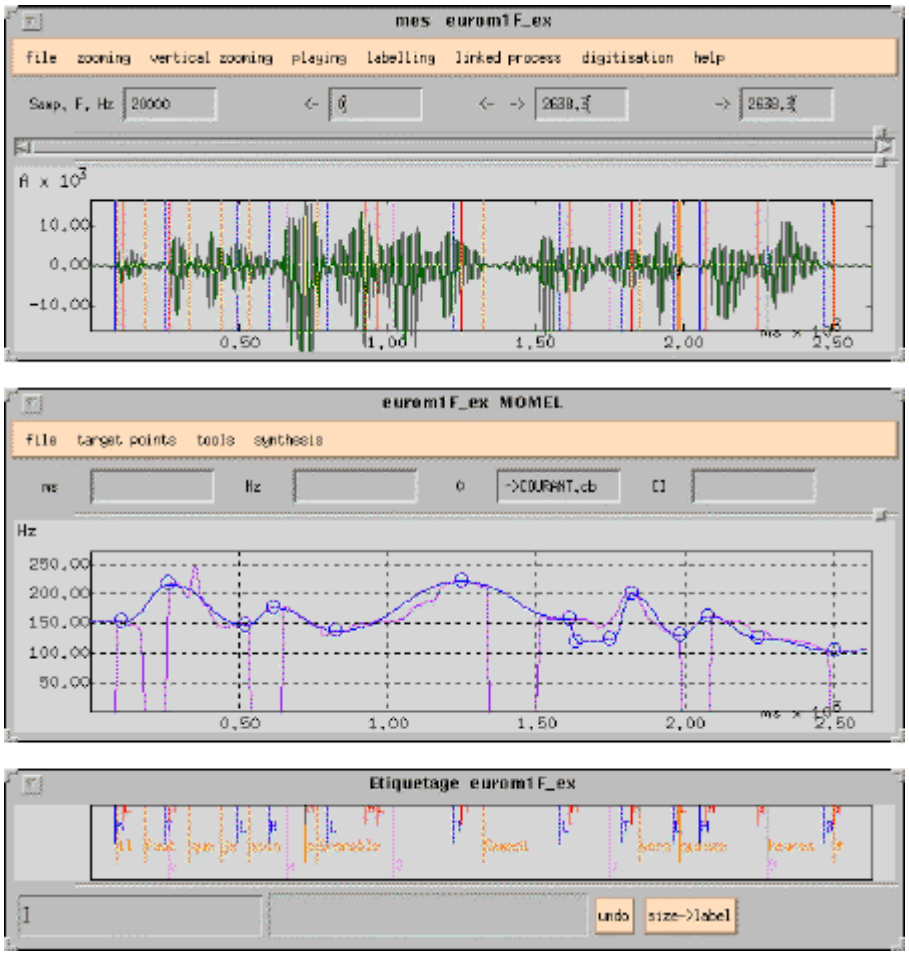
H 41575

D 44967

B 50117

M 54600

The following picture shows the visualization of transcription in the environment 'mes' (see below).



Markup language:

Symbolic labels time-aligned with the signal. Numerical values for f0 target points.

Annotation tools:

A tool for prosodic annotation using INTSINT (see picture above) has been developed in the framework of the MULTEXT project: 'mes' allows automatic annotation of speech signals from phonetic data (a stylized version of the F0 contour obtained by the MOMEL automatic stylization procedure).

A description of this tool is available at:

http://www.lpl.univ-aix.fr/valorisation/mes_signaix/

A.43 SAMSINT

SAMSINT (SAM system for Intonation Transcription) has been proposed by the SAM Prosody Working Group, as a computer-readable transcription system based on INTSINT.

Coding book:

Information not available.

Applications:

Information not available.

Evaluation:

Information not available.

Purpose and underlying approach:

As in the case of INTSINT, SAMSINT is based on the intonation model developed by Daniel Hirst and the group of the Aix-en-Provence University [Hirst 91, 94], [Hirst et al. ?].

[Llisterri 96a]:

”SAMSINT [...] was intended to be a computer-readable system for the transcription of intonation contours within defined intonation units. The system is based on INTSINT […], incorporating additional facilities and simplifications [Wells et al. 92]”.

List of phenomena annotated:

Prosodic boundaries:

[	initial intonation unit boundary
]	final intonation unit boundary

Phonetic cues of prosody:

F0 events

Local

T	top
B	bottom
+	higher
-	lower
^	upstep

!	downstep
>	same

Global

/	global rising in an intonation unit
\	global falling in an intonation unit

Examples:

Information not available.

Markup language:

Information not available.

Annotation tools:

Information not available.

A.44 E.9 IPO

The methodology for the study of intonation proposed by IPO (Institute for Perception Research, Eindhoven) has inspired many experimental works and synthesis implementations. The approach can be considered a reference in the field of intonation research, more for its general principles than for its representation scheme. The actual coding scheme has been applied by the authors to the modelling of Dutch intonation, and has also been adapted and applied to other languages.

Coding book:

The reference text both for general principles and notation is:

J. 't Hart, R. Collier, A. Cohen, "A perceptual study of intonation" [Hart 90].

Applications:

The actual coding scheme has been applied by the authors to the modelling of Dutch intonation, and has also been adapted and applied to other languages (English [Willems et al. 88], French [Beaugendre 94], Italian [Quazza 91], Mpur [Odé 97], German [Brindopke et al. 97]).

Evaluation:

Information not available.

Purpose and underlying approach:

IPO approach provides a framework for studying both the physical and linguistic aspects of intonation and "supports language-independent pre-theoretical description of speech melody allowing the development of new melodic categories" [Brindopke et al. 97]. The idea is that a model of intonation for a given language should be extracted from raw acoustic f0 data, by means of successive steps, first removing perceptually irrelevant details and finally getting to meaningful patterns related to linguistic function. Speech synthesis has a twofold role in the process, as an analysis tool allowing to assess the perceptual plausibility of the models, and as an application, where the acoustic content of the representation can be directly implemented.

The underlying theory of intonation represents the f0 curve as a sequence of pitch movements, superimposed on a general declination line, gradually lowering the pitch range through the utterance, with possible resets at phrase boundaries. The linguistically relevant pitch patterns are discovered by a direct analysis of f0 curves.

First the f0 curve is stylized with a sequence of straight lines representing a close copy of the original curve, perceptually identical when imposed on the original signal by means of resynthesis.

Then, segments in the stylized curves are classified and described according to four discrete parameters: direction (rise/fall), timing (early in the syllable/late/very late), rate of change (fast/slow), size (full/half). On this basis, a clustering of f0 segments identifies a standard set of pitch movements typical of the given language (speaker/domain/corpus). Pitch curves can so be standardized, i.e. described as sequences of standard pitch movements. When resynthesized, the standardized curve should be perceptually equivalent to the original one.

Further analysis would find out the typical and recurring configurations of pitch movements which carry some linguistic function: for example a 'pointed hat' marking a pitch accent or a 'flat hat' sounding as a sentence conclusion. A complete intonation model for a given language would discover the grammar according to which configurations combine into full intonation contours, realizing the basic intonation patterns of the language.

List of phenomena annotated:

The IPO methodology focuses on intonation only, providing different representations for pitch. Each step has its own coding of the f0 curve, from a detailed acoustic description up to phonetic/phonological representations.

Such codings presuppose a phonetic segmentation of the speech signal, or at least a segmentation into syllables, with respect to which pitch movements are aligned.

Acoustic description (stylization):

the curve is represented as a sequence of f0 straight segments, each measured in semitones of change, milliseconds of duration, alignment with syllable boundaries.

Phonetic description (standardization):

the curve is represented as a sequence of pitch movements which can be considered standard for the given language; each standard movement is characterized by four parameters:

direction	rise/fall
timing	early in the syllable/late/very late
rate of change	fast/slow
size	full/half

Standard movements for Dutch are distinguished into five types of rises, labelled 1, 2, 3, 4 and 5, and five falls, labelled A, B, C, D and E. Segments corresponding to lower and upper declination lines are labelled O and 0 respectively. Each syllable is assigned at least one label. If two or more movements occur on the same syllable, their labels are joined by "&".

Phonological description:

Once a 'grammar of intonation' has been defined for a given language, a more abstract labelling of the curve can be obtained in terms of pitch configurations and pitch contours.

In the same methodological framework, different notations have been adopted. For example, in the Dutch SPIN/ASSP Program [Heuven et al. 93] the following symbols have been used [Terken 93a]:

R	rise
F	fall
L	low level pitch
H	high level pitch
FF	gradual pitch fall
&	two movements on the same syllable
*	associated with accented syllable
%	associated with a boundary

The inventory of pitch contours for Dutch is represented as follows:

Transcription	IPO notation	Description
R*	1	prominence-lending rise (early in syllable)
F*	A	prominence-lending fall (early in syllable)
R*F*	1A	sequence of R and F associated with two successive accented syllables
R&*F	1&A	combination of prominence-lending rise and fall on the same syllable
R*FF	1D	prominence-lending rise followed by gradual falling pitch
LR%	O2	non-prominence lending rise (late in the syllable), starting from the baseline
R*H%	10	R* followed by high-level pitch until the end of the phrase

Examples:

Information not available.

Markup language:

Symbolic labels time-aligned with the f0 curve.

Annotation tools:

IPO developed its own tools for (manual) perceptual stylization and resynthesis. In this environment labels can be assigned to flags in the speech waveform, but the choice of labels is completely free and there is no well-formedness checker.

Several tools have been developed for pitch perceptual stylization, more or less related with the IPO approach (e.g. WinPitch www.winpitch.com).

Automatic stylization has been implemented too, with more or less sophisticated approaches. See for example [Coile et al. 94], where a piece-wise linear approximation of the f0 curve (in the log domain) is obtained starting from raw f0 data and, optionally, phonetic segmentation. See also [Mertens et al. 97], describing a sophisticated stylization algorithm based on a "tonal perception model".

A tool for (synthetic) speech manipulation explicitly implementing the IPO framework is Speech Maker [Leeuwen et al. 93], where the intonation contour can be represented in terms of IPO pitch movements, each controllable in its parameters (anchor, timing, duration, excursion).

A recent example of "an environment for labelling and testing of melodic aspects of spoken language" inspired by IPO methodology is the one described in [Brindopke et al. 97], implemented in C and integrated in EXPS/Xwaves. The tool "relies on the method of approximating the original f0 contour with a minimum set of straight lines", provides "labelling

facilities for model-based melodic description for German" and "supports language-independent pre-theoretical description of speech melody allowing the development of new melodic categories".

Tools for automatic labelling of pitch contours with IPO labels have also been implemented [Bosch 93a,b].

IPO is developing a system for automatic extraction and labelling of prosodic information. It takes the speech waveform as input, employs information about phoneme durations, identifies intonation phrases, accent locations, pitch accent types and boundary tones, and parameters for pitch range (baseline and topline parameters). The mapping between acoustic features and prosodic labels is defined in run-time readable files.

A.45 TSM

"TSM: Tonetic Stress Marks System" is a coding scheme based on the British School style of auditory intonation analysis [O'Connor et al. 73] and applied for the transcription of the SEC "Spoken English Corpus", created in a joint project by Lancaster University and IBM. The corpus has now been digitized and time-aligned: the 'machine readable' version is called MARSEC.

Information is available at http://midwich.reading.ac.uk/research/speechlab/marsec/marsec.html

Coding book:

G.Knowles,A.Wichmann, P.Anderson "Working with Speech: Perspective on research into the Lancaster /IBM Spoken English Corpus", London and NewYork, Longman, 1966 [Knowles et al. 66]

Applications:

MARSEC corpus: more than 50 texts from the BBC (different speakers, 30% female, RP accent, commentary, news broadcasting, etc.) amounting to about 52,000 words.

The corpus has been transcribed by two annotators.

Evaluations of scheme:

Information not available.

Purpose and underlying approach:

Based on the British School auditory intonation analysis ([Crystal 69], [O'Connor et al. 73]).

List of phenomena annotated:

Labels represent phrase boundaries and intonation contours.

The signal is phonetically segmented. Energy and f0 are automatically computed.

Prosodic annotation is inserted in the orthographic representation, time-aligned with the signal at beginning of accented syllables.

Two levels of intonation phrasing:

‖	major tone unit
|	minor tone unit

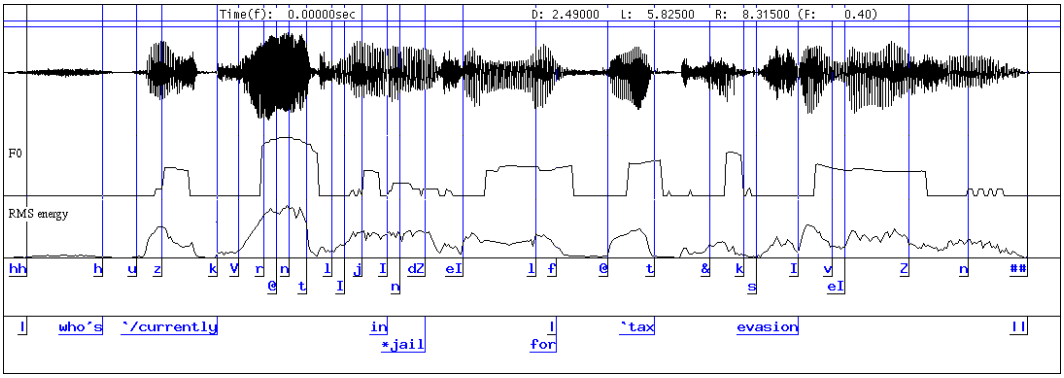
Each accented syllable is marked with a diacritic classifying the accent according to the following characteristics (describing the tone contour from the beginning of the syllable up to the next accented syllable or the tone unit end):

high/low (refers to the starting point of the tone, higher or lower than the previous pitch)

level/fall/rise/fall-rise/rise-fall (the shape of the contour)

A conversion has been attempted between TSM and ToBI [Roach 94].

Examples:



Markup language:

Prosodic labels are diacritics inserted in the orthographic stream and time-aligned with the signal at beginning of accented syllables.

Annotation tools:

Environment for (manual) labelling: Entropics/waves+

A.46 TILT MODEL

The TILT model has been proposed by Taylor [Taylor et al. 94] as a way of representing intonation, oriented both to speech synthesis and to intonation analysis. The model provides linguistics labels and quantitative parameters.

Coding book:

There is no coding book, instruction was given orally.

But see [Taylor 97] on http://www.cstr.ed.ac.uk/~pault/papers.html.

Applications:

TILT has been applied to the prosodic transcription of the Canadian DCIEM Maptask Corpus ([Bard et al. 95], http://www.cogsci.ed.ac.uk/hcrc/wgs/dialogue/dialog/maptask.html), the Boston Radio News Corpus [Ostendorf et al. 95] and the Switchboard Corpus [Godfrey et al. 92].

The annotators were 5 PhD students on intonation. The labelled material consisted of:

- DCIEM Maptask Corpus: a subset of 25 dialogues (2 hours)
- Boston Radio News Corpus: 34 stories (48 minutes)
- Switchboard: 1 hour subset of the 2000 spontaneous telephone dialogues

Evaluations of scheme:

Labelling consistency between the 5 labellers was tested with pairwise comparisons of their transcriptions [Taylor 97]. For each ordered pair of transcriptions, assuming the first as a reference, the correctness (number of events correctly identified) and accuracy (correct minus the percentage of false insertions) of the second was evaluated. The average correctness and accuracy were 81.6% and 60.4%, respectively. When ignoring minor accents, the average scores were 88.6% and 74.8%.

Manual labelling with TILT can be done using any suitable system which allows you to see a waveform and mark events at particular times. It has been noted that "labelling tilt events is much easier than labelling ToBI parameters" [Dusterhoff et al. 97].

Purpose and underlying approach:

The TILT model has been proposed by Taylor [Taylor et al. 94] as a refinement of its previous Rise/Fall/Connection model [Taylor 94] for a representation of intonation oriented both to speech synthesis and to intonation analysis. It defines a reversible function linking the f0 curve to its linguistic representation, providing means to automatically derive the representation from the curve and viceversa.

The f0 curve is seen as a sequence of *intonational events*, each linked to a syllabic nucleus. Events can be pitch accents or boundary tones. Each event is a movement in the f0 curve - a rise, a fall or a combination of both - which is described by:

- starting f0 value (Hz)

- duration
- amplitude (in Hz)
- shape, mathematically represented by its 'tilt', a value computed from the f0 curve
- starting point, time aligned with the signal and with the vowel onset.

List of phenomena annotated:

The labelling scheme is intended to represent the f0 curve. Labels are associated to intonational events, which can be accents or boundary tones.

Events can be automatically detected on the basis of f0 and energy information, or can be manually labelled and aligned to the signal. The quantitative description of the event (starting f0 value, duration, amplitude, tilt) is automatically derived from the f0 curve.

For manual labelling the following labels are defined:

sil	silence
c	connection
a	major pitch accent
fb	failing boundary
rb	rising boundary
afb	accent+falling boundary
arb	accent+rising boundary
m	minor accent
mfb	minor accent+falling boundary
mrb	minor accent+rising boundary
l	level accent
lrb	level accent+rising boundary
lfb	level accent+falling boundary

Examples:

Example xlabel file (segmentation file in the environment Entropics\xwaves).

The position of the accents and boundary tones was decided by the humans, and the numbers after "tilt:" were calculated automatically with reference to the F0 contour:

0.69333 26 c; tilt: 118.984

0.74000 26 sil; tilt: 0.000

0.86116 26 a; tilt: 118.984 17.020 0.121 1.000 0.000

1.13170 26 c; tilt: 136.004

1.28200 26 a; tilt: 112.858 0.311 0.150 -1.000 0.000

1.57008 26 c; tilt: 112.547

1.81056 26 afb; tilt: 107.899 13.914 0.240 -1.000 0.000

1.90001 26 sil; tilt: 93.985

Markup language:

Labelling is realized in the form of a segmentation file, each intonational event on a separate line, specified by a time coordinate, an ASCII label and a set of numeric values.

Annotation tools:

Manual annotation could be done using Entropics xwaves or any similar system.

An automatic event detector, based on HMM, is available [Taylor 97].

For each event, the quantitative parameters (duration, amplitude, tilt) are automatically computed from the f0 curve.

A.47 VERBMOBIL

Different coding schemes for prosody have been used in the VERBMOBIL Project. Here a Perceptual Scheme and a Syntactic-Prosodic Scheme are reviewed.

Information about prosodic labelling in VERBMOBIL can be found at:

http://sbvsrv.ifn.ing.tu-bs.de/prosody/verbmobil.html

Coding book:

M.Reyelt and A. Batliner, "Ein Inventar prosodischer Etiketten fur VERBMOBIL", Verbmobil Memo 33, 1994 [Reyelt et al. 94]

Applications:

Perceptual Scheme:

33 dialogues (about 2 hours)

480 sentences (20% of PHONDAT database, read sentences) - 5 annotators

Syntactic-Prosodic scheme:

7286 turns (about 150,000 words) - one annotator

Evaluations of scheme:

An evaluation has been performed on the PHONDAT material showing that utterances were "rather consistently labelled even by untrained listeners" [Reyelt 93]. The material consisted of sentences by 8 speakers. Labellers were 5.

Inter-labeller agreement (on 8 different speakers, min-max):

- phrase accent: 66%-79%
- secondary accent: 32%-44%
- phrase boundary: 67%-84%

Some opinions about the Perceptual Scheme (see (http://www.ims.uni-stuttgart.de/phonetik/joerg/stockholm/bserlmu.html): "the labelling of intonation is still difficult for the transcribers" . "Most partners in Verbmobil use only the functional and the break index tier". "The functional tier has several advantages:- it contains information usable for focus-analysis- it makes the two decisions (is a word accented? if, which pitch accent?) more transparent for the transcribers".

The Syntactic-prosodic Scheme has been judged easier, faster and more reliable [Batliner et al. 96]

Purpose and underlying approach:

The aim of the German VERBMOBIL project (http://www.dfki.de/verbmobil/) is to develop a system for automatic speech-to-speech translation in appointment scheduling dialogues. Prosody is studied in the project both in itself and as a cue to dialogue segmentation and to enhance syntactic parsing, to classify dialog acts, etc.

Different studies and experiments involving prosody have been conducted in the VERBMOBIL framework, where prosody has been represented in its acoustic aspects [Batliner et al. 97] or in its syntactic function [Batliner et al. 96]. The reference coding scheme for (auditory) prosodic labelling was developed at Braunschweig University, with the aim of providing a scheme usable by several project partners for a variety of purposes and usable also by transcribers with only little experience in prosodic labelling.

List of phenomena annotated:

The Perceptual Prosodic Labelling scheme represents phrasing, accents and intonation contours at a phonological level. For intonation contours, it uses a ToBI-like inventory consisting of H and L tones, while its more specific feature is a more abstract functional tier where prominence relations are explicitly marked in order to be more easily related with focus and discourse structure.

The label inventory splits into three tiers:

functional tier: *main accent, secondary accent, emphasized/contrastive accent, sentence modality*

break index tier: *(full) intonation phrase boundary, minor boundary, irregular boundary*

tone tier: *pitch accents and boundary tones*

Functional tier:

?	Question mark (several question types are labeled)
PA	Main accent (in each intonational phrase the *most prominent word* is labelled)
NA	Secondary accent (all other accents are secondary accents)
EK	Emphasized or contrastive accent

Break index tier:

B1	Normal word boundary	
B2	Minor (intermediate) phrase boundary	Weak intonational marking
B3	Full intonational phrase boundary	Strong intonational marking with or without lengthening or change in speech tempo
B9	Irregular boundary	Marks disfluencies at hesitations, repairs, etc.

Tone tier (ToBI-like, with additional distinctions for labelling spontaneous speech)*:*

Accents

H*	normal peak accent
L+H*	medium (or raised') peak. Starting with a low tone before the accented syllable the f0 rises to a high peak within the syllable.
L*+H	delayed peak, a H* accent that reaches high f0 in the syllable behind the accented one
L*	trough accent. Can be rising when followed by a H-H\% boundary.
!H*, L+!H*, L*+!H	downstepped accents
H+!H*	early peak. Fall before the accented syllable, often followed by a low boundary.

Boundaries

L-L%	terminal fall, boundary reaches the lower end of the speaker's pitch range
H-H%	question/continuation rise, high boundary reaches the high end of the speaker's pitch range
L-H%	low phrase tone with a pitch rise to mid or high level
H-L%	continuation fall (a fall from high to mid pitch, or a more level boundary at mid-high pitch)

Syntactic-prosodic scheme:

A different coding scheme has also been developed and used in the VERBMOBIL Project [Batliner et al. 96], more syntax-oriented labelling.

M3S	main/subord. clause
M3P	non sentential free element/phrase, elliptic sentence
M3E	extraposition
M3I	embedded sentence/phrase
M3T	pre/post-sentential particle with pause/breathing
M3D	pre/post-sentential particle without pause/breathing
M3A	syntactically ambiguous
M2I	constituent, marked prosodically
M1I	constituent, not marked prosodically
M0I	every other word

Examples:

Information not available.

Annotation tools:

A workstation for prosodic labelling is developed at Braunschweig University, including software for visualization and labelling (fish), using Tcl/Tk, resynthesis of the original speech signal with variation of F0 (according to the labelled pitch accents, for acoustic verification of the transcription), evaluation of the labelling, automatic pre-segmentation of word boundaries, potential phrase boundaries prediction.

Automatic labelling systems based on Multi-Layer Perceptrons have been implemented both for the Perceptual Scheme and the Syntactic-Prosodic one [Batliner et al. 96, 97].

A.48 KIM: Kiel Intonation Model

Developed at IPDS -Institut fur Phonetik und Digitale Sprachverarbeitung Kiel University.

Coding book:

K.J.Kohler, M Patzold, A.P. Simpson, "From Scenario to Segment: the Controlled Elicitation, Transcription, Segmentation and Labelling of Spontaneous Speech", AIPUK 29, 1995, Kiel University [Kohler et al. 95]

Applications:

(see [Kohler 96])

Kiel Corpus of Read Speech (598 sentences, 2 stories, recorded words: 31,374)

Kiel Corpus of Spontaneous Speech (82 dialogues, 25,603 words): 1/3 has been prosodically labelled

Evaluations of scheme:

Information about formal evaluations not available.

The scheme has been judged to reach "high level of consistency within and across segmenters" [Kohler et al. 95].

Purpose and underlying approach:

The Kiel intonation model KIM ([Kohler 95], [Kohler 97]) is intended to represent intonation in general, at a phonetic and phonological level, with special regard to German. The model is oriented both to prosodic research and to text-to-speech implementation.

List of phenomena annotated:

The model is articulated into the following domains: prosodic phrase boundaries, speech rate, f0 downstep, lexical stress, sentence stress, intonation contours (types of peaks and valleys), synchronization of pitch events with syllables.

For each domain in the KIM model, the relevant categories are represented by symbolic labels. The level of representation is phonological/phonetic. Prosodic labels (prefixed with &) are inserted directly into the phonetic annotation tier and describe quite precisely the morphology and alignment of intonation contours, although not quantitatively.

Prosodic labelling presupposes phonetic segmentation. At least word boundaries and accented vowels should be aligned with the signal. Prosodic labels are associated with word boundaries (in this case they are prefixed with #) or with vowels inside the word (in this case the prefix is $).

Prosodic phrasing:

PG1	when coinciding with syntactic clause boundaries
PG2	otherwise (phrase)
PG1< PG1>	for parenthesis
PG2/	for truncations, false starts
PG	for technical breaks ?

Speech rate (marked after the phrase boundary, if a rate change is perceived with respect to preceding phrase):

RP +	(rate plus) rate increase
RM	(rate minus) rate decrease

Downstep is considered a default, it is not marked, as well as reset at phrase boundaries. Exceptions:

-	before PG if reset is absent
+	before the accent digit, if reset is within the phrase
|	before the accent digit, in case of upstep

Lexical stress: integrated in the segmental notation of vowels (' for primary, " for secondary)

Sentence accent (attribute of the word; placed before the word if referring to the lexical-stressed vowel, otherwise placed before the accented vowel):

0	unaccented
1	partially accented
2	accented
3	reinforced

Intonation contours (marked between accents or at the end of a prosodic phrase):

.	falling contour, peak or hat (with three levels: steep, slight, level)
,	low, narrow rising contour, valley
?	high, wide rising contour, valley
.,	fall rise (low rise)
.?	fall-rise (high rise)

Peak and valley alignment (marked after the accent digit):

Peaks:

^	for centre of accented syllable nucleus
)	before the nucleus
(	late in the nucleus or after

Valleys:

]	early
[	non-early

Examples:

Information not available.

Markup language:

Prosodic labels are ASCII labels inserted directly into the phonetic annotation tier. Such tier is represented twice: firstly as a symbolic stream, then as a list of labels associated with their time alignment in the signal (segmentation).

All prosodic labels are prefixed with &. Those corresponding to word boundaries are also prefixed with #, those associated to vowels inside the word are prefixed with $.

Annotation tools:

Phonetic segmentation is performed manually with a Waveform editor ([Carlson et al. 85]).

Prosodic labels are manually inserted, aligned with segment boundaries, with the help of the environment Xassp developed at IPDS, which displays waveform, f0 curve and labels.

An automatic tool checks the formal consistency of labelling.

A.49 PROZODIAG (LUND)

The scheme has been developed in the Project "Prosodic Segmentation and Structuring of Dialogue", supported by the Swedish national "Language Technology Programme" and involving Lund University and KTH.

Information may be found at http://galaxy.ling.lu.se/projects/ProZodiag/ and in the papers: [Bruce et al. 95a,b, 96, 97a,b,c,d].

Coding book:

Information not available.

Applications:

The scheme has been applied to the "Waxholm Application Database" [Bertenstam et al. 95], http://www.speech.kth.se/waxholm/waxholm2.html

consisting of man-machine dialogues in a boat traffic information service, amounting to 198 scenarios and 1900 dialogue turns.

Evaluation of scheme:

Information not available.

Purpose and underlying approach:

"The object of study is the prosody of dialogue in Swedish in a language technology framework. The primary goal of the project is to increase our understanding of how prosodic aspects of speech are exploited interactively in dialogue - the genuine environment for prosody - and on the basis of this increased knowledge to be able to create a more powerful prosody model." [Bruce et al. 97b]

The adopted methodology is the following:

analysis of discourse/dialogue structure (independent of prosody)

prosodic analysis: 1) auditory analysis (prosodic transcription) 2) acoustic-phonetic analysis (f0 curve, waveform)

speech synthesis (model-based resynthesis, text-to-speech)

Dialogues are annotated with textual and prosodic information aligned with the original f0 contour, the 'fine tuned' contour (computed on the basis of : F0 register - baseline- and range of F0 movements, timing and slope) and the synthetic contour (based on the Lund model of Swedish intonation, [Bruce 77]).

List of phenomena annotated:

The model represents prosodic phrasing and intonation in Swedish.

A set of "discrete elements" is intended for a phonological (auditory) transcription of phrasing and intonation contours. A set of "gradational elements" gives a more precise acoustic description of intonation, allowing to compute a "fine-tuned contour" representing a close-copy of the original f0 curve.

Discrete Elements:

Tonal Labels aligned with the nucleus of the accented syllable:

accented I	HL*
" II	H*L
focussed I	(H)L*H
" II	H*LH,
" compound	H*L...L*H
juncture - initial	%L
" - "	%H
" - final	L%
" - "	H%
" - "	LH%

Grouping:

||	major boundary
|	minor boundary

Gradational elements:

For each phrase:

Register	0 baseline (f0 level of unaccented portions)	low/mid/high
Range	height of pitch movement (starting from baseline)	low/ mid/ high/ flat/ decreased

For each f0 movement: slope, timing

Non-intonational phenomena: duration, voice source characteristics, reduction

Examples:

Information not available.

Markup language:

Symbolic labels, organized in separate tiers: a tonal tier and a boundary tier, synchronized with discourse labelling (orthographic tier, discourse referent tier, textual segmentation tier).

Annotation tools:

Labels and signal information are synchronized in ESPS/Waves+

A.50 Göteborg

Coding scheme adopted in the Project: "Spoken Language and Social Activity", at the Göteborg University.

Coding book:

"Transcription Standard", J. Nivre, Semantics and Spoken Language, Department of Linguistics, Göteborg University [Nivre ?]

http://www.ling.gu.se/SLSA/

Applications:

Swedish Spoken Language Corpus (interviews, shop, meeting, phone, task-oriented dialogues)

Number of activities: 227

Number of words: 967,141

Evaluations of scheme:

Information not available.

Purpose and underlying approach:

Aim of the Project is to investigate spoken language in different social activities.

The theoretical framework is activity-based communication analysis.

Prosody has an indirect relevance, as a correlate of discourse structure.

List of phenomena annotated:

Prosodic markers may be inserted in the orthographic or phonetic transcription, representing global rhythmic or intonational features of speech or local events such as stress, lengthening and pauses.

Stress:

Emphatic or contrastive stress marked with uppercase letters

Lengthening:

marked with diacritic ":"

Pauses:

short (/) long (//) very long (///)

Properties of speech:

<high pitch>, <low pitch>, <quick>, <slow>, <loud>, <quiet>, etc.

Examples:

Information not available.

Markup language:

A transcription is divided into a header section and a body.

The body section is made of:

lines beginning with "$" contain the transcribed utterances, with possible prosodic diacritics;

information lines beginning with "@", for comments and properties of speech in angle brackets.

Annotation tools:

TransTool is a computer-tool for transcribing spoken language developed in accordance with the transcription standards used within the research program Semantics and Spoken Language at the Department of Linguistics, Göteborg University. *TransTool* is implemented in Tcl/Tk (Tool Command Language/Toolkit) a scripting language designed to be easy to embed into other applications with a window system toolkit for building graphical user interfaces.

Cross-level Schemes

A.51 BNC (British National Corpus)

Coding book (domain and application):

The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of current British English, both spoken and written.

The BNC Homepage is at http://info.ox.ac.uk/bnc

Number of annotators:

No information available.

Qualification of annotators:

Linguists.

Number of dialogues/turns/segments annotated:

Not counted.

Language of corpora annotated:

English

Evaluations of scheme:

Not conducted.

Underlying task:

Not available.

List of phenomena annotated:

Part of speech, function of punctuation marks

Examples:

```
<u who=FX8PS001>
<s n=14>
<w UNC>erm<c PUN>, <w ORD>first <unclear> <w CRD>twelve
<w NN2>weeks <w AJ0>pregnant <w CJS>so <w VM0>should <w PNP>I
<w VVI>mark <w PRP>at <w AT0>the <w AJ0-NN1>bottom <w AVQ-CJS>when
<w PNP>she <w NN2-VVZ>types <unclear><c PUN>.
<s n=15>
<w UNC>Erm <unclear> <w DT0>this <w PNI>one<c PUN>.
<event desc="end of recording">
</u>
```

Markup language:

SGMLrelated.

Existence of annotation tools (manual/automatic):

See:

CLAWS4: THE TAGGING OF THE BRITISH NATIONAL CORPUS

Usability (in given machine environments):

Can be used in machine environments.

Contact:

No information available.

Additional information:

Data can be queried in CQP/Xkwic.

A.52 Bonn Focus Research

Coding book (domain and application):

This study compares the annotation of American English and German.

Reference: Wolters, M.: Linguistic Annotation of Two Prosodic Databases. IKP Bonn

Number of annotators:

Multiple: machines and human annotators have tagged different aspects.

Qualification of annotators:

Linguists

Number of dialogues/turns/segments annotated:

AE: 443 dialogues

G: 11 dialogues

Language of corpora annotated:

American English and German

Evaluations of scheme:

See reference in 1.

Underlying task:

Prosodic encoding of topic structure.

List of phenomena annotated:

Part of speech, phrase levels, phonetic segmentation, morphological and syntactical structure, semantic, pragmatic information, coreference. focus, dialogue act,

Examples:

<dialogue id=3>

<sent id=1 typ=deq> <da typ=reqi>Do you like warm countries?<sent id=2 typ=st> <da ty=fn> No, <da typ=inf> I prefer cold countries.

Markup language:

SGML

Existence of annotation tools (manual/automatic):

None.

Usability (in given machine environments):

Can be used with SGML tools.

Contact:

Maria Wolters,

IKP

Poppelsdorfer Allee 47

53115 Bonn

mwo@ikp.uni-bonn.de

A.53 CHILDES (Child Language Data Exchange System)

Coding book (domain and application):

The domain of the CHILDES project is the exchange and dissemination of child language data.

Information can be found at the Child Language Data Exchange System home page: http://psyscope.psy.cmu.edu/childes

The database is steadily growing as it is an archive to which researchers from all over the world contribute to.

Number of annotators:

More than 100.

Qualification of annotators:

Linguists, psychologists

Number of dialogues/turns/segments annotated:

Uncountable.

Language of corpora annotated:

The majority of data is about English, but there are also data in Cantonese, Catalan, Danish, Dutch, Estonian, French, German, Greek, Hebrew, Hungarian, Italian, Japanese, Mambila, Mandarin, Polish, Portuguese, Russian, Spanish, Swedish, Tamil, Turkish and Welsh.

Also, there are some bilingual and language impairment corpora available.

Evaluations of scheme:

Not conducted yet.

Underlying task:

Most of the dialogues are taken from normal interaction of mother and child.

List of phenomena annotated:

Non-linguistic information: age, birth, coder, education, filename, id, language, socio economic status, sex, location, comment, time, duration

Linguistic phenomena: morphosyntax, intonation, overlaps, actions, speech acts, phonetics

Examples:

@Font: Monaco:9

@Begin

@Participants: MAN Manuela Target_Child, FAT Sergio Father

@Age of MAN: 1;7.8

@Birth of MAN: 24-JUN-1985

@Sex of MAN: female

@Filename: 870201s.fat

@Coding: CHAT 1.0 (1991)

@Date: 1-FEB-1987

@Time Duration: 8:41-9:21

@Situation: at home

@Activities: looking at pictures and playing with toys

@Location: Brighton, Sussex, England

@Coder: Paul Carter

@Tape location: videotape 4, 1097-1854

@Language: Spanish

*FAT: ### 0. [=! laughs]

%clk: 8:41

%tim: 00:00:16

*FAT: <ay come on!> [?]

*FAT: Manuela!

*FAT: <vamos a buscar> [?] # el equipo tuyo!

%eng: let's look for your stuff

%act: walks out of room

*FAT: habla [/] habla!

%eng: speak speak

*MAN: 0.

%gpx: playing with toys

*FAT: www.

%exp: humming a tune

%act: returns to room carrying papers and dragging box

*MAN: woof [?] yyy.

%pho: w U f ## b E n

%clk: 8:41

%tim: 00:00:41

Markup language:

SAM like.

Existence of annotation tools (manual/automatic):

CLAN (computerized language analysis) versions for various computer platforms. Most of the software is for the evaluation of the data collected, such as checking the files, obtaining frequency information, displaying the data etc.

Usability (in given machine environments):

Can be used with the CLAN software.

Contact:

See WWW page.

A.54 DAMSL (Dialogue Act Markup in Several Layers)

Coding book (domain and application):

The domain of DAMSL is the annotation of dialogues in terms of semantic/pragmatic and discourse features.

Reference of DAMSL can be found at http://www.dfki.de/dri

Number of annotators:

15

Qualification of annotators:

Linguists

Number of dialogues/turns/segments annotated:

18 dialogues

Language of corpora annotated:

English

Evaluations of scheme:

Not conducted

Underlying task:

Dialogue structures.

List of phenomena annotated:

Communicative-Status Information-Level Forward Looking Function

Statement

Info-Request

Influencing-addressee-future-action (Influence-on-listener)

Committing-speaker-future-action (Influence-on-speaker)

Discussion

Backward Looking Function

Agreement

Understanding

Answer

Information-Relations

Antecedents

Discussion

Examples:

```
Info-request                        utt1: u: How do I get to Corning?
Assert, Open-option, Answer(utt1)   utt2: s: Go via Bath.
 ...
Action-directive                    utt1: s: so I'm assuming you'll also
                                             be taking a tanker from
                                             [Corning](1)
Commit,Accept(utt1)                 utt2: u: [oh](1) okay
Repeat-rephrase(utt1)               utt3:    take a tanker there
Other-forward-function              utt4:    okay
Assert                              utt5:    so its two hours ...
...
Info-Request                        utt1: u: so that's
Assert, Answer(utt1)                utt2: s: loading the orange
                                             <sil> juice will take
                                             another hour
Accept(utt2)                        utt3: u: okay
```

Markup language:

Can be seen as equivalent to the SAM standards.

Existence of annotation tools (manual/automatic):

Dat system at http://www.cs.rochester.edu/research/trains/annotation

Usability (in given machine environments):

Can be used with dat.

Contact:

http://www.cs.rochester.edu/research/trains/annotation

A.55 Kiel Corpus Format

Coding book:

Description of the Kiel Corpus at http://www.ipds.uni-kiel.de.

The Kiel Corpus is a growing collection of read and spontaneous German which has been collected and labelled segmentally at the ipds since 1990. At present the Kiel Corpus available on CD-ROM comprises over four hours of labelled read speech on The Kiel Corpus of Read Speech Vol. I as well as two and a half hours of labelled spontaneous speech on The Kiel Corpus of Spontaneous Speech Vol. I and Vol. II.

Number of annotators:

Not known.

Qualification of annotators:

Phoneticians.

Number of dialogues/turns/segments annotated:

Kiel Corpus of Read Speech Vol. I: No dialogues, 4932 words in 598 sentences of 12 subcorpora, 48 speakers

Kiel Corpus of Spontaneous Speech Vol. I: 26 speakers, 31 dialogues, 525 turns, 9291 word-tokens (1099 word-types)

Kiel Corpus of Spontaneous Speech Vol. II: 16 speakers, 51 dialogues, 862 turns

Kiel Corpus of Spontaneous Speech Vol. III: 10 speakers, 35 dialogues, 597 turns

Language of corpora annotated:

German

Evaluations of scheme:

There was no evaluation.

Underlying task:

Appointment dialogues.

List of phenomena annotated:

Word boundaries, word-internal compound boundaries, pauses, breathing, sentence punctuation, level of certainty of coding, replacement, insertion, deletion of sounds, nasality, junctual creak/creaky voice, plosive release phase and aspiration. Additionally, for every word there is a standard pronunciation and the orthographical version available.

Examples:

```
signal eddy
type 0
comment k06be004.s1h
comment ]ber die Felder weht ein Wind.
comment oend
comment Q y: b 6+  d i:+  f 'E l d 6  v 'e: t  Q aI n+  v 'I n t  .
comment kend
comment c:  %Q -q y: b -h 6+  d -h i:+  f 'E %l d -h 6  v 'e: t -h  Q -
q aI n+
comment  v 'I n t -h  .
comment hend
font -misc-*-bold-*-*-*-15-*-*-*-*-*-*-*
separator ;
nfields 1
#
1.03644  -1  #c:
1.03644  -1  ##%Q
1.05231  -1  $-q
1.05231  -1  $y:
1.08756  -1  $b
1.13288  -1  $-h
1.14062  -1  $6+
1.21619  -1  ##d
1.25369  -1  $-h
1.26444  -1  $i:+
1.33125  -1  ##f
1.48325  -1  $'E
1.54706  -1  $%l
1.66856  -1  $d
1.68156  -1  $-h
1.68550  -1  $6
1.81281  -1  ##v
```

1.90712 -1 $'e:

2.04219 -1 $t

2.10550 -1 $-h

2.12688 -1 ##Q

2.17738 -1 $-q

2.17738 -1 $aI

2.27281 -1 $n+

2.34612 -1 ##v

2.43219 -1 $'I

2.52069 -1 $n

2.62125 -1 $t

2.69000 -1 $-h

2.76938 -1 #.

Markup language:

Kiel extension of Entropic eps/waves format.

Existence of annotation tools (manual/automatic):

Esps/xwaves, manual labelling. Also possible as a tool for the raw transcription - hough not used in this case - would be htk.

Usability (in given machine environments):

Esps/waves environment.

Contact:

Prof. Dr. Klaus Kohler

Institut für Phonetik und digitale

Sprachverarbeitung (IPDS) der

Christian-Albrechts-Universität zu Kiel

D-24098 Kiel

A.56 Partitur Format at BAS

Coding book (domain and application):

The coding coding book (html), (ps) can be found at http://www.phonetik.uni-muenchen.de/Bas/BasFormatsdeu.html#Partitur

The Partitur Format is intended for the annotation of speech data.

Number of annotators:

Impossible to say.

Qualification of annotators:

Phoneticians and linguists.

Number of dialogues/turns/segments annotated:

More than 1000 dialogues. Transliteration only.

Language of corpora annotated:

Mainly German but also American English and Japanese. There are some data on German dialects as well.

Evaluations of scheme:

Not conducted.

Underlying task:

Mos of the dialogues are appointment dialogues.

List of phenomena annotated:

Orthography, segments, prosody, dialogue annotation, syntax, semantics

Examples:

LHD: Partitur Version

REP: Aufnahmeort

SNB: Anzahl Bytes pro Sample

SAM: Abtastrate in Hz

SBF: Bytereihenfolge (Intel 01, Motorola 10)

SSB: Bitauflösung

NCH: Anzahl Kanäle

SPN: Sprecher ID

LBD:

Markup language:

SAM compatible

Existence of annotation tools (manual/automatic):

There are segmentation applications used at BAS but there is no other special software used.

Usability:

Used in Verbmobil.

Contact:

Bavarian Archive for Speech Signals

A.57 SABLE (A Synthesis Markup Language)

Coding book (domain and application):

SABLE is a standard for the control of speaking style attributes of written texts for the use in speech synthesis systems. Thus, it has not really been used for the annotation of text under linguistic investigation.

A quote:

Currently, speech synthesizers are controlled by a multitude of proprietary tag sets. These tag sets vary substantially across synthesizers and are an inhibitor to the adoption of speech synthesis technology by developers.

This SABLE markup language is being developed with the following goals in mind:

Enable markup of speech synthesis text input.

Internationalized: appropriate to a large number of languages.

Easy to learn and use: SABLE should not require specialized knowledge of speech synthesis, linguistics or markup languages, though users with such experience should be able to apply their knowledge..

Portability: provide application developers with a consistent mechanism for controlling synthesizers from different companies and on different platforms.

Tools: enable the creation of tools for use and control of speech synthesis: for example, software that generates SABLE text, SABLE editing tools, pronunciation and lexicon tools, SABLE parsers and verifiers..

Extensibility: SABLE should be able to evolve to support new features in future releases. SABLE should allow individual synthesizers to provide enhanced features without compromising the portability of SABLE text.

Documentation can be found at the SABLE Homepage at http://www.bell-labs.com/project/tts/sable.html

Number of annotators:

Not available.

Qualification of annotators:

Speech engineers, linguists.

Number of dialogues/turns/segments annotated:

Not available

Language of corpora annotated:

Not available

Evaluations of scheme:

The standard is currently under development.

Underlying task:

Not available

List of phenomena annotated:

Emphasis, break, pitch, (speech)rate, volume, links to audio files (*audio*), synthesis system specific calls (*engine*), setting of markers (*mark*), special pronunciation, text unit interpretation (dates, phone numbers, ordinal/cardinal, postal code,...)

Examples:

The following example shows the tags needed to tell the synthesizer how "2pm" is to be interpreted (as time string) etc.

At <SAYAS MODE="time">2pm</SAYAS> on <SAYAS MODE="date"
MODETYPE="YM"> 98/3</SAYAS> Mike will send <SAYAS
MODE="currency">$4000</SAYAS> to <SAYAS MODE="net"
MODETYPE="email">me@acme.com</SAYAS>.

Markup language:

SGML/HTML related.

Existence of annotation tools (manual/automatic):

No tools available.

Usability (in given machine environments):

For speech synthesis.

Contact:

See WWW page.

A.58 SAM STANDARDS

Coding book (domain and application):

STANDARDS of the ESPRIT "SAM" Project No 2589: Speech Input and Output Assessment Methodologies and Standardization at http://www.icp.grenet.fr/Relator/standsam.html

The application area of this scheme is very general and not especially dedicated to linguistic phenomena, rather it provides a set of description items for the documentation of technical properties of speech files.

Number of annotators:

Unknown.

Qualification of annotators:

Engineers and linguists.

Number of dialogues/turns/segments annotated:

None.

Language of corpora annotated:

European languages.

Evaluations of scheme:

Not done.

Underlying task:

Not available.

List of phenomena annotated:

identification of file, comments, sampling rate, recording situation, sample byte order, number of significant bit per sample, number of bit per sample, number of channels, speaker information, protocol information, labelling expert/system, date of labelling, etc.

Examples:

LHD: V4.0

FIL: label

TYP: orthographic

DBN:

VOL: EUROM.1

DIR: ENGLISH

SRC: DFS20014.SES

TXF: S2.TXT

CMT: Information about the recording session

SAM: 16000

BEG: 0

END: 431872

RED: 07/11/89

RET: 15:10:33

REP: I.C.P. Grenoble (FR)

SNB: 2

SBF: 01

SSB: 16

RCC: 1

NCH: 1

SPI: M, 39, French

PCF: SENTEN.DES

PCN: 1

CMT: Information about the labelling session

EXP:

SYS:

DAT:

lSPA:

CMT: Item: label start, end, input gain, min level, max level, string

LBD:

LBR: 0, 55551, 0, -5128, 4775, Decimal numbers are an aid in

EXT: adding up.

DSC:

LBR: 55552, 158975, 0, -7680, 8878, Monetary systems have

EXT: evolved to make use of this base ten notation.

DSC:

LBR: 158976, 223743, 0, -7123, 7562, France became the first

EXT: decimal country in Europe.

DSC:

LBR: 223744, 275199, 6, -12487, 13262, Germany's decision

EXT: followed eight years later.

DSC:

LBR: 275200, 361983, 6, -11965, 12451, Scandinavian States and

EXT: Russia changed in eighteen seventy-five.

DSC:

LBR: 361984, 431872, 6, -12902, 14320, Britain chose to have

EXT: decimal money only in nineteen seventy-one !

ELF:

Markup language:

The structure of the entries in general consists of single lines which have a menomonic and and information part. As can be seen in the example above, information of different kind can be put into one line, e.g., time information and orthography.

Existence of annotation tools (manual/automatic):

Not known.

Usability (in given machine environments):

The format is very reader-friendly but rather impractical for the use in machine environments.

Contact:
Institut de la Communication Parlée

46 Avenue Félix Viallet

38031 Grenoble CEDEX 1

FRANCE

A.59 Text Encoding Initiative (TEI)

Coding book (domain and application):

TEI Guidelines: http://www-tei.uic.edu/orgs/tei/p3/elect.html

TEI is a scheme only. TEI is mostly concerned with different kinds of written texts but has also guidelines of the coding of transcripts of speech, noise, and images.

Number of annotators:

Not available.

Qualification of annotators:

Not available.

Number of dialogues/turns/segments annotated:

Not available.

Language of corpora annotated:

Not available.

Evaluations of scheme:

Not conducted.

Underlying task:

Not available.

List of phenomena annotated:

Basically, various structural units (word, line, verse) of all kinds of written texts (dictionaries, drama, prose). It also offers description items for speech, graphs, tables, and graphics.

Examples:

```
<text><body><head>My Alba</head>
<lg type=free>
<l>Now that I've wasted</l>
<l>five years in Manhattan</l>
<l>life decaying</l>
<l>talent a blank</l>
</lg>
<lg>
<l>talking disconnected</l>
<l>patient and mental</l>
<l>sliderule and number</l>
<l>machine on a desk</l>
</lg>
<!-- ... -->
</text>
```

Markup language:

SGML.

Existence of annotation tools (manual/automatic):

Not available.

Usability (in given machine environments):

Intended for the use in machine environment.

Contact:

See WEBpage: http://www-tei.uic.edu/orgs/tei/p3/elect.html

A.60 TRAINS Dialogue Corpus

Coding book (domain and application):

Corpus of problem solving dialogues.

Homepage of the TRAINS Dialogue Corpus at

http://www.cs.rochester.edu:80/research/speech/dialogues.html

Number of annotators:

Unknown.

Qualification of annotators:

Linguists.

Number of dialogues/turns/segments annotated:

first collection: ?

1991: 16 dialogues, 1000 turns, 3419 utterances

1992 and 1993: 5900 turns, 55,000 words

Language of corpora annotated:

English.

Evaluations of scheme:

Not conducted.

Underlying task:

In the first wave, the use of prosody was a topic under investigation. Within all dialogues, specific problems had to be solved and the solution had to be described.

List of phenomena annotated:

Speaker, simultaneous speech, text spoken, duration of dialogue, number of turns, number of utterances.

Examples:

```
Dialogue 91-1.1
Total Time:     1'38''
Total Turns:      20
Total Utterances: 63

UU#     Speaker: Utterance
1.1     M: okay
1.2      : I have to
1.3      : ship a boxcar of oranges to Bath by 8 o'clock today
        ...
2.1     S: okay
3.1     M: um
3.2      : so
3.3      : let's see
3.4      : where are there _oranges_
4.1     S: the oranges are in the warehouse
4.2      : at Corning
5.1     M: oh okay
5.2      : and I see that there's a tanker car there
5.3      : oh we don't want a tanker car do we
5.4      : um
5.5      : I have to get a boxcar
5.6      : to Corning
 5.7     : and then I have to load it with oranges and eventually I
           have to get that to Bath
5.8      : by 8 o'clock
6.1     S: right
...
Dialogue: d92-1
Number of utterances files: 54
Length of dialogue: 189.448858
Estimated number of turns: 32
```

utt1 : s: hello <sil> can I help you

utt2 : u: yeah I want t- I want to determine the maximum number of boxcars of oranges <sil> that I can get to Bath <sil> by seven a.m. <sil> tomorrow morning

utt3 : so <brth> hm <sil>
so I guess all the boxcars will have to go through oran- <sil> through Corning because that's where the orange juice <brth> orange factory is

utt4 : so from Corning to Bath how far is that

utt5 : s: two hours

utt6 : u: and it's gonna take us also an hour to load <sil> boxcars right

utt7 : s: right + +

utt8 : u: + okay + so <sil> hm so <sil> every trip will take at least <sil> three hours <sil> then

utt9 : um

utt10 : s: right we can unload any amount of cargo onto a train in one hour

utt11 : so we can + <sil> do a maximum of three + boxcars in an hour

utt12 : u: + right <sil> okay +

utt13 : okay <sil> so I guess one thing we can do oh <brth> so <brth> I guess one thing is that we should see how many boxcars we can actually get to Corning in four hours

utt14 : um how far is it from Avon to Bath <sil> to Corning

utt15 : s: <click> <brth> that's six hours it's + shorter + through Dansville

utt16 : u: + okay +

utt17 : from Avon oh no but the thing is is that I was thinking if s- <sil> I was wondering if we could actually pick up those two boxcars which are at <sil> + Bath +

...

Markup language:

SAM/DAMSL/SGML related.

Existence of annotation tools (manual/automatic):

The Dialogue Transcription Tools at

ftp.cs.rochester.edu/pub/papers/ai/94.tn1.rev.Dialogue_transcription_tools.ps.gz

Usability (in given machine environments):

See 11.

Contact:

http://www.cs.rochester.edu/research/trains/home.html

A.61 Verbmobil II Conventions for Spontaneous Speech

Coding book (domain and application):

Lexicon of transliteration conventions for spontaneous speech in VERBMOBIL II at http://www.phonetik.uni-muenchen.de/VMtrlex2d.html

Orthographic transliteration of dialogue and conversation data.

Number of annotators:

No information available.

Qualification of annotators:

Linguists.

Number of dialogues/turns/segments annotated:

No information available.

Language of corpora annotated:

German

Evaluations of scheme:

No evaluation.

Underlying task:

Appointment dialogues.

List of phenomena annotated:

Lexical units ('words'), syntactical/semantical structure, non-verbal articulatoric production, sounds, pauses, acoustic overlaps, comments, special comments, turns, language, false starts, correction/repetition, breath, hesitation

Examples:

, guten Tag , mein Name ist <!1 is'> ~J"ansch . <"ah> wir hatten

bereits telefoniert<Z> , mein Name ~J<Z>"ansch , $J $"A $N $S $C

$H , wegen <:<#Mikrobe> eines:> <:<#Mikrobe Arbeitstreffens:> .

Markup language:

Sgml related.

Existence of annotation tools (manual/automatic):

No special tools required or available.

Usability (in given machine environments):

Possible.

Contact:

Susanne Burger

Additional information:

Information is in German, this scheme is for the transliteration of spoken data, does not provide information for the segmentation of data.