

Annotating Participant Reference in English Spoken Conversation

John Niekrasz and Johanna D. Moore

School of Informatics
University of Edinburgh
Edinburgh, EH8 9AB, UK
{jniekras, jmoore}@inf.ed.ac.uk

Abstract

In conversational language, references to people (especially to the conversation participants, e.g., *I*, *you*, and *we*) are an essential part of many expressed meanings. In most conversational settings, however, many such expressions have numerous potential meanings, are frequently vague, and are highly dependent on social and situational context. This is a significant challenge to conversational language understanding systems — one which has seen little attention in annotation studies. In this paper, we present a method for annotating verbal reference to *people* in conversational speech, with a focus on reference to conversation *participants*. Our goal is to provide a resource that tackles the issues of vagueness, ambiguity, and contextual dependency in a nuanced yet reliable way, with the ultimate aim of supporting work on summarization and information extraction for conversation.

1 Introduction

Spoken conversation — the face-to-face verbal interaction we have every day with colleagues, family, and friends — is the most natural setting for language use. It is how we learn to use language and is universal to the world's societies. This makes it an ideal subject for research on the basic nature of language and an essential subject for the development of technologies supporting natural communication. In this paper, we describe our research on designing and applying an annotation procedure for a problem of particular relevance to conversational language — *person reference*.

The procedure is a coreference annotation of all references to people, and the focus of our scheme is on distinguishing different types of *participant*

reference (references to the conversation's participants), the predominant type of person reference in face-to-face multi-party conversation. Participant reference is exemplified by the use of proper names such as *James* or most commonly by the pronouns *I*, *you*, and *we*.

Participant reference plays an essential role in many of the most important types of expressed meanings and actions in conversation, including subjective language, inter-personal agreements, commitments, narrative story-telling, establishing social relationships, and meta-discourse. In fact, some person-referring words are the most frequent words in conversation.¹

Perhaps contrary to intuition, however, interpreting person-referring expressions can be rather complex. Person-reference interpretation is strongly dependent on social, situational, and discourse context. The words *you* and *we* are especially problematic. Either can be used for generic, plural, or singular reference, as addressee-inclusive or addressee-exclusive, in reference to hypothetical individuals or non-human entities, or even metonymically in reference to objects connected to individuals (Mühlhäusler and Harré, 1990; Wales, 1996). In addition, these and many other issues are not simply occasional problems but arise regularly.

Consider the following utterance from the AMI corpus of remote control design meetings, which is typical of the corpus in terms of complexity of person-reference.

¹The words *I* and *you* are the most frequently used nominals in several conversational corpora, including Switchboard (Godfrey et al., 1992) and the AMI Meeting Corpus (McCowan et al., 2005). In the British National Corpus they are the two most common of any words in the demographic (i.e., conversational) subcorpus (Burnard, 2007), and Google's Web 1T 5-gram statistics (Brants and Franz, 2006) list *I* and *you* as more frequent even than the word *it*. The word *we* falls within the top 10 most frequent words in all of these corpora.

“Current remote controls do not match well with the operating behaviour of the **user** overall. For example, **you** can see below there, seventy five percent of **users** zap a lot, so **you**’ve got **your person** sunk back in the sofa channel-hopping.”

As this example demonstrates, person-referring expressions have many potential meanings and are often vague or non-specific. In this case, “the user” refers to a non-specific representative of a hypothetical group, which is referred to itself as “users.” The first use of “you” refers to the addressees, but the second use has a more ‘generic’ meaning whilst retaining an addressee-oriented meaning as well. The phrase “your person” refers to a specific hypothetical example of the “users” referred to previously.

1.1 Purpose of the Annotations

The annotation research we describe here aims at addressing the fact that if conversational language applications are to be useful and effective (our interest is primarily with abstractive summarization), then accurate interpretation of reference to the conversation’s participants is of critical importance. Our work looks at language as a means for action (Clark, 1996), and our focus is on those actions that the participants themselves consider as relevant and salient, such as the events occurring in a meeting that might appear in the minutes of the meeting. For our system to identify, distinguish, or describe such events, it is essential for it to understand the participants’ roles and relationships to those events through interpreting their linguistic expression within the dialogue. This includes understanding direct reference to participants and recognizing discourse structure through evidence of referential coherence.

Another aim of our research is to increase understanding of the nature of participant reference through presenting a nuanced yet reliable set of type and property distinctions. We propose novel distinctions concerning three main issues. The first distinction concerns vagueness and indeterminacy, which is often exploited by speakers when using words such as *you*, *they*, and *we*. Our aim is to provide a reliable basis for making an explicit distinction between specific and vague uses, motivated by usefulness to the aforementioned applications. The second distinction concerns an issue faced frequently in informal conversation, where words typically used to do person-referring are also commonly used in non-person-referring

ways. A principal goal is thus establishing reliable person/non-person and referential/non-referential distinctions for these words. The third issue concerns addressing roles (i.e., speaker, addressee, and non-addressee), which we propose can be a useful means for further distinguishing between different types of underspecified and generic references, beyond the specific/underspecified/generic distinctions made in schemes such as ACE (Linguistic Data Consortium, 2008).

1.2 Summary and Scope of Contributions

The work described in this paper includes the design of an annotation procedure and a statistical analysis of a corpus of annotations and their reliability. The procedure we propose (Section 3) is based on a simple non-anaphoric coreference-like scheme, modest in comparison to much previous work. The produced dataset (Section 4) includes annotations of 11,000 occasions of person-referring in recorded workplace meetings. Our analysis of the dataset includes a statistical summary of interesting results (Section 4.1) and an analysis of inter-coder agreement (with discussion of specific disagreements) for the introduced distinctions (Section 4.2).

Though our annotation procedure is designed primarily for multi-party spoken conversation, some of the central issues that concern us, such as addressee inclusion and vagueness, arise in textual and non-conversational settings as well. Our scheme therefore has relevance to general work on reference annotation, though principally to settings where social relationships between the participants (i.e., speakers/authors and addressees/readers) are important.

2 Related Annotation Schemes

Previous work on reference annotation has covered a wide range of issues surrounding reference generally. It is useful to categorize this work according to the natural language processing tasks the annotations are designed to support.

2.1 Schemes for anaphora and generation

Several schemes have been designed with the goal of testing linguistic theoretical models of discourse structure or for use in the study of discourse processing problems like anaphora resolution and reference generation. These schemes have been applied to both text and dialogue and label dis-

course references with a rich set of syntactic, semantic, and pragmatic properties. For example, the DRAMA scheme (Passonneau, 1997) and the GNOME scheme (Poesio, 2000; Poesio, 2004) include labels for features such as bridging relation type and NP type in addition to a rich representation of referent semantics. Other schemes label animacy, prosody, and information structure to study their relationship to the organization and salience of discourse reference (Nissim et al., 2004; Calhoun et al., 2005). Recent developments include the explicit handling of anaphoric ambiguity and discourse deixis (Poesio and Artstein, 2008).

Despite the depth and detail of these schemes, participant reference has not been their main concern. The annotations by Poesio et al. (2000; 2004) include dialogue source material, but the rather constrained interactional situations do not elicit a rich set of references to participants. The scheme thus employs simple default labels for words like *I* and *you*. The work by Nissim et al., (2004) is an annotation of the Switchboard corpus (Godfrey et al., 1992), which contains only two participants who are neither co-present nor socially connected. Participant reference is thus rather constrained. Other than labeling coreferentiality, the Nissim scheme includes only a single distinction between referential and generic instances of the word *you*.

2.2 Schemes for information extraction

In contrast to the schemes described above, which are mainly driven toward investigating linguistic theories of discourse processing, some reference annotation projects are motivated instead by information extraction applications. For these projects (which includes our own), a priority is placed on entity semantics and coreference to known entities in the world. For example, the objective of the Automatic Content Extraction (ACE) program (Doddington et al., 2004) is to recognize and extract entities, events, and relations between them, directly from written and spoken sources, mostly from broadcast news. The schemes thus focus on identifying and labeling the properties of entities in the real world, and then marking expressions as referring to these entities. Recent work in the ACE project has expanded the scope of this task to include cross-document recognition and resolution (Strassel et al., 2008). In the ACE scheme (Linguistic Data Consortium, 2008), per-

son reference is a central component, and in the broadcast conversation component of the corpus there is an extensive inventory of participant references. The annotation scheme contains a distinction between specific, underspecified, and general entities, as well as a distinction between persons and organizations.

Another closely related set of studies are four recent investigations of second-person reference resolution (Gupta et al., 2007a; Gupta et al., 2007b; Frampton et al., 2009; Purver et al., 2009). These studies are based upon a common set of annotations of the word *you* in source material from the Switchboard and ICSI Meeting corpora. The purpose for the annotations was to support learning of classifiers for two main problems: disambiguation of the generic/referential distinction, and reference resolution for referential cases. In addition to the generic/referential distinction and an addressing-based reference annotation, the scheme employed special classes for reported speech and fillers and allowed annotators to indicate vague or difficult cases. Our work builds directly upon this work by extending the annotation scheme to all person-referring expressions.

3 Annotation Method

Our person-reference annotation method consists of two main phases: a preliminary phase where the first names of the conversation participants are identified, and a subsequent person reference labeling process. The first phase is not of central concern in this paper, though we provide a brief summary below (Section 3.2). The primary focus of this paper is the second phase (Section 3.3), during which every instance of person-referring occurring in a given meeting is labelled. We provide more detail concerning the most novel and challenging aspects of the person-referring labeling process in Section 3.4 and present a brief summary of the annotation tool in Section 3.5.

3.1 Source Material

The source material is drawn from two source corpora: the AMI corpus (McCowan et al., 2005), which contains experimentally-controlled scenario-driven design meetings, and the ICSI corpus (Janin et al., 2003), which contains naturally occurring workplace meetings. All the meetings have at least four participants and have an average duration of about 45 minutes. In the AMI corpus,

the participants are experimental subjects who are assigned institutional roles, e.g. project manager and industrial designer. This helps to establish controlled social relationships within the group, but generally limits the types of person referring. The ICSI meetings are naturally occurring and exhibit complex pre-existing social relationships between the participants. Person referring in this corpus is quite complex and often includes other individuals from the larger institution and beyond.

3.2 Labeling Participant Names

The first phase of annotation consists of identifying the names of the participants. We perform this task for every participant in every meeting in the AMI and ICSI source corpora, which totals 275 unique participants in 246 meetings. Despite the fact that the participants' are given anonymized identifiers by the corpus creators, determining participants' names is possible because name mentions are not excised from the speech transcript. This allows identification of the names of any participants who are referred to by name in the dialogue, as long as the referent is disambiguated by contextual clues such as addressing.

To extract name information, the list of capitalized words in the speech transcript is scanned manually for likely person names. This was done manually due to the difficulty of training a sufficiently robust named-entity recognizer for these corpora. Proceeding through each meeting for which any participant names are yet unidentified, and taking each potential name token in order of frequency of occurrence in that meeting, short segments of the recording surrounding the occurrences were replayed. In most cases, the name was used in reference to a participant and it was clear from discourse context which participant was the intended referent. In the AMI meetings, 158 of 223 (71%) of the participants' first names were identified. In the ICSI meetings, 36 of 52 (69%) were identified. While these numbers may seem low, failure to determine a name was generally associated with a low level of participation of the individual either in terms of amount of speech or number of meetings attended. As such, the proportion of utterances across both corpora for which the speaker's name is identified is actually 91%.

3.3 Person-reference Annotation

The second, principal phase of annotation consists of annotating **person-referring** — instances

of verbal reference to people. The recognition of person-referring requires the annotator to simultaneously identify whether a referring event has occurred, and whether the referent is a person. In practice, this is divided into four annotation steps: markable identification, referent identification, functional category labeling, and co-reference linking. For non-specific references, there is an additional step of labeling addressing properties. For each meeting, annotators label every instance of person-referring in every utterance in the meeting, performing the steps in sequence for each utterance. Section 4 describes the set of meetings annotated. The UML diagram in Figure 1 depicts the formal data structure produced by the procedure.²

The first step is markable identification, which involves recognizing **person-referring expressions** in the transcript. Only expressions that are noun phrases are considered, and only the head noun is actually labeled by the annotator — the extent of the expression is not labeled. These identified head nouns are called **markables**. Note, however, that before human annotation begins, an automatic process identifies occurrences of words that are likely to be head nouns in person-referring expressions. The list of words includes all personal pronouns except *it*, *them*, and *they* (these are more likely to be non-person-referring in our dataset) and the *wh*-pronouns (not labeled in our scheme). It also includes any occurrences of the previously identified proper names. Some of the automatically identified words might *not* be person-referring. Also, there may be instances of person-referring that are *not* automatically identified. Annotators do not unmark any of the automatically identified words, even if they are not person-referring. The resulting set of manually and automatically identified words, which may or may not be person-referring, constitute the complete set of markables.

The second step is the labeling of **person referents**. Any people or groups of people that are referred to specifically and unambiguously (see Section 3.4.3 for details) are added by the annotator to a conversation **referent list**. The list is automatically populated with each of the conversation participants.

²The diagram may also be viewed informally as loosely reflecting a decision tree for the main annotation steps. A complete coding manual is available from the author's web site.

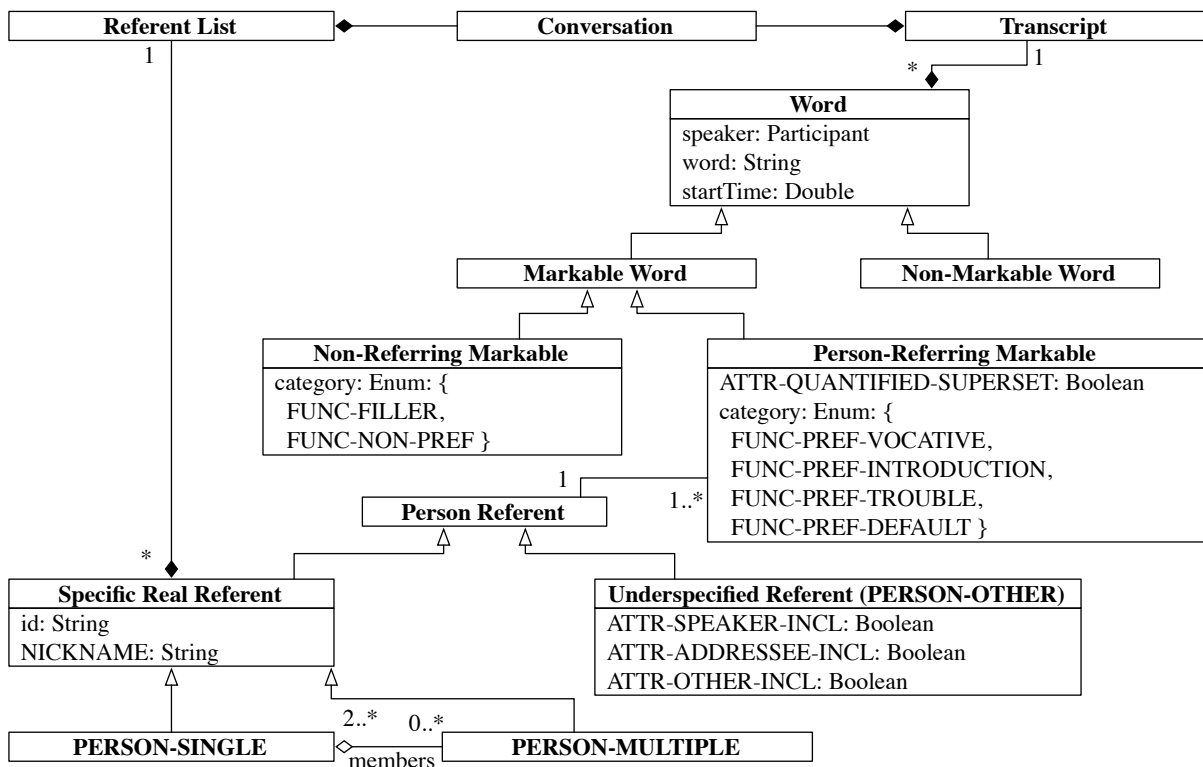


Figure 1: A UML diagram depicting the data structure used to represent and store the annotations.

The third step consists of labeling markables with a **functional category** (FUNC-*). The functional categories serve two main purposes. They are used to distinguish person-referring markables from all others (corresponding to the two main boxes in the diagram), and they are used to distinguish between specific dialogue purposes (the categories listed within the boxes, see Section 3.4.4).

The final step is to link the markables that were labeled as person-referring to the appropriate referent in the referent list. This is only done for specific and unambiguous referring. Otherwise, the referent is said to be **underspecified**, and instead of linking the markable to a referent, it is labeled with three binary **addressing inclusion attributes**. Inclusion attributes label whether the speaker, addressee, or any other individuals are included in the set of people being referred to, given the social, situational, and discourse context (details in Section 3.4.5).

3.4 Special Issues

3.4.1 Defining ‘person’ and ‘referring’

To be **person-referring**, an expression must satisfy two conditions. First, the expression’s primary contribution to the *speaker’s intended mean-*

ing or purpose must be either to identify, label, describe, specify, or address. These are the basic types of **referring**. Second, the referent being identified, labeled, etc., must be a **person**, which we define to include any of the following: a distinct person in the real world; a fictitious or hypothetical person; a human agent, perceiver, or participant in a described event, scene, or fact; a class, type, or kind of person, or representative thereof; a specification or description of a person or set of people; a (possibly vaguely defined) group or collection of any of the above; the human race as a whole, or a representative thereof.

If a noun phrase is used to do person-referring as defined, the associated markable is labeled with one of the four person-referring functional categories (FUNC-PREF-*). If a markable is not person-referring (either non-referring or referring to a non-person referent), it is labeled with the functional category FUNC-NON-PREF. The one exception to this is the use of a pre-defined list of common discourse fillers such as *you know* and *I mean*. When used as fillers, these are labeled with the non-referential FUNC-FILLER category.

3.4.2 Joint action and referring ‘trouble’

Annotators are asked to consider occasions of referring to be *joint* actions between the speaker and the addressee(s) of the utterance. The annotator assumes the role of an overhearer and considers as referring any case where the speaker’s *intended purpose* is to refer. If the instance of referring is not successfully negotiated between the participants (i.e., common ground is not achieved), but the speaker’s intended purpose is to refer, then the annotator marks this as `FUNC-PREF-TROUBLE`. This is used to identify problematic cases for future study.

3.4.3 Specific, Unambiguous Referring

Only the referents of *specific, unambiguous referring to a person in the real world* (`PERSON-SINGLE`) are included in the conversation referent list and made the subject of coreference annotation. References to more than one such individual can qualify (`PERSON-MULTIPLE`), but only if the members are *precisely enumerable* and qualify individually. The motivation for this distinction is to distinguish references that would be directly useful to applications. Coreference for underspecified references is not labeled.

3.4.4 Special Functional Categories

Two functional categories are used to distinguish special uses of person-referring for subsequent use in speaker name induction (the task of automatically learning participants’ names). The two categories are `FUNC-PREF-INTRODUCTION` and `FUNC-PREF-VOCATIVE`, which specify personal introductions such as “Hi, I’m *John*,” and vocative addressing such as “What do you think, *Jane*?” These categories are used only for proper names.

3.4.5 Addressing-based Inclusion Attributes

A major novelty in our annotation scheme is the use of addressing-based distinctions for underspecified referents. Rather than using the labels ‘generic’ or ‘indeterminate’, we employ three binary attributes (`ATTR-*-INCL`) that label whether the speaker, addressee or any other real individuals are members of the set of people referred to.

The use of this distinction is informed by the notion that addressing distinctions are of central importance to the recognition of joint activity type, structure, and participation roles. A generic pronoun, for example, will often have all three categories labeled positively. But as an example

of where this scheme creates a novel distinction, consider the phrase “You really take a beating out there on the pitch!”, where the speaker is a football player describing the nature of play to someone who has never played the game. This ‘generic’ use of *you*, used in an activity of autobiographical description, is intuitively interpreted as not including the addressee (`ATTR-ADDRESSEE-INCL=FALSE`) but including the speaker and others (`ATTR-{SPEAKER,OTHER}-INCL=TRUE`). These distinctions are hard to motivate linguistically yet critical to identifying useful properties relating to participation in the communicative activity.

3.4.6 Special or Difficult Cases

In some cases, an annotator can determine that a reference is specific and unambiguous for the participants but the annotator himself is unable to determine the identity of the referent. This is generally due to a lack of contextual awareness such as not having adequate video. In such cases, the annotator assigns a special `REF-UNKNOWN` referent.

Other difficult aspects of our annotation procedure are covered in the annotation manual, including handling of disfluencies, quantification, and identifying lexical heads.

3.5 Annotation Tool

The annotations were collected using a software tool we have designed for discrete event-based annotation of multi-modal corpora. The tool uses a simple, low-latency text-based interface that displays multiple streams of discrete events in temporal order across the screen. In our case, the events are time-synchronized words that are distributed to different rows according to speaker. The interface allows keyboard input only and is synchronized with the MPlayer playback engine.

4 Results and Analysis

4.1 Statistical summary

The dataset consists of approximately 11,000 individually annotated referring expressions in 16 experimentally-controlled, scenario-driven design meetings from the AMI corpus (McCowan et al., 2005) and 3 natural workplace meetings from the ICSI corpus (Janin et al., 2003). Figure 2 shows, for each grammatical type of referring expression, the frequency of occurrence of the five principal markable types, which are defined to consist of the two non-person-referring functional

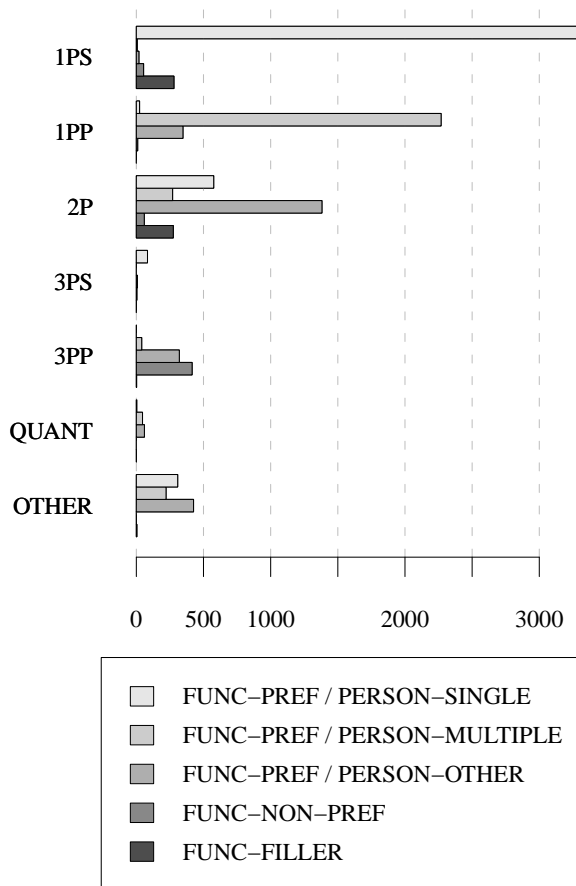


Figure 2: Frequency of occurrence of referring types for the whole corpus, by grammatical type of the referring expression.

categories (FUNC-NON-PREF and FUNC-FILLER), and a breakdown of person-referring according to the type of person referent: a specific individual (PERSON-SINGLE), multiple specific individuals (PERSON-MULTIPLE), or underspecified (PERSON-OTHER). The grammatical types include a grouping of the personal pronouns by grammatical person and number (1PS, 1PP, 2P, 3PS, 3PP), the quantified pronouns (QUANT), and a group including all other expressions (OTHER). Table 1 shows the relative frequency for the grammatical types and the most frequent expressions.

As is usually found in conversation, first-person and second-person pronouns are the most frequent, collectively comprising 82.0% of all person-referring expressions. Of particular interest, due to their high frequency and multiple possible referential meanings, are the 1PP and 2P categories (e.g., *we* and *you*), comprising respectively 24.6% and 23.7% of all person-referring expres-

Gram.	Freq. (%)	Ent. (bits)	Freq. words
1PS	33.7	.57	<i>I, my, me</i>
1PP	24.6	.67	<i>we, our, us</i>
2P	23.7	1.78	<i>you, your, yours</i>
3PS	.9	.66	<i>he, his, she</i>
3PP	7.2	1.25	<i>they, them, their</i>
QUANT	1.0	1.14	<i>everyone, everybody</i>
OTHER	8.9	1.57	<i>people, guys, user</i>

Table 1: A statistical summary of all the markables in the dataset by grammatical type (gram.), showing their frequency relative to all markables (freq.), the entropy of the referring type given the grammatical type (ent.), and a list of the most frequent examples (freq. words).

sions. In Table 1, we show the information entropy of the referring type, given the grammatical category. This measures the uncertainty one has about the type, given knowledge of only the grammatical type of the expression. The analysis reveals that second-person pronouns are a particularly challenging reference resolution problem, with a broad and relatively even distribution across referring types.

4.2 Reliability and Error Analysis

To show that our annotations are credible and suitable for empirical testing, we must establish that the subjective distinctions defined in our scheme may be applied by individuals other than the scheme developers. To do this, we assess inter-coder agreement between two independent annotators on four meetings from the AMI corpus, using Cohen’s Kappa (Cohen, 1960). Each of the decisions in the annotation procedure are assessed separately: markable identification, labeling referentiality, labeling specificity of person referents, and labeling addressing inclusion attributes. Because each decision depends on the previous, we employ a hierarchical assessment procedure that considers only instances where the annotators have agreed on previous decisions. This kind of multi-level assessment corresponds to that described and used in Carletta et al., (1997).

Markables The first annotation decision of interest is the identification of markables. Markables are either automatically identified occurrences of a pre-defined list of pronouns, or they are identi-

fied manually by the annotators. Agreement on this task, assessed only for manually identified words, was very good ($\kappa=.94$). Error analysis shows that the main issue with this decision was not determining lexical heads, but rather determining whether phrases such as “all age *groups*,” “the older *generation*,” and “the business *market*” should be considered as referring to *people* or not.

Person referentiality The next annotation decision is between person-referring and non-person-referring markables. For assessment of this choice, we measure agreement on a three-way categorization of the agreed markables as either FUNC-NON-PREF, FUNC-FILLER, or one of the FUNC-PREF-* categories. Agreement on this task was good ($\kappa=.77$). The only errors occurred on first- and second-person pronouns and between the FUNC-NON-PREF and FUNC-PREF-* categories. Error analysis suggests confusion tends to occur when pronouns are used with semantically light verbs like *go*, *get*, and *have*, for example in phrases such as “there *we* go” and “*you*’ve got the main things on the front.” As in the latter example, some of the difficult choices appear to involve descriptions of states, which the speaker can choose to express either from various participants’ points of view, as above, or alternatively without explicit subjectivity, e.g., “the main things are on the front.”

Specificity and cardinality The next choice we assess is the decision between referring specifically to a single person (PERSON-SINGLE), to multiple people (PERSON-MULTIPLE), or as underspecified (also referred to as PERSON-OTHER). Agreement on this choice was very good ($\kappa=.91$), though considering only the difficult 1PP and 2P grammatical categories (e.g., *we* and *you*), agreement was less strong ($\kappa=.75$). Note that due to the hierarchical nature of the scheme, evaluation considered only cases where both annotators labeled a word as person-referring. Errors on this decision often involved ambiguities in addressing, where one annotator believed a particular individual was being addressed by *you* and the other thought the whole group was being addressed. Another common disagreement was on cases such as “*we* want it to be original,” where *we* was interpreted by one annotator as referring to the present group of participants, but by the other as (presumably) referring to the organization to which the participants

belong.

Addressing inclusion attributes For the three inclusion attributes for underspecified referents (ATTR-*-INCL), agreement is calculated three times, once for each of the binary attributes. Agreement was good, though slightly problematic for addressee inclusion (speaker $\kappa=.72$; addressee $\kappa=.50$; other $\kappa=.66$). Disagreements were mainly for occurrences of *you* like the example of autobiography in Section 3.4.5. For example, “it’s *your* best friend” was used to explain why a dog is the speaker’s favorite animal, and the annotators disagreed on whether the addressee was included.

5 Conclusion

We have presented an annotation scheme and a set of annotations that address *participant reference* — a conversational language problem that has seen little previous annotation work. Our focus has been on eliciting novel distinctions that we hypothesize will help us to distinguish, label, and summarize conversational activities. We also address the issues of vagueness, ambiguity, and contextual dependency in participant referring.

Based on analysis of inter-annotator agreement, the major distinctions proposed by the scheme appear to be reliably codable. In addition, our statistical analysis shows that our dataset contains a wide variety of participant references and should be a useful resource for several reference resolution problems for conversation. Our novel method for distinguishing specific reference to real individuals appears to be very reliably codable. Our novel addressing-based distinctions for underspecified reference are less reliable but adequate as a resource for some dialogue structuring tasks.

Further work proposed for this task includes labeling a variety of conversational and non-conversation genres. Our immediate concern is to apply our annotations in the training and/or testing of machine learning approaches to discourse segmentation and abstractive summarization.

References

- Thorsten Brants and Alex Franz. 2006. Web 1T 5-gram, Version 1. Linguistic Data Consortium. Catalog ID: LDC2006T13.
- Lou Burnard, 2007. *Reference Guide for the British National Corpus (XML Edition)*. Research Technologies Service at Oxford University Computing Services.

- Sasha Calhoun, Malvina Nissim, Mark Steedman, and Jason Brenier. 2005. A framework for annotating information structure in discourse. In *Proceedings of the ACL Workshop on Frontiers in Corpus Annotation II: Pie in the Sky*.
- Jean Carletta, Stephen Isard, Anne H. Anderson, Gwyneth Doherty-Sneddon, Amy Isard, and Jacqueline C. Kowtko. 1997. The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–31.
- Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge.
- Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.
- George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. The Automatic Content Extraction (ACE) program: Tasks, data, and evaluation. In *Proc. LREC*.
- Matthew Frampton, Raquel Fernandez, Patrick Ehlen, Mario Christoudias, Trevor Darrell, and Stanley Peters. 2009. Who is "you"? Combining linguistic and gaze features to resolve second-person references in dialogue. In *Proc. EACL*.
- John J. Godfrey, Edward Holliman, and J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proc. ICASSP*, pages 517–520, San Francisco, CA.
- Surabhi Gupta, John Niekrasz, Matthew Purver, and Daniel Jurafsky. 2007a. Resolving "you" in multi-party dialog. In *Proc. SIGdial*, pages 227–230.
- Surabhi Gupta, Matthew Purver, and Daniel Jurafsky. 2007b. Disambiguating between generic and referential "you" in dialog. In *Proc. ACL*.
- A. Janin, D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, and C. Wooters. 2003. The ICSI meeting corpus. In *Proc. ICASSP*, volume 1, pages 364–367.
- Linguistic Data Consortium, 2008. *ACE (Automatic Content Extraction) English Annotation Guidelines for Entities, Version 6.5*. Downloaded from <http://projects.ldc.upenn.edu/ace/annotation/>.
- I. McCowan, J. Carletta, W. Kraaij, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, W. Post, D. Reidsma, and P. Wellner. 2005. The AMI Meeting Corpus. In *Proceedings of Measuring Behavior 2005, the 5th International Conference on Methods and Techniques in Behavioral Research*, Wageningen, Netherlands.
- Peter Mühlhäusler and Rom Harré. 1990. *Pronouns and People: The Linguistic Construction of Social and Personal Identity*. Blackwell, Oxford.
- Malvina Nissim, Shipra Dingare, Jean Carletta, and Mark Steedman. 2004. An annotation scheme for information status in dialogue. In *Proc. LREC*.
- R. Passonneau, 1997. *Instructions for applying discourse reference annotation for multiple applications (DRAMA)*.
- Massimo Poesio and Ron Artstein. 2008. Anaphoric annotation in the ARRAU corpus. In *Proc. LREC*.
- Massimo Poesio, 2000. *The GNOME Annotation Scheme Manual, Version 4*. University of Edinburgh, HCRC and Informatics.
- Massimo Poesio. 2004. Discourse annotation and semantic annotation in the GNOME corpus. In *Proceedings of the ACL 2004 Workshop on Discourse Annotation*, pages 72–79.
- Matthew Purver, Raquel Fernandez, Matthew Frampton, and Stanley Peters. 2009. Cascaded lexicalised classifiers for second-person reference resolution. In *Proc. SIGdial*, pages 306–309.
- Stephanie Strassel, Mark Przybocki, Kay Peterson, Zhiyi Song, and Kazuaki Maeda. 2008. Linguistic resources and evaluation techniques for evaluation of cross-document automatic content extraction. In *Proc. LREC*.
- Katie Wales. 1996. *Personal pronouns in present-day English*. Cambridge University Press, Cambridge.