Characteristics of the different usages of “You”
René Groothedde
University of Twente, The Netherlands
r.groothedde@student.utwente.nl

ABSTRACT
Sometimes when a speaker uses the English pronoun you he refers to an addressee, or a group of addressees, but other times the word is used in a generic, not in a deictic, sense. This research tries to find characteristics of these different usages of you in multi-party conversations in meetings.

Keywords
Generic, deictic-referential, plural, singular, you, discourse marker, conversation

1. INTRODUCTION
Many words in English have multiple meanings. When it is not exactly clear from the word itself or the form of the word what the meaning is, one has to look at other features of the word or the context of the word. While this is what native speakers normally do unconsciously, for a computer it is very difficult to find such features and to interpret these features in a correct way. It can be very important for a computer to understand what someone is saying, for example when a computer has to translate a text into another language or when a computer has to make a summary of a meeting.

1.1 Different Usages of you
The English pronoun you is a frequently used word in conversations and the meaning of the word depends on the context.

First there is a difference between deictic-referential and generic use of you. Deictic you is referring to a specific person or persons that are directly involved in the conversation (example 1 and 2). Generic you is also called impersonal or indefinite you and is referring to an arbitrary imagined experiencer (ex. 3). Instead of this use of you it is also possible to use the more formal word one. In Dutch, the deictic singular use of you can be translated to jij or the more polite way u, the deictic plural to jullie and the generic use to men. It is also possible to use je for both uses.

Second there is a difference in quantity of the people referred to by the deictic-referential you. You can refer to one specific person (ex. 1) or to a group of persons (ex. 2). Most of the time this group only consists of all or some of the other persons that are involved in the conversation but it is also possible that this is a group that is represented by one or more persons, for example a country (ex. 4) or a type of person.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

12th Twente Student Conference on IT, January 22, 2010, Enschede, The Netherlands.
Copyright 2010, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

(1) I think it’s good. You’ve done a good review.
(2) I don’t know if you guys have any questions.
(3) Often, you need to know specific button sequences to get certain functionalities done.
(4) You Russians are very good musicians.

There is also a difference between the collective and the distributive use of the plural you. Collective you means that you is referring to the group as a whole and distributive means that you is referring to every person of the group individual. This difference is not direct a property of you but more a property of a group and is difficult to understand.

In the literature there are a lot of opinions about what this difference is exactly [10]. In some cases it is clear which use it is, like in example (5) where every individual can take notes if he wants to, but no one is thinking about taking notes together. In example (6), it is clear that one piano is carried by two persons and not every single person is carrying his own piano. In the latter one, it is general knowledge that normally no one can carry a piano by its own.

In other cases it depends on the context. In example (7), when you is plural, it can mean that I will give the group as a whole one paper (collective) or that I will give every single person a paper (distributive). Sometimes the distinction is rather vague like in example (8) where both the group as a whole and every person individually is sleeping in the hotel.

(5) You can take notes if you want. (Distributive)
(6) You two have to carry a piano. (Collective)
(7) I will give you a paper.
(8) Tonight you will sleep in this hotel.

A special case of you is the discourse marker you know (ex. 9). The literal meaning of you know implies knowledge on the part of the hearer but according to Fuller, “researchers agree that you know functions to create focus on the information it frames.” [3] So it is not saying anything about the knowledge of the hearers.

(9) It’s not just, you know, noises like something hitting.

There are also other variants of you like your, yours, yourself and yourselves which also can be used in a generic and deictic-referential way.

1.2 Structure of a Conversation
In this paper we look at the different usages of you in conversations. A conversation can be marked up in several ways. One way is to annotate the transcription of a conversation according to speaker intention, what kind of thing each person is trying to achieve by what they say. Every single part of the conversation with its own intention is called a dialogue act.

In this paper we distinguish several types of dialogue acts. There are dialogue acts about the information exchange, acts about possible actions, comments on previous discussion, social acts like being-positive or being-negative, non-intentional acts e.g. backchannels, such as um and uh-huh, and attempts to maintain the floor.
An important feature of the dialogue act is the addressee. According to Jovanovic et al. [6], “the addressee is the one who is expected by the speaker to react on what the speaker says and to whom, therefore, the speaker is giving primary attention in the present act.” It is possible that the dialogue act is addressed to the whole group of participants of the meeting, to a particular subgroup of them, or to a single participant in particular.

Another way to segment a conversation is in terms of turns. A turn is a succession of dialogue acts of one speaker which are spoken directly after each other without being interrupted by another speaker.

A more abstract and complicated way to look at a conversation is in terms of floors. Edelsky [1] defined floor as “a psychologically developed, interactional space among the interactants”. A floor holder is the speaker that is the main speaker at that moment, the speaker that ‘has the floor’. This can change during the conversation and is not always clear.

There are more ways to segment a conversation. For example, you can look at the different subjects that are discussed or at the different types of activities, like giving an order, explaining something, giving an opinion etc. These activities can cover a number of dialogue act segments and sometimes even a number of turns. So they are often larger episodes of a conversation.

### 1.3 Research

The main research question in this paper is:

*What are the characteristics of the different usages of the English pronoun “you”?

In this research we focus on the characteristics that are distinguishing between generic and deictic usages of you, and between deictic singular and deictic plural you. We also look at the characteristics of you included in a discourse marker, like you know.

In order to find these characteristics we look at the correlation between the different usages of you and several features of the meeting during the usage of you.

These features include the type of the dialogue act, the addressee of the dialogue act, features of the meeting and agents, and lexical features

We expected that the type of the dialogue act and the addressee would have a strong relation with the different usages of you.

The paper is structured in the following way:

In the second section we take a look at the work of other people that is related to this research. In the third section we explain our research method and the used data. In the fourth section we look at deictic versus the generic usages of you. In the fifth section we discuss the plural versus the singular usages of you and in the sixth section the discourse marker. In the last section we give our conclusion.

### 2. RELATED WORK

In the past, Gupta et al. [4,5] used prosodic features, various lexical features and discriminative classifiers (support vector machines and conditional random fields) for differentiating between generic and deictic-referential you automatically. They found out that the context, prosodic and part of speech features they used, were not very useful for differentiating between generic and deictic-referential you. However, dialog act and sentential features were very useful, achieving an accuracy of 84% (baseline 55%) on two-party conversations and 75% (baseline 58%) on multi-party dialogue.

Recently, Frampton et al. [2] extended Gupta et al.’s method to include gaze direction. This gave a small improvement on the deictic vs. generic problem, achieving 79% accuracy. Purver et al. [8] used not only visual and discourse features, but also lexical features like n-grams. They also used Information Gain methods to examine the contribution of particular lexical features. Generic uses of you were more likely to appear in utterances containing words related to the main topic of the meetings and deictic-referential uses were more likely to appear in utterances containing words related to the management of the meetings. Some other words and n-grams were found which were correlated with different usages of you.

These researches focused on finding the best features for an automatic classifier. Most of them only mentioned which features were the best as input for the machine learning classifier but not why these features were the best. This is exactly what we are going to find out in our research.

### 3. DATA AND METHODS

For this research, we used sixteen meetings of the AMI Meeting Corpus [7]. In these scenario meetings, the participants play the roles of members of a design team whose task is to develop a prototype of a new television remote control. Each participant is given a different role to play. In each design project, there is a project manager (PM), a market expert (ME), a user interface designer (UI) and a industrial designer (ID). There are four phases in each project:

- Project kick-off
- Functional design
- Conceptual design
- Detailed design

We used four meetings of every phase from seven different projects. So totally there are twenty-eight different participants in this meetings.

The data set is annotated for a range of properties. For our research, we used the speech transcriptions, and the dialogue act annotations. The dialogue annotations included the addressee, and the dialogue act type. There are totally 12715 dialogue acts in all the sixteen meetings.

All these properties are annotated in the NXT XML format using the NITE XML Toolkit. For the you-usage annotations we used the annotations from Gupta et al. [5]. We had to translate these annotations from a plain text file to the NXT format using our own software written in Java. The types of you in table 1 are annotated by Gupta et al.

<table>
<thead>
<tr>
<th>Type</th>
<th>you</th>
<th>your</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISCOURSE</td>
<td>245</td>
<td></td>
<td>245</td>
</tr>
<tr>
<td>GENERIC</td>
<td>842</td>
<td>121</td>
<td>964</td>
</tr>
<tr>
<td>DEICTIC</td>
<td>794</td>
<td>117</td>
<td>921</td>
</tr>
<tr>
<td>PLURAL</td>
<td>288</td>
<td>23</td>
<td>313</td>
</tr>
<tr>
<td>SINGULAR</td>
<td>506</td>
<td>94</td>
<td>608</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1881</td>
<td>238</td>
<td>2130</td>
</tr>
</tbody>
</table>

The Stanford part of speech tagger [9] is used to get the part of speech of each word.

We also used floor annotations from one meeting. This meeting was annotated by four different persons of University Twente.
For each dialog act is annotated who the floor holder is at that moment. We used only the dialogue act with two or more annotations from different annotators and with totally agreement between the annotators.

In order to analyze the meetings, we made a framework-layer in Java over the standard Java API that was included with the NXT Toolkit. That made it possible for us to include the modeling of turns in this layer, using the following definition of turn: A turn is a sequence of dialogue acts of the same speaker with less than 0.5 seconds between them. There are totally 8445 turns in all the 16 meetings.

This framework is used to export the annotations to a Microsoft Access database. The advantage of this database over the XML Toolkit is that you can make queries of the whole database in an easy and often visual way. We exported the results of the queries to Microsoft Excel and used the statistical tools of Excel to perform statistical analysis.

In most cases, we used the Pearson's chi-square test with one degree of freedom to test if there is a relation between the type of the you usage and some other feature. We talk about a strong relation or something significant if p<0,001 and talk about a relation if p<0,05. In all of these cases, a 2x2 cross-table is used comparing a feature and all the other features of the same type with two types of you usages. In case of the discourse marker you know, all the other you types are grouped as not-discourse. We also used the Student's t-test but we will describe that explicitly when we use it.

4. DEICTIC VS. GENERIC

In this section, we only look at the characteristics which are differentiating between generic and deictic usages of you. We ignore the discourse marker and talk about that in section six.

In the next subsections the results are shown and at the end of each subsection we will discuss these results.

The goal of these meetings is to design a remote control. This means that there is a lot of talking about what the user wants and how the user will use the remote control. It is expected that the generic usage of you will be used to refer to the imagined end-user and that it will occur while talking about the main topic of the meetings.

4.1 Dialogue Act Type

There is a very strong relation between the type of the DA and the difference between generic and deictic you (table 2).

<table>
<thead>
<tr>
<th>Dialogue Act Type</th>
<th>GENERIC</th>
<th>DEICTIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assess</td>
<td>61</td>
<td>37</td>
</tr>
<tr>
<td>Social(Be-Positive/Negative)</td>
<td>7</td>
<td>68</td>
</tr>
<tr>
<td>Comment-About-Understanding</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Elicit-Assessment</td>
<td>5</td>
<td>80</td>
</tr>
<tr>
<td>Elicit-Comment-Understanding</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Elicit-Inform</td>
<td>24</td>
<td>125</td>
</tr>
<tr>
<td>Elicit-Offer-Or-Suggestion</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>Fragment</td>
<td>55</td>
<td>38</td>
</tr>
<tr>
<td>Inform</td>
<td>396</td>
<td>196</td>
</tr>
<tr>
<td>Offer</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Stall</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Suggest</td>
<td>89</td>
<td>142</td>
</tr>
<tr>
<td>Total</td>
<td>647</td>
<td>761</td>
</tr>
</tbody>
</table>

Generic usages of you do also occur in social acts. In most of these cases, the act is intended to be humorous, like in example 3.

(2) Wow. You're trendy.
(3) You can throw Homer when you're frustrated, doh.

When you occurs in a dialogue act marked as offer, it is more likely that this is a deictic usage.

When people do an offer, often there is also a person who is the indirect object of the offer. This is the case in 14 of the 24 deictic usages in an offer act. The you is referring to that person, like in example 4.

(4) I will send you a mail.

There is also a relation between the assessment dialogue acts and the generic usage of you. This make sense because most of the assessments in these meetings are expected to be about the topic and thus about the generic user (ex. 5) and not about the agents themselves (ex. 6).

(5) So maybe just having many functions in one button is not that good if you want to zap a lot.
(6) You made a very nice prototype.

The strongest relation between a you type and a dialogue act type is the relation between the inform acts and the generic usages of you. In 396 of the 647 generic usages of you the dialogue act is annotated as an inform act. In almost every of these acts, the given information is about the remote control (ex. 7). This confirms our expectation that the generic you is referring to the end-user of the remote control. However, there are also a lot of inform acts that contain a deictic usage of you. Most of these acts give information about the management of the meeting (ex. 8) or use the deictic you as a kind of discourse marker, like in example 9. Of course it is clear that all the listeners can see this, so you can see is not used to give real information.

(7) You can throw it to the television.
(8) So I will ask you all to do the same.
(9) As you can see, there's still lots of space on the actual remote control.
4.2 Addressee
A strong relation was found between the deictic usage of you and the addressing of an individual, and between the generic usage of you and the addressing of the group (Table 3).

Table 3. Addressee vs. you type

<table>
<thead>
<tr>
<th>Addressee</th>
<th>GENERIC PLURAL</th>
<th>DEICTIC SINGULAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual</td>
<td>118 33</td>
<td>304</td>
</tr>
<tr>
<td>Group</td>
<td>306 136</td>
<td>54</td>
</tr>
<tr>
<td>Unknown</td>
<td>74 12</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>498 181</td>
<td>388</td>
</tr>
</tbody>
</table>

Most of the deictic usages that are addressed to an individual are singular, so we will discuss the relation between deictic and individual in section 5.

As we saw before, most of the generic usages are used in a inform act, giving information about the topic. Most of the time this information is addressed to the group so that explains the relation between generic you and the addressing of the group. However, there are also generic uses that are addressed to an individual, like in example 10.

(10) But you need another sensor for that, right?

4.3 Lexical Features
We looked also at all the words in the dialogue act and their part of speech.

There is a strong relation between the question mark (?) and the deictic usage of you. Another relation that is related to this one is that between the deictic you and the occurrence of a verb directly before you, especially the verbs do, did and are. This all show us that there is also a relation between the deictic usage and question sentences. However, there are cases that there is a verb directly before a generic usage of you. In all these cases, you is not the grammatical subject of the sentence, like in example 11.

(11) I don’t think you need the power.

The nouns in the dialogue act say a lot about whether the you usage is deictic or generic. If the dialogue act contains nouns about the management of the meeting, like email, presentation, meeting, slide and project, the you usage is significant more likely to be deictic. If the dialogue act contains nouns about the topic of the meeting, such as button, channel, menu and battery, it is more likely to be generic.

If there already is a deictic personal pronoun in the dialogue act like me, us, I or we, it is likely that the you usage in the dialogue act also is deictic. In contrast, the personal pronoun it has a strong relation with the generic usage. It refers most of the time to the remote control or to parts of it, so it is more likely that in a dialogue act containing it, you will refer to the imagined user than to a real person.

There are two kind of verbs that are related with the generic usage of you. First there are the verbs that say something about the topic of the meeting, like turn, use, make, change, put. The other verbs are verbs like could, can, can’t. These verbs say something about the possibility of something. Almost all the verbs of the first kind are preceded by the verbs of the second kind. This confirms our expectation that the generic you is referring to the imagined user of the remote control.

Verbs that have a relation with the deictic usage of you are the verbs will, think, wanna, see, know, did, had, do, are.

4.4 Phase of the Meeting
For each meeting, we calculated if there was a significance relation between the meeting and the usage of generic or deictic you.

Three of the four meetings in the first phase(project kick-off) have a strong significant relation with the deictic usage and none of them with the generic usage.

Three of the four meetings in the third (Conceptual Design) phase have a significant relation with the generic usage and none of them with the deictic usage.

In the first phase the agents are getting to know each other and there is a lot of management things that need to be done, so they talk about real persons and use deictic you’s to refer to them.

In the third phase the agents will talk about the conceptual design. They talk a lot about the imagined end user of their product and use generic you to refer to that user.

4.5 Index turn in Meeting

![Figure 1. Relative index turn in meeting](image)

We looked at the relative place of the turn in the meeting (Figure 1). The first and the last five percent of the turns in the meetings are significant more deictic and less generic.

The reason why there are so less generic usages and so much deictic usages in the begin and the end of the meeting is because the generic you is related to the main topic of the meeting. A meeting never starts and end with speaking about the main topic but most of the time with some social acts like greeting and with management acts, using deictic you.

4.6 Role of the agent

There are in total twenty-eight agents and seven agents of each role. For every agent, we calculated if there was a significance relation between the agent and the usage of generic or deictic you.

Eight of the fourteen user interface and industrial designers had a significant relation with the generic usage of you and none of them had a significant relation with the deictic usage.

Five of the seven project managers had a significant relation with the deictic usage of you and none of them with the generic.

None of the marketing experts had a significant relation with either the generic or the deictic usage.
The user interface and industrial designers talk a lot about the remote control and also about the imagined user of it so they use the generic usage of you.

The project managers need to give orders and take care of what the other agents are doing so the project managers use deictic you to refer to the other agents.

### 4.7 Other Results

There is a significant difference in the number of words in the dialogue act between the deictic and generic you (t-test, p<0.001). The acts containing the generic you are longer on average than the acts containing the deictic you.

There is also a significant difference in the number of dialogue acts in the turn.

For every you usage we looked how many other usages there of the same type in the dialogue act. Generic usages occur significant more multiple times in a dialogue act. (t-test, p<0.0001).

Only three dialogue acts contain both deictic and generic usages (ex. 12).

(12) *Cause like you say, it's something you can put your hands on.*

There is a significant relation between the occurrence of the discourse marker you know in the dialogue act and the generic usage of you.

### 5. PLURAL VS. SINGULAR

In this section, we look at the characteristics that are distinguishing between the plural and the singular deictic usages of you. The generic usages and the discourse markers are ignored.

#### 5.1 Dialogue Act Type

A you usage in an elicit-inform dialogue act is more likely to be singular than plural (p<0.05). In this kind of dialogue act, the speaker wants to have some kind of information from someone else who is expected to have this information. Most of the time, you will refer to that person or persons, like in example 13.

(13) *Do you also do marketing?*

In contrast, a you usage in an elicit-offer-or-suggestion act is more likely to be plural (p<0.05). In this kind of dialogue act, the speaker expresses a desire for someone to make an offer or suggestion. In many cases the speaker will not know who is going to answer so you will refer to the whole group.

(14) *Do you have any other ideas?*

#### 5.2 Addressee

Like expected there is a very strong relation between the addressee and plural or singular you. Plural you is most of the time addressed to the group and singular you to an individual. There are also cases where it is annotated that singular you is addressed to a group and plural you is addressed to an individual. We think that in most of these cases the annotators of the type of the you usage didn’t agree with the annotators of the addressee. For example, we think that it is impossible to address dialogue acts like example 15 and 16 to a group while you is singular or visa versa.

(15) *Do you have any problems with that?*

(16) *Okay, see you later. Bye.*

However, in some cases the addressee is an individual and you is referring to a group that is represented by that individual, like in example 17.

(17) *Oh you guys are always the dampers on these projects, you industrial designers.*

### 5.3 Words in Dialogue Act

There is a strong relation between plural you and the words all and guys (ex. 18 and 19). These words are explicitly used to make clear that you is referring to more people.

(18) *So I will ask you all to do the same.*

(19) *What are you guys doing now?*

Another word that is related to the plural you is the verb see. In most of these cases it is part of you can see or as you can see.

Other parts that are correlated with the plural you are you will and thank you.

### 5.4 Dialogue Act Length

There is a significant difference in the number of words in the dialogue act between the plural and singular you (t-test, p<0.001). The acts containing the plural you are longer on average than the acts containing the singular you.

### 6. DISCOURSE MARKER

In the last section we look at the characteristics of the discourse marker you know. We do this by comparing the discourse markers with all the other types of you together.

We looked at the relation between the usage of the discourse marker you know and the agents. We found that there is a lot of variation between the agents. Some agents use the discourse marker a lot and some other never use it.

Looking at the native language of the agents, we saw that five of the eight agents that speak French as their native language had a significant relation with ‘not discourse’. Together they used only ten times the discourse marker.

We found a strong relation between the you usage not being the discourse marker and the elicit dialogue act. In only 13 of the 245 usages of the discourse marker, the type of the dialogue act is elicit.

If you occurs in a dialogue act that is annotated as stall, it is very likely that this is a discourse marker. The stall acts are used by speakers when they want to have the turn or keep the turn.

Of course there is a strong relation between the discourse marker and the occurrence of the verb know after you. But there are still 39 occurrences of you know that are not a discourse marker.

There is a relation between the occurrences of a pause before or after you know with the discourse marker usage of you.

If the sentence is a question it is not likely that the you usage is a discourse marker, because we found a strong relation between not being a discourse marker and the question mark or the verb do before you.
7. FLOOR
We looked at the floor-annotations to find a relation between the different you-usages and the floor-holder. There are not enough annotations to do real statistical tests or to draw significant conclusions. However, in all the 6 cases that an agent, who was not the floor holder at that moment, uses you, it is a singular deictic you usage.

We think that this relation is very likely to exist but there are more floor annotations needed.

8. CONCLUSION
A lot of characteristics are found that distinguish between the generic and deictic usages of you. Generic you is related to the informative, relative longer dialogue acts that often are about the topic of the meeting. These dialogue acts are often addressed to the group. Our assumption that the generic you is referring to an imagined user of the product explains most of the characteristics of the generic usage.

In contrast, the deictic usages occur more in shorter dialogue act and are less about the topic of the meeting but more about the management of the meeting. These dialogue acts are often addressed to an individual.

We didn’t find a lot of characteristics that are distinguishing between the plural and singular usages of you. However, as expected we found a very strong relation between the addressing of the dialogue act to the group/individual and the plural/singular usages. We also found some words like guys en all that are related to the plural usage of you. Including focus of attention features would probably lead to more results.

As expected, the most important characteristic of the discourse marker is the occurrence of the verb know behind you. Other characteristics are the pause before you or after know. The discourse marker almost never occurs in question sentences.

The difference between our research and the related research is that we have formulated some explanations for the found characteristics, especially for the characteristics of the generic usage. The other researches don’t see the generic usage as referring to a person, but we think it does refer to the imagined user of the product.

These results can be used for automatic classifying of you but also to detect the addressee of the dialogue act when the type of the you usage is known.

9. REFERENCES