WHAT DO(ES) YOU MEAN?
THE PRAGMATICS OF GENERIC SECOND PERSON PRONOUNS IN MODERN SPOKEN DANISH[1]

Torben Juel Jensen and Frans Gregersen

Abstract

In modern Danish, the most frequently used pronoun for generic reference is *man*, developed from the noun *man(d)* 'man'. Recently, though, the second person singular pronoun *du* has gained ground, in parallel to similar recent developments in other languages. A large-scale, longitudinal study of the LANCHART corpus of spoken Danish has documented a rise in the use of generic *du* in Copenhagen (and later in the rest of Denmark) during the period from the early 1970s, where generic *du* was practically non-existent, till the late 1980s where *du* comprised around 25% of all pronouns with generic meaning. However, recordings from the 2000s show that the use of *du* has peaked and is now decreasing or stabilizing at a lower level.

This article focuses on intra-individual and intra-conversational variation within the LANCHART corpus with the aim of uncovering the pragmatic effect of using *du* instead of other generic pronouns. All passages in the recordings have been coded according to macro speech act, activity type, type of interaction and genre as well as enunciation. The results of a statistical analysis using mixed models show a number of correlations as to the use of generic *du* (in comparison with *man*), and by and large support the claim that generic *du* is used as a resource for construing involvement, arguably by exploiting the ambiguity of *du* between a generic and a specific second person meaning. These quantitative results make up the point of departure for corroborating qualitative analyses of the discourse framing of the use of generic pronouns.

Keywords: Generic pronouns; Ambiguity; Danish; Socio-pragmatics; Language variation and change; Real-time studies; Discourse context analysis.

1. Introduction

This paper concerns the ambiguity of the second person pronoun in Danish between a second personal interpretation and a generic one. The paper comes from a sociolinguistic tradition, it is both quantitative and qualitative, and since it integrates sociolinguistics and pragmatics it may be seen as a contribution to the comparatively new discipline of socio-pragmatics (e.g. Fuller 1993).

1.1. *On ambiguity and variation*

Sociolinguistics – at least the Labovian variationist current within the mainstream – is concerned with variation. Ambiguity and variation are linked in a special way. Variation at the phonetic level does not entail any ambiguity of interpretation of cognitive content. But variation at any other level of language does indeed entail some ambiguity of interpretation as the variants are most often not semantico-pragmatically equivalent in all respects, and for precisely that reason there has been a great debate on the delimitation of the sociolinguistic variable (e.g. Lavandera 1978; Cheshire 1987). Ambiguity is furthermore unavoidable in any case where two variants are available for use within the same semantic domain but retain separate domains also. In the case under scrutiny here, the common semantic domain is that of generic reference to a person.

From an abstract point of view, Danish speakers at least since the 1970s have at their disposal both the option of using a second person singular pronoun *du* 'you' and an impersonal pronoun *man* (approx. 'one') when referring generically. In some cases *man* may substitute for the first person as well (Jensen 2009a), and of course in many cases *du* is used unambiguously to refer to the deictic second person. In some cases the second person pronoun may be used with reference to the addressee but including others, and in such cases we are at the foot of the stairway to generic reference. In this article, however, the uses with specific reference will be ignored and we shall concentrate on the domain of generic reference to take a look at the frequency, distribution and principled ambiguity of the use of *du* in contrast to *man*.

It has to be stressed at the outset that the facts presented here have already been interpreted once by the student assistants who assigned codes to the various instances. The coders were faced with the task of deciding whether a given instance of a *du* or *man* was in fact used generically or not. While it is in general clear which is which, it will become apparent in the qualitative analysis that in some cases this is no easy task. In accordance with the over-all theme of this issue we focus here on these ambiguous examples but this should not lead anyone to believe that they are in the majority in the corpus.

The paper is a corpus-based socio-pragmatic study. This means that we present data stemming from real speakers, not from constructed examples, and from real speakers' production of speech (not so-called second order data like questionnaires). Also, the paper views these speakers as individuals sociologically constrained by their position in society and by the social roles taken up during the interview situation (see below). We do not, however, exclusively exploit traditional speaker variables well-known from any sociolinguistic study such as Social Class and Gender, although these are known to be relevant for the variation at hand, cf. below section 4. Indeed, we have deliberately downplayed the influence of such language external variables in order to focus on the influence of pragmatic factors operationalized in the so-called Discourse Context Analysis, cf. section 3 below.

In order to make data comparable from various projects using different methodologies as to fieldwork (though mainly variants of sociolinguistic interviews) we have developed a Discourse Context Analysis. It is a commonplace in sociolinguistic research interested in intra-individual variation that use of variables varies according to discourse context. In the early literature, such as Labov (1966, 1972), and the literature springing from that tradition (cf. Schilling-Estes 2002), the contexts delimited were seen

as giving rise to either formal or casual passages in the data. Formality was seen as a consequence of monitored language, which was supposed to be the default mode of a sociolinguistic interview. In contrast, casual passages would be of the utmost interest since they would more appropriately manifest the speaker's vernacular, viz. his or her most systematic everyday use of language. Labov (1966, 1972) singled the following contexts out for special attention by both analytics and field workers:

- Speech outside the formal interview
- Speech with a third person
- Speech not in direct response to questions
- Childhood rhymes and customs
- The danger of death-question (and other questions trying to elicit emotional involvement)

(Labov 1972: 87-94)

In our analysis we have attempted to separate what we see as content cues (viz. 'childhood rhymes and customs' and 'the danger of death') from interactional cues ('Speech not in direct response to questions'), and we have generalized the interest in personal narratives such as those involving the danger of death (which fortunately is not so common in Denmark) to cover all the various everyday genres we could find in our data. As a result, we distinguish between 6 dimensions of coding and inside these dimensions between various (numbers of) categories as detailed in section 3 below, (cf. also Gregersen et al. 2009; Gregersen & Barner-Rasmussen 2011).

The Discourse Context Analysis (henceforth the DCA) was originally developed in order to make a diverse set of data (more) comparable and has subsequently been put to use in the analysis of phonetic variation both between speakers (Gregersen et al. 2009; Gregersen & Barner-Rasmussen 2011) and for the same speaker(s) in different situations (Gregersen et al. forthc.; Gregersen et al. ms.). In this paper we use the DCA to illuminate various linguistic factors that conspire to influence the use of *du* or *man*.

As will be documented below, it turns out that all the dimensions of the DCA are relevant in a statistical analysis of the use of one or the other variant. This is in itself a significant contribution to the analysis of the ambiguity of Danish *du*. In other words we hypothesize that speakers and interlocutors have at their disposal a repertoire attuned to the various dimensions of the DCA and that these dimensions enter into production. What specific form this communicative competence has, is until further analysis a moot question.

1.2. *Pronouns with generic reference in Danish*

In Danish, as well as in many other languages[2], the second person singular pronoun may

---

[2] According to Kitagawa & Lehrer 'impersonal' use of second person pronouns is documented in Chinese, English, French, German, Gulf Arabic, Modern Hebrew, Hindi and Italian (Kitagawa & Lehrer 1990). It is furthermore documented in Spanish (Cameron 1996; Coveney 2003; Kluge 2012) and Romanian (Coveney 2003: 172), Dutch (Berman 2004: 107; Tarenskeen 2010) and Finnish (Fremer 2000; Leino & Östman 2008). According to dialect dictionaries and records, generic *du* is documented in the dialects of Norwegian and Swedish but allegedly only in the northern part of Scandinavia including

not only be used with specific reference to the addressee (as in example 1 below) but also to refer to an undefined person or group of persons in general, i.e. with generic reference (as in example 2 below).

(1)   *vil **du** have lidt   mere kaffe*[3]
      will YOU have little   more coffee
      'would YOU like some more coffee?'

(2)   *hvis man ikke bruger kondomet   rigtigt så   kan **du** få*
      if   one not   use   the-condom correctly then can YOU get
      *børn   af det men **du**   kan også få   aids af det ikke*
      children of it   but YOU   can also get   AIDS of it   not
      'if you don't use the condom correctly YOU can get children but YOU can also get AIDS, right'

The second person pronouns in example 2 above do not refer specifically to the addressee but to a generalized person, as the speaker does not describe specific situations but facts about how the world is structured, according to him. *Du* is therefore also used in generalizations over experiences that do not include the addressee (at least at the descriptive level). This is illustrated by example 3 below as the speaker is actually talking about her own experiences of the difficulties of getting a new supply of stain for her kitchen table.

(3)   *og   så   er der også et problem at   det   trænger til en ny*
      and   then is there also a   problem that   it   needs   to   a   new
      omgang # men det   er laboratoriebejdse # og   det   kan
      round   # but it   is laboratory-stain # and   that   can
      du *simpelthen ikke få   mere*
      YOU simply   not   get   anymore
      'and then there's this problem that it [the kitchen table] needs a new coat of stain # but it's laboratory stain [industrial quality] and YOU simply can't get that anymore'

In contexts where generic *du* is occurring, the speaker is very often changing between *du* and *man,* also when speaking about the same referent. This is the case in example 2 above where the speaker talks about HIV prophylactics, and where the person not using the condom correctly must be the same as the one having the risk of getting children and AIDS. This indicates that *du* and *man* are semantically equivalent (cf. Zobel, this volume, for a discussion of the semantics of German *man* and *du*), at least as far as *descriptive* reference is concerned, but of course it also begs the question of why the speaker varies between the two pronouns.

       The use of second person pronouns for generic reference has been described by a number of linguists (e.g. Bolinger 1979; Kitagawa & Lehrer 1990; Blondeau 2001;

Norway, Northern Sweden (Västerbotten and Norrbotten) and the Swedish speaking areas of Finland (Norsk ordbok: II 203; Ordbok över Finlands svenska folkmål: 505; Lundeby 1996; Westerberg 2004: 49). Generic *du* (*tú/Þú*) is also known in modern Faroese and Icelandic (Føroysk Orðabók:II 1279; Íslensk Orðabók: 1843). Also see the list of languages in Siewierska (2004:212), based on reference grammars and monographic evidence.

    [3] All examples in this article are original excerpts from the LANCHART corpus.

Berman 2004; Hyman 2004; Malamud 2012). The general impression, not only in Denmark but also in other European countries (and francophone Canada), seems to be that the generic use of second person pronouns is new, or that it has at least increased significantly in language use in the last decades. In English, on the other hand, the second person pronoun *you* has been used with generic meaning for centuries. Quantitative work on variation and change in the use of second person pronouns versus other pronouns for generic reference includes studies of French, the first being Laberge's study of indefinite pronouns in Montreal French (Laberge 1976; Laberge & Sankoff 1980; cf. also Thibault 1991) and, more recently, of Danish (Beck Nielsen et al. 2009; Jensen 2009a; Maegaard et al. 2013). These studies find that the use of personal pronouns is indeed an area of grammar which exhibits much variation – probably, as (Coveney 2003: 164) points out, because of the close association of these pronouns with social relationships. Though the generic use of *du* is not completely new in Danish (Jensen 2009a: 90-91), there is no doubt that it has increased considerably since around the 1970s. The generic *du* has gained ground at the expense of the two traditional generic pronouns in Danish *man*, derived from the noun *man(d)* (≈ English *man*), and *en*, derived from the numeral *en* (≈ English *one*).

(4)   **man** *behøver bare at tage bussen forat høre at de unge taler*
      ONE only   needs to take   the-bus to   hear that the young   talk
      utrolig   dårligt
      incredibly   bad
      'ONE only needs to take the bus to hear that the young ones speak incredibly bad'

(5)   *de   bad faktisk* **en** *om   at vaske   fingre   hele tiden*
      they asked actually ONE   to wash   fingers   all   the-time
      'they actually asked ONE to wash one's hands all the time'

Man and *en* come close to a complementary distribution in contemporary Danish: *Man* is used exclusively in contexts where the pronoun functions as the grammatical subject, and *en* almost only where it does not function as subject (Jensen 2009a, 2009b). In the present study, the variable is therefore treated as binary: *du* versus *man/en*.

Earlier analyses of the LANCHART corpus (Jensen 2009a; Maegaard et al. 2013) have shown that the increased use of *du* started in Copenhagen in the 1970s and spread from there to the rest of the country at a pace dependent on the distance to Copenhagen. However, the use of *du* peaked in Copenhagen already around the late 1980s, and in the period from the 1980s to the 2000s it declines in Copenhagen, and later also in the rest of the country. During this period, the informants from outside Copenhagen therefore catch up with the Copenhageners. In the 2000s, the younger informants use less *du* than the older informants regardless of where they live, indicating that the use of *du* is in decline all over Denmark. Based on this and the patterning of other linguistic variables in the LANCHART corpus, Maegaard et al. conclude that Copenhagen is the sole centre of diffusion of nationwide changes in 20th century Danish. The orientation of speakers nationwide is centred on Copenhagen, and the linguistic adaptation taking place follows Copenhagen speech both in expansion and in attrition (Maegaard et al. 2013). A study of a panel of the informants from the LANCHART corpus shows that the change with respect to the use of generic pronouns does not only take place between the generations but also in an intra-individual, lifespan perspective: Around half of the informants change their language use from the 1980s to the 2000s, many of them radically. The

study does not show any difference between the younger and the older informants in this respect, though the older speakers seem to follow an earlier trend than the younger (Jensen & Maegaard 2010; Jensen forthc.).

Also within a synchronic or, more precisely, micro-diachronic perspective variation prevails. A study of speakers recorded in two different conversations within a short period of time shows that these speakers vary just as much with respect to the use of generic *du* between these two recordings as they vary in the lifespan study, where the conversations compared are recorded with a 20 years interval between them. Furthermore, a study of intra-conversational variation shows that the occurrences of *du* are far from equally dispersed over the duration of the conversation. In contrast, the results show that a speaker is far more likely to use *du* (instead of *man* or *en*) if another *du* occurs in the preceding context (within the last minute) (Jensen forthc.).

Though this "birds of a feather"-phenomenon (cf. Scherre & Naro 1991) may be caused partly by a purely psycholinguistically based priming effect, a subtle difference in the meaning of *du* and *man* may also explain the clustering as it makes *du* a more obvious choice in some discourse contexts than others. Qualitative analyses of a small subsection of the LANCHART corpus suggest that generic *du* has an interpersonal potential of construing involvement which separates it from *man*: *Du* is to a larger degree than *man* used as a resource for *enacting* something, that is, to show rather than to tell. Furthermore, *du* is more often than *man* used during assessment actions (Goodwin & Goodwin 1992), and here clusters of *du* most frequently appear at specific moments where speakers put an extra effort in establishing intersubjective common understanding (Beck Nielsen et al. 2009; also see Zobel, this volume). By the speaker's use of *du* the addressee is invited to see the phenomena in question from the inside, so to speak, that is from the viewpoint of the generalized person, while *man* is more or less neutral in this respect. A likely reason for this difference between *man* and *du* is that *du* retains some of its second person meaning also when used generically, not in a truth conditional sense but with respect to viewpoint (cf. Bolinger 1979; Kitagawa & Lehrer 1990; Ushie 1994; Berman 2004). This interpretation is corroborated by the results of the quantitative analyses which show that *du* is favoured in contexts where the addressee is included in the reference of the generic pronoun (Jensen 2009a; Maegaard et al. 2013).

When *du* is used generically two types of ambiguity may thus be involved. Firstly the type of ambiguity described above, which hinges on the retention of a second person meaning when referring generically. Using a concept coined by Langacker, the event or situation which the referent of the generic pronoun is predicated to be part of can be said to be "non-actual": an arbitrary instance with no particular location in time or reality 'conjured up' for purposes of conceiving how the world is supposed to work in general (Langacker 1997: 208). By using a generic pronoun the speaker is instructing the addressee to see the referent from a "structural" point of view: Even though it may in principle be possible to determine the pronouns precise extension, the whole point is that the predication illustrates how the world works in general, not the properties of specific persons or events. However, by using *du* as the generic pronoun, the speaker is at the same time employing a second person perspective.[4]

---

[4] Following Rubba (1996), Bettina Kluge suggests an analysis of the cognitive basis of generic seconds exploiting mental space and blending theory (Fauconnier 1994; Fauconnier & Turner 2002): "they can be understood as an invitation to imagine oneself in the position of the protagonist of a mental

The second type of ambiguity involved when using *du* generically is the ambiguity arising in contexts where it is not obvious to the addressee whether generic or specific second person reference is intended because both types would make perfectly sense (also see Kluge, this volume, on the scarcity of conversational misunderstandings and the advantages of ambiguous 2.sg.). We will analyse examples of such contexts in Section 5.

2. Data material: The LANCHART corpus

As mentioned before, the analyses presented in this article are based on the LANCHART corpus (Gregersen 2009). The LANCHART (LANguage CHAnge in Real Time) project is a real-time study of spoken language in Denmark in the 20th century. LANCHART exploits and expands a number of earlier dialectological and sociolinguistic projects which took place at a handful of different localities in Denmark in the 1970s and 80s (Kristensen 1977; Gregersen et al. 1991; Kristiansen 1991; Nielsen & Nyberg 1992, 1993; Jørgensen & Kristensen 1994). The corpus consists of the transcribed recordings from the earlier studies, as well as new recordings (carried out in 2005–2010) with the same informants who participated in the original studies but with new (young) interviewers, and recordings with new (even younger) informants in the same communities.

Since the analysis in Maegaard et al. 2013 documents that younger informants in general (and we have no reason to suspect that our interviewers are any different from their age mates) use less second person generics than older informants, there would in fact be less priming for that supposed strategy of intimacy in the new recordings. Counterbalancing this effect might be the fact that in the new recordings there was an explicit drive for intimacy which was not always the case in the old projects. It bears mention here that no informants were interviewed by the same interviewer in the two recording sessions.

The corpus is structured by generation: Generation 1 (born 1942–1963), generation 2 (born 1964–1973), and generation 3 (born 1987–1996). In each of these subgroups at each site the informants are distributed evenly with regard to gender and social class (see Gregersen 2009 for a detailed description of the LANCHART corpus).

The present study is based on a subcorpus comprising the transcribed recordings with informants from three of the six LANCHART communities: Copenhagen, the capital and linguistic norm centre of Denmark, Næstved, a regional centre in the southern part of Zealand, and Odder, a small town in the eastern part of Jutland (cf. Figure 1). These sites cover a broad spectrum of Danish communities with respect to population size, community type and to the geography of Denmark. The subcorpus

---

space. By means of a complex conceptual blending process, the interlocutor projects her fictive self in one input space, while a generic protagonist is posited in another input space. Both projections are mapped onto each other in the generic space, which in turn produces the conceptual blend. Subsequently, further actions can be projected into the resulting blend that has now been activated, producing a progression of conceptual blends…Previous approaches had supposed the existence of linguistic shifters that would shift a 2sg token's meaning away from an interpretation as a term of address to a generic one. Within mental space theory, at least some of these shifters can be reinterpreted as space builders, whereby speakers indicate to interlocutors that they are in the process of constructing a new mental space." (Kluge 2012: 305-306)

comprises 272 sociolinguistic interviews in which the 189 different informants produce 2.644.074 words.
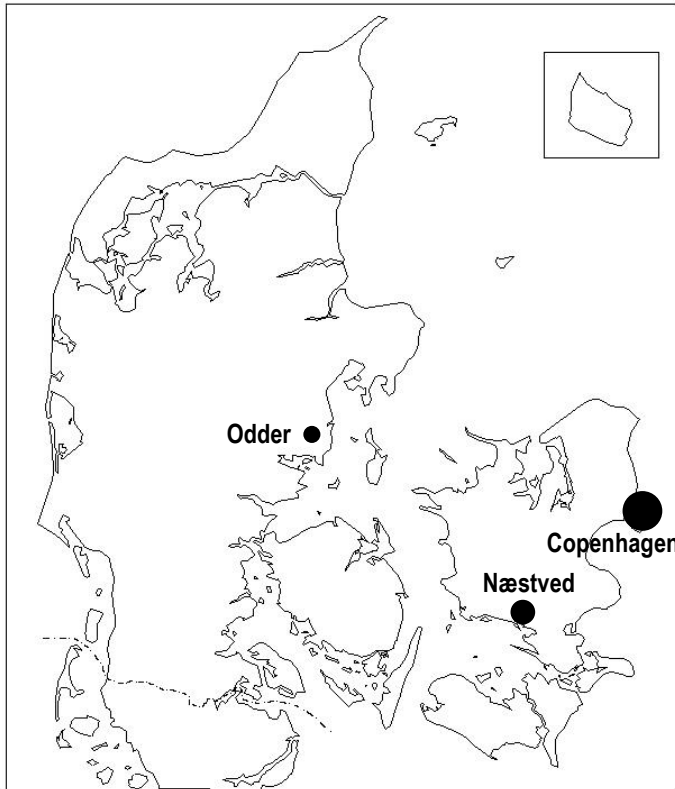
Figure 1: The LANCHART communities placed on a map of Denmark (Copenhagen, Næstved and Odder)

3. Coding

In this section, we will describe the coding procedure followed.

3.1. *Generic pronouns*

Occurrences of pronouns with possible generic reference (i.e. all forms of the lexemes *en*, *du* and *man*) in the corpus were tagged automatically. Afterwards these tokens were reviewed manually to exclude non-generic uses as well as tokens occurring in non-completed constructions (i.e. clauses interrupted before an interpretable intentional meaning has been expressed).

The generically used tokens were coded with respect to a number of syntactic and semantic factors that have been shown to influence the choice between *du* and other generic pronouns (Jensen 2009a; Maegaard et al. 2013). The first linguistic factor (Subject) is the clause subject, i.e. we determine whether the pronoun functions as the grammatical subject of the clause. The second factor (Ref2) is the addressee, i.e. we determine whether the addressee is included in the reference of the pronoun, as

established by a pragmatic analysis of the context. The third factor (Conditional) is the form of the construction, i.e. we determine whether the pronoun occurs in a conditional construction.

3.2. *Discourse context analysis*

Independently of the pronoun coding, the conversations in the corpus were coded with respect to discourse contexts as determined by a number of pragmatic variables. We distinguish two types of code: partial annotations and full annotations. Full annotations demand that *all passages be assigned to one or the other category making up the dimension*. Full annotations include the dimensions "Type of Speech Event", "Activity Type" and "Type of Macro Speech Act". Partial annotations, on the other hand, demand that *only passages which fulfil the criteria for one or another category of the dimensions are annotated*, and include the three dimensions of "Genre", "Interactional Structure" and "Enunciation".

The dimensions and categories coded for are thus (cf. Table 1 and 2 below):

- **Enunciation (Enunc)**: QILIR (a cover term for Quotation, Imitation, Language mention, Illustrative sounds and Reading aloud, i.e. all cases where the utterer is not totally responsible for the use of language manifested, cf. Goffman 1981; Bakhtin 1986; in this coding, these are not mutually distinguished features) versus Other (i.e. cases where the utterer is responsible). In order to avoid possible vicious circles, the DCA is performed solely on the basis of transcripts. Otherwise we would run the risk of using e.g. phonetic cues for QILIR which we would at a later stage want to argue were precisely caused by the QILIR category.
- **Type of Macro Speech Act (MSpAct)**: Exchange of information, Exchange of attitudes, Exchange of emotions, Speech accompanying action and Exchange of fiction.
- **Type of Speech Event (SpEvent)**: single person interviews versus more than one interviewee and group discussions (no interviewer present), cf. e.g. Kerbrat-Orecchioni 1997 on trilogues. The interviews are furthermore distinguished as to whether the interviewer and the interviewee know each other beforehand.
- **Genre**: Narrative, Specific Account, General Account, Soap Box[5], Gossip, Confidences, Reflections, and Jokes.
- **Type of Interactional Structure (IntStruc)**: Absence of asymmetry (only relevant when there is an interviewer and thus not coded for in group conversations), Reversal of interactional roles, Struggle for the floor, Informant(s) taking over, and Monologue.
- **Activity Type (AcType)**: Background interview, Conversation, Conversation with non-participant (equivalent to Labov's category 'Speech with a third person'), Elicited speech, Language attitude study, and Informants' Declaration of Consent form.

---

[5] Soap Box refers to a genre where a person starts to address more general, often political, issues as if addressing a general audience at a Hyde Park corner, hence the name (cf. Labov 2001: 91).

DIMENSION	CATEGORIES						
ENUNCIATION	Utterance(s) not the total responsibility of utterer (QILIR)				Utterance the responsibility of utterer (Other)		
MACRO SPEECH ACT	Exchange of Information (Inf)	Exchange of Emotion (Emo)	Exchange of Attitudes (Att)		Speech Accompanying Action (Act)	Exchange of Fiction (Fiction)	
SPEECH EVENT	Single person interview, interlocutors unknown to each other (SingleUkn)		Single person interview, interlocutors known to each other (SingleKn)		More than one interviewee, interlocutors unknown to each other (GroupUkn)	More than one interviewee, interlocutors known to each other (GroupKn)	
GENRE	Narra-tives (Narr)	General Account (GenAcc)	Specific Account (SpAcc)	Reflec-tion (Refl)	Soap Box (SoapB)	Confidenc es (Conf)	Gossip (Gossip) · Other
INTERACTION	Absence of Asymmetry (I4)		Monologue (I8)	Reversal of Roles (I5)	Struggle for the floor (I6)	Informants taking over (I7)	Other
ACTIVITY TYPE	Conversation (Conv)		Background Interview (SocBackgInt)	Language Attitude Study (LAttStudy)	Consent Form (ConsForm)	Conversation with non-participant (ConvNonP)	

Table 1: The CDA coding system before reduction of categories.

DIMENSION	CATEGORIES						
ENUNCIATION	Utterance(s) not the total responsibility of utterer (QILIR)				Utterance the responsibility of utterer (Other)		
MACRO SPEECH ACT	Exchange of Information (Inf)		Exchange of Emotion (Emo)		Exchange of Attitudes (Att)	Speech Accompanying Action (Act) Exchange of Fiction (Fiction)	
SPEECH EVENT	Single person interview, interlocutors unknown to each other (SingleUkn) Single person interview, interlocutors known to each other (SingleKn)				More than one interviewee, interlocutors unknown to each other (GroupUkn) More than one interviewee, interlocutors known to each other (GroupKn)		
GENRE	Narra-tives (Narr)	General Account (GenAcc)	Specific Account (SpAcc)	Reflec-tion (Refl)	Soap Box (SoapB)	Confidences (Conf) Gossip (Gossip)	Other
INTERACTION	Absence of Asymmetry (I4)		Monologue (I8)		Reversal of Roles (I5) Struggle for the floor (I6) Informants taking over (I7)	Other	
ACTIVITY TYPE	Conversation (Conv)		Background Interview (SocBackgInt) Language Attitude Study (LAttStudy) Consent Form (ConsForm)			Conversation with non-participant (ConvNonP)	

Table 2: The Discourse Context Analysis (DCA) dimensions and their respective categories after reduction. Categories merged in the final quantitative analysis are light grey, whereas categories excluded from the analysis are dark grey (cf. Section 4).

4. Results of the quantitative analysis

Figure 2 shows the distribution of *du* relative to the DCA categories before they were merged.

The bottom row, "Overall", shows that the mean share of *du* in the corpus is 16 %. However, as it is evident from the diagram, there is a very large variation between the different DCA categories. To give an example the share of *du* in the genre "Confidences" (Conf) is 27 % while it is only 11 % in "Specific accounts" (SpAcc).

The raw results may be misleading, though, in three respects. Firstly, some of the categories are very sparsely represented in the data. For example, only 6 tokens of generic pronouns occur in the Activity type "consent form" (ConsForm).

Secondly, looking at each factor one by one might cause us to over- or underestimate the effect of it, as much of the variation may be described better by some of the other factors (including the linguistic and sociolinguistic factors). Related to this is the problem of collinearity, i.e. when two or more of the independent variables are related. This is a problem in several cases with respect to the DCA categories. For example, the Macro speech act "Exchange of emotions" (Emo) only occurs during the Activity type "Conversation" (Conv).

A third, possibly misleading aspect of the simple picturing of proportions is that individual informants do not contribute equally with tokens. The contributions actually vary between 6 and 597 tokens (with a mean of 154) per speaker, partly due to the fact that some speakers are represented in both the old and the new recordings, while others (the generation 3 informants) are represented only in the new recordings. In other words, the number of tokens is unbalanced with respect to individuals, and this may cause us to under- or overestimate the external factors as some individuals might favour a linguistic outcome while others might disfavour it, over and above what the DCA category in itself (as well as gender, age, social class etc. of the informant) would predict (Johnson 2009).

We will therefore assess the impact of the DCA factors described above via multivariate analyses, more specifically generalized linear mixed effects models, including the speaker (Participant) as a random factor/effect. A generalized linear model is a formalised mathematical model of the relationship between the (binary) dependent variable, in this case *du* vs. *man/en*, and the independent variables, such as the factors shown in Figure 2. The term 'mixed effects' refers to the fact that this type of modelling in addition to the so-called 'fixed effects' or 'non-random variables' such as gender and time of recording, also includes 'random effects', such as the individual speaker. The latter effects are characterized by being non-repeatable: They are sampled randomly from populations of speakers (i.e. they are intended to represent a larger population), and if we were to replicate the sampling we would have to choose other speakers. This type of statistical model hence takes into account the non-repeatable effect of the individual speaker by assigning a baseline mean (called an intercept) to each speaker with respect to the dependent variable, as an adjustment for the fact that the behaviour of individuals cannot be expected to be completely determined by the (non-random) social and linguistic factors. A mixed effects model is therefore more conservative than other types of multivariate analyses, as the social and linguistic factors (the fixed effects) are only chosen as significant when they are strong enough to rise above the inter-speaker variation (Baayen 2008; Johnson 2009).
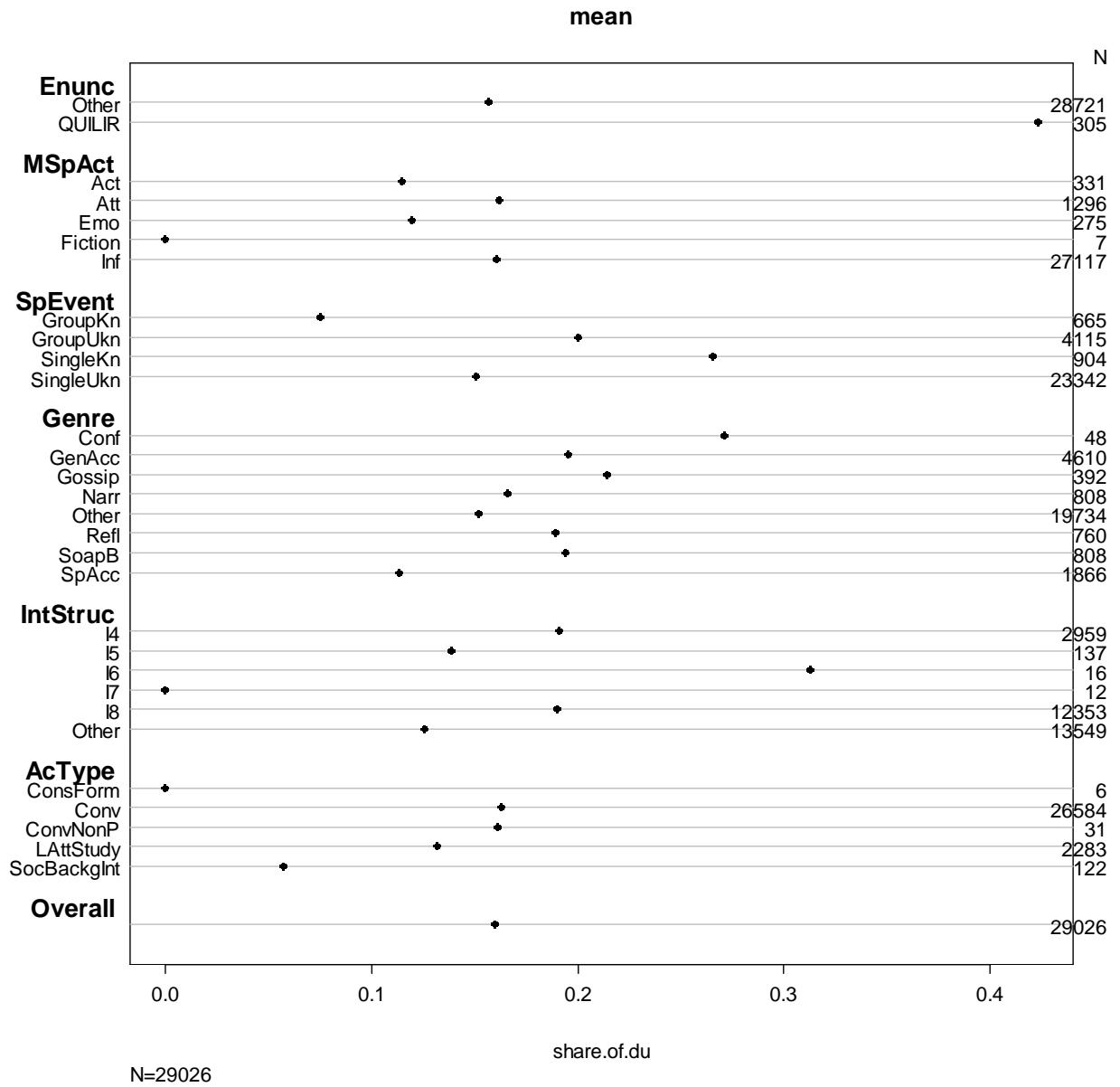
Figure 2: Distribution of *du* on DCA-variables (each dimension, i.e. Enunc etc., adds up to the total 29026)

As was mentioned above, it turned out that some of the DCA categories are very rare and unevenly distributed in the data. This is a problem when we want to perform a multiple regression analysis as it causes collinearity between the factors. The problems were solved either by merging the problematic category with another category within the same dimension if that makes sense, or, if that is not possible, by simply excluding the passages coded as belonging to the category in question from the analysis (cf. Table 3).

| DIMENSION	CATEGORIES						
ENUNCIATION	QILIR				Other		
MACRO SPEECH ACT	Inf		Emo		Att		Excluded: Act & Fiction
SPEECH EVENT	Single (SingleKn + SingleUkn)				Group (GroupKn + GroupUkn)		
GENRE	Narr	GenAcc	SpAcc	Refl	SoapB	Gossip (including Conf)	Other
INTERACTION	Symmetric (I4)		Monologue (I8)		Excluded: I5, I6, I7	Other	
ACTIVITY TYPE	Conv		Other (SocBackgInt + LAttStudy + ConsForm)			Excluded: ConvNonP	

Table 3: Mergers and exclusions of DCA categories in order to avoid collinearity, cf. Table 1 and 2.

The DCA factors were added to the model which according to prior analyses of the LANCHART corpus describes the variation in the data best (Maegaard et al. 2013). This model contains the linguistic factors SUBJECT, REF2 and CONDITIONAL in addition to the extra-linguistic factors GENDER, SOCIALCLASS, BIRTHYEAR, LOCALITY and Time of recording (TOREC) (some of them interacting with each other) as fixed effects and the individual speaker (PARTICIPANT) as a random effect.

Through the analysis and comparison of possible models and the use of model criticism (Baayen 2008) we reached the model which most accurately describes the variation in the data given the factors we included in the study (cf. Appendix).[6] Figure 3 shows the model effects as regards the DCA variables, i.e. the probability of *du* estimated by the model as a function of a (statistically significant) factor when all the other factors are kept constant.[7]

---

[6] Using R's glmer-function (R version 2.15.1, Package lme4 version 0.999999-0).

[7] The plots are made with the plotLMER.fnc-function (Package *languageR* version 1.4). The predicted probabilities presupposes that all the other factors are kept at their default levels which are Subject=No, Ref2=No, Conditional=No, Locality=Copenhagen, ToRec=1980s, Socialclass=MC, Gender=Female, MSpAct=Att, SpEvent=Group, Genre=GenAcc, IntStruc=Monologue, AcType=Conv.
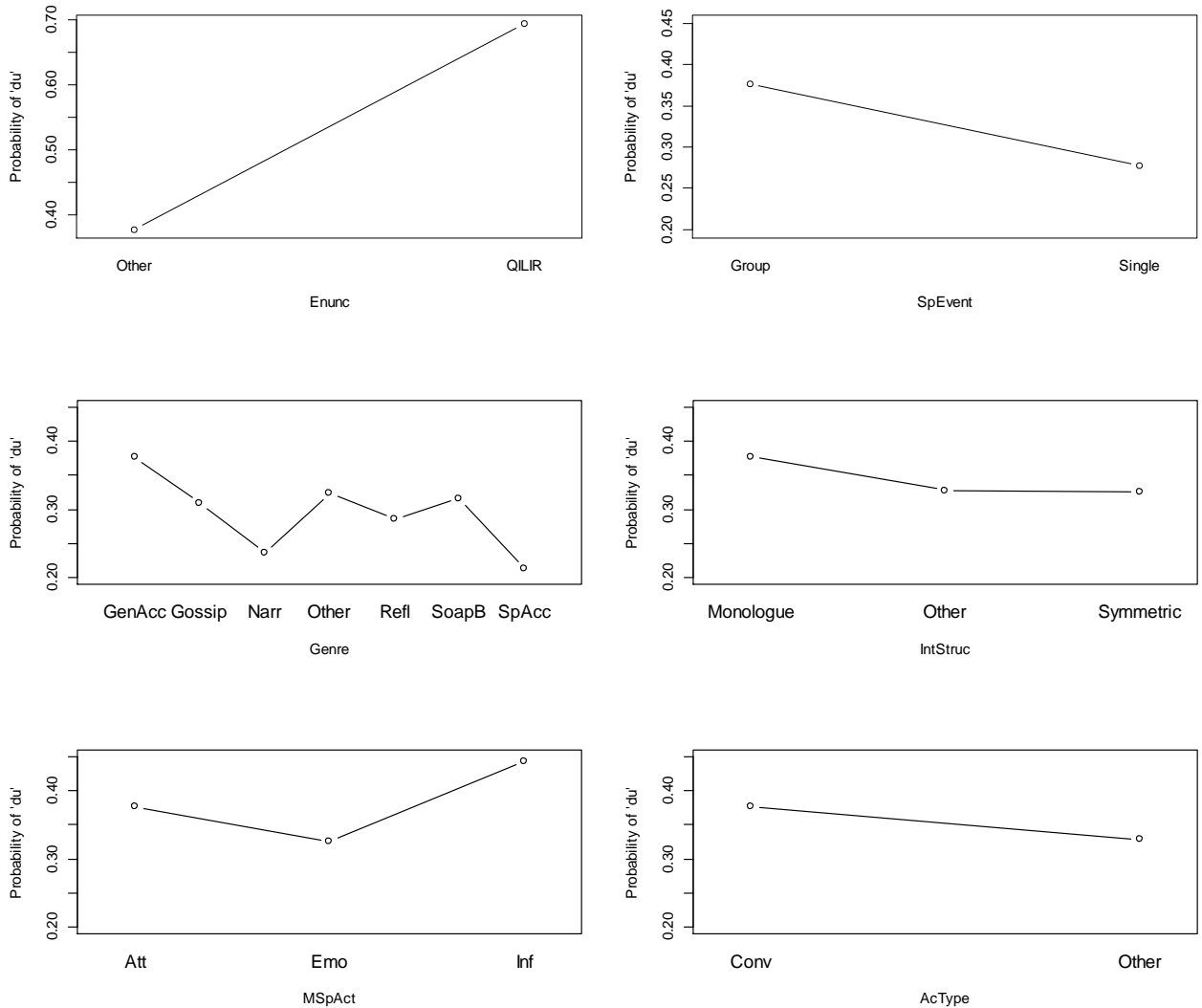
Figure 3: Partial effects of DCA variables according to best model

The analysis shows that all the six DCA variables are statistically significant with respect to explaining the variation in the data. In addition, all the fixed effects and interactions of the best model without the DCA variables are still chosen as significant in the new, more elaborate model. This means that the DCA variables influence the choice of generic pronoun even when taking into account that the use of *du* is changing during the period studied, that it is distributed unevenly with respect to geographical origin, gender, social class and age of the informant, and that the choice of pronoun is influenced by the syntactic function, the type of reference and whether it occurs in a conditional construction or not. In other words, such pragmatic factors as the type of interaction and the genre used by the interlocutors contribute significantly to determining the use of one variant rather than another.

The variable with the greatest effect is Enunciation (ENUNC; please notice that the scale in this particular diagram in Figure 3 is different from the rest). The model shows that passages in which the speaker is "Quoting, imitating, using illustrating sounds or

reading aloud" (abbreviated QILIR in the diagram) have a higher tendency for generic *du* than other passages seen as a whole. We will discuss this effect in Section 5.

With respect to the variable SPEVENT the results show that "Group conversations" (Group) are more disposed for *du* than "Single person interviews" (Single). Though further studies are definitely needed in order to establish this, it may be interpreted in line with Kluges finding that most tokens of generic seconds are found in conversations categorized as communicatively intimate: "Knowing each other well and having knowledge of each other's life circumstances apparently makes interactants confident to use generic seconds, as it helps to prevent misunderstanding." (Kluge 2012: 307; cf. also Kluge this volume). Another interpretation, which may indeed complement the former one, is that in group conversations participants may use other signals to avoid misinterpretations since there are simply more second persons present leaving the *du* option more available.

The results as regards GENRE show that the seven categories can be reduced to three groups of genres with respect to their disposition for generic *du*: "General account" (GenAcc) is the genre which is most disposed for generic *du*, while a cluster including "Narratives" (Narr) and "Specific account" (SpAcc) is the least disposed. The cluster containing the categories "Gossip", "Reflections" (Refl), "Soap box" (SoapB) and passages outside any genre as defined in the DCA (Other) lies in between the two. Inside the clusters, categories are not statistically significantly different from each other. On the face of it this is not the result we expected.

The categories in the Genre dimension may be ordered in descending order of intimacy in various ways, but at the one end would be "Confidences" (most intimate), possibly grouped with "Gossip" (as it is in this analysis) and "Narratives", and at the other end of the cline "Reflections" and in particular "Soap box" (as the least intimate categories). If the *du* was a sure sign of intimization, we would expect this to be the order, but it is not. "Confidences" does indeed seem to be more disposed to *du* than any of the other Genre categories (see Figure 2), but coupled with the much more frequently occurring category "Gossip" it ends up being neutral. "Narratives" and "Specific accounts", the least likely to call forth a *du* share the teller's personal involvement and only differ as to whether the events reported on are presented to the addressee as highly reportable (in which case we have a "Narrative"). They differ from "General accounts" in their reference to specific events. Since "Generalized accounts" are already concerned with generic stuff, our interpretation of this result is that this category needs the *du* to dramatize events - or to make them more relevant for the addressee. "Narratives" are by definition dramatic whereas "Specific accounts" are presented as neutral or even bordering on boring, both of which might explain their resistance to *du*. Furthermore, the low disposition for *du* in narratives and specific accounts may be explained by the fact that they are both at the specific end of the specific-generic continuum with respect to the situations described (cf. section 5). The generic pronouns in these contexts therefore often relate to generalisations over the informants' past experiences or present conditions which do not - as far as the descriptive level is concerned - include the addressee (i.e. in most cases the interviewer). Though it is certainly not impossible to use *du* even though the addressee is not included in the reference at the descriptive level (cf. example 3), the model shows that *du* is dispreferred in such contexts. This interpretation is corroborated by the fact that the most "generic" genre, General account, constituting descriptions of recurring events,

routines, or general processes and courses of events, is the context which is most disposed for *du*.

Also with respect to Interactional structure (INTSTRUC) the number of categories can be reduced, as there is no significant difference between "Absence of asymmetry" (Symmetric) and passages with no distinct interactional structure (Other) seen as a whole. However, the model shows that "Monologue" is more disposed for *du* than passages with other types of interactional structure, seen as a whole. The reason for this may be straightforward. There is no competition for the floor, here and no ambiguity may result from using the *du*, hence it may be preferred whenever a dramatization or intimization is needed.

Similarly, the dimension Macro speech act (MSPACT) displays a binary pattern with respect to *du*. There is no significant difference between the categories "Exchange of attitudes" (Att) and "Exchange of emotions" (Emo) but they are both less disposed for *du* figuring as the generic pronoun than passages coded as "Exchange of information" (Inf). The data set is heavily biased towards the category "Exchange of Information". This category dwarfs the other two. Even so, the explanation for its being more prone to call forth a *du* again may lie in its neutrality. In the other two categories we are concerned with emotional stuff already.

Finally, the results with respect to Activity type (ACTYPE) shows that passages coded as "Conversation" (Conv) are more disposed for *du* than other passages (i.e. "Signing of consent form", "Language attitude study" and "Social background interview") seen as a whole. This is simple. Finally, we have the venerable contrast of formal and casual passages and this fits the bill perfectly, the *du* being more likely in the more casual passages.

In the next section, we have performed a detailed, qualitative analysis of two passages from the same interview. The choice of passages was guided by the results of the quantitative analysis in that we have selected two passages which hold a number of categories constant so that we may hone in on the ambiguity of the second person pronoun.

5. Results of a qualitative analysis

There is an acknowledged difference between quantitative and qualitative approaches to such phenomena as the one we are studying here. Through quantitative analysis of a large number of occurrences we may discover statistical patterns in the data. Such statistical patterns may be seen as probabilities of use. As such, they state that the probability for us as data miners to find a *du* used generically is greater if the person is recorded in the 2000s than in the 1980s, if the utterance is one which raises the possibility of the utterer not saying this with full responsibility herself (QILIR) and if she is performing a General Account (cf. Section 4). Note that we speak of probabilities, and that this does not involve a claim to be able to interpret why at a single point in time a given person uses a *man* or a *du*. It should not be overlooked, however, that the coders who categorized the instances of *man* or *du* in the first place, indeed performed a number of qualitative analyses of whether this was actually a generic use of the pronoun or not, taking the full context into account. The quantitative results thus rest on prior qualitative analyses. On the other hand, the ever relevant (qualitative) question of *Why*

that now? cannot be answered in a quantitative analysis except by pointing to the relevant factors determining differences of probability. Thus, the challenge for a qualitative analysis, any qualitative analysis, is precisely the in-depth scrutiny of specific passages of language use in order to answer the question of *Why that now?*

In the present case where there are only two options to be explained, the qualitative analysis may feed into the quantitative one by generating hypotheses as to the different semantic values of a *du* and a *man*. It does so, on the basis of the analysis of single examples, and the generated hypotheses subsequently have to be tested in the explanation of the probabilistic patterns disclosed in the data. The previous qualitative-quantitative analysis by Beck Nielsen, Fosgerau and Jensen (2009) has already generated one such hypothesis about the use of the *du* instead of the *man*: we suggest that it is used to construe (more) involvement (cf. Section 1.1).

The opposite direction is for the quantitative analysis to inform the qualitative one. We have chosen to present the quantitative results before the qualitative ones since we believe that we may fruitfully use the first to frame the latter. In choosing which examples to analyse qualitatively we have to be certain that we find 'the right ones'. From a bird's eye view any example will do, but in fact we always run the risk of finding something which is idiosyncratic, accidental or the like if we do *not* let the quantitative analysis guide the selection of informants and examples. In this case, we looked for an informant who was using instances of *du* alternating with *man* in quotations, i.e. passages where we could not be sure who was responsible for the utterances. In addition to the Speech Event and Activity Type we wanted to hold both the Macro Speech Act and the Genre constant. We found an interview from Odder which made it possible to analyse a passage in detail where the speaker included quotes or reported speech in her performance of the Genre of General Account. From the same interview we picked a previous passage where we are outside Genres but where we find both the *du* and the *man*. The Speech Event (Single person interview), the Activity Type (Conversation) and the Macro Speech Act (Exchange of Information) are all held constant.

The two examples analysed below are thus taken from the same interview. The interview is both typical of the best interviews in the LANCHART corpus in that the informant 'opens up', and somewhat atypical in the frequency of the *du* variant. But exactly for the latter reason it is ideal for exemplification of the ambiguity of *du* in context.

The interview is a single person interview from Odder, one of the Jutland sites (where the overall use of *du* is lower than in Copenhagen and Næstved). The informant, who is a 34 year old trained caretaker (i.e. belongs to the Working Class), was born and bred in the area by parents from the same Social Class. They were, in their turn, also from the area. In these respects she conforms to the demands of traditional dialectology for a Non-Mobile person as the ideal informant. She is well integrated into her community.

The informant is previously unknown to the interviewer, who is a young female student research assistant from the LANCHART Centre. But the interviewer has at the time of the recording already carried out a few interviews at the site and knows her way around. She has apparently talked to the informant's mother (she refers to this conversation on the tape) and very soon in the interview displays her knowledge of the informant's family relations. She is also able to navigate quite skilfully in local

geography, e.g. as to the schools available in Odder. This all adds up to make the first section of the interview, which, as is normally the case, is about background data, easy, and it probably inspires confidence on the part of the informant.

We have to stress that the use of the third person plural pronoun which in earlier periods of Danish has been used to mark distance and hence politeness (i.e. as the V pronoun) is never used in these sociolinguistic interviews. The Danish speech community has since the 1960s distanced itself considerably from the traditional situation described in the classic paper by Brown and Gilman (1960). There is therefore no possibility of using this marker as a sign of the development of intimacy during the interview. Nevertheless, we submit that there are other indications that the relationship between interviewer and interviewee develops during the interview session so that it rather quickly becomes that of a conversation between 'intimate strangers' (Albris 1991).

The informant was interviewed by two dialectologists in the original Odder study in 1986-88 when she was 15. As with a number of interviews with young informants in this study, the ratio of two adults (who were 53 and 59 years old, respectively) to one teenager makes for a less intimate atmosphere. This may or may not be the reason why at that point in time the informant did not have any instances of *du* at all, but 19 instances of *man*. In the new study, carried out in 2005, there is a marked difference (which is general for the site, cf. section 1.1 above and Maegaard et al. 2013) since she now has 40 instances of *du* versus 121 instances of *man*. The informant has apparently either acquired the possibility of using a *du* for generic reference at some point in time between 1986 and 2005 or did not feel the need for it in the original interview session. Or, a third possibility, is now talking to a person whom she considers to belong in her own age segment (the interviewer is 25 years old) and accordingly is more prone to create intimacy in this particular way. Be that as it may, with a *du* percentage of 25 she places herself in the upper half of the population sampled, cf. section 4 above.

In a single person interview there is an asymmetry between the interviewer's use of the second person pronoun and the informant's. For the interviewer, the *du* 'you' almost invariably refers to the person s/he is addressing, i.e. the informant. The interviewer will most of the time be asking questions about the informant's life and life history, hence the lack of ambiguity on her part. Not so with the informant. She has the option of using the second person pronoun both to refer to the interviewer (which is quite rare unless you would include weakly appealing set phrases such as *du ved* 'you know') or with a more or less general reference. We have only coded the informant's use of generic pronouns.

The interview follows the path of most single person LANCHART style interviews: Following the Activity Type of Background interview (where the probability of a generic *du* is lower than in the Activity Type of Conversation), the informant is prompted to tell her life history with particular emphasis on key turning points. In this case, the informant has been through a divorce and is now living with another man than the one who is the father of her first two children. First, the interviewer introduces the theme of difficulties in a question focussing on the fact that the new husband lived with his parents when they fell in love. This first attempt at breaking default formality squanders on the appearance of the oldest child, the son, which gives rise to what Labov would call Speech to a Non-Participant; we capture this with the Activity Type 'Conversation with non-participant'. The interviewer has to start

over and next focuses on the children only to be interrupted anew by a telephone call. The son answers the call but he has to ask his mother for an answer to the caller. Again, these interruptions are coded as different Activity Types (and excluded from the quantitative analysis, cf. Section 4).

In the next passage the interviewer after a brief discussion of what went wrong in the informant's marriage wants to know what it is like to get a new boy-friend. This is a good example of the most frequent Macro Speech Act, Exchange of Information. This is the Macro Speech Act which according to the statistical analysis is most prone to call forth a *du*:

(6) *1   Interviewer: men er det altså er det er det svært den der overgang mellem altså*
'But is it right is it is it difficult this here transition between right'

2   når **man** *skal præsentere (p) en ny mand (p)*
'when ONE is to present a new man/husband'
3   Informant: for børnene
                                        for the children
'But isn't it difficult the transition when YOU have to present a new man/husband to your children.'

The informant is here prompted to answer in the generic mode by the interviewer's prior use of the generic pronoun. This is a politeness strategy used with the utmost delicacy (cf. Stewart 1995). The interviewer wants to hone in on a subject that may be rather emotionally charged, i.e. presenting a new man and possible husband (in Danish the noun *mand* may mean both 'man' and 'husband') to children who already have another father, and thus exploits the possibility of going generic. It would be difficult for the interviewer to use *du* in this context without being misunderstood as referring (unambiguously) specifically to the informant. Having formulated the question herself, the informant completely grasps this and recognizes the troubling nature of the question by starting with a very audible sigh, in the transcript rendered by a 'hh', thus acknowledging that this is in fact a dilemma:

(7) *1 hh jamen det er jo*[8] *svært at finde ud af om det er hvor hurtigt* **man** *skal gøre det*
'[sigh] yeah it is difficult to find out whether it is how quickly ONE/YOU may do it'

2 altså fordi at det er for **du** *har jo allermest lyst til at gøre det hurtigt fordi* **du** *vil gerne*
'right because that it is because YOU would most like to do it quickly because YOU want to'

3 være sammen med **din** *nye kæreste T*[9] *meget ikke også og men* **du** *T så omvendt*
'be together with YOUR new boyfriend H very much right and but YOU H then on the other hand'

4 så skal **du** *også tænke at det kan heller ikke hjælpe noget at* **du** *bare [uninterpretable]*[10]

---

[8] We have not been able to find a good translation equivalent for the discourse particle 'jo' in Danish which indicates that information presented (in this case by the informant) may be seen as common knowledge (in this case for both interviewer and informant).

[9] T Indicates hesitation; the English equivalent is H.

'YOU have to think that it cannot help that YOU just [uninterpretable]'

*5 og så går der en måned og så er **du** ikke kærester med ham mere og så*
'and then a month goes by and then YOU are not friends with him any longer and then'

6 det kan jo heller ikke hjælpe noget
'that cannot either be helpful right'

'[sigh] It is difficult to say how quickly YOU should do it because what YOU would most like to do is to do it quickly because YOU want to be together with your new boyfriend a lot but YOU on the other hand YOU have to think about that it can't help if YOU just [rush into things] and then when a month has gone by YOU do not see him any longer - that will not be great either'

The informant's first generic pronoun is a *man*, maybe because of the generic pronoun used by the interviewer (which can be interpreted as a case of priming, cf. Neely 1977; Bock 1986). But the next one and in fact the next seven generic pronouns are second person pronouns. Why does the informant use the second person instead of *man*? Our interpretation is in line with the prior analysis in Beck Nielsen et al. (2009) that what she is doing is to build *rapport* (cf. Fosgerau 2007: 262ff with references). The informant recognizes that this is indeed tricky and personal stuff (cf. the initial deep sigh), but by using the second person pronoun she indicates her willingness to answer questions of this kind. The reason the *du* may be used in this way is that it retains some of its second person meaning, in this case: It could be you, i.e. the interviewer.

This passage can be explained as one indication that the informant 'opens up' to the interviewer. That this option of opening up is not present every time someone uses a generic pronoun may be demonstrated by a passage somewhat later when the informant is describing the present state of 'persons who are accepted to live at residential homes for the elderly'. Here a *du* cannot be used: It would not be possible for the informant to include the addressee, a young female, in the group of elderly people. Accordingly, only the *man* is used for generic reference.

The quantitative analysis in section 4 above clearly indicates that the one discourse Context dimension that has by far the largest effect is that of Enunciation. Passages which may be seen as quotations (bear in mind that the Discourse Context Analysis is performed on the basis of transcripts) have a far higher probability of containing a generic *du* than passages which may be ascribed to the utterer only.

A brief note on genericity may be useful here. In theoretical linguistics literature genericity is often conceptualized in terms of quantification as in: *All lions have a mane* or indefinite plural NPs as in: *Lions have manes*. Thus, genericity is seen as being about general statements which hold true for all possible worlds. As such it is in contrast to specificity as either instanced by proper names or by determinate NPs (e.g. Carlson 1982). In everyday conversation we may define a situation or state of events as specific if the situation is pinned down in time and space. Quotations may be an instrument to do this or rather, they may by themselves indicate specificity as in: "I go 'ha I never thought of that' and he goes 'well you'd better start thinking now!'" This would normally be understood as referring to a particular incident when two persons (one of them the

---

[10] Our guess is that the uninterpretable part may be something like *kaster dig ud i det* 'jump right into it'.

utterer, the other a male third person) said precisely these words. But in our data it turns out that genericity is not a question of either totally specific or totally generic but rather a cline leading from one to the other. Whenever we want to typify recurring events or refer to the fact that more than one event of the same type have occurred, we may use a habitual or a plural to indicate a measure of genericity, i.e. that this should be seen as typical or general behaviour. Yet, we may still want to keep the scenic elaboration by using quoted speech – this time however as a typified quotation so to speak. The exact words are a condensation of the meaning taken from the many instances and are thus not to be seen as direct quotes. In the following we shall see why that leads to typical ambiguities which illuminate the choice of pronoun.

The context is as follows: In a previous passage, the theme of young caretakers of the elderly was introduced. Two distinctions are essential here: One is the distinction between staff and trainees: Staff will be there permanently (if they do not themselves choose to change jobs, it is up to them) whereas trainees come and go since they are only there for the in-service training period. For this brief period, they are under the guidance of the permanent staff who act as their supervisors.

The other relevant dimension is that of age. Although the informant is the youngest of the permanent staff, she presents herself as very different from the young trainees whom she views as pupils to be taught the professional stance. She thus comes across as being quite critical of the work attitudes of the trainees. In the passage under analysis, she is explaining how difficult it is to educate the young trainees to be responsible caretakers of the elderly:

(8) *1 Og det er ikke sjovt for de stakkels vejledere jeg under dem det ikke da ha*
　　'And it is not fun for the poor supervisors I do not envy them that right certainly ha'

　　2 hvordan de så prøver at forklare dem at det kan altså ikke hjælpe noget
　　'how they then try to explain them that it can certainly not help anything'

　　*3 hvis **du** vil det her så må **du** så må **du** komme til tiden og **du** må **du** skal*
　　'if YOU want this here then YOU must then YOU must be on time and YOU must YOU have to'

　　*4 når **du** går ud og ind fra en stue så skal **du** gøre så er der nogle ting **du** lige skal tjekke ikke også*
　　'when YOU go out and enter a room then must YOU then there are some things YOU just have to check, right'

　　*5 altså **du** kan heller ikke lade toiletstolen stå med urin i eller mere vel*
　　'right YOU cannot either let the toilet chair stand with urine in it or more you know'

　　*6 det tager **man** altså ud det skal ikke stå derinde og og rydde op altså der det prøver de*
　　'that ONE/YOU do take out it cannot stand there inside and and clean up right there that they try'

　　7 og der er nogen gange så kan kan de jo også når de har fået det at vide mange gange
　　'and there are some times when they can can so right when they have been told a lot of times'

　　8 og andre gange så lykkes det ikke rigtig hvor de nogen gange siger til dem

'and at other times they do not quite succeed where they at times will say to them'

*9 I du burde måske tænke på om det ikke var noget andet **du** skulle prøve først her*
'you you ought perhaps to wonder whether it was something else YOU should try first here'

'And it is not fun for the poor supervisors I sure do not envy them this task ha how they try to explain to them [i.e. the young trainees] that it is not a good idea if YOU (*du*) want this then YOU (*du*) will have to YOU (*du*) will have to arrive on time and YOU (*du*) will have to YOU (*du*) must when YOU (*du*) enter a room and when YOU (*du*) leave it then YOU (*du*) will have to then there are certain things YOU (*du*) have to check right eh YOU (*du*) cannot either leave the toilet chair full of urine or more than that YOU (*man*) take that out it cannot be left in there and and clean up there right that's what they are trying to [teach them] and sometimes they can actually do it when they have been told a lot of times and at other times they do not quite succeed where they will sometimes say to them you (plural) you (singular) ought to think about whether there wasn't something else YOU (*du*) would rather want to try first'

We are here presented with a prime example of referential ambiguity. The passage is probably to be understood as follows: The informant describes a scene where supervisors instruct trainees to take care of the elderly. In actual fact the situations are probably exclusively one-on-one, one supervisor training one trainee. But please note that here the supervisors are in the plural. This is our reason for thinking that the informant has condensed a number of one-on-one situations between trainee and supervisor to 'a generic scene'.

However that may be, the interviewer is invited to imagine that the supervisors speak to the young trainees, i.e. the passage has to be coded as containing messages which are not the sole responsibility of the utterer. We cannot know whether the trainees are addressed as singular addressees in a specific situation (which would mean that the second person pronoun is both quoted and specific) or whether the *du* is actually a generic pronoun and the situation(s) is/are viewed as generic, viz. including what the job as a caretaker will involve. The discourse context coders at the LANCHART Centre coded the passage as a General Account in the Genre dimension and another group of discourse context coders concluded that the whole passage was quoted. Furthermore, the Grammar coders (a totally different group of coders) coded all the instances of *du* as generic. In so far they have disambiguated the ambiguity.

This ambiguity, we argue, is, however, precisely the point. The passage is alive because we are there with the trainees and the supervisors, and we are there any and every time the educational advice is said or meant: It is general advice, valid in any training situation. At the same time it could be seen, again along the line of the previous interpretation by Beck Nielsen, Fosgerau and Jensen (2009), as construing the relationship between the interviewer and the interviewee as one of alignment in critique of the young 'others': 'They are like this and I don't like it; you wouldn't be like that, would you?'

At the end of the passage, the scene changes again. First (in line 6) we get the first instance of the specialized generic pronoun *man*. The utterance in focus here is about how to behave in the presence of a full toilet chair: "ONE/YOU (*man*) take that out it cannot be left in there". Either, the utterance is meant to be a more general rule, applicable in any case where you are faced with a full toilet chair. The use of *man* may in that case be interpreted as a move away from the specific scene with its ambiguity of

whether the informant is actually quoting a specific person who addresses specific trainees to a level of general rules of behaviour given to an imaginary public of trainees. If we follow this interpretation the answer to the question of "Why *man* now?" is simply that the rhetorical move is from the ambiguous specific scene to the unmistakeably general one. This would in fact mean that it would include both the informant and the addressee more clearly and as such mean a return to the immediate conversation: This is how anyone ought to behave, including me and you[11].

As becomes evident from the transcript, the passage we have analysed so far does not close off the rhetorical structure. The informant has an even better punch line in store. This appears in lines 7-9 when the informant underlines that such behaviour (as depicted indirectly in the rebukes from the scenic generalizations) should make trainees think again: Do they really want to work at a home for the elderly? Here the adverbial *nogen gange* 'sometimes' makes the interpretation unambiguously non-specific (line 8). The quotation itself initially hovers almost imperceptibly between the use of the plural 'you' (Dan. *I*) and the singular 'you' (Dan. *du*) but soon opts for the singular. Whether this *du* is the generic pronoun or not, again hinges on the interpretation of the incidence of quotation as repeated instances of advice but there is no doubt that we have reached the punch line: Think twice if you cannot behave!

In this short passage, we have seen why quoted passages may contain indications which invite both a generic interpretation (plural of the supervisors and the trainees) and in principle a specific one (the use of the second person pronoun as referring to the addressees in the imagined situation(s)) and thus that the ambiguity of the *du* is principled: Quotations tend towards specificity whereas the use of *du* outside of quotations may be generic. We suggest that such passages which contain both indications of genericity and (indications of) quotation may be termed *generic scenes* (a special type of mental space) and may make up a new category. This category would capture repeated incidences where someone would utter more or less the same things. They would thus be placed between the specific incidence where a quotation is construed as being the exact words coming out of some one individual's mouth and the generic episode where no direct quotations could be involved.

6. Conclusion

We started out by showing that there is variation in Danish as to how to express generic reference to persons. The two options (considering the two forms *man* and *en* as allomorphs) are neither in principle nor in actual language use distributed evenly. By far the most common generic pronoun in Danish is *man*. This is specialized and unambiguous as a generic pronoun except that in some (very specialized) cases it may be used to refer to the first person singular (or even the second person, but that is extremely rare), and such cases have not been included in the study at all. Not so with the second person pronoun *du*. This is both used with reference to the addressee in a speech event and by extension to an unspecified or loosely specified group including the addressee and by a further extension to a group excluding the addressee. The use of *du* therefore involves two types of ambiguity: a principled one hinging on the retention of a

---

[11] We owe this observation to Don Kulick.

"second person meaning" also when referring generically, and a pragmatic, contextually determined, one caused by the fact that the addressee from case to case needs to determine whether generic or specific reference of *du* is intended. For the genericity to be interpreted as such, there have to be indications in the context. Such markers of genericity are various and we have not entered into any discussion of which indicators have been used when our coders have interpreted the numerous examples coded as generic or not. The point is that in any case, when using *man*, there is a default interpretation of genericity. In functional terms: That is why the word is there.

By the same token, we have to view the occurrence of the generic variant of *du* (which to all appearances is new compared to the generic *man*) as enabling the utterer to do something new. We have argued, along the lines of the previous analysis in Beck Nielsen, Fosgerau and Jensen (2009) that the value of the *du* is that it colours the genericity with a nuance of involvement which probably has to do with the original default reference to the second person. The use of *du* invites the addressee to involve him- or herself in the generic mental space while *man* is more or less neutral in this respect.

The quantitative analysis has attempted to unravel the many factors which alone or combined influence the probability of choosing a *du* or a *man*. If the above interpretation was the only determinant of the choice, the apparent variation in the data would be based on (cognitive) semantics: When some informants use *du* while others use *man* in what seems to be the same context, it is simply because they mean different things. Differences between e.g. men and women or the changes from the old to the new recordings would therefore be a case of semantic variation or change (Hasan 2009). Another, and in our opinion more likely, interpretation is that the difference with respect to the interpersonal potentials of *du* and *man* is not an invariantly coded semantic feature of the pronouns. It is certainly motivated by the strongly and probably invariantly conventionalized second person meaning of the prototypical use of *du*, but there is no compelling reason to think that the semantic-pragmatic meaning of generic *du* is the same to all the speakers in our data, or even to the same speaker at two different points in time. Also the influence of the different DCA factors is difficult to reduce to a simple dichotomy such as the one between involvement and abstract genericity. We cannot exclude the possibility that further analyses might show that the variation explained by the DCA variables is actually based in functional differences between *man* and *du*, but we believe it to be more likely that differences in the social meaning of the two variants are decisive. In the spirit of this general interpretation we offer the results of the mixed effects model as a contribution to the new field of socio-pragmatics: We show that a handful of different discourse factors together do in fact influence the probability of which generic pronoun is used, also when taking more narrow linguistic as well as sociolinguistic factors into account.

Finally, we do not intend to disambiguate all discourse. We have shown that the ambiguity of pronouns may cause listeners to involve themselves fully not only in the interpretation of specific discourse passages but also in picturing for themselves what actually went on in the scenes evoked by the use of a specific instance of *du*. So, what do you mean? or rather: What does *du* mean? That was the question. We have provided some answers.

Appendix

Factor	Level	Estimate	p
Intercept		-0.708591	0.024558
Subject	Yes	-1.620080	<2e-16
Ref2	Yes	0.187509	0.000313
Conditional	Yes	0.471570	5.46e-12
Locality	Næstved	-1.016264	0.005755
	Odder	-1.133398	0.001084
ToRec	2000s	-0.058253	non-sig
Socialclass	WC	0.241947	non-sig
Gender	Male	0.257067	non-sig
Birthyear (centered)		-0.041750	3.44e-09
Subject by Locality	Yes:Næstved	0.036788	non-sig
	Yes:Odder	0.437275	0.019887
Locality by Ref2	Næstved:Yes	0.440526	4.64e-06
	Odder:Yes	0.179642	non-sig
Conditional by ToRec	Yes:2000s	0.184817	0.027890
ToRec by Socialclass	2000s:WC	0.178294	0.033672
ToRec by Gender	2000s:Male	0.287632	0.001144
Locality by ToRec	Næstved:2000s	0.251911	0.025493
	Odder:2000s	0.244021	non-sig
Enunciation	QILIR	1.325709	< 2e-16
MSpAct	Emo	-0.227364	non-sig
	Inf	0.274632	0.002949
SpEvent	Single	-0.453667	1.13e-06
Genre	Gossip	-0.300008	0.029741
	Narr	-0.668039	3.25e-09
	Other	-0.233272	5.37e-06
	Refl	-0.414686	0.000351
	SoapB	-0.266904	0.022306
	SpAcc	-0.800703	< 2e-16
IntStruc	Other	-0.212894	1.38e-06
	Symmetric	-0.225979	0.000393

AcType	Other	-0.211353	0.004602

Number of observations: 28514, Participants: 189
Random effect: Participant (intercept), Variance 2.0693, Std. deviation 1.4385
The goodness of fit of the model is acceptable with a concordance statistic C of 0,80 and a Somers' D_{xy} of 0,60

References

Dictionaries:
Norsk ordbok - Ordbok over det norske folkemålet og det nynorske skriftmålet (1966-) Oslo: Det Norske Samlaget.

Ordbok över Finlands svenska folkmål (1982-) Helsingfors: Forskningscentralen för de inhemska språken.

Føroysk orðabók (1998) Tórshavn: Føroya Fródskaparfelag & Fródskaparsetur Føroya.

Íslensk orðabók (2002) Reykjavik: Edda.

Albris, Jon (1991) Style analysis. In F. Gregersen, and I.L. Pedersen (eds.)*, The Copenhagen study in urban sociolinguistics vol 1*. Copenhagen: C.A. Reitzel, pp. 45-106.

Bakhtin, Mikhail M. (1986) The problem of speech genres. In C. Emerson, and M. Holquist (eds.)*, Speech genres and other late essays*. Austin: University of Texas Press, pp. 60-102.

Beck Nielsen, Søren, Christina Fogtmann Fosgerau, and Torben Juel Jensen (2009) From community to conversation – and back. Exploring the interpersonal potentials of two generic pronouns in Danish. *Acta Linguistica Hafniensia* 41: 116-142.

Berman, Ruth A. (2004) Introduction: Developing discourse stance in different text types and languages. *Journal of pragmatics* 37: 105-124.

Blondeau, Hélène (2001) Real-time changes in the paradigm of personal pronouns in Montreal French. *Journal of Sociolinguistics* 5.4: 453-474.

Bock, J. Kathryn (1986) Syntactic persistence in language production. *Cognitive psychology* 18: 355-387.

Bolinger, Dwight (1979) To catch a metaphor: You as norm. *American Speech* 54: 194-209.

Brown, Roger, and Albert Gilman (1960) The pronouns of power and solidarity. In T.A. Sebeok (ed.), *Style in Language*. Cambridge: MIT Press, pp. 253-276.

Baayen, R. Harald (2008) *Analyzing Linguistic Data*. Cambridge: Cambridge University Press.

Cameron, Richard (1996) A community-based test of a linguistic hypothesis. *Language in Society* 25.1: 61-111.

Carlson, Greg N. (1982) Generic terms and generic sentences. *Journal of philosophical logic* 11: 145-181.

Cheshire, Jenny (1987) Syntactic variation, the linguistic variable and sociolinguistic theory. *Linguistics* 25.2: 257-282.

Coveney, Aidan (2003) 'Anything *you* can do, *tu* can do better': *Tu* and *vous* as substitutes for indefinite *on* in French. *Journal of Sociolinguistics* 7.2: 164-191.

Fauconnier, Gilles (1994) *Mental spaces. Aspects of meaning construction in natural language*. Cambridge: Cambridge University Press.

Fauconnier, Gilles, and Mark Turner (2002) *The way we think. Conceptual blending and the mind's hidden complexities*. New York: Basic Books.

Fosgerau, Christina Fogtmann (2007) *Samtaler med politiet: Interaktionsanalytiske studier af sprogtestning i danske naturalisationssamtaler*. Ph.D. thesis, Department of Nordic Studies and Lingustics, University of Copenhagen.

Fremer, Maria (2000) Ve e du då - generisk du hos ungdomar och vuxna talare. In U.-B. Kotsinas, A.-B. Stenström, and E.-M. Drange (eds.)*, Ungdom, språk og identitet*. København: Nordisk Ministerråd, pp. 133-147.

Fuller, Janet (1993) Hearing between the lines: Style switching in a courtroom setting. *Pragmatics* 3: 29-43.

Goffman, Erving (1981) Footing. In E. Goffman (ed.), *Forms of talk*. Philadelphia: University of Pennsylvania Press, pp. 124-159.

Goodwin, Charles, and Marjorie Harness Goodwin (1992) Assessments and the construction of context. In A. Duranti, and C. Goodwin (eds.)*, Rethinking context: Language as an interactive phenomenon*. Cambridge: Cambridge University Presss, pp. 85–117.

Gregersen, Frans (2009) The data and design of the LANCHART study. *Acta Linguistica Hafniensia* 41: 3-29.

Gregersen, Frans, Jon Albris, and Inge Lise Pedersen (1991) Data and design of the Copenhagen study. In F. Gregersen, and I.L. Pedersen (eds.)*, The Copenhagen study in urban sociolinguistics vol 1*. København: C.A. Reitzels Forlag, pp. 5-43.

Gregersen, Frans, and Michael Barner-Rasmussen (2011) The logic of comparability: On genres and phonetic variation in a project on language change. *Corpus Linguistics and Linguistic Theory* 7: 7-36.

Gregersen, Frans, Søren Beck Nielsen, and Jacob Thøgersen (2009) Steeping into the same river twice: On the discourse context analysis in the LANCHART project. *Acta Linguistica Hafniensia* 41: 30-63.

Gregersen, Frans, Jens Normann Jørgensen, and Janus Spindler Møller (forthcoming) Sideways. 5 methodological studies of the sociolinguistic interview. Unpublished ms. under revision before resubmittance.

Gregersen, Frans, Torben Juel Jensen, and Nicolai Pharao (forthcoming) Comparing speech samples: On the challenge of comparability in panel studies of language change in real time. In Isabelle Buchstaller, and Suzanne Wagner (eds.), *Panel studies of Language Variation and Change*. London: Routledge.

Hasan, Ruqaiya (2009) On semantic variation. In J.J. Webster (ed.), *Meaning in society and in sociolinguistics*. London/Oakville: Equinox, pp. 41-72.

Hyman, Eric (2004) The indefinite *you*. *English Studies* 85.2: 161-176.

Jensen, Torben Juel (2009a) Generic variation? Developments in the use of generic pronouns in late 20th century spoken Danish. *Acta Linguistica Hafniensia* 41: 83-115.

Jensen, Torben Juel (2009b) Standardisering, globalisering og lokalisering - generiske pronominer i Vinderup. In H. Hovmark, I. Stampe Sletten, and A. Gudiksen (eds.)*, I mund og bog*. København: Københavns Universitet, pp. 157-169.

Jensen, Torben Juel (forthcoming) Generic 'du' in time and context. A study of intra-individual variation and change**.** In D. Duncker, and P. Bettina (eds.), *Creativity and Continuity. Perspectives on the Dynamics of Language Conventionalization*.

Jensen, Torben Juel, and Marie Maegaard (2010) Stability and change in adult speech. A real time panel study. Paper read at *Sociolinguistics Symposium 18*, at Southampton.

Johnson, Daniel Ezra (2009) Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3.1: 359–383.

Jørgensen, Jens Normann, and Kjeld Kristensen (1994) *Moderne sjællandsk*. København: C.A. Reitzel.

Kerbrat-Orecchioni, Catherine (1997) A multilevel approach in the study of talk-in-interaction. *Pragmatics* 7.1: 1-20.

Kitagawa, Chisato, and Adrienne Lehrer (1990) Impersonal uses of personal pronouns. *Journal of Pragmatics* 14: 739-759.

Kluge, Bettina (2012) *Referential ambiguity in interaction. Establishing generic reference with second person pronouns in the Romance languages*. Habilitationsschrift, Universität Bielefeld.

Kristensen, Kjeld (1977) Variationen i vestjysk stationsby-mål. *Dialektstudier* 4.1: 29-109.

Kristiansen, Tore (1991) *Sproglige normidealer på Næstvedegnen*. Ph.D. dissertation, Institut for Nordisk Filologi, University of Copenhagen.

Laberge, Suzanne (1976) The changing distribution of indefinite pronouns in discourse. In R.W. Shuy, and A. Shnukal (eds.)*, Language use and the uses of language*. Washington: Georgetown University Press, pp. 76-87.

Laberge, Suzanne, and Gillian Sankoff (1980) Anything you can do. In G. Sankoff (ed.), *The social life of language*. Philadelphia: University of Pennsylvania Press, pp. 271-293.

Labov, William (1966) *The social stratification of English in New York city*. Washington: Center for applied linguistics.

Labov, William (1972) *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.

Labov, William (2001) The anatomy of style-shifting. In P. Eckert, and J.R. Rickford (eds.)*, Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press, pp. 85-108.

Langacker, Ronald (1997) Generics and habituals. In A. Athanasiadou, and R. Dirven (eds.)*, On conditionals again*. Amsterdam: John Benjamins Publishing Company.

Lavandera, Beatriz L. (1978) Where does the sociolinguistic variable stop? *Language in Society* 7.2: 171-182.

Leino, Pentti, and Jan-Ola Östman (2008) Language change, variability, and functional load. Finnish genericity from a constructional point of view. In J. Leino (ed.), *Constructional Reorganization*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 37-54.

Lundeby, Einar (1996) Om *man* og *en* og *du*. In S. Språknämnden (ed.), *Språket lever - festskrift til Margareta Westman*. Stockholm: Svenska Språknämnden, pp. 137-144.

Maegaard, Marie, Torben Juel Jensen, Tore Kristiansen, and Jens Normann Jørgensen (2013) Diffusion of language change: Accommodation to a moving target. *Journal of Sociolinguistics* 17.1: 3-36.

Malamud, Sophia A. (2012) Impersonal indexicals: One, you, man, and du. *The Journal of Comparative Germanic Linguistics* 15.1: 1-48.

Neely, James H. (1977) Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of experimental psychology: General* 106.3: 226-254.

Nielsen, Bent Jul, and Magda Nyberg (1992) Talesprogsvariation i Odder kommune. I. Lokalsprog og rigsmål i sociolingvistisk belysning. *Danske Folkemål* 34: 45-202.

Nielsen, Bent Jul, and Magda Nyberg (1993) Talesprogsvariation i Odder kommune. II. Yngre og ældre rigsmålsformer i sociolingvistisk belysning. *Danske Folkemål* 35: 249-348.

Rubba, Jo (1996) Alternate grounds in the interpretation of deictic expressions. In G. Fauconnier, and E. Sweetser (eds.)*, Spaces, worlds and grammars*. Chicago: Chicago University Press, pp. 227-261.

Scherre, Maria Marta Pereira, and Anthony J. Naro (1991) Marking in discourse: "Birds of a feather". *Language Variation and Change* 3: 23-32.

Schilling-Estes, Natalie (2002) Investigating stylistic variation. In J.K. Chamber, P. Trudgill, and N. Schilling-Estes (eds.)*, The handbook of language variation and change*. Oxford/Malden: Blackwell, pp. 375-401.

Siewierska, Anna (2004) *Person*. Cambridge: Cambridge University Press.

Stewart, Miranda M. (1995) Personally speaking...or not? The strategic value of *on* in face-to-face negotiation. *French Language Studies* 5: 203-223.

Tarenskeen, Sammie (2010) *From you to me (and back). The flexible meaning of the second person pronoun in Dutch*. MA thesis, Department of Linguistics, Radboud University Nijmegen.

Thibault, Pierette (1991) La langue en mouvement: Simplification, régularisation, restructuration. *LINX: Linguistique Institut Nanterre-Paris-X* 25: 79-92.

Ushie, Yukiko (1994) Who are you? And what are you doing? Discourse and pragmatic functions of the impersonal pronoun you in conversational narratives. *Ochanomizu University Studies in Art and Culture* 47: 127-147.

Westerberg, Anna (2004) *Norsjömålet under 150 år*. Uppsala: Kungl. Gustav Adolfs Akademien för svensk folkkultur.