GRAMMATICALIZATION OF *A GENTE* AS A CLUSTER OF CHANGES: EVIDENCE FROM APPARENT AND REAL TIME STUDIES

Ana Maria Stall ZILLES – Universidade do Vale do Rio dos Sinos/Universidade Federal do Rio Grande do Sul

ABSTRACT

This is a study of the grammaticalization of a gente ('the people') in Brazilian Portuguese. This NP is replacing first person pronoun nós. The findings of both apparent and real time studies suggest that two parallel semantic changes are going on, the generic reference change being more advanced than the specific reference change. This is consistent with the concept of grammaticalization as a set of changes. This kind of result is not obtainable by apparent time studies alone, illustrating the importance of convergent use of real and apparent time data. Also the leaders in each change seem to be different, pointing to the importance of investigating the social embedding of these sets of changes.

KEY-WORDS

grammaticalization of a gente; set of inter-related changes; social embedding in grammaticalization.

1. Introduction

Meillet (1912: 131) presented his definition of grammaticalization as "the attribution of a grammatical character to a previously autonomous word in the beginning of the XX century. Nevertheless, grammaticalization as an area of investigation only started to develop in the 1970s and 1980s. Grammaticalization studies were predominantly diachronic at first, but soon expanded to include synchronic and typological investigations" (Diewald & Wischer, 2002: ix).

Until recently, investigators were mainly concerned with conceptual and methodological issues or with the description of grammati-

calization phenomena in single linguistic items, but now there is an increasing interest in understanding the role of linguistic contexts in grammaticalization processes (Wischer & Diewald, 2002). Furthermore, some studies (Serrano, 1996; Romaine & Lange, 1998; Torres Cacoullos, 2001; Blondeau, 2001; Lopes, 2001 e 2003; Zilles, 2002; Omena, 2003) have investigated grammaticalization in the light of sociolinguistic theory about language variation and change, proposing to study it as a change (or set of changes) in progress and trying to capture its linguistic and social embeddings.

This paper is organized as follows. First grammaticalization is characterized as a process involving several interrelated changes which can be analyzed as several variable rules, some simultaneous, some sequenced. Then, focusing on Brazilian Portuguese, a cluster of ongoing changes is briefly described, including the grammaticalization of the new personal pronoun *a gente* 'we', competing with historical *nós*. Since this phenomenon has been extensively studied, its distribution across Brazil is also summarized, to give an overview of its development. Next the methodology of the present investigation is described, characterizing samples, data collection, variables and analytical procedures. The section on the results deals first with three different analyses: apparent time study, real time panel study and real time trend study, in which all tokens of *a gente* in subject position are compared with all tokens of *nós*. Then a separate analysis of referential and generic usage of *a gente* is proposed, in order to discuss grammaticalization as an interrelated set of changes. In the conclusion, results and some theoretical and methodological questions are critically evaluated.

1.1. Grammaticalization as a set of changes

According to Heine (2003: 579), the set of changes that constitute grammaticalization can be minimally described as including four interrelated mechanisms,

i. desemanticization (or 'bleaching', semantic reduction): loss in meaning content;

ii. extension (or context generalization): use in new contexts; (pragmatics)

iii. decategorialization: loss in morphosyntactic properties characteristics of the source forms, including the loss of independent word status (cliticization, affixation);

iv. erosion (or 'phonetic reduction'), that is, loss in phonetic substance.

The idea that grammaticalization involves a set of interrelated changes is also present in its conceptualization as a cline, meaning that "forms do not shift abruptly from one category to another, but go through a series of gradual transitions (…) that tend to be similar in type across languages" (Hopper and Traugott, 1993:6). Thus the prototypical cline would be the progression from a content word, to a grammatical word, to a clitic, to an inflectional affix, to zero or loss, conceived as a pathway along which forms evolve over time, or as a **continuum**.

An important aspect of grammaticalization is that it is usually conceived as being linguistically motivated and highly embedded in the linguistic system. As Diewald (2002: 117) says:

> "the decisive factors for the triggering and continuation of a grammaticalization process are not to be found exclusively in the grammaticalizing items themselves, but also in changes in related linguistic categories and subsystems".

In this respect, one could think of **clusters of changes** that may – but do not have to – take place in a language, maybe even with one change triggering or contributing to other(s).

In this paper I argue that these mechanisms of change can be analyzed as different variable rules. Some may trigger others, but the process can also remain stable for long periods of time. This has in-

teresting implications: if grammaticalization involves several interrelated changes, it may be the case that each has its own social embedding, its own leading group, its own pace, etc.

A controversial point about grammaticalization is whether it is unidirectional or not. Looking at unidirectionality from a sociolinguistic perspective may help to clarify the issue. For example, from the social embedding of a process we might be able to show that a linguistic item was not going constantly in the predicted direction because of the interplay of social forces associated with prestige or stigma, allegiance, contact, etc.

Another controversial point about grammaticalization is whether the changes in the set are synchronized or not (Croft, 1990: 242). The idea that each change in a grammaticalization process progresses gradually, involving several stages, makes sense, but the further conception that they are synchronized seems to me to be sociolinguistically unacceptable. Based on what we know of different variable rules in action in the same community (for instance, Bortoni-Ricardo, 1985, and Guy, 1981), it is possible to think that each change may be led by different speakers, from different generations or social groups, or from different social histories. For example, upward social mobility may be especially important, in terms of people adhering to linguistic forms that are perceived as having prestige in order to gain cultural capital. Thus it may be crucial that we test the synchronization of interrelated changes by looking at them in the speech of the same speakers in a community, in order to see whether the general tendencies are comparable for all the changes and whether or not the same speakers are leading every change involved.

1.2 Cluster of changes in Brazilian Portuguese

As Brazilian Portuguese takes its own course away from European Portuguese, several morphosyntactic changes are in progress, affecting both the paradigm of personal pronouns as well as subject-verb agreement. These changes include, among others, the introduc-

tion of new pronouns and a related overall reduction in the use of verbal agreement or a variable reduction in the inflected forms still in use (Faraco, 1996; Vitral, 1996; Omena, 1996a, 1996b; Menon, 1996; Ramos, 1997; Kato, 1999; Duarte, 2000; Zilles, Maya & Silva, 2000; Lopes, 2001, among many others).

The first step in this process may have been -the loss of 3rd person singular inflection in Vulgar Latin. Later a new second person singular pronoun developed in Portuguese, which also affected the verb system. This pronoun arose from the grammaticalization of the address form *Vossa Mercê* (meaning 'Your Mercy', 'Your Grace'), whose first record dates from 1331 (Faraco, 1996: 58). It was coined in the Middle Ages and was used exclusively for the king, but in the XIV century it was already being used among nobles and in the XV century among the bourgeoisie, as can be seen in the plays of Gil Vicente. *Vossa Mercê* underwent several stages of phonological reduction, attested in writing: *vossa mercê > vossamecê > vosmicê > você*. As a second person pronoun, *você* ends up being used either alongside the original pronoun *tu*, or in many Brazilian dialects, replaces *tu* entirely. The pronoun *você* also developed a corresponding plural form *vocês*, imposing another change in the system: the distinction between singular and plural is not provided anymore by different lexical items (*tu* - *vós*), but by morphology (-s ending). In both singular and plural, the new pronouns trigger third person verbal inflections, thus leading to what Marilina dos Santos Luz (Faraco, 1996:54-5) appropriately labeled as "the revolution of the third person" (already unmarked, since the corresponding Latin inflection had already been lost).

This "revolution of the third person" is extended still further with the gramaticalization of *a gente* as first person plural pronoun. Also deriving from a nominal expression, which took 3rd person agreement on the verb, the new pronoun also has the effect of reducing verbal morphology.

1.3 The new personal pronoun *a gente*

There are several pieces of evidence, therefore, that Brazilian Portuguese is undergoing grammatical reorganization. One of these processes is the pronominal usage of *a gente*, originally a full NP meaning 'the people', which is increasingly being used as a first person plural pronoun, and sometimes as first person singular pronoun as well. In examples (1) and (2), from the Varsul corpus, the new form alternates with the old pronoun *nós*. It is interesting to note that in these examples the speakers starts using one of the forms and then switch to the other[1].

---

(1)
E: E condução, como é que é?
F: Condução, bom, **nós**$_i$ não temos condução própria. (hes) Então **a gente**$_i$ depende do ônibus. (POA02, l.250-52)

I: And transportation, what's it like?
S: Transportation, well, **we** don't have our own [car]. (hesitation) So **we** depend on the bus.

---

(2)
então **a gente** às vezes quando estava chovendo, **a gente** se falava de longe, **nós** brincávamos muito também naquele terreno baldio (POA30, l.09-11)

---

so **we** sometimes when it was raining, **we** used to talk from afar, **we** would often play also in that empty lot

---

It is not surprising that a word like 'people' was the source for

---

1 Although both directions of switching (*nós* to *a gente* and *a gente* to *nós*) have been found, I have not investigated their frequencies, which might be interesting to understand how the change progresses.

this change. According to Castilho (1997:37) and Heine and Kuteva (2002: 232-33), languages tend to have generic nouns like *man*, *people* and *person* as sources for indefinite pronouns. Well known examples are the indefinite *Mann*, in German; the formerly 3rd person indefinite *on* in French (derived from the noun *homme*, 'man'), now also used as 1st person plural; *homem/ome*, meaning 'man', in Old Portuguese; and European Portuguese *pessoa*, meaning 'someone' ("A pessoa não deve preocupar-se").

The shift affecting *a gente* probably began in the XVI century with the decline in the use of *homem/ome* ('man') and the rise in the use of *a gente* as an indeterminate expression with generic meaning. According to Lopes (2001:140-41), the grammaticalization of *a gente* has been slow and gradual, involving intermediate 'stages' in which the noun *gente* lost the syntactic feature [+plural] and crystallized as a singular NP (definite article + noun) with collective and thus generic semantic interpretation. These 'stages' could also be understood as leading what Heine (2003: 579) labeled as extension (to non-referential, generic contexts) and decategorialization (loss of morphosyntactic properties).

Another important aspect of this grammaticalization process has to do with a change in gender agreement. Several authors (Schmitz, 1973:640; Menon, 1996:626; and Lopes, 2001:137) have shown that while as a noun, *gente* has feminine gender and requires feminine modifiers, as a personal pronoun agreement is with the referent. Consider first the examples of the noun *gente* in two NP: *muita gente* and *gente bonita*. They mean respectively 'a lot of people' and 'beautiful people'. The quantifier *muita* and the adjective *bonita* are marked feminine forms in Portuguese, as opposed to the unmarked masculine forms *muito* or *bonito*. Nevertheless, as a pronoun, the form *a gente* can appear both with masculine and feminine adjectives or nouns, the selection depending on the gender of the speaker and accompanying referent(s), like the pronouns for first and second person

both singular and plural[2]. In other words, this is to say that *a gente* has acquired at least some of the morphosyntactic properties of BP personal pronouns. Consider the examples in (3), meaning 'X is/are tired', with gender (feminine ending –a) and number (plural ending – s) agreement in the adjective 'tired'.

(3)

Singular	Masculine referent	*Feminine referent*
1st p.s	Eu estou cansado (unmarked)	Eu estou cansada (marked with –a)
2nd p.s.	Tu estás cansado	Tu estás cansada
	Você está cansado	Você está cansada
Plural	*Only masculine or both masculine and feminine referents*	*Only feminine referents*
1st p.p.	Nós estamos cansados	Nós estamos cansadas
	A gente está cansado (no plural agr.)	A gente está cansada (no plural agr.)
2nd p.p.	Vocês estão cansados	Vocês estão cansadas

1.4 Semantic change: from generic to referential [+specific]

Most important for our purposes in this paper is the change in the meaning of *a gente*. Stemming from the generic meaning it had developed in the XVI century (Lopes, 2003), by the second half of the XIX century, - *a gente* was being used to make reference to the **speaker**, which is a crucial step to become a personal pronoun. In example (4), *a gente* is used to make reference to the *speaker and a*

---

2 Note however that there is no plural ending in the predicates constituted by adjectives or nouns when the subject is *a gente*, although the meaning of this pronoun is still collective. This is why the usage of *a gente* for 1st person singular leads to explanations in terms of mitigation: it is not a mighty 'I', but and 'I' based on others. This is the pronoun of choice for soccer players, for example, who care to enhance that their performance is the product of a collective work.

group of other specific referents. This example is from a short story published in 1893, by Arthur Azevedo, a well-known Brazilian playwright and journalist of the time. The story depicts a familiar scene where father, mother and two children are talking. One of the children had asked the meaning of the word *plebiscite* to the father, and everybody was waiting for his answer (which in fact he did not know). Impatient with his delay, the mother compels him to answer:

(4)

Vamos: se sabe, diga o que é plebiscito! Então? **A gente** está esperando! Diga! (Azevedo 1893, reproduced in Rebelo 1967, p.179)

-Come on: if you know, tell us the meaning of plebiscite! So? **We** are waiting! Say it!

The change from generic meaning to specific, to personal meaning leads to the fact that today *a gente* may be used to mean the following:

a) **generic reference**: the speaker and everybody else (be it understood as anyone or everyone), as in example (5).

(5)

a gente não vê mais isso porque ninguém pode brincar na rua, né?

one doesn't see this anymore because nobody can play on the streets, right?

b) **exclusive reference**: the speaker and a clearly defined group of persons (e.g. the family), excluding the interlocutor, as in example (4) above;

c) **inclusive reference**: the speaker (interviewee) and the interlocutor (interviewer) or audience, as in example (6), from Borges (2004: 95);

(6)

E: esse lugar aí que tu disse, eu não sou de sair muito.

I: ué, mas se tu quiser **a gente** pode, pode sair.

R: This place there you mentioned, I don't go out very often

S: well, but if you want to **we** can , can go out

> d) reference to the **speaker** alone: first person singular, as in the example '*a gente* está zangad**o'** ('I am angry') where *a gente* is used by a male speaker to refer to himself, with masculine singular marking of the predicate adjective (Schmitz, 1973: 640), instead of the feminine form that would be required by the noun *gente*.

The fact that *a gente* may have these different interpretations is well established in the literature (for example, Lopes, 1998), although in general there has been no separate analysis according to reference[3]. Nevertheless, under the theoretical perspective that characterizes grammaticalization as a set of changes, it is essential to perform separate analyses, once contrasting generic *a gente* with generic *nós*, then contrasting referential *a gente* with referential *nós*. Both are choices made by the speakers and may be conceived as two separate variable rules: a) one dealing with the alternate forms of expressing subject indetermination (*se, a gente, nós*, null subject with 3rd person plural ending on the verb, null subject with 3rd person singular ending); b) the other dealing with the alternate forms of referring to first person plural with specific reference (either inclusive or exclusive). This separate analysis is particularly relevant to the question of whether there are social differences in the usage of *a gente* with one or the other interpretation. The existence of such differences would reinforce the idea that each change in the set that constitute

---

3 Borges's doctoral dissertation is an exception, since he presents a separate analysis of the referential (specific) tokens of *a gente* (Borges, 2004:160).

grammaticalization is (or may be) implemented by different speakers. The present paper investigates this possibility, which was suggested in Zilles (2002) and Zilles (2005).

In order to understand the grammaticalization of *a gente*, it is also important to characterize its implementation in the country as a whole, even if the results in the literature do not distinguish the generic from the referential use. In the next section we present an overview of the phenomenon, and discuss some issues regarding the social characteristics of the samples investigated.

1.5 National distribution of *a gente*

The change from *nós* to *a gente* can be identified throughout Brazil. In Figure 1, all the data are from the 1970s, except for the two columns on the right, whose data are from the 1990s. It is important to note that they all represent the speech of the cultural elite in the country, as a result of the samples analyzed by the respective authors: a) Leite & Callou (2002:54), considered data from Porto Alegre (South), São Paulo (Southeast), Rio de Janeiro (Southeast), Salvador (Northeast), and Recife (Northeast), collected by the project called NURC[4] in the 1970s; b) Duarte (1997) considered data from Rio de Janeiro (Southeast), collected by NURC in the 1990s.

---

4 NURC (Cunha, 1985:23) was a research project developed in the 1970s to study the urban patterns of speech considering only informants with a university degree. Its complete denomination was 'Projeto de Estudo Conjunto e Coordenado da Norma Lingüística Oral Culta de Cinco das Principais Capitais Brasileiras", linked to the "Proyecto de Estudio Coordinado de la Norma Linguistica Oral Culta de las Principales Ciudades de Iberoamérica y de la Península Ibérica".
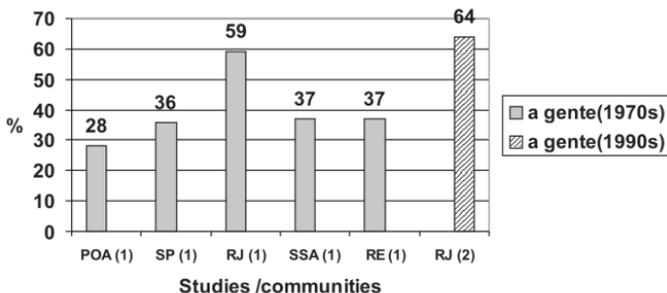
Figure 1 – Percentages of *a gente* as used by the cultural elite in the 1970s and the 1990s

Legend:

(1): POA: Porto Alegre (South); SP: São Paulo (Southeast); RJ: Rio de Janeiro (Southeast); SSA: Salvador (Northeast); and RE: Recife (Northeast); data from NURC, collected in the 1970s (Leite & Callou, 2002)

(2): RJ: Rio de Janeiro (Southeast); data from NURC, collected in the 1990s (Duarte,1997)

Taking a closer look at the results in Figure 1, let us start from the left. The first pair of columns refers to Porto Alegre, the southernmost state capital. Note that in this report by Leite and Callou (2002), Porto Alegre shows the lowest figure for *a gente* – only 28% (a result the present analysis will contradict, probably because of sampling differences). Next there is São Paulo, the largest city in Brazil, the most complex community in sociolinguistic terms, with a lot of ethnic and social diversity. In that sample, the rate of *a gente* was 36%. Next comes Rio de Janeiro, supposedly the leaders in this change, according to the graph by Leite and Callou, and the only city in their study which already had a majority of *a gente* use in the 1970's. Going to the NE, Salvador and Recife display identical rates

of *a gente*, very similar to São Paulo.

These results show that this change was quite advanced in the 1970s, because what we see in Figure 1 is that the cultural elite[5] was already using the new form very frequently. So Schmitz's observation (Schmitz 1973), that this form was not confined to lower social classes, and could also be found in novels and short stories as well as in the speech of people from all social classes is made clear in Figure 1. And if we think that the elite could be still avoiding the new form to a certain point, we can guess that it was even more frequent in the speech of the other social classes.

The two columns to the right of Figure 1 show a somewhat higher level of *a gente* in the 1990s, in Rio. Since the speakers in the 1990s were also interviewed in the 1970s (the "recontato" study), this difference, if it does not reflect sampling differences, could mean speakers have continued to increase their usage of *a gente* in that community. We will discuss a related issue in our data.

The further progress of the change is suggested by several other studies with data from the 1980s and 1990s, showing still higher rates of use. This is what Figure 2 shows, with data from three different regions in the country: the Southeast (data from Rio de Janeiro, by Omena and Braga, 1996); the Northeast (data from João Pessoa, by Fernandes, 1999), and the South (data from Florianópolis, by Seara, 2000; and from Jaguarão and Pelotas, by Borges, 2004). In these studies, samples are not restricted to the cultural elite of the country, so that these results are more representative of the speech communities investigated.

---

5 Of course, the fact that *a gente* is being so much used by the cultural elite of the country leads us to say that this new form has prestige (at least covert, unconscious prestige) in speech, but not in writing.
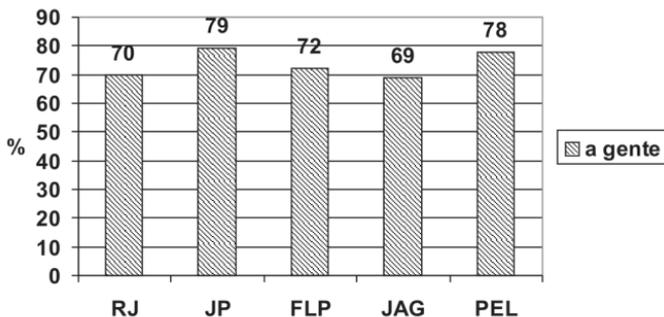
Figure 2 – Percentages of *a gente* as used in stratified samples (data from the 80s, 90s and 2000)
Legend: RJ: Rio de Janeiro, Southeast (Omena and Braga, 1996)
JP: João Pessoa, Northeast; VALPB corpus (Fernandes, 1999)
FLP: Florianópolis, South; VARSUL corpus (Seara, 2000)
JAG: Jaguarão, South; BDS-Pampa corpus (Borges, 2004)
PEL: Pelotas, South; VarX corpus (Borges, 2004)

In sociolinguistic terms, these samples have been methodologically improved and are stratified for level of formal education or social class, age groups and gender. What strikes us most in these results are the similarities among the communities, showing that this is a highly advanced change (or set of changes) all over the country. Omena (2003) conducted a real time study of *nós* and *a gente* in Rio de Janeiro, using both the panel and trend approaches. Since her procedures in devising the analysis are different from the ones adopted in the present study, it will not be possible to compare the results in detail. Some of her findings, however, are briefly discussed here.

In the trend study, Omena (2003:66) performs separate analyses of the data in the two samples (one collected in the 1980s, and the

other in 2000) and finds the same proportion of *a gente*: 78% and 79%. The author interprets this as evidence of stability in the community.

Consistent with this interpretation, the author believes her results by age groups are evidence of age-grading, with younger speakers using more of the newer pronoun whereas the older speakers (or as the speakers grow older) use more the old pronoun. In order to discuss this interpretation, in Table 1 we reproduce her results in the trend study of *a gente* for each sample as analyzed separately.

Table 1 – Distribution (Ns and %) of *a gente* by age groups in two samples, data from Rio de Janeiro (Omena, 2003: 66)

	1980 Sample N/Total & Percentages	2000 Sample N/Total & Percentages
7-14 years old	103/116 = 89%	99/105 = 94%
15-25 years old	473/543 = 87%	211/227 = 93%
26-49 years old	271/369 = 73%	208/251 = 83%
50 years old or above	154/267 = 58%	250/385 = 65%
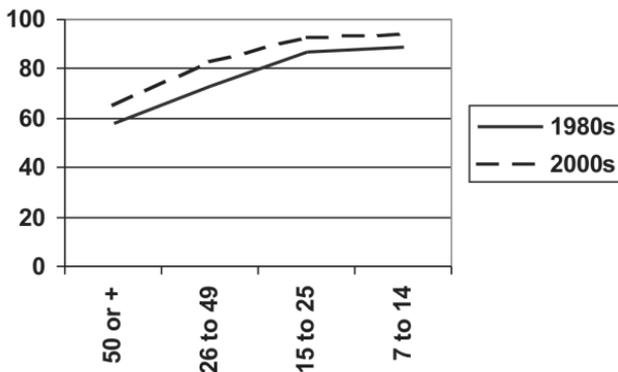
Figure 3 – Percentages of *a gente* by age in Rio (Omena, 2003)

Both Table 1 and Figure 3 show that systematically, in all age groups, there is more use of *a gente* in the 2000 sample than in the 1980s sample (increases ranging from 5% (children), to 6% (youth), to 10% (adults) to 7% (senior adults) respectively). Considering the historical and demographical details presented above, it seems quite plausible to interpret these results as a *nearly completed change* (Labov, 1994: 83) for this community, since younger speakers in 2000 show 94% and 93% of the new pronoun (some of them never using *nós* at all!). In other words, having reached such a high level of usage, the change may be stabilizing now. The similar percentages of usage of *a gente* that Omena reports for the two samples appear in this break-down to be an artefact of different distributions across the age groups: the earlier sample draws over half of its data from the younger two age groups, which use *a gente* most, while the more recent (2000) sample has two thirds of its tokens from the two older age groups, who use *a gente* less. This distributional difference has the effect of obscuring in the pooled totals for each sample the general trend towards more use of *a gente*.

Comparing older and younger speakers in both samples, it seems reasonable to consider that this is a generational change[6] **(each new generation pushes the change further)**, since both lines in the graph increase as age decreases. Unfortunately Omena only presents the results of separate Varbrul runs, one for each decade, so we do not know whether age group is a significant factor group or not if the data had been analyzed together.

Indirectly supporting the interpretation of a generational change, there seems to be individual stability in the samples. Comparing each age group in the 1980s with the "next" age group in 2000, there is always a small decrease (never an increase) in the use of *a gente*.

Although the above alternative explanations offered for Omena's re-

---

6 See section 2 below for an overview of this model of changes.

sults in the trend study are quite plausible, there is another point to be considered. According to the theoretical perspective adopted in this study, there are two changes going on at the same time, so there could be both stability and change at the same time. Omena's results for reference suggest that this may indeed be the case. In the trend study the author shows that, with determined reference, the percentage of *a gente* was only 67% in the 1980s but rose to 80% in 2000, an increase of 13 percentage points! This is even more remarkable when we note that with indeterminate reference the use of *a gente* decreases from 85% in the 1980s to 79% in 2000 (Omena, 2003:68). Besides this contrast, there is an interesting inversion in factor weights: in the 1980s, it was indeterminate reference that favored *a gente* (.53 factor weight against .44 for *nós*); in 2000, it is determined reference that favors *a gente* (factor weight [7] of .61 against .43 for *nós*). These results support the hypothesis discussed in this paper. The change which was in fact progressing in the final decades of the XX Century is the *extension* that converts *a gente* into a full personal pronoun, with specific (determined) reference. As for the older change, it is reasonable to think that having already reached very high levels of usage, now it can either become stable (and never reach completion), or it may be retreating, with the parallel advancement in the use of another pronoun for the generic meaning (perhaps *você* or *tu*[8], similar to what Laberge and Sankoff (1980) describe in Canadian French where *on* has become a full personal pronoun and *tu* usage as a pronoun with generic reference has increased).

2. Methodology

The use of apparent time data has proven to be a powerful technique in the study of language change. This approach uses the age distribution of a variable in a synchronic sample as a window on the

---

[7] Note, however, that this factor group was not selected as significant by Varbrul.
[8] This possibility is under investigation by Paulo Ott Tavares and Ana M. S. Zilles, considering data from Porto Alegre and adopting real-time panel and trend approaches.

past, assuming that older speakers provide evidence about how language was used at an earlier time. But despite its utility, it requires confirmation, wherever possible, by real-time evidence, drawn from different points in time.

In order to interpret real-time data, Labov (1994:83) proposes a model of how individuals and communities change or not. He says that at least the four distinct patterns are possible: *stability*, *age-grading*, *generational change*, and *communal change*. They can be better understood through Figure 4, reproduced from Labov (1994:83):

	Individual	Community
1. Stability	Stable	Stable
2. Age-grading	Unstable	Stable
3. Generational change	Stable	Unstable
4. Communal change	Unstable	Unstable

Figure 4 – Patterns of change in the individual and the community

In view of this model, one has to decide which approach or approaches to use, in order to understand the behavior of the individuals and the community. According to Labov (1994:84), a panel study

> will detect conditions where the individual either changes or is stable: *age-grading* and *communal change*. But a panel study by itself will not differentiate between these two, or between *stability* and *generational change*, since it provides no view of the community except through the behavior of the same individuals.

On the other hand, also according to Labov (1994:84),

> Since a trend study includes two studies of apparent time, it will both detect unstable behavior of individuals and distinguish stable from unstable communities, differentiating all four of these patterns.

Thus he concludes by stating that trend studies are the best possible approach to language change.

This paper compares apparent time data (collected in the

1990s) with real time data drawn from a corpus of recorded materials collected 20 years earlier. The real time data are of two forms: a panel study which tracks a set of individuals across the 20 year period, and a trend study, comparing matched individuals in the two time periods.

2.1 Data

Materials analyzed here are from two sources: (i) the VARSUL data base (a corpus of sociolinguistic interviews focusing on urban language variation in southern Brazil collected in the 1990s) and (ii) the NURC data base (a corpus of interviews collected in the 70s under the Norma Culta project, with the objective of providing materials for a description of the spoken language of highly educated speakers in five different communities in the country). Data analyzed here are all from Porto Alegre.

The two corpora are not thoroughly comparable, since there have been different procedures in data collection. In the NURC corpus, the interview focused on specific topics such as for a) movies, TV, radio and theatre; b) professions; c) the human body; d) leisure and social life; e) institutions; etc. (Hilgert, 1997), whereas in the VARSUL corpus the interview focused on the speaker's experiences, memories, expectations, etc. (a sociolinguistic interview). Despite these different procedures and associated stylistic differences, the two corpora have been compared because they enable real time study. Results to be presented below seem to support this comparison.

2.2 Sample in the apparent time study

The apparent time study considered the speech of 39 informants (as shown in Figure 5) stratified according to gender, two age groups (from 25 to 49 years old and above 50 years

old), and four levels of education. These are the criteria established by VARSUL.

Sample:	Genders:	Age groups:	Levels of education:
39 Sociolinguistic Interviews from	males – 19	younger (25-50) – 17	Elementary – 10
Porto Alegre, RS, recorded in the 1990s	females – 20	older (> than 50) – 22	Intermediate – 8 Secondary – 9 Post-secondary – 12

Figure 5 – Sample in the apparent time study of *a gente* in Porto Alegre, Brazil, in the 1990s

2.3 Sample in the real time, panel study

Only speakers with post-secondary education are considered in the panel. This paper looked at the materials of 15 speakers, but effectively compares the speech of 13 individuals who were recorded in both the 1970s and the 1990s. There are 7 males and 6 females in this sample.

2.4 Sample in the real time, trend study

The sample for this analysis is made up of 36 speakers, 18 males and 18 females, divided in two age groups: younger (25 to 44 years old) and older speakers (45 to 69 years old). All of them have secondary or post-secondary education. 20 were recorded in the 70s, 16 in the 90s.

2.5 Variables

The *dependent variable* includes two variants: the old pronoun *nós* and the new one, *a gente*, in subject position. Examples 1,2, 5, and 6 show that these two forms can alternate or not in the same discourse sequences.

Although there were several independent linguistic variables investigated in our coding system, only reference is being reported here. The factors in this group were defined as follows: a) generic

reference; b) exclusive reference, meaning the speaker and other(s), except the hearer; c) inclusive reference, meaning the speaker and the hearer; d) only the speaker. Cases which were considered ambiguous[9] were coded as generic, for two reasons: this was one possible reading, maybe the most obvious one; and this has been the widespread use of 'a gente' for centuries, so *a priori* it was considered the most probable. To deal with problems in data distribution (very few cases of *a gente* meaning just the speaker; very few cases of inclusive reference because of the interview format), the four categories were subsequently rearranged in two: generic (the original factor a) and specific (which corresponds to b, c, and d).

The *independent social variables* were defined according to what had previously been established by the NURC and VARSUL corpora. In the apparent time study, they were: gender, age group, and level of formal education; in the two real time studies, only gender and age group were considered, because all the informants in the 1970's data collection had post-secondary education, so we had to keep this parameter in the 1990s.

3. Results
3.1 Apparent time

In this apparent time analysis only **subject** forms were investigated. The two variants are the old personal pronoun *nós* 'we' and the new pronoun *a gente*. The general distribution is shown in Figure 6. With 69% of *a gente*, the speakers in Porto Alegre are using it almost as much as those investigated in Rio (70%) and Florianópolis (72%). The difference between Porto Alegre and João Pessoa, though, may be due to differences in the social stratification of the sample.

---

9 This may not have been the best solution for ambiguous cases, because with this coding, one cannot refine the analysis by excluding them to check if there is a quantitative difference or not.
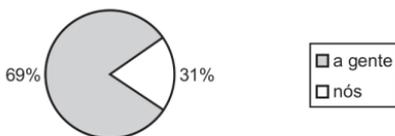
Figure 6 - Percentage of *nós* and *a gente* in Porto Alegre, 1990s (Varsul data)

The linguistic factor group of reference was selected as significant in a Varbrul analysis. The results in Table 1 show that generic use is still strongly associated with *a gente*, which is compatible with what Hopper (1991: 22) calls Persistence:

> The Principle of Persistence relates the meaning and function of a grammatical form to its history as a lexical morpheme. This relationship is often completely opaque by the stage of morphologization, but during intermediate stages it may be expected that a form will be polysemous, and that one or more of its meanings will reflect a dominant earlier meaning.

Table 2 – The use of *a gente* by *reference* in Porto Alegre, in the 1990s: Ns, % and weights (Varsul data); apparent time

Reference	N/Total	%	Weight
Generic	717/931	77	.63
Specific	620/1013	61	.38
Totals	**1337/1944**	**69**	**Input: .85**

Although these results are as expected, later we will show that there is more to them than can be inferred by apparent time analysis alone.

Looking now at Table 3, the apparent time data also show a significant age-grading in the population examined in the 1990s: younger speakers use *a gente* at a significantly higher rate than older speak-

ers: 78% versus 65%, factor weights of .66 and .42 respectively. Also significant was the factor group gender, with females leading this change with a factor weight of .55 (against .41 for males).

Table 3 – Significant social factor groups for the use of *a gente* in Porto Alegre, Brazil, in the 1990s: Ns, % and weights (Varsul data); apparent time

	N/Total	%	Weight
Gender			
Males	422/678	62	.41
Females	915/1266	72	.55
Age groups			
25-50	480/618	78	.66
50-70	857/1326	65	.42
Totals	1337/1944	69	Input: .85

Since the apparent time approach assumes that the individuals are stable, our next step will be a real-time panel study to check whether this is true for our sample. Unfortunately, our panel will include only about ¼ of the 39 individuals in the apparent time sample (as explained in 2.3).

3.2 Real time, panel study

For the panel study, we plotted the speakers in a scattergram, according to their usage in the two different time periods. To the extent that they fall on a straight line, we conclude that they are generally stable in their usage. Such a test is shown in Figure 7, for 12 of the speakers (one person in the panel had only one token in the 1970s *corpus*, which did not permit a quantitative analysis.) For the most part, the distribution of the speaker points is systematically linear: most speakers have similar values on both axes.
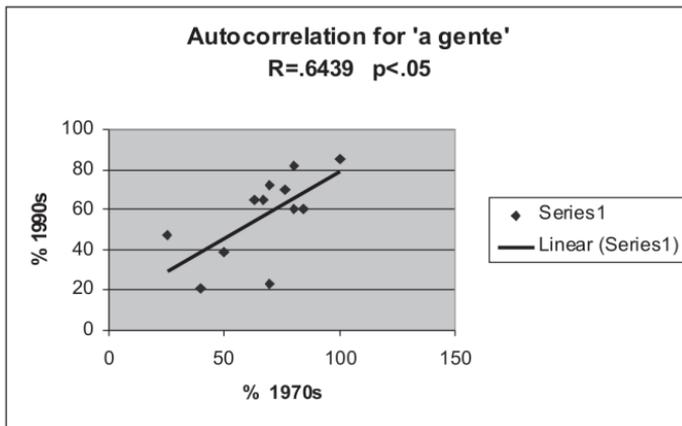
Figure 7 – Autocorrelation of speakers' *a gente* usage in the 1970s and 1990s (NURC and Varsul data)

As noted above, a crucial assumption of the apparent time method is that individuals remain relatively stable in their usage during their adult life. This assumption is partly supported in the present data, in that the autocorrelations of the panel study speakers for the two different points in time are quite high (r=.64).

Based on the results of the panel study, the general stability of the individuals in the apparent time study can be taken for granted. The trend study to be presented next will help us interpret the age-grading pattern shown in Table 3 above.

3.3 Real time, trend study

In the real time trend study we found overall 65% usage of *a gente* (991 out of 1533 tokens), and an input of .65. The results in Table 4 strongly support that this is an ongoing generational change. The rates of use in data collected in the 1970s are significantly lower

than the usage of *a gente* in the 1990s: respectively 56% and 72%, factor weights of .31 and .67. Other social parameters such as age and gender are also suggestive of change in progress: younger speakers and females lead the process.

Table 4 – Significant social factor groups in the use of *a gente*, trend study of two samples (1970s and 1990s) analyzed together (NURC and Varsul data)

	N/Total	%	Weight
Age groups			
Younger	633/865	73	.64
Older	358/668	54	.32
Decade/interview			
1970s (NURC)	403/721	56	.31
1990s (VARSUL)	588/812	72	.67
Gender			
Males	387/654	59	.46
Females	604/879	69	.53
Totals	**991/1533**	**65**	**Input: .74**

In order to examine in detail the results concerning decades and age groups, we performed the crosstabulation shown in Table 5.

Table 5 – Cross-tabulation between age groups and decades in relation to the use of *a gente*; trend study (NURC and Varsul data)

	25-50 years		50 or more years	
	N/Total	%	N/Total	%
1970s	357/529	67	46/192	24
1990s	276/336	82	312/476	66
Difference		**15**		**42**

The percentages in the cross-tabulation suggest that the individuals are stable, but the community is changing, as predicted by the model of generational change. Considering that at least 20 years had elapsed between the first and the second data collections, virtually all the younger informants from the 70s would fall into the older co-

hort in the 90s. As predicted, their percentages are the same: 67% and 66% respectively, which demonstrates they are stable.

On the other hand, table 5 also shows evidence that the community is unstable. Comparing first those who were young in the 1970s with those who were young in the 1990s, the use of *a gente* has increased in 15%: from 67% to 82%. Then comparing those who were older in the 1970s with those who were older in the 1990s, there is a difference of 42% in the use of *a gente*: from only 24% to 66%. So this cross-tabulation supports the interpretation that a generational change is going on in this community. This change can be represented by considering year of birth and ignoring the fact that data had been collected in two different points in time. This is displayed in figure 8.
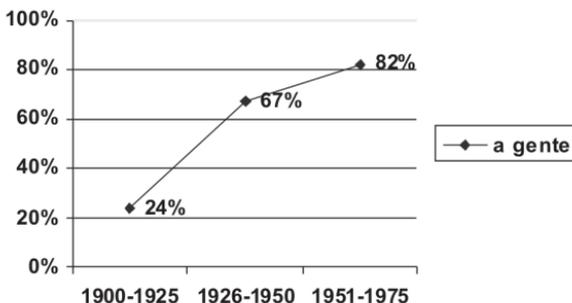

Figure 8 – *A gente* usage according to year of birth and assuming the stability of the individuals

Looking now at the cross-tabulation between reference and decades, in the trend study, it is possible to say that our data support the distinction of two changes going on at this time: a) a *midrange change*[10], related to the

---

10 Labov (1994:79-82) postulates five stages of change in the Philadelphia vowel system, naming them from the oldest to the youngest as follows: completed changes; changes nearing completion; midrange changes; new and vigorous changes; and incipient changes. We are using his terms in order to describe the difference between the two changes in the grammaticalization of *a gente*.

use of *a gente* with generic reference; and b) a *new and vigorous change*, related to the use of *a gente* with specific reference, as a full personal pronoun.

Table 6 – Cross-tabulation between reference and decades in relation to the use of *a gente*; trend study (NURC and Varsul data)

	Generic N/Total	%	Specific N/Total	%
1970s	385/660	58	18/61	30
1990s	263/366	72	325/446	73
Difference		**14%**		**43%**

According to Table 6 there has been a huge increase of *a gente* with specific reference (more than the double): from 30% to 73%, revealing that it was mainly in this context that the change has progressed. Compare this with the difference of only 14% observed in the generic usage of *a gente*. Such differences lead us to the last step: separate analyses according to reference.

3.4 Real time panel study, separate analyses according to reference

Conducting two separate VARBRUL analyses of *a gente* versus *nós* according to reference, different factor groups were associated with the generic and the specific use in the panel study[11]. The fact that different factor groups have been selected as significant seems to support the idea that there are in fact two variable rules or two different changes going on, one more advanced than the other. The details are as follows:

a) comparing only the tokens with specific reference, the only two significant factor groups selected by VARBRUL were **age groups** and **decade**.

b) comparing only the tokens with generic interpretation, the sig-

---

11 In these Varbrul runs, there were two interviews of each of the 13 informants in the panel, totaling 26 interviews analyzed once for generic use, then for specific use.

nificant factor groups were linguistic constraints **subject in previous clause**, and **proximity between S and V,** plus the social factor group **gender**.

The results is (a) are very interesting because they reinforce the idea that this is a new and vigorous change in which the most important factor groups are both related to the course of time, and seem to correspond to the acceleration phase of this change. In contrast, the results in (b) show both linguistic and social constraints, as usually happens in an ongoing, more advanced, change. In Table 7 the details of data distribution are presented for both decades and both changes.

Table 7 – Distribution of *a gente* versus decades; two separate Varbrul runs according to generic and specific reference; panel study speakers only

	Generic N/Total	%	Specific N/Total	%
1970s	338/499	68	18/55	33
1990s	320/479	67	148/288	51
Input	.68		.49	

The first thing to note in Table 7 is that the percentage of *a gente* in the generic usage is the same in both decades, whereas there is an increase of 18% in specific use if the 1990s are contrasted to the 1970s. So in one change (with the generic interpretation), the individuals in the panel are stable, as predicted by the apparent time construct. Nevertheless, they are NOT stable during their adult life in relation to the other change (with the specific interpretation). Why is this so? One possibility is that this change is newer and these elite speakers avoid less the new personal pronoun in the 1990s, in comparison to the 1970s, perhaps because in the meantime this new personal pronoun has acquired covert prestige and, at the same time, its usage enables speakers to skip verbal agreement (and the inconveniences of the strong stigma associated with lack of verbal agreement).

Relating these results to the mechanism of *extension* (as explained

in section 1), through which *a gente* is now being used in new contexts and acquiring new meanings as a fully personal pronoun, it seems justifiable to say that, in the set of changes of this grammaticalization process, this is in fact a new (semantic) change. The trend study with separate analyses to be presented next will help us understand what is going on in the community.

3.5 Real time, trend study, separate analyses according to reference

To test this possibility, we also did separate analyses with the trend data. The results shown in Table 8 support the distinct treatment of specific and generic reference tokens. First, note that the factor group *age* is selected in all four runs as statistically significant. Second, comparing separate analyses of the two types of reference, in **four** distinct Varbrul runs, there is a larger **difference** in input values in the 1970s as compared to the 1990s: specific reference lags generic reference by .29 in the 70s, but by only .16 in the 90s.

Table 8 – Input values for *a gente* usage; four separate Varbrul runs according to decades vs. generic and specific reference

	Generic	Specific	Difference
1970s	.59	.30	**.29**
1990s	.78	.62	**.16**
Difference	**.19**	**.32**	

Another important piece of evidence in Table 7 is that in the 1970s the factor weight of .30 disfavored the use of *a gente* with specific reference, whereas in the 1990s the weight of .62 favors *a gente* with specific reference[12].

---

12 We find here the same inversion of factor weights noted by Omena (2003). And although it seems true that the change is more advanced in Rio than in Porto Alegre, it is important to remember that the samples studied are very different: Omena's samples are socially stratified, whereas ours include only speakers with post-secondary education, a social group who usually avoids or uses less of the newer forms.

This finding also support the hypothesis that there are two parallel semantic changes going on, the generic reference change, which started in the XVI century and is much more advanced, by means of which the NP *a gente* became an indefinite pronoun, and the specific reference change, by means of which it is becoming a personal pronoun, which may have started in the XIX century and accelerated in the last three decades of the XX century. We can only speculate about the reasons for this acceleration: it is here conceived as a result of deep demographic and socio-cultural changes all over the country, whose main impacts have been to dramatically invert the distribution of the population in rural and urban areas and increase the total population by a factor of 4 times (from 40 million in the 1950s to 170 million in the 1990s).

4. Final remarks

Despite the limitations of the samples studied here, there are some important points to be highlighted.

First, the hypothesis of distinct semantic changes proposed here, one more advanced than the other, is consistent with the concept of grammaticalization as a set of inter-related changes. Although these changes may overlap in time, they may also show different paces, which in turn may be due to related processes of changes going on in the language. Recall that grammaticalization processes are highly linguistically motivated and their progress may depend on the advancement of other related changes that trigger them. In this specific case, the related changes concerning the loss of the null subject parameter and the loss of verbal agreement are important forces driving *a gente* to become a personal pronoun because it facilitates both of these preferences at once: using *a gente* simultaneously avoids first person plural verbal inflection and null subject occurrences.

Second, this kind of result is NOT obtainable by apparent time

studies alone. Remember that in our apparent time analysis, we could only say that generic reference was still a more favorable context for *a gente*. In my view this illustrates the importance of the convergent use of real and apparent time data, wherever possible.

Third, there is an important implication here: the alternation between *nós* and *a gente* with both generic and specific reference should be redefined as two variable rules: the first, in which *a gente* would be contrasted to all the other forms with generic meaning, including *nós*; the second, where *a gente* would be only contrasted to the specific usages of first person plural *nós* and its null form.

Fourth, the leaders in each change may be different; as a consequence, it seems quite relevant to investigate the social embedding of the grammaticalization processes as sets of changes

Last, since our real time data include only the speech of people with high levels of formal education, it would be interesting to test the hypothesis of two separate changes using real time data from a stratified and more representative sample of the community.

References

ASSIS, Rosa Maria. 1988. *Variações lingüísticas e suas implicações no ensino do vernáculo: uma abordagem sociolingüística*. Ilha do Desterro. v. 20:p.59-81.

BLONDEAU, H. 2001.*Real-time changes in the paradigm of personal pronouns in Montreal French*. Journal of Sociolinguistics. 5/4: p.453-474.

BORGES, Paulo R.S. 2004. *A gramaticalização de 'a gente' no português brasileiro: análise histórico-social-lingüística da fala das comunidades gaúchas de Jaguarão e Pelotas*. Porto Alegre: UFRGS. Tese de Doutorado.

CROFT, W. 1990. *Typology & Universals*. Cambridge: CUP.

DIEWALD, G. & WISCHER, I. 2002. Introduction. In Wischer, I. & Diewald, G. *New reflections on Grammaticalization*. Amsterdam/

Philadelphia: John Benjamins. p.ix-xiv.

DUARTE, Maria Eugênia L. 1997. *Aspectos do sistema pronominal do português falado nas regiões Sudeste e Centro-Oeste*. In Anais do XI Encontro Nacional da ANPOLL. João Pessoa, junho de 1996.

_____. 1998. *O português do Brasil no contexto das línguas românicas*. Paper presented at the XXIIᵉ Congrès International de Linguistique et Philologie Romanes. Université Libre de Bruxelles, juillet 1998.

HEINE, B. & KUTEVA, T. 2002. *World Lexicon of Grammaticalization*. Cambridge: CPU.

HEINE, B. 2003. *Grammaticalization*. In Joseph, Brian & Janda, Richard D. (eds.) *The Handbook of Historical Linguistics*. Oxford: Blackwell. p. 575-601.

HILGERT, José Gaston. (org.) 1997 *A linguagem falada culta na cidade de Porto Alegre: materiais para o seu estudo*. Passo Fundo: EDIUPF; Porto Alegre: Ed. da Universidade/UFRGS.

HOPPER, Paul. 1991. On some principles of grammaticalization. In E. Traugott & B. Heine, eds. *Approaches to Grammaticalization*. Amsterdam: J. Benjamins. Vol. 1. p.17-35.

LABERGE, S. & SANKOFF, G. 1980. Anything YOU can do. In Sankoff, Gillian. *The social life of a language*. Philadelphia: University of Pennsylvania Press. p.271-293.

LABOV, William. 1994. *Principles of Linguistic Change*. Vol. 1: Internal Factors. Oxford: Blackwell.

LEHMANN, Christian. 1995. *Thoughts on Grammaticalization*. München, Lincom Europa.

LEITE, Yonne & CALLOU, Dinah. 2002. *Como falam os brasileiros*. Rio de Janeiro: Jorge Zahar Editor.

LOPES, Célia R.S. 1998. *Nós e a gente no português falado culto do Brasil*. *DELTA*. vol. 14 (2): p.405-422.

_____. 2001. O percurso de *a gente* em tempo real de longa duração. In *Para a história do português brasileiro*. Vol. II, Tomo I – Primeiros estudos. p.127-148. São Paulo: Humanitas Publicações/FFLCH/USP.

_____. 2003. *A inserção de a gente no quadro pronominal do Português.* Madrid: Iberoamericana.

MEILLET, A. 1912. L'évolution des formes grammaticales. In: Meillet, A. *Linguistique historique et linguistique générale.* Paris: Champion. p. 130-148.

MENON, Odete P. S. 1994. *Analyse sociolinguistique de l'indétermination du sujet das le Portugais parlé au Brésil, a partir des données du NURC/SP.* Université de Paris VII. Doctoral dissertation.

_____. 1996. *A gente*: um processo de gramaticalização. *Estudos lingüísticos.* XXV: p.662-28.

OMENA, Nelize P. de. 1996. As influências sociais na variação entre *nós* e *a gente* na função de sujeito. In *Padrões sociolingüísticos: análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro,* ed. Giselle M. de Oliveira e Silva & Maria Marta P. Scherre, 309-323. Rio de Janeiro: Tempo Brasileiro: Depto. de Lingüística e Filologia, UFRJ.

_____. 2003. A referência à primeira pessoa do plural: variação ou mudança? In PAIVA, Maria da Conceição & DUARTE, Maria Eugênia L. (orgs.) *Mudança lingüística em tempo real.* Rio de Janeiro: Contra Capa Livraria. p.63-80.

_____ & BRAGA, Maria Luiza. 1996. *A gente* está se gramaticalizando? In: Macedo, Alzira T. et al, eds. *Variação e discurso.* Rio de Janeiro: Tempo Brasileiro.

ROMAINE, S. & LANGE, D. 1998. The use of *like* as a marker of reported speech & thought: a case of grammaticalization in progress. In: Jenny Cheshire & Peter Trudgill, eds. *The Sociolinguistics Reader,* vol. 2, Gender & Discourse. London: Arnold.

SCHMITZ, John Robert. 1973. The linguistic flexibility of 'a gente' in Portuguese. *Hispania*, [Sept. 1973]. p. 639-644.

SEARA, Izabel Christine. 2000. A variação do sujeito nós e a gente na fala florianopolitana. *Organon*, vol. 14 (28/29):p.179-194.

SERRANO, Maria José. *Accounting for morpho-syntactic change in Spanish: the Present Perfect case.* University of Pennsylvania Working Papers in

Linguistics. 3,1:p.51-61.

TORRES Cacoullos, Rena. 2001. From lexical to grammatical to social meaning. *Language in Society.* 30:p.443.478.

TRAUGOTT, E. 1989. From less to more situated in language: The unidirectionality of semantic change. In Adamson, S. et al. (eds.) *Papers from the Fifth International Conference on English Historical Linguistics.* Amsterdam: John Benjamins.

WISCHER, I. & DIEWALD, G. 2002. *New reflections on Grammaticalization.* Amsterdam/Philadelphia: John Benjamins. p. ix-xiv.

ZILLES, Ana M. S. 2000. *Variable subject-verb agreement with first person plural in spoken Brazilian Portuguese.* Paper presented at NWAV 29, East Lansing, Michigan, October 2000.

_____. 2002. *Grammaticalization of a gente in Brazilian Portuguese.* In Johnson, Daniel Ezra & Sanchez, Tara, eds. University of Pennsylvania Working Papers in Linguistics. v. 8, n.3:p.297-310.

_____. 2005. *The development of a new pronoun: the linguistic & social embedding of a gente in Brazilian Portuguese.* Language Variation & Change. 17 (1):p.19-53.

_____ & MAZZOCA, P. da R. 2000. *O papel dos fatores sociais no processo de gramaticalização de 'a gente'.* Paper presented at the IV CELSUL, UFPR, Curitiba, Brazil. November 2000.