The development of a new pronoun: The linguistic and social embedding of *a gente* in Brazilian Portuguese

A N A M . S . Z I L L E S
Universidade Federal do Rio Grande do Sul

A B S T R A C T

The Portuguese NP *a gente*, meaning "the people," is undergoing grammaticalization and is acquiring characteristics of a personal pronoun, increasingly replacing first-person plural *nós*, meaning "we," in speech. In Brazilian Portuguese, this process seems to be correlated with a number of other ongoing morphosyntactic changes. In this study I compare data from Southern Brazil on the use of *a gente* in the 1970s and the 1990s. Quantitative analyses are conducted in terms of two methodological approaches: apparent-time and real-time studies. In the real-time analysis, two kinds of studies are discussed: a trend study, with two comparable groups of speakers, and a panel study, with the same speakers compared longitudinally. The linguistic and social embedding of this process is discussed in terms of the Labovian classification of changes as being "from above" or "from below."

The term grammaticalization was apparently used for the first time in 1912, by Meillet, who defined it as "the attribution of a grammatical character to a previously autonomous word" (Meillet, 1912:131). Nevertheless, grammaticalization as an area of investigation started to develop only 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). Nowadays, the term is used to refer to a process of interrelated changes, as well as "to the degree of grammatical function a linguistic item has on a scale between purely lexical and purely grammatical meaning" (Diewald & Wischer, 2002:ix).

Until recently, investigators were mainly concerned with conceptual and methodological issues or with the description of grammaticalization phenomena in single linguistic items, but now there is an increasing interest in understanding

the role of linguistic contexts in grammaticalization processes. Furthermore, some studies (Blondeau, 2001; Lopes, 2001; Romaine & Lange, 1998; Serrano, 1996; Torres Cacoullos, 2001) have begun to investigate 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. In so doing, one aspect researchers have to consider is whether grammaticalization is a special kind of change or not. If it is special, then one consequence might be that it does not show the same tendencies as other changes do in terms of social embedding.

To answer this question, it is not sufficient to analyze written texts from different times, or even spoken-language data extracted from *corpora* that are dissociated from the speakers and the sociohistorical contexts in which the data were collected. Rather, we must address the social, linguistic, and diachronic distribution of the phenomenon under investigation. So, this article is an attempt to explore these issues in connection with one case of grammaticalization drawn from the reorganization of the pronominal system of Brazilian Portuguese; and to deal with it not only by describing the linguistic features and contexts of the change, but also by looking for possible associations between them and the social characteristics of the speakers and the sociohistorical context.

The article is organized as follows. First, I discuss some important points about the theoretical framework of grammaticalization. Next, I consider the Labovian model of change types – change from above and change from below. Then, I describe the variable that I have investigated, and summarize the diachronic background of the change, highlighting its connections to other ongoing related changes. Finally, the core of the article will be the results concerning both the linguistic and social embeddings of this change, using three different methodological approaches: an apparent time study, a real time panel study, and a real time trend study. Although there is substantial evidence supporting the interpretation that this grammaticalization process is a change from below, its social embedding is somewhat masked by the fact that it involves a set (or cluster) of interrelated changes that appear to overlap in time. At the same time, there is ample evidence regarding its linguistic embedding, which conforms to the expectations of the theory of grammaticalization: "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" (Diewald, 2002:117).

GRAMMATICALIZATION

In terms of the theoretical framework, the first question to be addressed is whether or not grammaticalization is a distinctive type of change. To my knowledge, this question has not yet had a satisfactory answer. The effort that has been put into establishing principles of grammaticalization (Heine & Reh, 1984:269–282; Hopper, 1991; Lehmann, 1995) has shown that the same kinds of linguistic processes

(attrition, semantic change, coalescence, etc.) also occur in other kinds of changes. If there are no exclusive principles or features that distinguish grammaticalization from other kinds of change, what is special about it? Hopper (1991) and Traugott (1989) believed that grammaticalizing changes are *not* different from other changes, meaning that, for example, if grammaticalization involves a semantic change, this is not different from other semantic changes; or if it involves a phonological change, this is not different from other phonological changes.

Although there seems to be no reason to deny that, in essence, each kind of change involved in any grammaticalization process is not, in itself, different from any other similar but isolated change, there are two further points to highlight. First, grammaticalization may be special in that it involves a set of interrelated changes, and second, grammaticalization (at least in a narrow sense of the word) seems to be unidirectional.

Grammaticalization as a set of interrelated changes

One way of conceiving grammaticalization is to consider the *continuum* of changes that define it as a set of different processes affecting an item over time. This is apparent in the following definition by Croft (1990:230): "Grammaticalization is the process by which full lexical items become grammatical morphemes. (…) Phonological, morphosyntactic and functional (semantic/pragmatic) changes are correlated: if a lexical item undergoes a certain kind of morphosyntactic change, it implies corresponding functional and phonological changes."

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, transitions that tend to be similar in type across languages" (Hopper & 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, in terms of an arrangement of forms along an imaginary line with a fuller, lexical element at one end, and a reduced, grammatical element at the other.

That grammaticalization represents a correlation of changes over time raises the question of whether its various processes are synchronized or not (Croft, 1990:242). The hypothesis that each kind of change in a grammaticalization process progresses gradually or involves several stages makes sense, but the further conception that they are synchronized seems to me to be too idealized and sociolinguistically unacceptable. If we think of grammaticalization in terms of several distinct but simultaneous changes, based on what we know of different variable rules in action in the same community (e.g., Bortoni-Ricardo, 1985; 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, to see whether the general tendencies are comparable for all the changes and whether or not the same speakers are leading every change involved.

Grammaticalization also seems to be special in that it appears to be linguistically motivated and highly embedded in the linguistic system. As Diewald (2002:117) said: "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). As I expect to show in this article, this may well be the case in Brazilian Portuguese, as it takes its own course away from European Portuguese. Several morphosyntactic changes are going on in Brazilian Portuguese that affect both the paradigm of personal pronouns and subject–verb agreement. These changes include, among others, the introduction of new pronouns and a related overall reduction in the use of verbal agreement. Several hypotheses have been proposed in the literature about the initial causes of this complex of changes, from change in word order to loss of verbal agreement. In this article I will not concentrate on this issue; these other parallel changes will be mentioned only in connection to the new pronoun that is the focus of attention here.

Grammaticalization as unidirectional

The idea of changes having direction is not new in the literature (see, e.g., the discussion about the unidirectional principles of chain shifting in Labov, 1994, ch. 5). In grammaticalization studies, it is often said that this process is unidirectional, meaning that the reverse sequence is impossible. It is also said that it is cyclic, meaning that the return to the original state is effected by a different grammaticalization process. So grammatical morphemes originate from lexical items, disappear through loss, and reappear when new words become grammatical morphemes (Croft, 1990:230). But unidirectionality is a highly controversial issue. In this respect, Heine (2002:97) recently wrote that "a number of examples contradicting the unidirectionality principle have been pointed out. (. . .) Still, as acknowledged by most scholars who have identified exceptional cases, such examples are few compared to the large number of cases that conform to the principle."

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, and so forth. Furthermore, it is often observed that a grammaticalization process can involve long periods of stability in the intermediate stages, or become halted and "never" reach the endpoint of complete loss. This also might be interpretable in terms of social evaluation, prestige, and other factors. It is easier to understand this if we think of grammaticalization as a set of

interrelated changes: The direction is possible, but not compulsory; after some of the changes have happened, it may take a long time (indeed centuries!) until new ones develop. Language is not a self-governed mechanism; it is the result of social practices developed by socially organized individuals in interaction. The same can be said about changes; the individuals and the social groups change the language.

Now, all of this raises an important question. If grammaticalization has the properties just described, how does it fit into the Labovian classification of changes as being "from above" (those that involve conscious or at least subconscious imitation of an external prestige norm) or "from below" (those that involve unconscious, spontaneous development internal to the speech community)? Or more specifically, is grammaticalization sensitive to or driven by social factors? (If so, how could it be unidirectional?) Or is it linguistically motivated and controlled in a way that overrides social processes and pushes it relentlessly forward?

I will return to the Labovian model of variation and change in a moment, but for the present it is worth noting this: If we try to answer these questions by treating the whole process of grammaticalization at once, or as a single change, we may not find a solution. Putting all the interrelated steps in a grammaticalizing change together may obscure the role of speakers as they engage in or resist the different processes. I think that the Labovian methodology can only be useful in this respect if we analyze the linguistic and social embedding of each separate subchange one at a time. Having done this, we will then have to put them together in a coherent way.

LABOV'S MODEL OF CHANGES

Labov's work, echoed by that of a number of other researchers, distinguishes two distinct sociolinguistic types of change within a community of native speakers. (There are also other proposals from other researchers[1] about change types that involve nonnative speakers or situations of extreme language contact, but these are not discussed here.) Each of the change types is seen as having characteristic social distributions and embeddings, and they may also be differentiated by their linguistic embedding. Labov's terms for the two types are "change from above" and "change from below" (Labov, 1994:78).

Change from above is viewed as linguistic change imported into a speech community from elsewhere, ordinarily as a prestige model, and most speakers are more or less consciously aware of it. It should have no linguistic motivation, because it is socially driven. Such changes will be sporadic with respect to the linguistic system they are being incorporated into, and not infrequently they involve reversals of previous directions of change.

The classic example of change from above is the recovery of post-vocalic or coda (r) in New York City (Labov, 1966). New York City accompanied the 18th–19th century change that vocalized or deleted /r/ in this position in southern dialects of England as well as much of the Atlantic coast of English-speaking

North America, so that by the end of the 19th century, the city was thoroughly "r-less." However, the neighboring dialects in Pennsylvania and New Jersey, upstate New York, and southwestern Connecticut are all r-pronouncing, so that the City was left as a linguistic island on this feature. Furthermore, for complex historical reasons, the dialect of New York City has long had very low status in North America, being the object of stigmatization and ridicule from other dialect regions. Beginning sometime around World War II, New Yorkers began to readopt postvocalic /r/, importing it from the adjacent higher-status dialects as a prestige feature.

In the New York case and elsewhere, we find a typical pattern for the social distribution of change from above. First, they are almost invariably led by the highest status group, which ordinarily includes the most-highly educated speakers. Second, as prestige features, changes from above are ordinarily favored in more formal styles. Third, like any ongoing change, we expect to find that younger speakers use the form more than older speakers, but in changes from above, the peak age group is normally young adults rather than adolescents, because the young adults, being in a more demanding position in the linguistic market, tend to be more sensitive to the prestige demands of the wider community. Fourth, women tend to lead these changes, just as they lead most linguistic changes.

By contrast, the other change type – change from below – typically involves different motivations and different social distributions. Change from below is seen as a spontaneous development, emerging from within the speech community, not imported or modeled on developments elsewhere. In its initial stages at least, speakers generally have little conscious awareness of the existence of the change. These are the changes that are often described in historical studies as having linguistic motivations (e.g., chain shifts, assimilations, etc.), but they are also construed by Labov as being driven by social motivations, as well (such as local identity or solidarity).

The typical social distribution of changes from below is as follows. First, these changes are never led by the highest status group, rather, the leaders of these changes are typically the lower-middle-class or the upper-working-class (in Labov's view, these are the social groups with the strongest investment in local identity). Second, these changes start out with no social evaluation and consequently no stylistic variation, although at later stages they may develop stylistic variation, depending on whether or not they receive some social evaluation. Third, younger speakers use these changes more than older speakers, but the age-peak is ordinarily in adolescents rather than young adults. And along the gender dimension, women usually lead these changes, although there are attested cases where men lead.

BACKGROUND ON THE VARIABLE

As previously mentioned, Brazilian Portuguese seems to be undergoing grammatical reorganization. One of these processes results in the pronominal usage of

a gente, originally a full NP meaning "the people," which is increasingly being used as a personal pronoun (Omena, 1996a). In example (1), from the VARSUL corpus[2], this new form alternates with the old pronoun *nós*.

(1) POA02,[3] l. 250–252

E E condução, como é que é?	I And transportation, what's it like?
F Condução, bom, *nós$_i$ não temos* condução própria. (hes) Então *a gente$_i$* *depende* do ônibus.	S Transportation, well, we don't have our own [car]. (hesitation) So we depend on the bus.

It is not surprising that a word like "people" was the source for this change. According to Castilho (1997:37) and Heine and Kuteva (2002: 232–233), languages tend to have generic nouns as sources for personal pronouns. Words like *man*, *people*, and *person*, probably for semantic reasons, are good candidates for grammaticalization as indefinite pronouns. Well-known examples are the indefinite *Mann*, in German; the formerly third-person indefinite *on* in French (derived from the noun *homme*, "man"), now also used as first-person plural; *homem/ ome*, meaning "man," in Old Portuguese; European Portuguese *pessoa*, meaning "someone" ("A pessoa não deve preocupar-se"); Swahili *mtu* "person," used as an indefinite pronominal in existential expressions; and so forth.

The shift concerning *a gente* probably began in the 16th 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. Although it is not clear yet why *homem/ome* turned out to be dispreferred at that time, it is easy to recognize that its disappearance corresponds to the final stage of its previous unidirectional process of grammaticalization, that is to say, loss. The emergence of *a gente* represents the renewal of the process and corresponds to a new cycle. What is also clear in this respect is that both forms competed for a certain time (Lopes, 2001) or, in other words, there was variation before change or as change was getting underway.

As expected, the grammaticalization of *a gente* was slow and gradual, and involved an intermediate stage 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 (Lopes, 2001:140–141).

According to several authors (Lopes, 2001:137; Menon, 1996:626; Schmitz, 1973:640), another aspect of this process of interrelated changes has to do with gender agreement. As a noun, *gente* has feminine gender and requires feminine modifiers as in *gente bonita* ("beautiful people"). The adjective *bonita* is a marked feminine form in Portuguese, as opposed to the unmarked masculine form *bonito*. Nevertheless, as a pronoun, *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.

During the second half of the 19th century, use of the NP *a gente*, referring to the *speaker and a group of other specific referents*, one of the steps on its way toward becoming a personal pronoun, was already noted. Example (2) is an excerpt

from a short story published in 1893 by Artur Azevedo, a well-known Brazilian playwright and journalist of that time. In a scene of family life, father, mother, and two children are talking. One of the children asks the father about the meaning of a word (which in fact he does not know), and everybody is waiting for his answer. Impatient with his delay, the mother compels him to reply:

(2) Vamos: se sabe, diga o que é   Come on: if you know it, tell us the
plebiscito! Então? *A gente* está   meaning of plebiscite! So? We are
esperando! Diga! (Azevedo, 1967:179)   waiting! Say it!

Nowadays, as a personal pronoun, *a gente* is used to refer to first-person plural, with the meanings shown in (3).

(3) a. the speaker and the interlocutor or audience (the so-called inclusive usage)
     b. the speaker and a clearly defined group of persons (e.g., the family), excluding
        the interlocutor
     c. the speaker and everybody else (the generic meaning, be it understood as anyone
        or everyone in a group)

Besides these semantically collective references, *a gente* is sometimes used with the first-person singular reference, although many of these cases may seem somewhat ambiguous, depending on context and the verb tense of the clause (the present and imperfect tenses being more likely to have a generic interpretation than the preterite). Nevertheless, there are unequivocal cases showing that *a gente* may convey the meaning of first-person singular, such as the following example from Schmitz (1973:640): *a gente* está zangad**o** ("I am angry"), in which *a gente* is used by a male speaker to refer to himself, with masculine singular marking of the predicate adjective.

Another clear example, from my data, is presented in example (4).

(4) Gosto muito de churrasco, [...]   I like barbecue a lot, I really like
gosto mesmo de churrasco, mas como   barbecue, but since I normally eat
normalmente todos os domingos eu almoço lunch at my mother's every Sunday,
na mãe, e no sábado eu almoço em casa, o and Saturday I eat at home, the
churrasco sobra para quando nós vamos a barbecuing is left for when we go to
Canela. Nós temos uma casa em Canela e   Canela. We(1pl) have(1pl) a house
lá, todas as vezes que vamos, o   in Canela and there, every time we(Ø)
churrasquinho é feito. E esse churrasco é *a* go(1pl), there's a barbecue. And this
gente mesmo que faz, mas o mérito não é barbecue, it's *a gente* (= 'I') myself
do assador e sim da carne. (NURC 127,   who makes it, but the quality comes
l. 95–104)   not from the cook but from the meat.

Parallels with "você"

These observations show that *a gente* has acquired the semantic properties of a personal pronoun and support the view that it is undergoing grammaticalization. Significantly, this process parallels a previous development of the second-person singular pronoun *você* in Portuguese. This pronoun arose from the grammatical-

	Old pronominal system	Agreement	Emerging pronominal system	Agreement
P1	Eu	cantO	Eu	cantO
P2	Tu	cantaS	Você / Tu	canta∅
P3	Ele / ela	canta∅	Ele / ela	canta∅
P4	Nós	cantaMOS	Nós / A gente	canta∅
P5	Vós	cantaIS	Vocês	cantaM
P6	Eles / elas	cantaM	Eles / elas	cantaM

FIGURE 1. The pronominal paradigms for subject position and corresponding verb forms in Portuguese (old system) and in Brazilian Portuguese (emerging system).

ization of the address form *Vossa Mercê* (meaning "Your Mercy," "Your Grace"), whose first record dates from 1331 (Faraco, 1996:58). Presumably, it was coined during the Middle Ages and was first used exclusively for the king, but in the 14th century it was already being used among nobles, and in the 15th century among the bourgeoisie, as can be seen in the plays of Gil Vicente. *Vossa Mercê* underwent several stages of phonological reduction, which are 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, it 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).

These changes concerning the second-person pronouns had a great impact on the corresponding subject–verb agreement, leading to the progressive loss of second-person endings in favor of third-person forms (because, as NPs, *Vossa Mercê/você* were third-person singular forms, taking corresponding third-person singular verb agreement). This process was very appropriately labeled "the revolution of the third person" by Marilina dos Santos Luz (Faraco, 1996:54–55). This expression, therefore, refers to a sociolinguistic process leading to a chain of morphological and syntactic changes, which are still in progress in the language, particularly in Brazil (for practical reasons, European Portuguese is beyond the scope of this study).

This "revolution of the third person" is extended still further with the pronominalization of *a gente*. The integration of both *você/vocês* and *a gente* in the subject pronoun paradigm has far-reaching consequences for both the pronominal and verbal agreement systems. Deriving from nominal expressions, which took third-person agreement on the verb, the new pronouns have had the effect of reducing verbal morphology from six different forms to only three, as shown in Figure 1, where the old and the emerging systems are contrasted.Observing the right column in Figure 1, we distinguish: (a) the marked first-person singular ending *canto*; (b) the marked second-person and third-person plural endings: *vocês/eles cantam*; (c) the unmarked, generalizing form for second-person and

third-person singular, as well as for first-person plural forms: *você/ele/a gente canta*. Note that *você* is the widespread P2 pronoun, but *tu* is still used in several areas (in the South, North, and Northeast of the country) with variable agreement in the verb.

Note, furthermore, that there are still other changes going on in the verbal system, which lead to further reduction in the inflectional paradigm. Among other studies, Guy (1981) has shown that there is a lot of variability in second-person and third-person plural agreement, so that it is possible to get productions like *vocês/eles canta∅*. The ultimate end of all these changes seems to be the loss of person marking in the verbal system, with the correlated decline of null subject and rise in the use of full pronouns in subject position, as shown, for example, in Duarte (2000). Also when *nós* is used there is great variability in agreement (Zilles, 2000; Zilles, Maya, & Silva, 2000). Besides the standard inflection *-mos*, there are two nonstandard forms in the VARSUL corpus: *-mo* and ∅ ending, the latter being mainly restricted to contexts in which the target form would be a word with antepenultimate stress: *nós cantávamos* (antepenultimate stress) ∼ *nós cantava* (penultimate stress, no personal ending)

Phonological reduction of a gente *and* você

Now let us turn our attention from the linguistic embedding of the change to the specifics of the grammaticalization of *a gente*. As we saw previously, grammaticalizations typically entail several interrelated processes. Two such processes are evident in the data at hand: (a) the first is the recategorization of *a gente* as a personal pronoun, which can be described, in terms of a variable rule, as the alternation between *nós* and *a gente*; (b) the second and further process involves phonological reduction of the grammaticalized element, and can be analyzed as the alternation between *a gente* and *a'ente*, with deletion of the fricative segment.

Let me begin with some remarks about the phonological reduction of *a gente* to *a 'ente*. As far as I know, there is no evidence of a general phonological process going on in the language that would have the effect of deleting this initial fricative segment. The reduction seems to be restricted to this specific lexical item, leading to the conclusion that it is happening as a result of the grammaticalization process, which implies higher frequency of use and conventionalization (Zilles & Mazzoca, 2000; Zilles, 2002).

I have found two prior references to this reduction in the literature, both about data collected in the 1970s. One is by Menon (1996), who referred to the speech of highly educated people from São Paulo, and the other is by Guy (1981:111), who studied the speech of illiterate people from Rio de Janeiro. Unfortunately, neither of the authors quantified the process in detail. Guy said: "the only place where [h] occurs with any frequency as a sibilant allophone in non-syllable-final position is in the word *gente*, which varies in pronunciation between [žẽtši], [hẽtši], and [ẽtši]." But even without quantification, these mentions of reduction in the 1970s are useful historical information for our purposes, because we also have data from that decade, as we will discuss later.

And once again, it is interesting to look at the parallel development of *você*. According to Vitral (1996) and Ramos (1997), this form, already dramatically reduced from its original source *Vossa Mercê*, is currently undergoing further phonological reduction, presenting two variants: *ocê* or simply *cê* in the singular, and *ocês* or *cês* in the plural. Interestingly enough, there is no general loss of this initial fricative segment /v/ in the language either. Besides reduction, there is an important syntactic constraint that must be mentioned. The reduced forms *ocê/ cê* occur almost only in the subject position, though *ocê* may be combined with a preceding preposition, as, for example, in a sentence like *isso é pr'ocê* (with contraction between the preposition and the pronoun) instead of *isso é para você*, meaning "this is for you." When it appears in the object position (which is pre-scriptively disfavored), it generally occurs without phonological reduction.

Apparently, therefore, the grammaticalization of *você* is still underway, with new stages evolving that may be supplying the language with a clear distinction between subject and object pronouns – a distinction greatly affected by other ongoing changes, such as the use of third-person nominative pronouns in accusative position and the parallel, almost total, loss of the third-person accusative clitics (*o/a, os/as*).

The phonological reductions of both *você* and *a gente* seem to help "reorganize" the grammar. And if it turns out to be the case that the reduction of *a gente* to *a'ente* is also constrained to the subject position, it is possible to surmise that another change may be on its way, in the same direction as observed for French, in which subject pronouns cliticize as the verbal morphology is being lost (Duarte, 2000; Vitral, 1996). Whether or not this reduction is indeed cliticization, or per-haps a change from a strong to a free weak form, as proposed in Kato (1999), is unfortunately beyond the scope of the present study.

Long-term diachronic studies

Let us now consider some of the evidence from the literature concerning the diachronic development of *a gente*. One work of note is an extensive, long-term diachronic study of *a gente* in European and Brazilian Portuguese conducted by Lopes (2001). This work reveals the early path of grammaticalization of *a gente*, but it does not provide much information about the social history of this change. Note that this is in marked contrast with the development of *você*. There were clear social motivations for the introduction of *Vossa Mercê* into the language and for its further conventionalization as a personal pronoun, but we still do not know what social forces (if any) favored the introduction of *a gente* into the pronominal system.

A more localized diachronic investigation of the use of *a gente* in southern Brazilian drama, from 1890 to 1990, was conducted by Borges (2001). In this study, all of the playwrights investigated were from the South of Brazil and their plays talked about life in this region, with local characters and issues. This, of course, is highly relevant for this article, in that the data that we have analyzed are from this same region. In his research, Borges addressed two main questions:
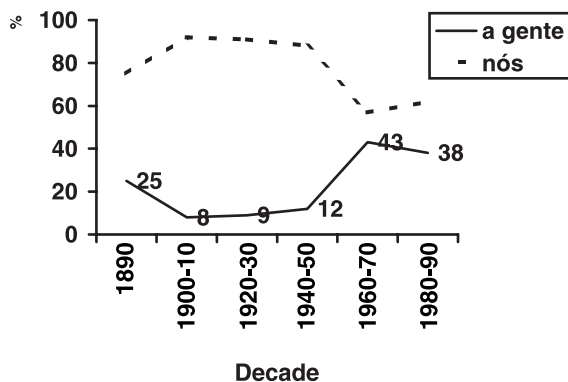
FIGURE 2. Percentage of *nós* and *a gente* in RGS's theatrical plays from 1890 to 1990 (from Borges, 2001).

(1) Is *a gente* used in these plays as a subject referential pronoun? (2) Is there any evidence that this use increases with time? Both questions had affirmative answers. Borges then suggested that his results are evidence that this change accelerated during the 1960s and the 1970s, as can be seen in Figure 2.

Figure 2 shows that in the 1960s and 1970s there is a great increase in the use of *a gente*, reinforcing the idea that during these decades the change was speeding up. Other works on the subject have also made essentially the same claim. For example, Omena (1996b:313–320) suggested that the shift to *a gente* accelerated rapidly in the 1960s and 1970s, and the 1973 article by Schmitz reinforced the impression that *a gente* became very salient at that time (the 1970s), suggesting that the change was being intensified or was already very advanced.

The timing of this apparent acceleration of the change is strikingly coincident with a substantial social transformation in Brazil affecting the demography, geography, and socioeconomic structure of the country. This transformation involved industrialization, migration from rural to urban areas, huge technological developments in communication, the development of an urban working class, and a progressive increase of enrollment in public schools, among many other aspects.

Now, the results by Borges (Figure 2), come from written literature; therefore they could also be interpreted as a reflection of a change in the literary norms, accepting more the use of forms that are closer to the spoken language. But, because other authors, like Omena (1996b) and Schmitz (1973), also mentioned evidence from the spoken language, this suggests that, in fact, the change was clearly in progress at that time.

Popular origins

Another point that emerges from the literature on *a gente* concerns the social origins and social distribution of this change. Schmitz (1973) offered some insights on this point. He claimed that the word *gente* had various uses "in modern Por-

tuguese in both Brazil and the Peninsula," and presented several examples extracted from modern Portuguese and Brazilian prose fiction (Schmitz, 1973:639). Reviewing the description of this word in standard reference grammars available at the time, he found three main ways of considering it. In the first, it was classified as an *address form*, probably insofar as the grammarians were already recognizing that this word was being used to include the speaker as part of its reference. In the second and most common interpretation, it was treated as an indefinite form. But it is the third way that is most interesting for our purposes. Schmitz said: "The third way of interpreting *gente* is either to consider it as part of *popular speech*, uttered only by 'a boca do povo' [the mouth of the people], and hence unimportant, or to ignore completely the existence of the word" (1973:639).

This observation offers us a sense of the possible social origins of this change, namely, that it originated in popular – as opposed to elite – usages. In Labovian terms, this could indicate a "change from below." This possibility gains further support in other studies. For example, Assis (1988) described a rural dialect in which popular usage strongly favors *a gente* over *nós*. The study deals with a rural community in the state of Minas Gerais, in central Brazil; the informants were all illiterate, of both genders and several age groups, who were born and lived in the area. The author reported that speakers favor what she analyzed as "indeterminate forms," even in contexts where first-person singular forms would be expected, as shown in (5).

(5) Entrev.: Dona Antônia, o que que a senhora faz em casa? Como é o seu trabalho? A.: Ah! *Eu* faço café, faço armoço, *a gente* lava vazia, óia menino, assim… (p. 76)

Interv.: Dona Antônia, what do you do at home? What is your work? A.: Oh! I prepare breakfast, lunch, *a gente (I)* wash pots, take care of children, like this…

Assis reported that in plural contexts, where *a gente* is equivalent to *nós*, there is a very strong preference for *a gente*, with 86% (65/76) usage. As we shall see from other studies, this is a very high rate of use. Most importantly, however, is that it comes from a rural, illiterate population, hence offering more evidence to support the idea that this change had a popular origin, in accordance with Schmitz's observations.

Schmitz's study is actually very useful to us for two other reasons. First, because the earliest grammar he reviewed is from the very beginning of the 20th century (1907), it suggests that the grammaticalization of *a gente* must have already been underway in popular speech in the 19th century. Second, by analyzing data from prose fiction, he showed that this usage was misrepresented or ignored by traditional grammarians. This is very clear in the following quotation: "Granted that *a gente* is not always preferred by Portuguese or Brazilians on formal occasions as in religious services, political speeches, and academic conferences, the word is, however, employed by *both educated and untutored* speakers of Portuguese in *semiformal and informal occasions*. It is not the province of only marginal elements of society as traditional grammarians tend to imply; *a gente* is used by *all social classes*" (Schmitz, 1973:640). Hence, although perhaps popular in origin,
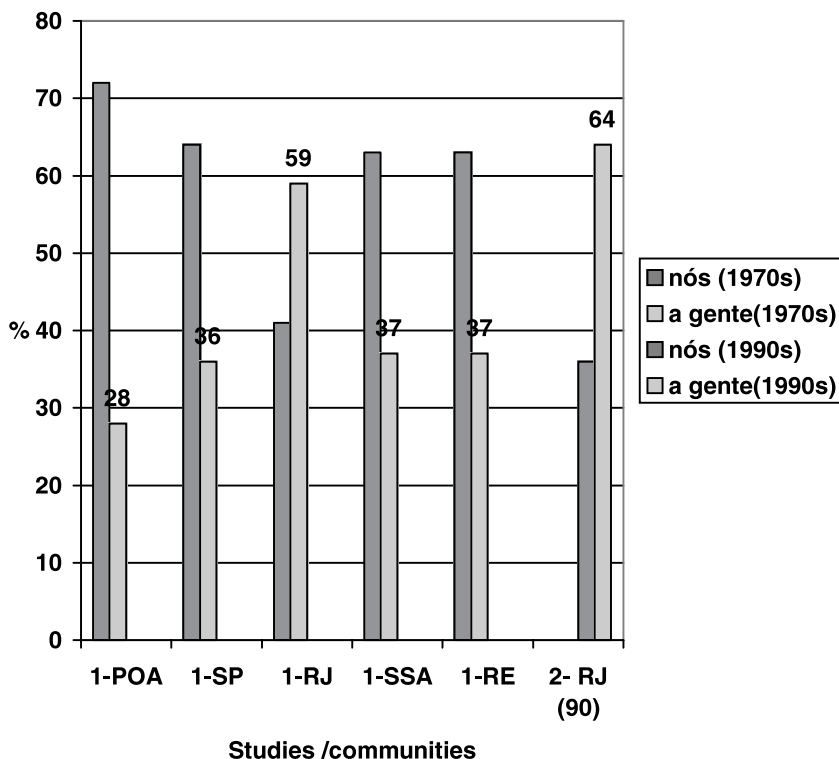
FIGURE 3. Percentages of *nós* and *a gente* (figures added) as used by the cultural elite in the 1970s and the 1990s. 1: POA: Porto Alegre (South); SP: São Paulo (Southeast), RJ: Rio de Janeiro (Southeast); SSA: Salvador (Northeast); and RE: Recife (Northeast) – data from NURC (1970), Leite & Callou (2002:54); 2: RJ: Rio de Janeiro (Southeast) – data from NURC (1990), Duarte (1996:505).

Schmitz explicitly argued that the form had penetrated all levels of Brazilian society by the dates covered in his study.

National distribution

Finally, to conclude our background review of the variable, it is interesting to note that the change seems to be going on all over the country, as becomes clear in the distribution of *nós* and *a gente* found by other researchers, shown in Figures 3 and 4.

In Figure 3, all the data are from the 1970s, except for the last two columns, on the right, which are from the 1990s. They all represent the speech of the cultural elite in the country, meaning people with post-secondary education. Starting from the left to the right, the first pair of columns refers to Porto Alegre, the southernmost state capital in the country. Note that in this report by Leite and Callou

(2002), Porto Alegre showed the lowest figure for *a gente* – only 28%, a result that the present analysis will contradict. One possible explanation for such a low rate may be that the authors considered the speech of just six persons per city, thereby perhaps misrepresenting the facts because of sample fluctuation. Next, in the Southeast, there is São Paulo, the largest city in Brazil, and probably 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%, thus higher than Porto Alegre. Next comes Rio de Janeiro, also in the Southeast, supposedly the leaders in this change, according to the graph by Leite and Callou, and the only city in their study that already had a majority of *a gente* use in the 1970s. Moving to the Northeast, Salvador and Recife display identical rates of *a gente*, very similar to São Paulo. Although there may be sample limitations in this study, it is important to recognize that it shows this change was going on in the 1970s, even in the speech of the cultural elite. In this respect, Schmitz's observation that this form was not confined to lower social classes is made clear in Figure 3. One last comment is made now, about the right-hand columns of Figure 3. They show a somewhat higher level of *a gente* in the 1990s, in Rio, in comparison to the 1970s: 64% versus 59%. Because the speakers in the 1990s were also interviewed in the 1970s (the "recontato" study), this difference, if it does not reflect sample differences, could mean that speakers continue to increase their usage of *a gente* in that community. At present, because the details of the two studies were not all available, it can only be treated as an open question: Have the individuals changed or are the samples discussed somewhat different?

Further progress of the change is suggested by several other studies with data from the 1980s and 1990s, showing still higher rates of use and implying that the change is both ongoing and becoming very advanced. This is what is shown in Figure 4, with data from three different regions in the country: Rio de Janeiro, in the Southeast (Omena & Braga, 1996), João Pessoa, in the Northeast (Fernandes, 1999), and Florianópolis, in the South (Seara, 2000). Here, in comparison with the data in Figure 3, the samples have been methodologically improved and are stratified by level of formal education, age group, and gender. The higher rate of *a gente* in João Pessoa may result from the fact that this is the only sample that includes illiterate speakers as well as people with formal education. According to Fernandes (1999:333), illiterate speakers are the ones who most favor its use. Nevertheless, what strikes us most are the similarities among the communities, again suggesting that this is a highly advanced change all over the country.

What is a gente?

The historical evidence presented earlier supports the view that this is an ongoing change, in fact, it appears very advanced all over the country wherever sociolinguistic studies have been done. The literature reviewed shows that it may have started in the 19th century. It is also clear that *a gente* received negative evaluation, because it was associated with people of lower social status. This suggests it is a change from below.
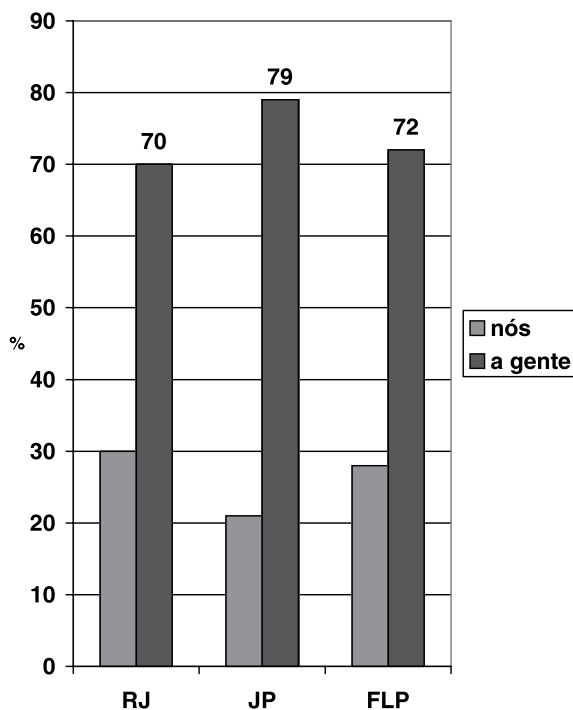
FIGURE 4. Percentages of *nós* and *a gente* as used in stratified samples from the 1980s and 1990s. RJ: Rio de Janeiro (Southeast) (Omena and Braga, 1996); JP: João Pessoa (Northeast) VALPB's corpus (Fernandes, 1999); FLP: Florianópolis (South) VARSUL's corpus (Seara, 2000).

In what follows, I will present the results of my study in three sections:

1. apparent time data from the 1990s,
2. a real-time panel study considering the speech of the same individuals in the 1970s and 1990s, and
3. a real-time trend study, comparing two groups of speakers, one from the 1970s, the other from the 1990s.

For this investigation I am using materials from the VARSUL database (a corpus of sociolinguistic interviews focusing on urban language variation in southern Brazil) and from the NURC database (a corpus of interviews collected in the 1970s 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, a large metropolitan center and capital of the state of Rio Grande do Sul, the southernmost state of Brazil.

Sample	Gender	Age groups	Level of education
39 Socioling. Interviews Porto Alegre, RS Recorded in the 1990s	Males – 19 Females – 20	Younger (25–49) – 17 Older (> than 50) – 22	Elementary – 10 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.

Apparent time, data from the 1990s

To develop the apparent time study, I analyzed the speech of 39 informants (as shown in Figure 5) stratified according to gender (males and females), age group (25–49 years old and above 50 years old), and level of education (elementary, intermediate, secondary, and post-secondary education). These are the criteria established in the VARSUL database.

The distribution of the speakers is not completely balanced in this sample because some of the interviews of the VARSUL database are still being transcribed or coded. But, in any case, each category is represented by more than the desirable minimum of five speakers, and the sample is quite proportional for each social-factor group. Data collection was done through automated search (using *Interpretador* ©Engesis), as well as careful reading of the transcriptions. The collection and coding of data was conducted with the assistance of several undergraduate students at UFRGS.

This analysis focuses only on subject forms. The two variants of the dependent variable 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%), as shown in Figure 4, but there is a difference between Porto Alegre and João Pessoa, where Fernandes found 79% usage of *a gente*. This may result from differences in sampling, because in João Pessoa, Fernandes included more informants from lower social levels, in contrast to the VARSUL sample.

First let us consider the results of the *linguistic factor groups*. Verbal agreement was one of the independent variables analyzed. Results show categorical use with *a gente* of the unmarked verb form (i.e., the historically third-singular
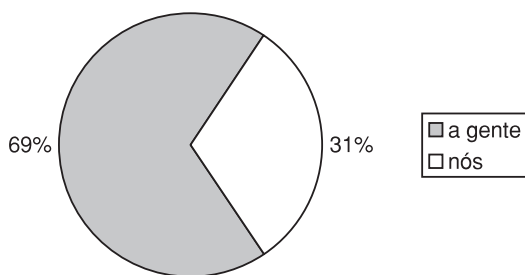
FIGURE 6. Percentage of *nós* and *a gente* in Porto Alegre, Brazil, in the 1990s (VARSUL database).

form), as in example (1) previously presented. There is only one token of *a gente* followed by a verb with first-person plural ending, spoken by a male who has been to high school and belongs to the younger age group. He said:

(6) 'e pra se entrá numa faculdade também não é fácil, né? então *a gente 'stamos aí*, né?'

The meaning of this is "it's not easy for someone to get into a faculty, so there I am" (literally: there the people are, using the verb *estar* + first-person plural inflection). This is a crystallized expression (possibly crystallized in comic TV shows) meaning "I'm ready/open for what may come." By saying this, he is making fun of the difficult situation, joking on the fact that he could not change it easily. The conclusion is twofold. First, this kind of agreement is not productive at all in the corpus analyzed here; second, the fact that the replacement of *nós* with *a gente* is so advanced (69%) that it clearly has the effect of dramatically reducing the usage of first-person plural verbal inflection. One should add to this reduction the fact that there is some variability in agreement when *nós* is chosen as the subject pronoun, with 6% of the verbs having the zero ending (which is equivalent to the third-person singular form) (87/1395 tokens). All of these tokens with the zero ending fall into two categories: (a) target verb forms[4] with antepenultimate stress, which are avoided by choosing a zero ending form with penultimate stress; (b) rare verb forms like the Future Subjunctive (e.g., *fizermos* "make") or the Inflected Infinitive (e.g., *fazermos* "make"), which are also avoided. The conclusion from this sample is that agreement is categorical with *a gente* and variable with *nós*. For this reason, the factor group "agreement" had to be excluded from further quantitative analysis.

I also investigated the effect of clause-level word order, thinking that *a gente*, being a new pronoun in the language, would be much more associated with subject-verb (SV) than with verb-subject (VS) order. I thought this might be the case because of the well-documented change in the language that has confined VS order to just 5% of sentences, these being mostly restricted to intransitive verbs (Zilles, 2000) or more specifically to inaccusative existential verbs (Coelho,

TABLE 1. *Significant linguistic factor groups for the use of* a gente *in Porto Alegre, Brazil, in the 1990s (Ns, percentages, and weights from VARSUL's data)*

	N/Total	%	Weight
Proximity of subject and verb			
SV	1076/1591	68	.48
SXV	246/330	75	.58
Reference			
Generic	717/931	77	.63
Referential	620/1013	61	.38
Subject in previous clause			
Nós	14/161	9	.02
Null + 1st p. pl. verb	12/52	23	.05
a gente	421/432	97	.88
Null $_{(coref)}$ + unmarked verb	77/105	73	.33
Totals	1337/1944	69	Input: .85

2000:89). Because these contexts also tend to disfavor personal pronouns as subjects (Zilles, 2000), it was tentatively hypothesized that only the old pronoun *nós* would appear in VS order. This was exactly what happened. There were no tokens of *a gente* in VS order; as for *nós*, there were only 6 tokens out of 1927 (0.3%). With this result in mind, I redefined this factor group to test only the proximity of the subject in relation to the verb in SV order, working with two factors. These factors are (1) the subject is adjacent to the verb, and (2) the subject is not adjacent to the verb. Intervening material ranges from clitics like *me, te, se*, to negation *não*, to adverbs like *sempre* "always" and *nunca* "never", to adverbs and adverbial phrases of different kinds, like *certamente* "certainly" and *com certeza* "for sure." This group was selected by VARBRUL as significant and so is included in Table 1, where I present the results for all the selected linguistic factor groups.

One reason to explore the proximity between subject and verb is the possibility that *a gente* may be changing to a free weak form or undergoing cliticization, with adjacency therefore being increasingly required. This, however, can only be done with the simultaneous analysis of the reductions of *a gente*, which I will only briefly discuss, on the basis of an analysis with a smaller sample (Zilles, 2002). In the present analysis, as shown in Table 1, this variable is examined with a different tentative purpose, that is, to test the effect of intervening material on the selection of *a gente*. The results show that *a gente* is favored when the subject is distant from the verb.

Nevertheless, this result is not necessarily definitive, because, for example, I have observed that there is twice as much use of *a gente* with negation. This suggests that the coding must be refined in order to identify the forces operating in these contexts and explain the significance of this group.

TABLE 2. *Use of* a 'ente *according to syntactic function and reference; Porto Alegre, Brazil (VARSUL database)*

Factor groups	N / Total	%	Weight
Syntactic position			
Subject	190/1117	17	.54
Prepositional Object	2/87	2	.13
Object	0/10	0	–
Reference			
Generic	108/775	14	.48
Reference group	79/477	17	.51
1st. p. s.	11/31	35	.74

Note: Compare with Zilles, 2002.

The second factor group in Table 1 is reference. In this case, we notice that generic use is still strongly associated with *a gente*. This is compatible with grammaticalization processes and corresponds to what Hopper (1991:22) called 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."

Anyway, the new meaning (*a gente* as used to make reference to specific entities, including the speaker) already corresponds to 61% of the tokens in this category. If we associate this result with the parallel reduction of *a gente* to *a'ente*, the semantic change appears more clearly. In the previous study of this reduction (Zilles, 2002:306), considering the speech of 32 of the present 39 informants, the general distribution showed that 15% (198 out of 1289 tokens) of reduced forms presented deletion of the fricative segment. The results for reference in this analysis were very interesting, in that the reduced form is significantly favored for referential meanings, as can be seen in Table 2, reproduced from Zilles (2002:307).

These results support the idea that this reduction is a new stage in the grammaticalization process. The old generic meaning is disfavored, while the actual referential pronominal meanings are favored.[5] The reduced form is mostly favored when reference is for first-person singular, but this result must be taken with caution, because the number of tokens is still too small to enable a firm conclusion at this point.

Table 2 also shows relevant results about the syntactic position of the reduced form. Out of 198 tokens of *a'ente*, 190 are in subject position[6] (Zilles, 2002:306). This is fully consistent with the constraint already observed for the reduced forms of *você*. Reinforcing this, there are only two tokens preceded by a preposition and none in object position. (Of course object pronouns, being a knockout, were not included in the VARBRUL analysis, but they are cited in Table 2 because of the importance of this constraint.)

These comments about the phonological reduction of *a gente* also show the complexities of studying a grammaticalization process in which several changes may happen simultaneously. These interrelated processes may affect the quantitative results of each other, but until now, the only way I could think of to deal with this problem is to analyze the speech of the same speakers for these several processes as several variable rules and then try to compose an integrated picture of them altogether. Hence, the present analysis is but a part of what needs to be done.

Now, returning to the apparent-time analysis, the last group of linguistic factors in Table 1 addresses the question of whether or not the token under consideration is affected by the formal realization of the subject of the preceding clause. The purpose of this factor group was to investigate whether there is a discourse-level perseveration effect, so that having produced one instance of, say, *a gente*, a speaker is more likely to continue using the same form in a sequence of clauses with the same subject, as opposed to switching to *nós* (which might possibly be interpreted by hearers as implying a different referent or a change in discourse grounding). Only those cases in which the preceding syntactic subject was either *nós* or *a gente* or their respective nulls were included, amounting to only one third of the data. The objective of verifying that the selection of *a gente* in the previous clause would favor its maintenance was attained. It has a very strong effect: 97%. Given that the same tendency appears with *nós*, these results were interpreted in terms of the general discourse principle of maintaining the reference to the same subject/topic for a number of clauses. Of course, the fact that the pronoun is repeated in two consecutive sentences has to be further investigated, because it could be the result of another general tendency in the language – to use only overt pronouns. In particular, I intend to investigate the syntactic relationship between the clauses involved, given that the repetition of *a gente* could be more favored in the subordinate clauses because of the new tendency in Brazilian Portuguese to use overt pronouns instead of nulls. According to Duarte (1996:506), "The evidence that we [the Brazilians] are moving in the direction of using overt pronouns is [evident] in their use in constructions where the null subject is obligatory in Romance Languages of the pro-drop group, like European Portuguese, Spanish and Italian. This is exemplified by structures... which exhibit subordinate clauses with correferential subjects." [Duarte's examples include sentences connected with *quando* "when," *porque* "because," *mesmo que* "even though," etc.]

So it turns out to be very important to carefully study the relationship between clauses and the selection of overt or null subjects. To do so, I have already started to include first-person plural null subjects in the analysis. Although the kinds of clauses are not specified yet, these preliminary results of a partial sample show interesting tendencies. As can be seen in Table 3, percentages are the reverse for overt and null subjects, suggesting that the introduction of *a gente* in the language also has a great impact by reducing the percentage of null subjects.

Another interesting aspect of the selection between *nós* and *a gente* may be described as a constraint on *a gente* according to which it cannot be combined with the plural quantifier *todos/todas* (all), unlike all the other pronouns with

TABLE 3. *Distribution of subject pronouns* nós *and* a gente, *and respective nulls*
(adapted from Aires & Zilles, 2002)

| | *Nós* | | *A gente* | | |
	N	%	N	%	Total
Overt pronouns	358	32	760	68	1.118
Nulls	327	62	200	38	527

plural reference, including the old pronoun *nós*, as well as the second- and third-person plural pronouns (*todos nós / todas nós; todos vocês / todas vocês; todos eles / todas elas*). Nevertheless, in the vernacular, *a gente* (as well as the other plural pronouns) can be combined with the invariable form *tudo*, as shown in examples (7) to (10).

> (7) era um carro aberto assim, ele [o presidente Getúlio Vargas] passava de pé abanando ainda pr*a gente*, a coisa mais linda. E *a gente tudo* com bandeirinha abanava pra ele. (RSPOA16, l. 1103–1104; 2 tokens of nós, 89 of *a gente*).
> ...it was an open car like that, he (the President) was standing waving to us, it was so beautiful. And we all with little flags would wave to him.
>
> (8) *nós* íamos *tudo* atrás dele (POA49, l. 898)   we would all go after him
> (9) *todos nós* somos médium (POA19, l. 1057)   all of us are mediums
> (10) *nós todos* somos assim (POA45, l. 1535)   we all are like that

This constraint is probably related to the fact that *a gente* is not syntactically plural but still has a collective, generic interpretation. This may be the reason why *nós* is also preferred when speakers use numerals in the predicate, as in examples (11) and (12), or when the quantifier *muitos* ("many") is used, as in (13).

> (11) Nós somos nove. (POA06, l. 413)   we are nine
> (12) Agora nós somos três irmãs. (POA49, l. 384)   now we are three sisters
> (13) como nós éramos muitos... (POA49, l. 258)   since we were many

This constraint is interpreted as another instance of the Principle of Persistence, whereby "traces of its original lexical meanings tend to adhere to it, and details of its lexical history may be reflected in constraints on its grammatical distribution" (Hopper, 1991:22). The word *gente*, as a noun, had a generic, mass or noncount meaning, which persists in the course of its grammaticalization. Nevertheless, its cooccurrence with *tudo* demonstrates its intrinsically plural reference, distinct from the singular pronouns, which cannot be so modified (*eu tudo, *você tudo).

The last linguistic characteristic to be mentioned is related to gender agreement. In (14) to (16) I present examples from the VARSUL corpus, which are evidence that agreement with *a gente* is governed by the gender of the referent(s), and not by the gender of the word *gente* anymore. In (14), a male speaker was

talking about his childhood and how children would offer their seats to adults in buses or streetcars. He said:

(14) *a gente*, por exemplo, era *guri*, *a gente*   we, for example, were children, we
vinha do colégio e coisa, entrava uma pessoa   were coming back from school and
idosa… (POA07, l. 468–470)   so on, an old person would come into
                                                  (the bus)…

In (14), the noun *guri* is a regional word for "boy"; it is morphologically unmarked for gender and used if males are included in the referents, in opposition to the morphologically feminine form *guria* "girl," to be chosen by a female speaker if all the referents she would have in mind were females. Other interesting examples involve adjectives (15) or passive constructions (16).

(15) (…) *a gente* está muito *cínico*, assim   *we* are very *cynical* (masculine),
muito na retaguarda. (RSPOA31, l. 900–905)   really kind of hanging back.

This example comes from an interview with a female speaker who is talking about the electoral campaign of the candidates of the Workers' Party. She uses the adjective *cínico* "cynical" in the unmarked masculine form, not the feminine form *cínica*, because she is referring to everybody who votes for this party. The next example, also from a female speaker, is a passive construction, which requires gender agreement in the participle form of the verb according to the gender of the subject.

(16) *a gente* foi criado assim   we have been raised (masculine) like
(POA06, l. 1229, female speaker)   that (female speaker referring to her
                                     family, which included male referents)

These examples, concerning a change in gender agreement, are thus a very important piece of evidence to attest to the integration of *a gente* into the pronominal system. The first- and second-person pronouns, because they can be used to refer to any speaker or hearer, do not have intrinsic gender; *a gente* had feminine gender when it was a noun, but in becoming a first-person pronoun, it lost this trait. These examples also give us a further sense of the complex integration of this change in the language.

Now that we have considered the most salient linguistic characteristics in the grammaticalization of *a gente*, it is clear not only that it is strongly associated with other changes, but also that it involves several less visible, internal processes that are semantic, syntactic, and phonological. Nevertheless, it is not clear yet if they all happen in the same way and at the same time in different communities where the grammaticalization is underway. To answer this question, these processes must be studied in their social context. One may conjecture that even if the "same" change (e.g., the use of *a gente*) is going on in several communities at the same time, in each of them the social evaluation it receives may be somewhat different, and this may delay or advance the process. So I turn now to the discussion of the social embedding of *a gente* in one community, hoping that it will

TABLE 4. *Significant social factor groups for the use of* a gente *in Porto Alegre, Brazil, in the 1990s (Ns, percentages, and weights from VARSUL's data)*

	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
Level of Education			
Elementary	128/238	54	.33
Intermediate	411/559	74	.55
Secondary	315/425	74	.50
Postsecondary	483/722	67	.52
Totals	1337/1944	69	Input: .85

illuminate the question of what is the role of the speakers in a grammaticalization process. To do so, I present first the results of the social factor groups in the apparent-time study. Table 4 contains the significant social factors for the use of *a gente* in the apparent-time study with data from the 1990s.

Results for the social variables in Table 4 support the idea that this is a change in progress led by female speakers (note the difference in weight, females favor the use of *a gente* much more, with .55, than males, at only .41). Results for age groups also indicate that this is a change in progress, because younger speakers lead, with a very robust difference in weight: .66 as opposed to only .42. Note also that the input probability is very high (.85), indicating this change is already well advanced. Nevertheless, the results for level of education are somewhat puzzling, because less-educated speakers in our sample disfavor the new form (contradicting the idea that it is a change from below), whereas the others not only favor it, but seem to be very uniform in this respect.

If we look at the weights and percentages for the three higher educational levels in Table 4 (intermediate, secondary, and postsecondary education), we see that there is not much difference between them. The chi-square test showed that they were not significantly different, and therefore should be amalgamated. But instead of simply doing this, I investigated the possibility of interaction between gender and level of education by doing a cross-tabulation, as shown in Figure 7.

Results show that, in fact, there is interaction between gender and level of education: the line for females is fairly flat, whereas for males it is quite bumpy. It is also evident that the difference between the two genders is neutralized at the higher level of education, as women show a small reduction at this level. Recalling that the difference between males and females in the three upper levels of education was not significant, it is the behavior of males with the lower level of education that strikes us as different. They seem to resist or avoid the use of *a*
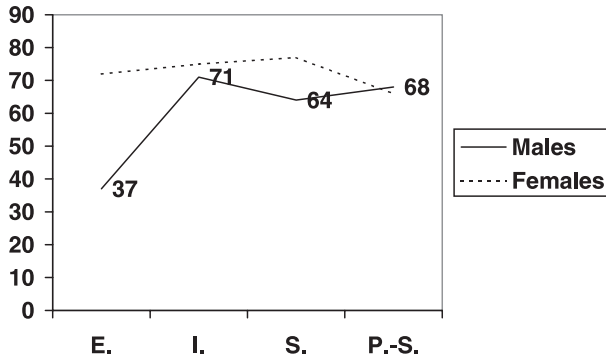
FIGURE 7. Gender versus level of education in the use of *a gente* in the 1990s (Porto Alegre, VARSUL database).

gente, with a usage rate of only 37%. Although this may be a result of sampling error, and needs further investigation, I can also speculate that this might be an issue of identity. If the use of *a gente* is perceived by this group as a female feature, males in this lower social level may wish to avoid it to present themselves as different from females. This result is consistent with findings in other sociolinguistic studies; for example, Guy et al. (1986) found exactly this pattern, of low-status males lagging markedly in a female-led change. In addition to further investigation of speakers from the lower social levels, a study of attitudes could probably be of great help to explain these results.

Now, to understand the inversion between males and females with post-secondary education, a parallel analysis was made, this time considering only this higher social group. The analysis included 20 speakers, 10 males and 10 females, subdivided in three age groups: four young adults (25–49), eight older adults (50–69), and 8 seniors (in their 70s or older). Note that this last group was not included in the apparent time study to avoid bias towards having a lot more people from the higher social level in the sample. A cross-tabulation of the use of *a gente* by the speakers in this new sample also showed interaction between gender and level of education, although we can say that to a certain extent they are all participating in this change, as can be seen in Figure 8.

According to Figure 8, females are behind males in the two extremes, but only the oldest female speakers seem to resist or avoid the use of *a gente*, showing less than 50% of the new form. Despite a general tendency towards increasing the use of *a gente*, younger males are the ones to show near-completion, whereas women seem to lag behind, with a percentage that is quite similar to a "national" average (some point between 70% and 80%). One possible explanation for this difference between males and females in the younger group could be that, perceiving the situation as stable and the change as completed, males of the younger generation could now be leading the reduction of *a gente*, marking themselves as different from women. Results in this direction of variation on top of change have
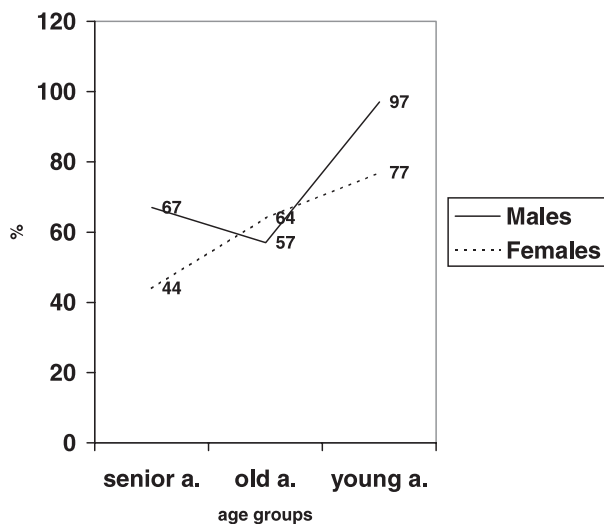
FIGURE 8. Gender versus age in the use of *a gente* by speakers with postsecondary education in Porto Alegre, Brazil, in the 1990s (VARSUL database).

already been mentioned, and further investigation in the community is underway. Another possibility is that females at this level of education could be diminishing their use of *a gente* because people are now showing some overt social awareness of it, and negative evaluations have appeared in newspapers, TV shows, and schools.

Considering again the results for the larger sample, it is worth noting that the VARSUL corpus only includes educated speakers with at least four to five years of formal education, so the lower social levels are probably underrepresented in this analysis. To understand the behavior of the male speakers having the lower level of education, it would be crucial to have more data from lower social groups. This would also help to answer the question as to whether in this community this is a change from below. Despite these limitations, evidence presented here does not seem to support the idea that this is a change from above, because speakers with postsecondary education show lower overall rates of *a gente*, except for the younger speakers.

Although the results of the apparent-time study with the larger and the reduced samples point in the direction of a change in progress by detecting a differentiation of generations, one cannot say for sure whether this is a case of age-grading without change (individuals are unstable, but the community is stable), or a case of generational change in which individual speakers are stable throughout their lifetimes, but each generation increases the use of the variable (Labov, 1994:83–85). To disentangle these alternatives, a panel study can tell us whether individuals are stable or changing. This is our focus in the next section.

TABLE 5. *Distribution of* nós *and* a gente *in the speech of 13 speakers recorded in the 1970s and in the 1990s (Panel study)*

Speakers (gender/age in 1970)	1970		1990		χ^2 calc.	$p < .05$ Significant?
	N/Total	%	N/Total	%		
a (M, 28)	1/1	100	36/53	68	(n.a.)	
b (F, 41)	78/97	80	107/130	82	0.03	
c (F, 46)	7/10	70	10/43	23	6.13	Yes
d (F, 48)	26/31	84	57/95	60	4.91	Yes
e (M, 29)	28/42	67	28/43	65	0.01	
f (M, 41)	27/27	100	28/33	85	2.67	
g (M, 47)	14/20	70	13/18	72	0.04	
h (M, 45)	32/40	80	26/43	60	2.89	
i (M, 34)	20/40	50	22/56	39	0.69	
j (M, 46)	39/51	76	44/63	70	0.33	
k (F, 39)	55/87	63	65/100	65	0.01	
l (F, 28)	5/20	25	15/32	47	1.65	
m (F, 59)	23/58	40	6/29	21	2.33	

Real-time panel study

In this real-time panel study only speakers with postsecondary education are considered, because these are the only ones for whom data from an early time period (the 1970s) are available at present.[7]

In this analysis, I compare the speech of the same 13 speakers as recorded in the 1970s and in the 1990s. There are seven males and six females in this sample. The chi-square tests in the last two columns in Table 5 compare, for each person, their usage of *a gente* in the 1970s with their usage in the 1990s, to find out whether they had remained stable or shown any statistically significant change in their behavior.

There were two people in the panel who exhibited significant changes: two elderly women who used significantly less *a gente* in the 1990s – speaker "c" (POA44), who is 73 years old, and speaker "d" (POA49), who is 75 years old. But note that they move *against* the historical direction of this change, towards less, not more, use of *a gente*. Because they *decrease* their use of the new pronoun, they do not challenge the conception that individuals in this panel study are stable in respect to this change. They are not following the community, which is using more and more *a gente*.

The same question can be addressed in another way, by plotting 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 would conclude that speakers are generally stable in their usage. Such a test is shown in Figure 9, for 12 of the speakers (one person in the panel had only one token in the 1970s corpus, which did not permit a meaningful quantitative analysis). For the most part, the distribution of

Autocorrelation for 'a gente'
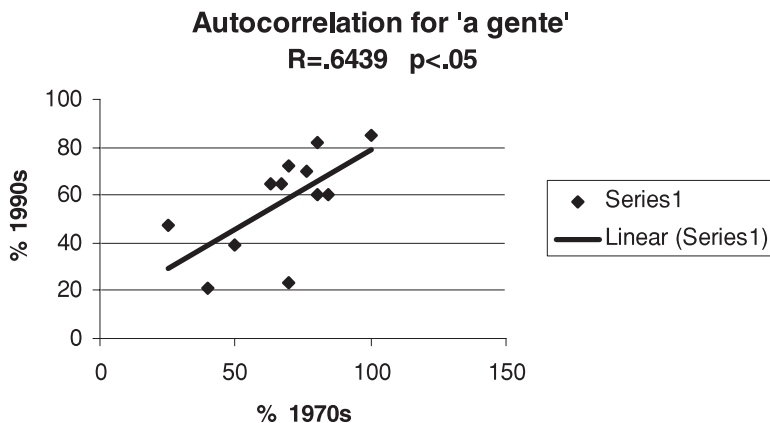### R=.6439   p<.05

FIGURE 9. Autocorrelation of speakers' *a gente* usage in the 1970s and 1990s.

the speaker points is systematically linear: most speakers have similar values on both axes.

The appropriate statistical measure of stability in this approach is the *r*-correlation statistic (in this case, correlating an individual's usage of *a gente* in the 1970s with his or her usage in the 1990s). The *r* value obtained for this data was $r = .644$, which for 12 speakers is significant well beyond the .05 level, meaning, in this case, that the data support the hypothesis that a person's rate of use of *a gente* in the 1970s is closely associated with their own rate in the 1990s.

A more detailed consideration of the speakers in Figure 9 is also revealing. In general, all the speakers are plotted very close to the trend line, except for a clear outlier corresponding to the point on the lower right side of the graph, a speaker who reduced her use of *a gente* from 70% (7 out of 10 tokens) to 23% (10 out of 43 tokens). As previously shown, this person – "c" in Table 5 – significantly reduced her use of *a gente*. The next lower point corresponds to speaker "m" (POA46), who also reduced her usage of *a gente* in the 1990s, though the difference was not significant. This case is particularly interesting, because this person overtly condemns the use of *a gente* in the very beginning of the interview and avoids it until the last quarter of the recording.

The other speaker who had a significant difference in the chi-square test is speaker "d" (POA49), who nevertheless was placed among those who always use *a gente* more than 50% of the time. This shows that even if she reduced the use of the form from one interview to the other, she is not very different from the rest of the speakers in this group.

In general, Figure 9 shows two groups of speakers, both rather stable in terms of the values on both axes: three of them always use *a gente* 50% or less of the time (plotted in the lower area) while the other eight always use *a gente* more than 50% of the time (shown in the upper area). Considering the fact that the great majority of the speakers do not change in their use of *a gente*, at least across the

TABLE 6. *Significant social factor groups in the use of* a gente, *trend study of two samples (1970s and 1990s) analyzed together (NURC and VARSUL databases)*

	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

25-year age span investigated here, and by recalling the generational differences presented earlier, we may say there is strong evidence to support the interpretation that this is, in fact, a generational change, an ongoing alteration of the grammar of the community across time. In other words, it seems likely that the age differences observed in apparent time are a reflection of a generational change, not an age-grading pattern that each individual will retrace in the course of their lifetime.

Real-time trend study

Although in the apparent-time study our speakers with postsecondary education lagged somewhat behind the others, it is clear that they are indeed participating in this change. The panel study then showed that most of them are quite stable from the 1970s to the 1990s. Now, to make sure this is a generational change (individuals do not change, but each new generation increases the use of the variable), a trend study is needed. In this methodology the researcher conducts an analysis of two comparable samples from the same community, collected at two different times, to illuminate the scene. This is our last step here.

The sample for this analysis is made up of 36 speakers, 18 males and 18 females, divided into two age groups: younger (25–44 years old) and older speakers (45–69 years old). All of them have secondary or postsecondary education; 20 were recorded in the 1970s, 16 in the 1990s.

In this analysis *a gente* occurred with a frequency of 65% (991 out of 1533 tokens). The linguistic factor groups that were significant here are the same as those in the apparent-time study: proximity between subject and verb, subject in the previous clause, and reference. Their effect is less strong, but goes in the same direction, so in what follows, I concentrate on the social factor groups whose results are shown in Table 6.

The results in Table 6 are very informative. Two factor groups show very strong effects: age group and decade. According to the weights (note the internal difference of 32 and 36 points!), younger speakers and the 1990s are highly favorable factors to the use of *a gente*. This evidence, together with the results of the panel study, supports the interpretation that this change *is* generational. As for gender, once more females appear in the lead, but the effect is smaller than for the other groups, in that males follow females closely.

Besides this, another conclusion emerges from Table 6. Considering just the data from the 1970s, it is also possible to demonstrate that the general distribution in Porto Alegre is not as described by Leite & Callou (2002), presented in Figure 3. According to them, there was only 28% usage of *a gente* in Porto Alegre in the 1970s, but Table 6 shows that it is much more similar to what has been observed in the speech of the cultural elite in other communities, where there is 56% usage of *a gente* (403/721). This huge difference in results (twice as much use of *a gente* as Leite & Callou have reported) is probably explained by differences in sampling: they analyzed just 6 speakers in each community, whereas the sample here includes 20 speakers.

Finally, I conducted another VARBRUL run considering just the data from the 1970s (20 informants) to check for age differences then. Results show there is apparent-time gradation in the 1970s, with a very robust difference: young speakers use 67% (357/529 tokens) of *a gente*, whereas older speakers use only 24% (46/192 tokens). This difference supports the idea that this change was accelerated in the 1970s, in conformity to what has been shown in Figure 2 for theatrical plays. Note also that the age difference is much smaller in the apparent-time study of the 1990s, where younger speakers show 78% usage (480/618 tokens), whereas older speakers show as much as 65% usage (857/1326) of *a gente*. All of these findings, together with the results from the panel study, consistently support the hypothesis that this is a generational change.

CONCLUSIONS

The social distribution of *a gente* in this sample from Porto Alegre can be summarized as follows.

1. Concerning the education-level distribution, results showed a fairly flat curve for women, and a bumpy curve for men, in which the men with the lowest levels of education used *a gente* the least. How does this fit with Labov's model of change types? First, it is noteworthy that both male and female speakers separately, and the pooled data as a whole, show a peak rate of use at the middle educational levels, suggestive of change from below. The gender split at the bottom end of the educational scale (where there is a marked decline from the peak for male usage), is consistent with a number of other studies of changes from below, such as Guy et al. (1986). This may indicate that males with the lowest level of education are resisting the use of *a gente*, perhaps perceiving the new pronoun as a feminine trait (remember this change is led by women) or monitor-

ing their speech during the interview (thus being more formal and using more *nós*). In any case, it is clear that the data do not show the marked social stratification, with peak usage in the highest status groups, associated with Labovian changes from above like New York City (r). So it looks reasonable to reject the change from above scenario for these data.

However, the results in Figure 7 also do not show the pronounced peak in the middle that would be expected in a Labovian change from below. Why should that be the case? One possible reason is that we are using level of education to infer social status, which may be inappropriate or inaccurate. Labov's scales of social class also take into account a speaker's occupation and income. Another possibility is that the lower end of the social class scale is underrepresented in our sample, thus the expected patterns do not come out clearly (for example, all of our subjects had at least a primary level of formal education, whereas Brazilian society at large includes many unschooled and illiterate individuals). In either case, it would mean that our sample is not entirely representative, so we should avoid drawing conclusions without looking at other social characteristics, as well. A third possibility is that changes from below do not absolutely follow the Labovian model; Kroch (1978), for example, proposed a model in which peak usage may be found lower down the social scale. Again, sample limitations seem to preclude conclusions. In contrast, one could claim that the reason for this mismatch between results and theoretical models is that this case involves a special kind of change – grammaticalization – that is not socially motivated. To deal with this question, we must first examine the other results in relation to the other social criteria under consideration.

2. Concerning age groups, younger speakers always use more *a gente* than older ones, which is consistent with both change from above and change from below. The main difference between Labov's two types lies in whether the peak usage occurs in young adults or in adolescents. Unfortunately, our corpus does not yet include any adolescents, so we cannot distinguish between the two models on the basis of the available data (although impressionistic observation suggests that this age group uses *a gente* and also its reduced form even more than adults), suggesting change from below.

3. Concerning gender, in general we found that women lead the process, except in the group with the highest level of education, where gender difference is neutralized. Investigating this group in more detail, we found that there is also interaction with age, with senior and younger women using less *a gente* than men. We also found that one female speaker in this senior group clearly condemns the use of *a gente* and tries to avoid it during most of her interview. These results, then, are consistent with change from below and suggest that a negative evaluation of the new form is emerging, in accordance with the behavior of highly educated women in this sample. Gender differences, and this overt negative social evaluation of the new form, then, contradict the idea that there is no social motivation and embedding for this grammaticalization process.

4. To say that this is a change from below, we must also consider the question of the reversibility of the change. Examining data from the 1970s and 1990s, we

observed that the individuals are stable, but each new generation is increasing the rate of *a gente*, so there seems to be no evidence of reversibility. However, the reduced rate identified for the younger female speakers with the highest level of education could suggest that this group may be beginning to withdraw from using the form, but this is just a conjecture. Reversibility seems out of the question for two other reasons: One is the evidence of reduction to *a'ente*, suggesting a new stage in the grammaticalization process; the other is the connection between this process and other changes going on in the language, related to loss of verbal agreement and null subjects.

5. Concerning stylistic variation, there is no direct evidence available, but we have no reasons to believe *a gente* is used *more* in *more careful styles*; if anything, prescriptivist evidence indicates that the historical form *nós* would be favored in more careful styles, as it is in writing (despite data from prose fiction collected by Schmitz). Indirectly related to this, but probably contributing to the increment of *a gente*, there is the social stigma on lack of verbal agreement. Agreement is a common issue in Brazil. For example, teachers who work in public schools very often mention it as the worst "problem" of their students, and people in general talk about it. In one of our interviews a woman who is 68 and has elementary education also talks about it as the first aspect in speaking the language correctly.

(17) E: [O que que a senhora]- o que que pra senhora assim é falar português corretamente?
I: [What is it for you]- what is it for you to speak Portuguese correctly?

F: A concordância.
S: Agreement.

E: A concordância?
I: Agreement?

F: A concordância, né? É muito importante, né? (POA16, l. 1127–1132)
S: Agreement, right? It's very important, isn't it?

In this respect, *a gente* provides a safe way of avoiding the heavy stigma associated with omitted agreement, that is, given the choice between making a mistake in agreement and using a nonstandard but generalized new pronoun, people prefer the second option. Therefore, this is evidence that there is social motivation for this change in progress. In fact, since the second half of the 19th century, with the imposition of the standard written language, lack of agreement has been socially constructed/represented as something really bad (typical of uneducated lower social classes) that should be avoided at any cost! So ideology has been forging a favorable context for the use of *a gente*.

So, to sum up, there is no evidence that this is a change from above. Indeed, what external prestige dialect would it be coming from if it were? This is difficult to say, because Brazilians defer to no one in the Portuguese-speaking world (in fact, they think the Portuguese of Portugal sounds strange or funny), and Porto-Alegrenses apparently defer to no one in Brazil. Besides this, in our sample the use of *a gente* does not have a peak in the speech of the highest status informants, and there is no stylistic favoring of the innovation. We conclude, rather, that *a gente* is a spontaneous innovation that has emerged from within the speech community.

Overall, then, our results are more consistent with the Labovian change from below. The sociohistorical evidence cited from other studies suggests this, and the little differentiation by level of education that does appear in our data is more consistent with this model than with change from above. But we have not found strong evidence in the educational distribution of the Labovian class pattern with a peak in the middle of the scale. Further research on adolescents, different speech styles, speakers of lower social levels, and a more accurate social categorization of the informants are important steps to understanding this change. Nevertheless, one thing that *is* very clear in our results is that *a gente* is strongly embedded in the linguistic system, tied in with several other changes in the language. In fact, the grammaticalization of *a gente* is itself a whole set of interrelated changes, which makes it difficult to study each separate step in isolation, and which may have the effect, at least in part, of masking the social embedding. There are a number of linguistic processes going on at the same time that all promote the expansion of *a gente*. These include the long-term syntactic drift away from agreement marking and towards use of overt, preverbal subjects, possibly cliticized subjects; the semantic shift in the meaning of the form; the phonological reduction of the form; and so on. We may not know where these processes will end, but we certainly can tell what direction they have taken over the last 30 years, and where they are heading at present. As for the future, *a gente vai ver* – we (both generic and specific) will see!

NOTES

1. For example, in Guy (1990) there is a discussion of the sociolinguistic types of change.
2. See the section "What is *a gente*?" for details about this database.
3. POA is an abbreviation for Porto Alegre, the city where data in this analysis have been collected. See the section "What is *a gente*?" for more information. The number following POA, 02, identifies the informant and is followed by an indication of the line(s) in which the token occurred.
4. See examples in the last paragraph of the section titled "Parallels with 'você'".
5. When contrasting *nós* and *a gente*, results for reference are the opposite: *a gente* is favored with generic meaning (529/676, 78%, weight of .66) and disfavored with specific, referential meanings (512/807, 63%, weight of .37).
6. Six tokens were excluded from this analysis of syntactic function for other reasons.
7. This limitation refers to Porto Alegre, the community studied in this article. See Lopes (2003) and Omena (2003) for real-time studies with data from Rio de Janeiro.

REFERENCES

Aires, Kátia M. L., & Zilles, Ana M. S. (2002). *Referência à 1ª pessoa do plural em Flores da Cunha e Porto Alegre*. Porto Alegre: Paper presented at the XIII Salão de Iniciação Científica da UFRGS.
Assis, Rosa M. (1988). Variações lingüísticas e suas implicações no ensino do vernáculo: Uma abordagem sociolingüística. *Ilha do Desterro* 20:59–81.
Azevedo, Artur. (1967). Plebiscito. In Rebelo, Marques (ed.), *Antologia Escolar Brasileira*. (1st ed.). Rio de Janeiro: MEC-FENAME. 178–180. (originally published in 1893, in *Contos fora da moda*)
Blondeau, Hélène. (2001). Real-time changes in the paradigm of personal pronouns in Montreal French. *Journal of Sociolinguistics* 5(4):453–474.
Borges, Paulo R. S. (2001). *Variação e mudança envolvendo o uso de "a gente" em dez peças de teatro escritas por autores gaúchos: Uma viagem de 1896 até 1998*. Unpublished paper.
Bortoni-Ricardo, Stella Maris. (1985). *The urbanization of rural dialect speakers: A sociolinguistic study in Brazil*. Cambridge: Cambridge University Press.

Castilho, Ataliba T. de. (1997). A gramaticalização. *Estudos lingüísticos e literários* 19:25– 64.

Coelho, Izete L. (2000). Sobre a natureza do verbo monoargumental na ordem V PN. *Organon* 14(28/29):75–93.

Croft, William. (1990). *Typology and universals*. Cambridge: Cambridge University Press.

Diewald, Gabriele. (2002). A model for relevant types of contexts in grammaticalization. In I. Wischer & G. Diewald (eds.), *New reflections on grammaticalization*. Amsterdam: Benjamins. 103–120.

Diewald, Gabriele, & Wischer, Ilse. (2002). Introduction. In I. Wischer & G. Diewald (eds.), *New reflections on grammaticalization*. Amsterdam: Benjamins. ix–xiv.

Duarte, Maria Eugênia L. (1996). Aspectos do sistema pronominal do português falado nas regiões Sudeste e Centro-Oeste. In *Anais do XI Encontro Nacional da ANPOLL*, João Pessoa, Brazil. 504–509.

Duarte, Maria Eugênia L. (2000). O português do Brasil no contexto das línguas românicas. *Actes du XXIIᵉ Congrès International de Linguistique et Philologie Romanes. Vol. 2: Les nouvelles ambitions de la linguistique diachronique*. Tübingen: Niemeyer.149–156.

Faraco, Carlos A. (1996). O tratamento *você* em português: Uma abordagem histórica. *Fragmenta* 13:51–82

Fernandes, Eliene A. (1999). Nós/a gente: Variação estável ou mudança em progresso? In Maria Elias Soares & Maria do Socorro S. de Aragão (eds.), *Anais da XVI Jornada de Estudos Lingüísticos*. Fortaleza: UFC/GELNE. 331–334.

Guy, Gregory R. (1981). *Linguistic variation in Brazilian Portuguese: Aspects of the phonology, syntax, and language history*. Doctoral dissertation, University of Pennsylvania, Philadelphia.

Guy, Gregory R. (1990). The sociolinguistic types of language change. *Diachronica* 7(1):47– 67.

Guy, Gregory R., Horvath, Barbara, Vonwiller, Julia, Daisley, Elaine, & Rogers, Inge. (1986). An intonational change in progress in Australian English. *Language in Society* 15:23–52.

Heine, B. (2002). On the role of context in grammaticalization. In I. Wischer & G. Diewald (eds.), *New reflections on grammaticalization*. Amsterdam: Benjamins. 83–101.

Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

Heine, B. & Reh, M. (1984). *Grammatical categories in African languages*. Hamburg: Helmut Buske.

Hopper, Paul. (1991). On some principles of grammaticalization. In E. Traugott & B. Heine (eds.), *Approaches to grammaticalization* (Vol. 1). Amsterdam: Benjamins. 17–35.

Hopper, Paul & Traugott, Elizabeth. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.

Kato, Mary A. (1999). Strong and weak pronominals in Brazilian Portuguese. *Probus* 11(1):l–37.

Kroch, Anthony. (1978). Toward a theory of social dialect variation. *Language in Society* 7:17–36.

Labov, William. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

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.

Lopes, Célia R. dos S. (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*. São Paulo: Humanitas Publicações / FFLCH/USP. 127–148.

Lopes, Célia R. dos S. (2003) *A inserção de 'a gente' no quadro pronominal do português*. Madrid: Iberoamericana.

Meillet, Antoine. (1912). L'évolution des formes grammaticales. In *Linguistique historique et linguistique générale*. Paris: Champion. 130–148.

Menon, Odete P. da S. (1996). *A gente*: Um processo de gramaticalização. *Estudos lingüísticos* 25:622– 628.

Omena, Nelize P. de. (1996a). A referência à primeira pessoa do plural. In G. M. de Oliveira e Silva & M. M. P. Scherre (eds.), *Padrões sociolingüísticos: Análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro*. Rio de Janeiro: Tempo Brasileiro, Depto. de Lingüística e Filologia, UFRJ. 183–216.

Omena, Nelize P. de. (1996b). As influências sociais na variação entre *nós* e *a gente* na função de sujeito. In G. M. de Oliveira e Silva & M. M. P. Scherre (eds.), *Padrões sociolingüísticos: Análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro*. Rio de Janeiro: Tempo Brasileiro, Depto. de Lingüística e Filologia, UFRJ. 309–323.

Omena, Nelize P. de. (2003). A referência à primeira pessoa do plural: Variação ou mudança? In M. C. Paiva & M. E. L. Duarte (eds.), *Mudança lingüística em tempo real*. Rio de Janeiro: Contra Capa Livraria Ltda. 83–60.

Omena, Nelize P. de, & Braga, Maria Luiza. (1996). *A gente* está se gramaticalizando? In A. T. Macedo et al. (eds.), *Variação e discurso*. Rio de Janeiro: Tempo Brasileiro.75–83.

Ramos, Jânia. M. (1997). O uso das formas você, ocê e cê no dialeto mineiro. In D. da Hora (ed.), *Diversidade lingüística no Brasil*. João Pessoa, Brazil: Idéia. 43–60.

Romaine, Suzanne, & Lange, Deborah. (1998). The use of *like* as a marker of reported speech and thought: A case of grammaticalization in progress. In J. Cheshire & P. Trudgill (eds.), *The sociolinguistics reader, Vol. 2: Gender and discourse*. London: Arnold. 240–277.

Schmitz, John R. (1973). The linguistic flexibility of '*a gente*' in Portuguese. *Hispania* (September):639–644.

Seara, Izabel C. (2000). A variação do sujeito *nós* e *a gente* na fala florianopolitana. *Organon* 14(28/29):179–194.

Serrano, Maria José. (1996). Accounting for morpho-syntactic change in Spanish: The Present Perfect case. *University of Pennsylvania Working Papers in Linguistics* 3(1):51–61.

Torres Cacoullos, Rena. (2001). From lexical to grammatical to social meaning. *Language in Society* 30:443–478.

Traugott, Elizabeth. (1989). From less to more situated in language: The unidirectionality of semantic change. In S. Adamson et al. (eds.), *Papers from the Fifth International Conference on English Historical Linguistics*. Amsterdam: Benjamins. 497–517.

Vitral, Lorenzo. (1996). A forma CÊ e a noção de gramaticalização. *Revista de Estudos Lingüísticos* 4(10):115–124.

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.

Zilles, Ana M. S. (2002). Grammaticalization of 'a gente' in Brazilian Portuguese. *University of Pennsylvania Working Papers in Linguistics* 8(3):297–310.

Zilles, Ana M. S., Maya, Leonardo Z., & Silva, Karine Q. (2000). A concordância verbal com a primeira pessoa do plural em Panambi e Porto Alegre, RS. *Organon* 14(28/29):195–219.

Zilles, Ana M. S., & Mazzoca, Patricia da R. (2000). *O papel dos fatores sociais no processo de gramaticalização de 'a gente'*. Paper presented at the IV Círculo de Estudos Lingüísticos do Sul, Universidade Federal do Paraná, Curitiba, Brazil.