



Open Library of Humanities

## ***As pessoas*, a quasi-pronominal human impersonal construction in European Portuguese?**

**Pekka Posio**, University of Helsinki, Finland, FI, [pekka.posio@helsinki.fi](mailto:pekka.posio@helsinki.fi)

---

The grammaticalization of nouns meaning ‘man’ or ‘person’ into impersonal pronouns is a well-documented phenomenon found across the world. In Europe, such impersonal pronouns are characteristic of languages with obligatory subject expression like Germanic languages and French (Siewierska, 2011). However, European (EP) and Brazilian Portuguese (BP) also display impersonal uses of noun phrases with the lexeme *peessoa* ‘person’ as recently discussed by Posio (2017, 2021) and Amaral and Mihatsch (2019, forthcoming). The present paper analyses the use of the plural form *as pessoas* ‘the people’ in a sociolinguistic interview corpus of spoken EP, comparing it with similar data from Peninsular Spanish (Fernández Juncal, 2005) and with dialectal data from the CORDIAL-SIN corpus (Martins, 2000–). In terms of frequency, *as pessoas* occurs over 30 times per 10,000 words in the EP interview data, in contrast to just over 3 occurrences in the CORDIAL-SIN and less than one occurrence of the cognate *las personas* in the Spanish data. In addition to discussing the high token frequency and other properties associated with grammaticalization, the paper examines the referential and quantificational properties of *a gente* and explores the hypothesis that *as pessoas* might be a sociolinguistically conditioned alternative for the impersonal pronoun *a gente* ‘people; us’, considered ‘popular’ in normative EP. However, this hypothesis is not supported by the data: while sociolinguistic factors do affect the frequency of use of *a gente* and *as pessoas*, the two forms are functionally different in that *a gente* typically includes the speaker in its referential range and *as pessoas* implies a speaker-exclusive reading.

---



## 1. Introduction

The grammaticalization of nouns meaning ‘man’ or ‘person’ into impersonal referential devices known as *man*-impersonals is a pattern documented in many languages (see, e.g., Giacalone Ramat & Sansò, 2007; Siewierska, 2011). While the existence of such overt impersonal pronouns has been related to non-pro-drop properties of the languages where they are found (Siewierska, 2011, p. 80), similar developments based on the lexeme *pessoa* ‘person, human being’ are also found in Portuguese and have recently been discussed by Posio (2017, 2021) and Amaral and Mihatsch (2019, forthcoming), who have mostly focused on the singular forms of the noun, i.e., the noun phrases *uma pessoa* ‘a person’ and *a pessoa* ‘the person’. These authors suggest that the grammaticalization of *a/uma pessoa* resembles the development of the noun phrase *a gente* ‘the people’ into an impersonal referential device and further into a first-person plural pronoun. The grammaticalization of *a gente* has been documented extensively in Brazilian Portuguese (e.g., Lopes, 2003; Zilles, 2005) and, to a lesser extent, in European Portuguese (see, e.g., Pereira, 2003; Posio & Vilkuna, 2013).

Although considered by Amaral and Mihatsch (2019) as a part of the larger family of constructions including other grammaticalizing constructions like *o pessoal* literally ‘the personnel’ and *o povo* ‘the people, the folk’, the plural form *as pessoas* ‘the people’ has received considerably less attention in previous research than the singular forms. In colloquial, spoken European Portuguese, *as pessoas* is a very frequently used item that seems to have developed some special properties that may distinguish it from regular uses of lexical noun phrases. At first sight, the most striking property of *as pessoas* is its usage frequency: in the data analysed for the current paper (see Section 2), it is used by speakers of European Portuguese ca. 30 times per 10,000 words, i.e., 30 times more frequently than its cognate *las personas* in a comparable corpus of spoken Peninsular Spanish (see Section 2). Such a high frequency could be indicative of a grammaticalization process towards a human impersonal pronoun. I will adopt the definition of impersonalization provided by Gast and van der Auwera (2013, p. 124) as “[t]he process of filling in a predicate with a variable ranging over sets of human participants without establishing a referential link to any entity from the universe of discourse”. If the high-frequency item *as pessoas* is used for impersonalization, it could be considered a human impersonal construction on a par with the singular uses of the lexeme *a pessoa* and *uma pessoa* (Amaral & Mihatsch, 2019; Posio, 2017, 2019, forthcoming).

The present paper sets out to explore the uses of *as pessoas* in spoken European Portuguese to determine whether it could be considered a human impersonal construction or a pronoun, or a noun used frequently in pronoun-like functions, i.e., quasi-pronoun (Gast & van der Auwera, 2013, p. 123). This research question is approached through quantitative analysis and qualitative, pragmatic scrutiny of the occurrences of *as pessoas* in the data. If the form is susceptible of grammaticalization, we could expect to find a large proportion of uses with a fixation of the form

(i.e., lack of modification, lack of variation in the use of the definite article) and regularization of the functions (e.g., stabilization of referential properties to either include or exclude the speaker) over time. In order to corroborate the eventual diachronic tendencies, the oral corpus used for the main analysis will be complemented with diachronic corpus data (see Section 2).

In addition to the grammaticalization hypothesis and diachronic development of the form *as pessoas*, the present paper will also examine the sociolinguistic distribution of the form. Since the collective noun *a gente* ‘the people’ that is used as an impersonal or first-person plural referential device in colloquial speech has traditionally been condemned by normative grammars and teachers and is consequently expected to be avoided by educated or upper-class speakers, one could hypothesize that the form *as pessoas* ‘the people’ be used by such speakers as a sociolinguistically conditioned variant of *a gente* (see Section 3). This hypothesis will be tested by comparing data from corpora representing elderly speakers of rural dialects with little formal education (CORDIAL-SIN) and urban speakers with a university-level education (*Português Falado no Porto*; see Section 2).

This paper is structured as follows: in Section 2, I present the data. Section 3 introduces the constructions being analysed and previous research on them. Section 4 presents the analysis, and Section 5 discusses the results in the light of previous research from the grammaticalization of noun phrases into pronouns and pronoun-like referential devices. Section 6 concludes the study and presents some ideas for future research.

## 2. Data

The main source of data for the present study is *Português Falado no Porto*, a 77,000-word sociolinguistic interview corpus collected by the author in the town of Porto in Northern Portugal in summer 2014 (Posio, 2021). The corpus contains 16 interviews of speakers with a relatively high socioeconomic status, all of whom had been born and were living in the city of Porto or its surroundings at the time of the interviews. All informants had either completed or were carrying out university studies at the time of the recordings. The interviewers were four female students of linguistics who were approximately 20 years old at the time of the data collection and spoke the local variety of European Portuguese. The informants’ ages range from 22 to 69 years. The interview situations were relatively familiar, given that the informants were friends or acquaintances of the interviewers, or in some cases members of the extended family or friends of their parents. This is reflected, for instance, in the use of the familiar second-person singular address form that was used by all informants towards the interviewers. The interviewers also used the second-person singular with most informants, except for three informants who were addressed in third-person singular.

In order to provide a point of comparison from a closely related language, I used the *Habla culta de Salamanca* corpus (Fernández, 2005) containing 14 sociolinguistic interviews and approximately 74,000 words. The informants of this corpus are native speakers of Peninsular

Spanish living in the town of Salamanca (Castile and León in North-Western Spain) who have completed a university degree and who were between 30 and 72 years of age at the time the recording. The interviewer was a 40-year-old female university teacher. The level of formality and the conversational topics are very similar to the Porto corpus, which makes the two corpora a good match for comparisons between the European varieties of Portuguese and Spanish.

For a sociolinguistic comparison, data from the *CORDIAL-SIN Syntax-Oriented Corpus of Portuguese Dialects* (Martins, 2000–) was used. This corpus consists of data obtained during dialectological interviews carried out in Portugal (including mainland, Madeira and Azores) between 1974 and 2000. The data included in this corpus come from 42 locations and contain 600,000 words. In contrast to the *Português Falado no Porto* corpus, the informants of CORDIAL-SIN are elderly, rural people with little formal instruction. The purpose of the interviews, as well as the discourse topics, are also very different from the other corpus: in CORDIAL-SIN, the focus is on collecting dialectal data, and the interviews evolve around themes like agriculture, fishing, and baking bread. As will be seen in the examples, this also affects the types of reference established by the informants in their answers to the interviewers' questions.

For the diachronic analysis, I used the historical subcorpus of *Corpus do Português* (Davies & Ferreira, 2006–) that contains 45 million words from the 13<sup>th</sup> to the 20<sup>th</sup> century (the data from the 13<sup>th</sup> to the 15<sup>th</sup> century is from Portugal, the subsequent centuries combine Portuguese and Brazilian data). The historical data do not represent any particular genre but rather a wide selection of available literary (both fiction and non-fiction) texts from these centuries. In these data, I have focused on the comparison of frequency and formal properties of *as pessoas*.

### **3. Competing constructions with collective plural reference: a gente and first-person plural**

The grammaticalization of the noun phrase *a gente* into an impersonal pronoun with collective reference and –subsequently– into a first-person plural pronoun has been studied in both European and Brazilian Portuguese (e.g., Pereira, 2003; Zilles, 2005 and references therein). The use of this referential device seems to be more limited and more oriented towards the impersonal uses in EP: Posio (2012) found 177 occurrences of the morphological first-person plural form (either with or without an expressed subject pronoun) and 147 occurrences of *a gente* in data consisting of 25,000 words of spoken EP, but only 14 of the occurrences of *a gente* had a primarily first-person plural interpretation. In BP, both the frequency of use of *a gente* in comparison with the first-person plural verb forms and pronouns and the first-person plural interpretation (as opposed to the generic or impersonal one) is generally higher than in EP. For instance, Zilles (2005, p. 47) found that in speech corpora from Porto Alegre, Brazil, the use of *a gente* instead of the first-person plural (verb forms ending in *-mos* and the subject pronoun *nós*) was the most frequent option used to express first-person plural reference, and that the proportion of *a gente* had risen from 56% in the 1970s to 72% in the 1990s. Zilles (2005) argues that the increasing use of *a*

*gente* is a case of Labovian “change from below”, as the construction has popular origins and the change is being led by young and female speakers.

In EP, the use of *a gente* referring to a group including the speaker is often considered ‘uneducated’ or pertaining to ‘popular speech’ or simply ignored in grammars and schoolbooks (see Pereira, 2003, pp. 4–6 for an overview). The sociolinguistic stigma attached to the use of *a gente* may explain how the speakers in the *Português Falado no Porto* corpus use this item. Given that speakers tend to monitor their speech in formal settings (such as sociolinguistic interviews) and all speakers in the corpus have received higher education, they may avoid the use of this item and prefer to use the “proper” first-person plural, or resort to other referential devices when a more impersonal or generic reference is intended. In this sense, using the expression *as pessoas* might be a strategy to avoid the use of the stigmatized *a gente*.

In addition to *a gente*, the noun phrase *as pessoas* may also compete with the first-person plural if indeed the reference of *as pessoas* includes the speaker. As argued in Posio (2012), in spoken EP the morphological first-person plural (i.e., verb forms ending in *-mos* and the subject pronoun *nós* and the oblique pronoun *nos*) is referentially highly flexible and can be used to establish reference to any number of people including the speaker, ranging from the speaker and the addressee to the whole humankind. If, on the other hand, *as pessoas* is used with a speaker-exclusive reference, it may be considered a variant of the impersonal or generic use of the third-person plural, a frequently used impersonalization strategy in spoken European Portuguese (Posio & Vilkuna, 2014).

While null-subject languages typically make greater use of verbal and pronominal human impersonal constructions than noun-based strategies like *man*-impersonals (Siewierska, 2011), there is evidence to the effect that Portuguese (particularly Brazilian, but also European; see Posio, 2021) does not behave like a canonical null subject language with regard to the expression of pronominal subjects or the use of nominal impersonalization strategies. According to Posio (2012a, 2012b, 2021), European Portuguese has considerably higher rates of expressed subject pronouns than Peninsular Spanish, and it also possesses a range of referential expressions with expressed nominal subject that tend to grammaticalize towards pronouns (e.g., *você* ‘you’, deriving from *vossa mercê* ‘your mercy’, and other address forms like *o senhor* ‘the mister; you’; collective noun phrases like *a gente*, *a malta* ‘the gang; us’, *o pessoal* ‘the personnel; us’; the singular *a/uma pessoa* ‘the person; one, I’, and so on). Thus, if indeed the plural noun phrase *as pessoas* presents signs of an ongoing grammaticalization –including increasing token frequency and diminishing type frequency, semantic bleaching, and loss of phonetic contents– this development could be related with a more general tendency in the language to fill in the subject slot in constructions where a null pronoun would be expected in a more prototypical null subject language.

#### 4. Analysis

In this section, the data are analysed first from the perspective of the distribution of *as pessoas* and other functionally overlapping referential devices (Section 4.1). Section 4.2 is dedicated to the

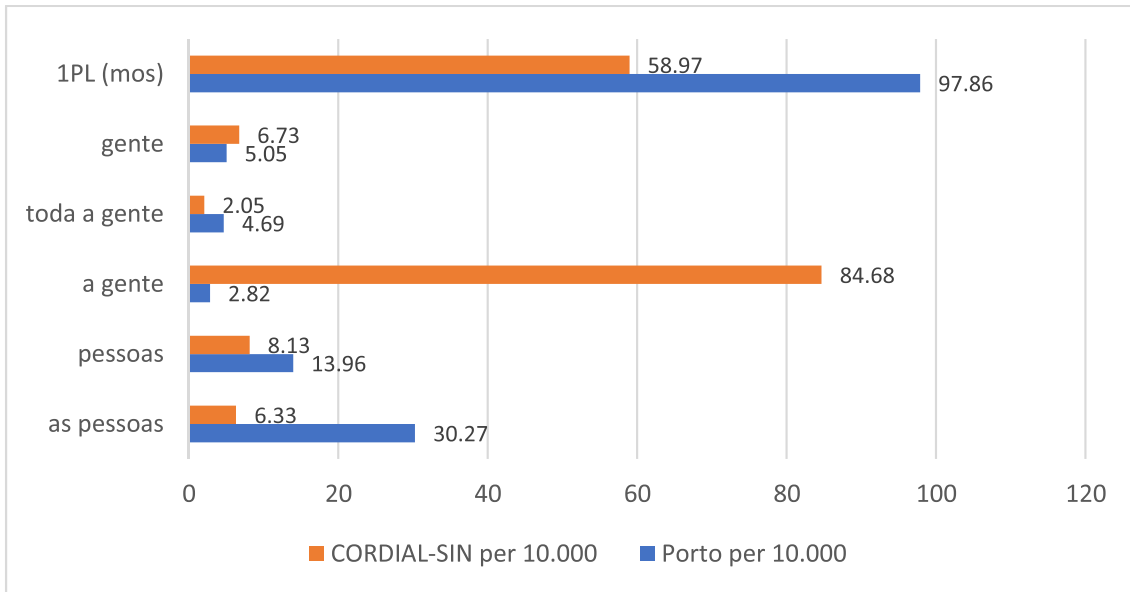
scrutiny of quantificational and referential properties of *as pessoas*. Finally, section 4.3 provides a brief overview of the historical development of the distribution of *as pessoas* in diachronic data.

#### 4.1 Distribution of plural collective referential expressions in the data

**Table 1** and **Figure 1** present an overview of the raw and normalized frequencies of the different plural referential expressions analysed in the present paper, i.e., the first-person singular, *a gente* and *as pessoas*, in the *Português falado no Porto* (PFP) corpus and the CORDIAL-SIN corpus. As can be observed, morphological first-person plural verb forms are much more frequent in the PFP corpus, but what is even more striking is the 30 times higher frequency of *a gente* in CORDIAL-SIN in comparison with PFP. The form *as pessoas*, on the other hand, is almost five times more frequent in PFP than in CORDIAL-SIN. Part of the distributional differences between the corpora can be attributed to the different discourse types and topics treated in the interviews: in CORDIAL-SIN, typical questions asked by the interviewers are of the type “How were things done or called in the past?”, and the informants are more likely to use collective or impersonal referential expressions like *a gente* in their answers, while the perspective in PFP is more individual: for instance, the informants are asked what made them choose their field of study or what they think of their home town. Nevertheless, the speakers in PFP also use plural referential expressions, as can be seen in the high incidence of the first-person plural and *as pessoas*, suggesting that the distribution of *a gente* is, indeed, sociolinguistically conditioned. However, this is not true of the lexeme *gente* on its own: its frequency of use is very similar in the two corpora, and the locution *toda a gente* ‘everybody’ is even more frequent in PFP than in CORDIAL-SIN.

	PFP	PFP per 10.000	CORDIAL-SIN	CORDIAL-SIN per 10.000
<i>1PL (-mos)</i>	834	97,86	6443	58,97
<i>gente</i>	43	5,05	389	3,56
<i>toda a gente</i>	40	4,69	123	1,13
<i>a gente</i>	24	2,82	4973	45,51
<i>pessoas</i>	119	13,96	532	4,87
<i>as pessoas</i>	258	30,27	336	3,08
<i>Total words in corpus</i>	85226		1092632	

**Table 1:** Normalized frequencies of first-person plural, (*a gente* and (*as pessoas* in PFP and CORDIAL-SIN.



**Figure 1:** Normalized frequencies of first-person plural, (*a gente* and *as pessoas* in PFP and CORDIAL-SIN.

Interestingly, the form *a gente* is used only 25 times in the PFP corpus, and all these uses are found in the speech of male informants aged around 50 years: in fact, almost half of the occurrences (12/25) are produced by the same speaker, for whom the use of *a gente* seems to constitute an idiolectal or stylistic feature. An example from this speaker is provided in (1), where *a gente* is used as a first-person plural expression.

- (1) *eu acho que: para mim fazia-me bem o silêncio gosto muito do silêncio, e a gente tem pouco,*  
 (1.2) *eh: gostava de fazer se calhar um ou dois anos num:, (0.7) num convento. (1.1) num*  
*mosteiro, (1.7)*<sup>1</sup>

‘I think that for me silence would do good I like silence a lot, and **we** [= *a gente*] have little [silence], (1.2) eh I’d like to do one or two years in (0.7) in a convent, (1.1) in a monastery, (1.7)’

PFP, male, 47 years

The example (2) shows how *a gente* is used coreferentially with *nós* ‘we’, while at the same time establishing a contrast with *as pessoas*. In the discourse context, the speaker is referring roughly to ‘people visiting Porto as tourists’ and ‘locals, people from Porto’. This example also illustrates a use of *as pessoas* that was classified as speaker-exclusive in our analysis (see Section 4.2).

<sup>1</sup> The examples follow the transcription conventions of the corpus. The colon (:) indicates lengthening of the preceding sound; the comma (,) indicates an ascending intonation and the full stop (.) a descending one. The numbers between brackets refer to silences longer than 0.1 seconds, which is indicated by (.). The English translation follows the original as closely as possible.



- (2) *se nós não temos vinho para oferecer portanto e e ou ou pelo menos se as pessoas não souberem que a gente tem vinho para oferecer. (0.6) e:: e a saberem divulgar houve também um trabalho de marketing de divulgação dessas marcas no estrangeiro*  
 ‘If **we** [= *nós*] don’t have wine to offer and then and and or or at least if **people** [= *as pessoas*] don’t know that **we** [= *a gente*] have wine to offer. (0.6) and [they] know how to publicize it there was also an effort to market to publicize these brands abroad’  
 PFP, male, 30 years

Unlike *a gente*, both the noun phrase *as pessoas* and the plural form *pessoas* without the definite article are considerably more frequent in the PFP corpus than in CORDIAL-SIN. Just how frequent is the use of *as pessoas* in comparison with other data, and could such usage frequency constitute a piece of evidence of an ongoing grammaticalization process – or, at least, a shift from “minor” to “major” use patterns in terms of Heine and Kuteva (2006, p. 56)? To provide a point of comparison, I examined the use of the cognate noun *personas* in Peninsular Spanish in the *Habla culta de Salamanca* corpus (Fernández Juncal, 2005). In the Spanish data, the noun phrase *las personas* ‘the people’ is used only seven times in the 77,000-word corpus (yielding a normalized frequency of 0.91 times per 10,000 words), but only two of these uses are unmodified, i.e., there is no adjectival or clausal modifier, which corresponds to a normalized frequency of just 0.26 occurrences per 10,000 words.

The infrequent use of the plural *personas* in Spanish could naturally be due to the existence of the collective noun *gente* that can be considered a suppletive plural form: indeed, the first definition of *gente* in the dictionary of the Real Academia Española is ‘pluralidad de personas’, i.e., ‘plurality of persons’ (RAE, s.v. *gente*). Since the noun phrase *la gente* does not have similar first-person plural uses as *a gente* in Portuguese, nor is it equally stigmatized, one could expect Spanish *la gente* and Portuguese *as pessoas* to be distributionally and functionally similar. However, a further analysis shows that the use of the lexeme *gente* is not very frequent in the *Habla culta de Salamanca* corpus: there are 169 occurrences, i.e., a normalized frequency of 21.95 per 10,000 words, but only one 53 of these occurrences (i.e., 31%) are of the definite noun phrase *la gente* without adjectival or clausal modifiers. Thus, the normalized frequency of *la gente* is 6.89 times per 10,000 words, which is considerably lower than the normalized frequency of *as pessoas* in the PFP corpus, i.e., 30.27 per 10,000 words; see **Figure 1**).

The rest of the uses of *gente* in the Salamanca corpus are modified by adjectives or relative clauses, have no determiner, or are quantified expressions (e.g., *la mayoría de la gente* ‘the majority of people’); some occurrences are found in collocations like *la gente mayor* ‘the elderly’ or *la gente joven* ‘the young’, the translation equivalents of which are infrequent in the Portuguese data. Thus, the striking difference in the frequency of *as pessoas* vs. *las personas* cannot be attributed to the use of *gente* as a suppletive plural form instead of *las personas* in the Salamanca data.



## 4.2 Quantificational and referential properties of *as pessoas*

Can *as pessoas* be considered a human impersonal (quasi)pronoun, similarly to *people* (Gast & van der Auwera, 2013, p. 132)? A definitional property of human impersonal pronouns (or other constructions expressing human impersonality) is their incapacity to introduce referents into discourse (Gast & van der Auwera, 2013). Consequently, one cannot refer back to human impersonal expressions like English *one* or Germanic *man* using personal pronouns like ‘he’ or ‘she’. In Posio (2021), I found this to be the case for *a/uma pessoa* in European Portuguese; however, see Amaral and Mihatsch (2019, p. 164, for Brazilian Portuguese). The same holds for *as pessoas* in my data: there are only three cases where the speakers refer back to an unmodified occurrence of *as pessoas* using a personal pronoun (out of 119 cases of *as pessoas* in total; see **Table 1**). In these cases, one could argue that *as pessoas* is not a universal impersonal expression but rather refers to a contextually defined group of people, for instance ‘those people who have been offended by *praxe* [initiation rituals for first-year students in some Portuguese universities]’, as in (3).

- (3) *ehm: (.) as pessoas (.) acham que foram ofendidas na praxe, mas aguentaram até ao fim. (0.4) depois quando são elas a praxar (.) vão fazer passar os outros por um: (.) uma vida negra entre aspas.*  
 ‘ehm (.) **people** [= *as pessoas*] (.) think that they were offended in *praxe* but they endured until the end, (0.4) after that when **they** are the ones doing *praxe* (.) they are going to make the others go through (.) hell quote–unquote’  
 PFP, male, 31 years

On the other hand, the noun phrase *as pessoas* is frequently repeated within the same turn or stretch of discourse when it does not constitute a contextually delimited group. For instance, in (4) *as pessoas* could be interpreted as more universal, concerning ‘everyone, all people in this hypothetical situation’.

- (4) *aquilo é algo que marca, que: que: que de alguma forma aproxima as pessoas, (0.6) que: (.) que:, (0.3) que: que transmite eh: uma uma sensação de união das pessoas e que as pessoas se, (0.4) se unem em torno de:, (0.4) de de: da:, (0.4) dali do curso (.)*  
 ‘that is something that marks that that that somehow unites **people**, [= *as pessoas*] (0.6) that (.) that, (0.3) that that transmits eh a a feeling of a union between **people** [= *as pessoas*] and that **people**, [= *as pessoas*] (0.4) are united around, (0.4) of of of the, (0.4) of the program (.)’  
 PFP, male, 30 years

Another characteristic of human impersonal constructions is that they can be further classified into universal and existential ones (Gast & van der Auwera, 2013, p. 150). In the first category, the referential range is, in principle, unlimited and concerns all human beings, while in the second category, there is a contextually restricted set of humans that are relevant as potential

referents. The same construction – such as the impersonal third-person plural – can receive both interpretations depending on the context: for instance, *In Spain they eat late* would concern ‘anyone’ in Spain and thus be considered a universal reading, while *They have changed the tax laws again* has a vague existential reading called “corporate” reading, as it implies the existence of a specific group of people (i.e., the government) who has the capacity to change tax laws, and a sentence like *They’re knocking on the door* would be classified as a specific existential reading, since it implies the existence of one or several people whose identity is not known (Siewierska & Papastathi, 2011, p. 581).

*As pessoas* clearly does not project existential readings, as can be shown by the oddity of examples like *\*as pessoas estão a bater à porta* ‘people are knocking on the door’ or *\*as pessoas mudaram a legislação tributária* ‘people have changed the tax laws’. The use of *as pessoas* (as well as the use of *people* in the English translations) would imply a universal reading or a lexical, contrastive reading in which ‘people’ is understood as the counterpart of, e.g., ‘elites’ or ‘politicians’. However, even in the case of universal readings of human impersonal expressions, the referential range of the constructions can be contextually restricted, as occurs in the case of universal third-person plural impersonals with locative expressions (e.g., *In Spain they eat late*) or third-person plural impersonals with a corporate reading (e.g., *They have changed the tax laws again*; Siewierska & Papastathi, 2011, p. 581 and references therein). This is also true of *as pessoas*: in the data, it is very rarely completely universal but rather points at a contextually restricted set of potential referents. Nevertheless, *as pessoas* is incompatible with the corporate reading (Amaral & Mihatsch, 2019, p. 155).

Since human impersonal constructions do not establish reference to clearly identifiable individuals, Posio and Vilkuna (2013) propose the term *referential range* to be used of the set of potential referents of these constructions. The referential range, just like reference in the case of personal pronouns, can either include or exclude the speaker. In fully universal readings, the speaker is always included by virtue of being human, but when the referential range is restricted by the discourse context, there are three possibilities: the inclusion of the speaker can be either (1) left unspecified or ‘open’, the speaker can be (2) explicitly included, or (3) explicitly excluded from the referential range. If our starting hypothesis that *as pessoas* is a sociolinguistically conditioned variant of *a gente* was correct, we could expect to find predominantly speaker-inclusive uses when *as pessoas* occurs on its own, i.e., not modified by an adjectival or clausal complement.

Based on a qualitative scrutiny of the discourse context, I classified the occurrences of *as pessoas* in the PFP corpus into ‘open’ ones, i.e., those where the speaker may potentially be included in the referential scope, and into exclusive ones, i.e., those where the speaker is explicitly excluded from the reference. In practice, the latter category consists of clauses where there is a first-person singular or plural referring to the speaker that is portrayed as clearly distinct from

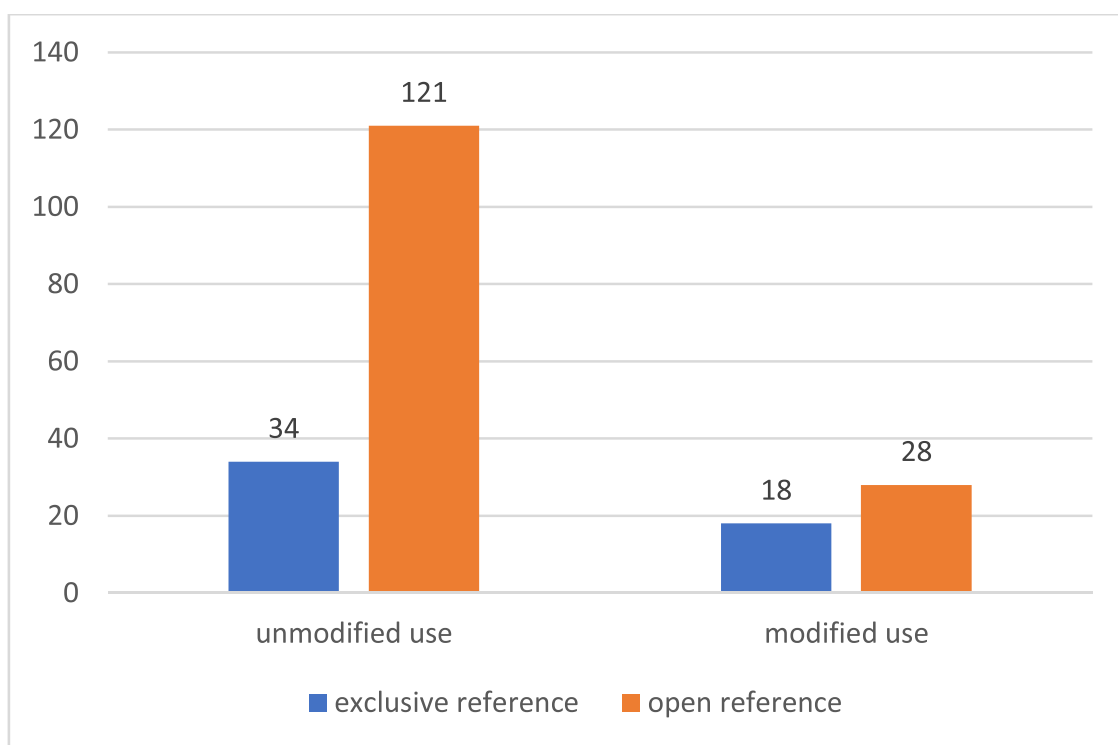
the reference of *as pessoas*, as in example (5) where the use of *eu acho que* ‘I think that’ and *eu vejo* ‘I see’ project the speaker in the role of an outside observer looking at ‘people’. There were no cases where the referential range of *as pessoas* would be explicitly inclusive with regard to the speaker (i.e., not permitting an exclusive interpretation).

- (5) *mas eu acho que um: eu vejo as pessoas hoje muito; eu vejo as pessoas absorvidas demais com o estudo. (0.7) eu acho que as pessoas às vezes passam ao lado a juventude por causa do estudo. (0.4)*

‘but I think that um I see **people** [= *as pessoas*] today very, I see people [= *as pessoas*] too absorbed into studying. (0.7) I think that **people** [= *as pessoas*] sometimes skip being young because of studies. (0.4)’

PFP, male, 47 years

**Figure 2** shows the distribution of the speaker-exclusive and referentially open (i.e., potentially speaker-inclusive) uses of *as pessoas* for unmodified uses (i.e., where *as pessoas* occurs with no adjectival or clausal modifiers) and for modified uses (i.e., where there is an adjectival or clausal modifier). As can be observed, the unmodified uses of *as pessoas* are over three times more frequent than the modified ones, and the unmodified uses are more likely to include the speaker in their referential range than the modified ones.



**Figure 2:** Referential range and modified vs. unmodified uses of *as pessoas*.

In other words, when *as pessoas* is used without an adjectival or clausal modifier, it is more likely to have open reference potentially including the speaker in its referential range. However, it is clearly not a variant of *a gente* or the first-person plural in the sense that, unlike *a gente*, it does not suffice to establish speaker-inclusive reference on its own, but rather only in contexts where this interpretation is based on contextual cues. On the contrary, a speaker-exclusive interpretation easily emerges whenever a contrast between the first-person singular or plural is created in the discourse context, as can be observed in the example (6).

- (6) – *acho que se fala mal.* (0.3)  
 – *porquê.*  
 – *há muita:; (1.0) há muita gente, (0.9) que: diz muitos palavrões mesmo grosseiros, (1.7) as pessoas dizem muitos palavrões eh: em duas palavras dizem um palavrão (.) grosseiro, acho que não (.) não se justifica.*  
 – ‘I think people [= *se*] talk poorly. (0.3)’  
 – ‘why’  
 – ‘there are (1.0) there are many people [= *gente*] (0.9) that say many swearwords that are really vulgar, (1.7) people [= *as pessoas*] say many swearwords eh in two words they say one swearword (.) a vulgar one, I think that it’s not (.) it’s not justified.’  
 PFP, male, 51 years

The notion of perspective has also been used to distinguish between types of human impersonals. It should not be confused with speaker-inclusive or speaker-exclusive referential range. While Gast and van der Auwera (2013, p. 151) regard speaker-inclusive uses of human impersonals as invariably having an internal perspective, they argue that speaker-exclusive uses can express both internal and external perspectives. For instance, in the example (7), borrowed from Moltmann (2010, p. 448), the sentence a (with the impersonal pronoun *one*) expresses the internal perspective and the sentence b (with the impersonal quasi-pronoun *people*) expresses the external perspective, even though *John* is logically included in the referential scope of both *one* and *people*.

- (7) (a) *John found out that **one** gets sick when one eats these mushrooms.*  
 (b) *John found out that **people** get sick when they eat these mushrooms.*  
 (Moltmann, 2010, p. 448; emphasis added)

The use of an adjectival, clausal or locative modifier restricts the referential scope of *as pessoas* to a more specific group. In example (8), the speaker pertains to both groups being established with *as pessoas*, i.e., ‘people in Portugal’ and ‘people from Porto’ and the clause is thus classified as having ‘open’ reference, i.e., the speaker may be included in the referential range. However, despite the logical inclusion of the speaker in the groups referred to by *as pessoas*, the clause

expresses a judgment given from an outsider perspective – again, looking at ‘people’ from a certain distance.

- (8) *de uma forma geral eu acho que as pessoas (.) em Portugal têm uma ideia, (1.0) eh: (0.4) das das pessoas do Porto e: e: e um estereótipo,*  
 ‘in general I think that **people** [= *as pessoas*] (.) **in Portugal** have an idea, (1.0) eh (0.4) of of **people** [= *as pessoas*] **in Porto** and and and a stereotype’  
 PFP, male, 30 years

In conclusion, the qualitative analysis of the occurrences of *as pessoas* in the *Português falado no Porto* corpus indicates that, while the most frequent uses of the construction are unmodified and have a universal referential range, in practice the use of the construction implies an outsider perspective to the events depicted or judgments made by the speaker. In other words, *as pessoas* does not appear to be a stylistically or sociolinguistically conditioned variant of *a gente*, which implies the inclusion of the speaker in the referential range, but is rather used to allude to ‘other people’, i.e., not including the speaker. In other words, *as pessoas* clearly conveys an external perspective (in the sense of Moltmann, 2010, p. 448), even if its referential range includes the speaker, as in example (8).

### 4.3 Historical development of *as pessoas*

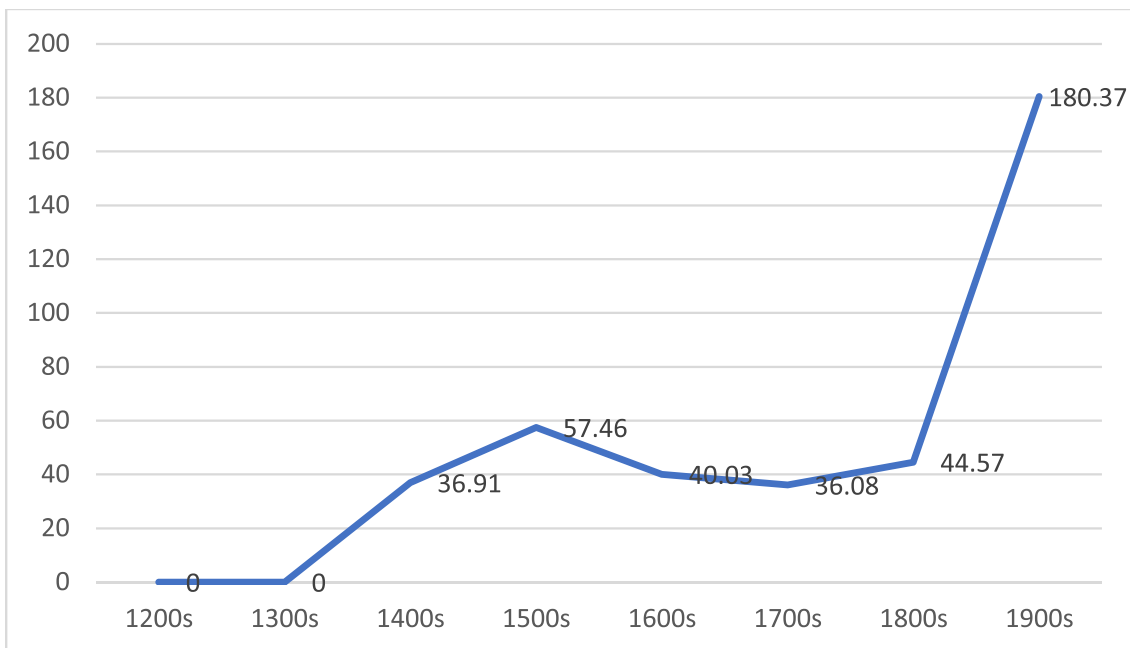
If *as pessoas* is indeed developing into a human impersonal pronoun, its frequency of use should increase over time. Using the historical data in *Corpus do Português* (Davies & Ferreira, 2006–), I controlled for the normalized frequency of the item in different centuries. The results are shown in **Figure 3**. Although the plural word form *pessoas* appears already in the 13<sup>th</sup> and 14<sup>th</sup> century, the first occurrences of *as pessoas* are from the 15<sup>th</sup> century. Their frequency remains relatively stable until the 19<sup>th</sup> century, with a slight increase in the 16<sup>th</sup> century, but increases considerably in the 20<sup>th</sup> century. As shown in **Figure 4**, this increase is due to the inclusion of oral data in this century, since the oral section of the 20<sup>th</sup> century stands out with as high a frequency as 660.49 occurrences per million words. **Figure 5** further shows that the construction is considerably more frequent in European than Brazilian Portuguese (in the *Corpus do Português* data from the 20<sup>th</sup> century there are 2009 occurrences in 10,506,703 words in European Portuguese and 1646 occurrences in 10,271,022 words in Brazilian Portuguese; Log Likelihood = 28.35, significant at the level of  $p < 0.01$ ).

If the grammaticalization hypothesis is correct, in addition to the increasing overall frequency of the lexeme, the proportion of unmodified uses of *as pessoas* (see example (9)) should increase over time, as its use as a quasi-pronominal human impersonal construction becomes more widespread. As shown in **Figure 6**, the proportion of unmodified uses does show a gradual increase over centuries, with the exception of the 19<sup>th</sup> century. Again, the inclusion of oral data

in the 20<sup>th</sup> century boosts the frequency of unmodified *as pessoas* so much that almost three out of four occurrences of *as pessoas* are unmodified (i.e., do not contain adjectival or clausal modifiers; see example (10)).<sup>2</sup>

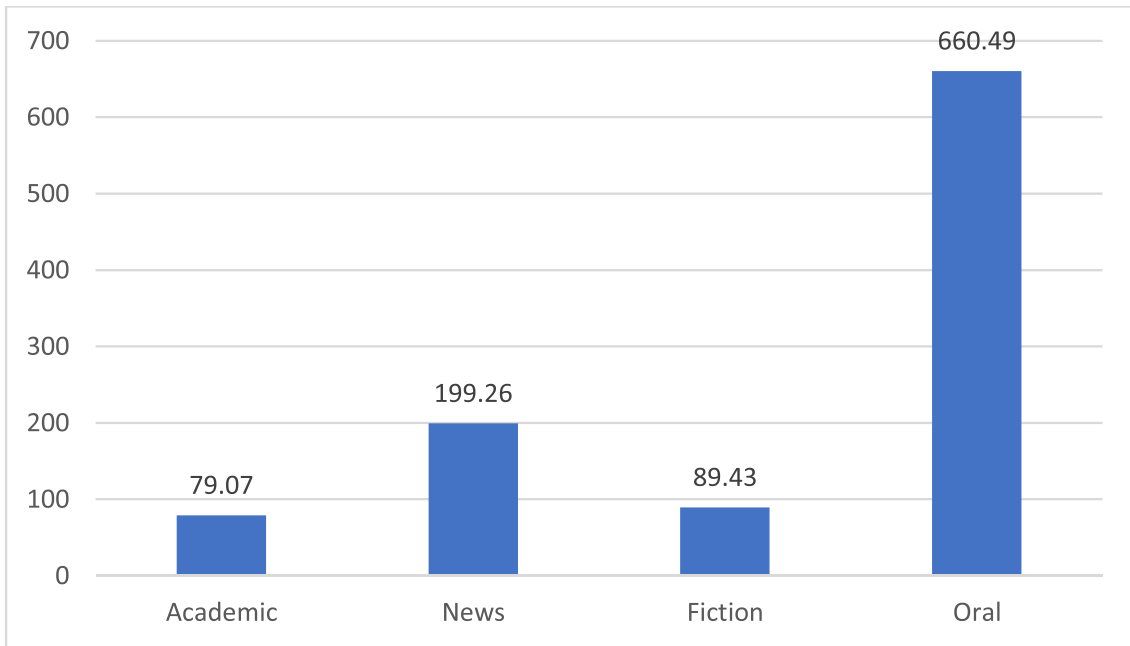
- (9) *A delegacia continuava silenciosa e as pessoas sentadas pelas cadeiras não ousavam entreolhar-se.*  
 ‘The office remained silent and people who were sitting on the chairs did not dare to look at each other.’  
*Corpus do Português, 19:Fic:Br:Barreto:Caminha*
- (10) *Tem de haver um controlo em isto, porque depois as pessoas não compram.*  
 ‘There has to be a control on this, because otherwise people don’t buy.’  
*Corpus do Português, 19N:Pt:Público*

The drop in the frequency of unmodified uses in the 19<sup>th</sup> century could be due to a change in the proportion of text types included in the corpus, as the data from this century is to a large extent drawn from novels, in contrast to the previous centuries that include letters and sermons, and to the 20<sup>th</sup> century that has the widest scope of different genres.

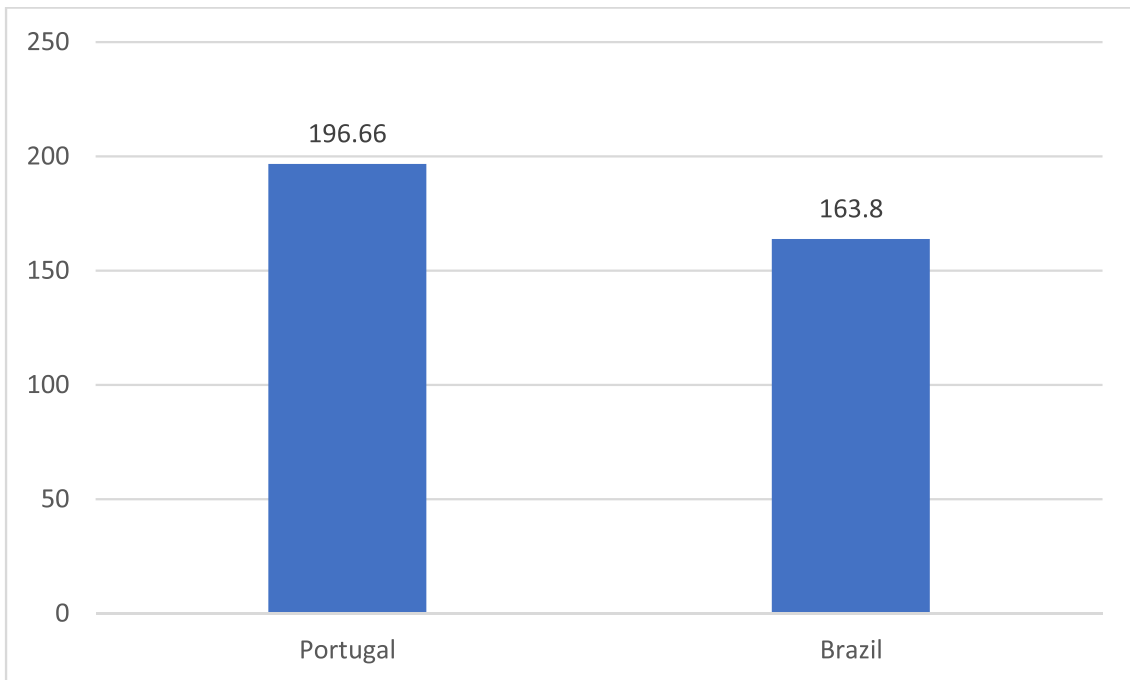


**Figure 3:** Normalized frequency (per one million words) of *as pessoas* in Corpus do Português historical data, by century.

<sup>2</sup> Due to large amount of data from the 1800s and 1900s, a random sample of 200 occurrences was used to calculate the proportion of unmodified uses in these centuries.

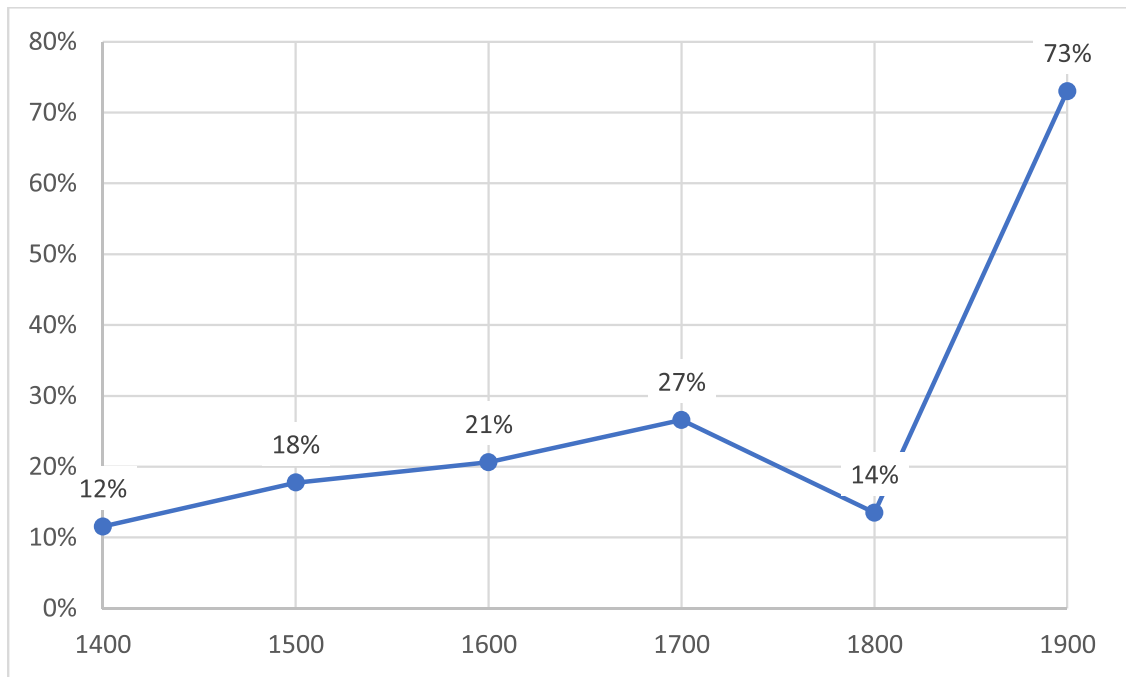


**Figure 4:** Normalized frequency (per one million words) of *as pessoas* in different genres in *Corpus do Português*.



**Figure 5:** Normalized frequency (per one million words) of *as pessoas* in Portugal and Brazil in *Corpus do Português*.





**Figure 6:** Proportion of unmodified uses of *as pessoas* in historical data.

## 5. Discussion and conclusions

The main aim of this paper has been to scrutinize the uses the noun phrase *as pessoas* in spoken European Portuguese to determine whether it could be considered a human impersonal construction or a (quasi)pronoun used for impersonalization. In addition, we have analysed the distribution of this construction in two corpora representing different data types (sociolinguistic interview vs. dialectal interview) and speakers with different sociolinguistic profiles, and we have reviewed the diachronic developments in the use of this construction in historical data.

To answer the first question, *as pessoas* clearly falls under the category of impersonalization as defined by Gast & van der Auwera (2013, p. 136), consisting of “the process of filling an argument position of a predicate with a variable ranging over sets of human participants without establishing a referential link to any entity from the universe of discourse”. Considering another definitional criterion of human impersonal pronouns discussed by Gast & van der Auwera (2013, p. 134), the incapacity to introduce discourse referents and the subsequent inability to be referred to by personal pronouns, *as pessoas* seems to represent an intermediate category between a noun phrase with a vague reference (that can be referred to by the personal pronoun *elas* ‘they.FEM’ as in example (3)) and fully impersonal uses, where the form *as pessoas* is repeated instead of pronominal reference (as in example (4)).

All in all, anaphoric pronominal references to *as pessoas* are rare in my data, and there are no long referential chains with *as pessoas* functioning as the topic. However, together with

the fact that *as pessoas* does accept modification by relative clauses or adjectives – despite mostly occurring unmodified, a tendency that seems to be increasing in the historical data – the possibility of using personal pronouns for anaphoric reference to *as pessoas* suggests that it does not constitute a grammaticalized impersonalization strategy but rather a “major pattern” in Heine and Kuteva’s (2006) terms, i.e., a pattern that is used in a wide range of contexts and that is consistently associated with a grammatical function, in this case human impersonalization. This contrasts with the singular forms *a pessoa* ‘the person’ and *uma pessoa* ‘a person’ that are more grammaticalized and more limited with regard to modification and being referred to by personal pronouns (Amaral & Mihatsch, 2019; Posio, 2021). However, the very high frequency of *as pessoas* in the spoken European Portuguese data analysed for this study suggests that it has a special status as a referential device in this language variety.

Regarding the referential range of *as pessoas*, we have seen that it can be used in both speaker-inclusive and speaker-exclusive ways, even though the speaker is logically included in the referential range if no restrictions arise in the context, for instance due to contrastive readings or locatives specifying the referential range. However, even if the speaker is encompassed in the reference potential of *as pessoas*, the construction represents external perspective (Moltmann, 2010) to the events being depicted. This property differentiates it from *a gente*, a referential device with both human impersonal and personal interpretations, that is inherently speaker-inclusive. Thus, unlike hypothesized in the beginning of the study, *as pessoas* is not a sociolinguistically conditioned variant of *a gente*. The comparison of the normalized frequencies between the *Português falado no Porto* and CORDIAL-SIN corpora nevertheless suggest that frequent use of *a gente* is typical of elderly, rural speakers with less studies and a lower socioeconomic status whereas *as pessoas* is more characteristic of younger, urban speakers with university-level studies and a higher socioeconomic status. Obviously, the different distribution of the constructions in the two corpora could also depend on the different data types they represent – dialectal interviews vs. sociolinguistic interviews – although it is not clear why reference to ‘people’ would be characteristic of the latter type rather than the former.

Finally, the analysis of larger-scale corpus data from *Corpus do Português* (Davies & Ferreira, 2006–) shows that *as pessoas* is characteristic of spoken data and more frequently used in European than Brazilian Portuguese. There is also some evidence to the effect that the frequency of the construction in texts, as well the frequency of unmodified occurrences of *as pessoas*, may be increasing over the centuries. However, these observations should be taken as highly preliminary due to the inconsistencies in the types of texts included in the corpus from different centuries. A topic to be addressed in future research is the development of *as pessoas* as part of a larger set of quasi-pronominal referential devices including the singular forms *a pessoa* and *uma pessoa* but also *a gente* and other referential expressions grammaticalized from noun phrases that are characteristic of Portuguese unlike other related language varieties, including Peninsular Spanish.

---

## Competing Interests

The authors have no competing interests to declare.

---

## References

- Amaral, E., & Mihatsch, W.** (2019). Incipient impersonal pronouns in colloquial Brazilian Portuguese. In P. Herbeck, B. Pöll & A. C. Wolfsgruber (Eds.), *Semantic and syntactic aspects of impersonality* (pp. 149–185). Hamburg: Buske.
- Amaral, E., & Mihatsch, W.** (Forthcoming). Portuguese *a pessoa* and *uma pessoa*. Emerging inclusive impersonals. In P. Posio & P. Herbeck (Eds.), *Referring to discourse participants in Ibero-Romance languages* (Open Romance Linguistics). Berlin: Language Science Press.
- Davies, M., & Ferreira, M.** (2006–). *Corpus do Português: Historical Genres* [data set]. Retrieved August 12, 2021 from <http://www.corpusdoportugues.org/hist-gen/>.
- Fernández Juncal, C.** (2005). *Corpus de habla culta de Salamanca* [Corpus of educated speech of Salamanca]. Burgos: Instituto Castellano y Leonés de la Lengua.
- Gast, V., & van der Auwera, J.** (2013). Towards a distributional typology of human impersonal pronouns, based on data from European languages. In D. Bakker & M. Haspelmath (Eds.), *Languages across boundaries – Studies in the memory of Anna Siewierska* (pp. 119–158). Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9783110331127.119>
- Giacalone Ramat, A., & Sansò, A.** (2007). The spread and decline of indefinite *man*-constructions in European languages. An areal perspective. In P. Ramat & E. Roma (Eds.), *Europe and the Mediterranean as Linguistic Areas. Convergences from a Historical and Typological Perspective* (pp. 95–131). Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/slcs.88.07gia>
- Heine, B., & Kuteva, T.** (2006). *The Changing Languages of Europe*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199297337.001.0001>
- Lopes, C. R. dos S.** (2003). *A inserção de a gente no quadro pronominal do português* [Insertion of *a gente* into the pronominal paradigm of Portuguese]. Madrid: Iberoamericana, Frankfurt am Main: Vervuert Verlag. DOI: <https://doi.org/10.31819/9783865278494>
- Martins, A. M.** (coord.) (2000–). *CARDIAL-SIN: Corpus Dialectal para o Estudo da Sintaxe / Syntax-oriented Corpus of Portuguese Dialects* [data set]. Lisbon: Centro de Linguística da Universidade de Lisboa. Retrieved on February 10, 2021 from <http://www.clul.ulisboa.pt/en/10-research/314-cardial-s>
- Moltmann, F.** (2010). Generalizing detached self-reference and the semantics of generic *one*. *Mind & Language*, 25(4), 440–473. DOI: <https://doi.org/10.1111/j.1468-0017.2010.01397.x>
- Pereira, S. M.** (2003). *Gramática Comparada de a gente: variação no Português Europeu* [Comparative grammar of *a gente*: variation in European Portuguese] (Unpublished masters' dissertation), Faculdade de Letras da Universidade de Lisboa.
- Posio, P.** (2017). Entre lo impersonal y lo individual. Estrategias de impersonalización individualizadoras en el español y portugués europeos [Between impersonal and individual. Individuating strategies of impersonalization in European Spanish and European Portuguese]. *Spanish in Context*, 14(2), 209–229. DOI: <https://doi.org/10.1075/sic.14.2.03pos>

**Posio, P.** (2021). *A pessoa and uma pessoa: Grammaticalization and functions of a human impersonal referential device in European Portuguese*. *Journal of Portuguese Linguistics*, 20(1), 2. DOI: <https://doi.org/10.5334/jpl.254>

**Posio, P.** (2021). *Português Falado no Porto*. [Portuguese spoken in Porto] Unpublished sociolinguistic interview corpus. University of Helsinki.

**Posio, P., & Vilkuna, M.** (2013). Referential dimensions of human impersonals in dialectal European Portuguese and Finnish. *Linguistics*, 51, 177–229. DOI: <https://doi.org/10.1515/ling-2013-0006>

**Real Academia Española.** *Diccionario de la lengua española*, 23.<sup>a</sup> ed. [Dictionary of the Spanish language] Online version 23.5. <<https://dle.rae.es>> Accessed on April 28, 2022.

**Siewierska, A., & Papastathi, M.** (2011). Third person plurals in the languages of Europe: Typological and methodological issues. *Linguistics*, 43(2), 575–610.

**Zilles, A. M. S.** (2005). The development of a new pronoun: The linguistic and social embedding of *a gente* in Brazilian Portuguese. *Language Variation and Change*, 17, 19–53. DOI: <https://doi.org/10.1017/S0954394505050027>

