
RESEARCH PAPER

A pessoa and *uma pessoa*: Grammaticalization and functions of a human impersonal referential device in European Portuguese

Pekka Posio

Department of Languages, University of Helsinki, Helsinki, FI

pekka.posio@helsinki.fi

The grammaticalization of nouns meaning ‘man’ or ‘person’ into impersonal pronouns (so-called *man*-impersonals) has been related to languages with obligatory subject expression, like Germanic languages and French. Albeit regarded as a null-subject language, European Portuguese (EP) presents a similar development: the noun phrases *uma pessoa* ‘a person’ and *a pessoa* ‘the person’ are used in ways that resembles *man*-impersonals. The present paper aims at shedding light to the semantics, pragmatics and morphosyntactic properties of the *pessoa* constructions in EP and discuss their status as potentially grammaticalized or grammaticalizing referential devices. It is argued that the emergence of these constructions is favored by pragmatic and morphosyntactic tendencies found in EP: avoidance of direct person reference and a preference for expressed subjects even in contexts permitting null subjects.

Keywords: European Portuguese; grammaticalization; impersonality; pronouns; person; reference

1 Introduction

The present paper examines the use of the noun phrases *uma pessoa* and *a pessoa*, literally ‘a person’ and ‘the person’, respectively, in European Portuguese (EP). The data come from a corpus of sociolinguistic interviews collected in the town of Porto (see Section 2). Both noun phrases are used in a way that resembles the use of human impersonal pronouns like Germanic *man* and French *on* (see Section 1.2). An example of a generic reading of *uma pessoa*, translatable as *you* or *one* into English, is given in example (1).

- (1) *quando uma pessoa vai para um país estrangeiro, também vai à capital*
when a person goes to a country foreign, also goes to.the capital
‘When you go ~ one goes to a foreign country, you also go ~ one also goes to the capital.’ (female, 51)

The research questions tackled in the present paper are twofold. On the one hand, I examine the semantic and pragmatic properties of these constructions, whether there are any differences between the indefinite (*uma pessoa*) and the definite (*a pessoa*) variant, and how they receive their pragmatic interpretation in the contexts of occurrence. On the other hand, I assess to what extent these constructions are grammaticalizing and to what degree they are similar to so-called *man*-impersonals found in other languages (see Section 1.2).

The paper is structured as follows: Section 1.1 discusses the concept of human impersonality and human impersonal constructions in general, and Section 1.2 focuses on the case of so-called *man*-impersonals; Section 1.3 further looks at the data available on the historical development of *man* and *person* constructions in Portuguese. Section 2 presents the data analyzed in this paper. Section 3 presents the results of the analysis, divided into a qualitative scrutiny of the referential properties of *uma pessoa* and *a pessoa* (Section 3.1) and a survey of the grammaticalization of these construction (Section 3.2). Section 4 presents the discussion of the results, divided into an evaluation of the eventual grammaticalization of the constructions (Section 4.1), the pragmatic tendencies related to their use (Section 4.2) and their relation with the null subject properties of EP (Section 4.3). Section 5 summarizes briefly the conclusions of the paper.

1.1 Human impersonality as a functional domain

Before embarking on the discussion of the *pessoa* constructions, a brief review of the framework of human impersonality and human impersonal pronouns is in order. The term *impersonality* is used in various different and sometimes contradictory ways. While syntactic impersonality typically refers to lack of subject argument in a clause or lack of subject agreement on the verb, impersonality in a semantic or pragmatic sense is defined as reduction in referentiality or lack thereof. These two notions are independent from each other: although semantically impersonal constructions may also be syntactically impersonal, they do not need to be. Terms such as *reference impersonality* (Siewierska 2011) or *human impersonality* (Cabredo Hofherr 2008) are used to highlight the two defining characteristics of the construction type where the *pessoa* constructions pertain: reduction in referentiality of the subject (i.e. first argument of the verb) and humanness of the intended referent (the referents cannot be interpreted to be non-human, unless they are personified in the discourse context). Human impersonal pronouns (and other referential devices¹) can be used either impersonally in the narrow sense (e.g. *They have raised the taxes again*, where the speaker may have an intended referent in mind but does not specify who they are) or with a generic reading (e.g. *They say there's dragons guarding the high security vaults*, where the speaker does not think of any particular people; the example is from Siewierska & Papastathi 2011). Thus, the notion of human impersonals subsumes the properly impersonal uses as well as the generic uses of the constructions, occasionally even specific uses (see Section 1.2). As a working definition of human impersonals, Gast & van der Auwera's (2013: 124) characterization of impersonalization as "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" is adopted in the present paper.

Although human impersonal referential devices do not specify who exactly is or are the intended referent(s), they typically restrict the choice of possible referents, i.e. the *referential range* of the pronoun or construction (Posio & Vilkuna 2013). Thus, for example, third person plural impersonals (like the pronoun *they* in the previous examples) do not usually include the speaker nor the addressee in their referential range; depending on the degree of grammaticalization of the impersonal uses of this person form in a given language, they may imply that the referent is plural, or allow for both singular and plural interpretations. Second person singular impersonals generally present the

¹ I use the term *referential device* (Kibrik 2011) to refer to free pronouns, bound person markers and other elements used to express reference to other entities in discourse. The term is useful for our purposes, in particular because it subsumes pronouns and noun phrases potentially grammaticalizing into pronoun-like elements: it is questionable whether *a/uma pessoa* can be called a "pronoun", but the uses discussed in the present paper are clearly characterizable as "referential devices".

referent as singular and allow for speaker-oriented readings. Such interpretations may be further specified by the presence of lexical items in the immediate context (Siewierska & Papastathi 2011).

While prototypical personal pronouns in their canonical uses (e.g. *I* referring to the speaker or *you* referring to the addressee) establish either deictic or anaphoric reference to entities in the physical world or antecedents in the discourse universe, human impersonal referential devices receive their interpretation through inference rather than reference. Thus, when a human impersonal referential device seems to receive a specific interpretation, such as referring to the speaker, this interpretation depends on a pragmatic inference based on contextual clues or discourse conventions. Inferential relations in discourse are “motivated by the hearer’s drive to make the speaker’s discourse coherent” (Koenig & Maurer 1999: 228).

1.2 Man-impersonals: an areal feature of European languages?

In addition to personal pronouns like *you* and *they*, another frequent source of human impersonal constructions are words meaning ‘man’ or ‘person’. The grammaticalization of these nouns is found in several unrelated languages (Heine & Kuteva 2002: 232), although most studies on this phenomenon have focused on so-called *man*-impersonals in European languages (see e.g. Giacalone Ramat & Sansò 2007; Siewierska 2011). Within Europe, the core area of *man*-impersonals coincides with so-called Charlemagne area comprising German and French, and extends also to Mainland Scandinavian languages. The *man*-impersonals found in this area are fully grammaticalized into impersonal pronouns that have both generic and episodic uses, as exemplified by (2) and (3) from French. Even specific uses, where the intended referent is a first, second, or third person participant (Siewierska 2011: 64–65), can be found (4).

- (2) *On ne vit qu’une fois.* ‘One only lives once.’; generic use, i.e. not anchored into any specific point in time
- (3) *On a volé mon vélo.* ‘They have stolen my bike.’ (cf. van der Auwera, Gast, Vanderbiesen 2012); episodic use, referring to an action situated in a specific point in time.
- (4) *Avec Jean, on ira au théâtre ce soir.* ‘Jean and I will go to the theatre tonight.’ (Creissels, to appear); specific use, referring to a first-person discourse participant

In the periphery of the man-impersonal area, including e.g. Faroese and Icelandic, Celtic languages, West and South Slavic and Albanian, *man*-impersonals are reported to occur only in generic clauses, but not in episodic ones (Giacalone Ramat & Sansò 2007; Siewierska 2011: 70). According to Siewierska (2011: 70), man-impersonals are not found outside the aforementioned area, i.e. they do not exist in “East Slavic, Lithuanian, Rumanian, Greek, Sardinian, Galician, Portuguese nor in Basque”, and the sporadic generic uses of the nouns ‘man’ or ‘person’ attested in these languages lack grammaticalized status.

As remarked by Siewierska (2011: 80), the core area of *man*-impersonals coincides with languages where the expression of pronominal subjects is obligatory, i.e. languages pertaining into the non-pro-drop paradigm, whereas languages with optional subject expression (i.e. pro-drop languages) make use of other impersonalization strategies such as third person plurals (3PL) or second person singulars (2SG). Another generalization can be established with regard to the distribution of reflexive impersonal constructions, like Portuguese and Spanish *se*-constructions: in languages with obligatory subject expression and *man*-impersonals, such reflexive constructions have only restricted uses, as in French

where the reflexive *se*-construction is limited to so-called middle voice constructions (e.g. *Ce plat se mange avec une cuillère*. ‘This dish is eaten with a spoon.’) but is not possible in generic or impersonal sentences.

Within Romance languages, a particularly interesting point of comparison for Portuguese is provided by Spanish. While Portuguese and Spanish have very similar verbal and pronominal systems, they present differences precisely in the domain of human impersonal constructions. In Spanish, there is no evidence of any grammaticalization processes resembling the *pessoa* constructions examined in this paper. In contrast, the impersonal or generic use of the second-person singular is very frequent in Peninsular Spanish, while its use in EP is more restricted. Posio (2017) compares the corpus analyzed in the present paper, collected in Porto, Portugal (see Section 2), with a comparable sociolinguistic interview corpus from Salamanca, Spain, finding that the frequency of the impersonal second-person singular is almost five times higher in the Spanish data (9.69 occurrences per 10,000 words in the Salamanca corpus, as opposed to 1.84 occurrences in the Porto corpus). In addition, while all informants in both the Porto and the Salamanca corpus use the second-person singular as an address form in reference to the interviewer, in the Porto corpus only half of them use it impersonally, while in the Salamanca corpus all informants use the impersonal second-person singular to some extent. Thus, despite the availability of the second-person singular as an address form, it remains significantly less frequent as an impersonalization strategy in the Porto corpus than in the Salamanca corpus. Unfortunately, there are no larger-scale comparable speech corpora available from Portuguese and Spanish representing communicative situations where the second-person singular is used for addressing the interlocutor that would permit evaluating this observation in the light of more data. However, since the second-person singular constitutes a direct reference to the addressee, it could be the case that its use for impersonalization is avoided due to it being a potentially face-threatening act, in particular if there are significant differences in the age or social status of the speakers (Carreira 2005; Posio 2017).

In addition to the impersonal uses of the second person singular, Spanish also features the human impersonal pronoun *uno* ‘one’ that has no equivalent in Portuguese. The distribution of *uno* and the impersonal second person singular in different varieties of Spanish seems to reflect similar pragmatic and cultural preferences that could explain the existence of the *pessoa* constructions in EP. In some varieties of Spanish, like the Andean variety, the impersonal pronoun *uno* ‘one’ is preferred over the second person singular (Guirado 2011), whereas in other varieties like Peninsular Spanish *uno* is gradually becoming obsolete and is being replaced by the second person singular (Cameron 1993; Blanco Canales 2004: 282; Guirado 2011; Posio 2017). The complementary distribution of *uno* and the second person singular has been correlated with the choice of address forms: those dialects where the second person singular is less common a choice for addressing the interlocutor also show a less frequent use of the second-person singular as an impersonalization strategy (Guirado 2011). Similarly, the avoidance of the second person singular in EP, in particular in formal contexts, might be related with the preference for other means of expressing human impersonality, like *a/uma pessoa*. However, more research would be needed to understand the distribution and pragmatic constraints of the use of the second person singular and *pessoa* in different varieties of Portuguese.

1.3 Impersonal uses of ‘man’ and ‘person’ in Portuguese

Although most modern Romance languages do not have *man*-impersonals in their present stage, impersonal uses of the noun ‘man’ were found in old Spanish, Portuguese, Catalan and Italian (Meyer-Lübke 1900: 109; Barrett Brown 1931; Lopes 2003; Giacalone Ramat & Sansò 2007; see examples (5)–(6)). One of the main arguments for analyzing such uses

as pronominal – and not just generic uses of the noun – is the lack of determiner and modification (for Portuguese, see Lopes 2003). In the aforementioned languages, the use of the impersonal subject pronoun *hom* (cf. the noun *home* ‘man’) persists only in Catalan, where it is characteristic of formal, written registers (7).

- (5) Old Spanish: *Con ellos ombre non puede beuir.*
‘One cannot drink with them.’ (*Arcipreste de Talavera* 243, Barrett Brown 1931: 269)
- (6) Old Portuguese: *pode homem chegar; por segredos que homem não conhece*
‘one can arrive’; ‘because of secrets that one doesn’t know’ (Meyer-Lübke 1900: 109).
- (7) Contemporary Catalan: *Hom pot evitar fàcilment aquests problemes.* ‘One can easily avoid these problems.’ (<https://www.enciclopedia.cat/EC-GEC-0188199.xml>, consulted on 6/12/2018).

In Portuguese and Spanish, such impersonal uses of the noun ‘man’ were lost by the 15th century (Barrett-Brown 1931; Giacalone Ramat & Sansò 2007). This decline of *man*-impersonals has been attributed to the increase in the use of other human impersonal constructions such as the reflexive-based impersonals or, in the case of Spanish, the pronoun *uno* ‘one’, although the directionality of the correlation is not completely clear (Giacalone Ramat & Sansò 2007). Among present-day Romance languages, only French has maintained the generalized use of the subject pronoun *on*, deriving from Latin *homo* ‘man’, in both spoken and written genres, to the extent that it has become the most frequent way of referring to first-person plural in spoken registers (Fonseca-Greber 2003). The extended use of the impersonal *on* in French has been attributed to contact influence from Germanic languages due to extensive cultural contacts within the Holy Roman Empire (Giacalone Ramat & Sansò 2007).

Despite the loss of *homem* as an impersonal subject, modern Portuguese has two other noun phrases denoting ‘humans’, viz. *a gente* ‘the people’ and *a/uma pessoa* ‘a/the person’ that have developed uses as human impersonal referential devices. While the use of *a gente* is a relatively well-studied phenomenon, especially in Brazilian Portuguese (BP), where it has become the most frequent way of referring to first-person plural, akin to French *on* (Lopes 2003; Zilles 2005), the impersonal uses of *pessoa* have received less attention in previous research. However, the NP (noun phrase) *a pessoa* ‘the person’ and *uma pessoa* ‘a person’ have been mentioned sporadically in the literature as alternatives to other impersonal and personal referential devices (Dias 1918: 89; Nunes 1919: 265) or as indefinite pronouns (Stolz 1991: 12). More recently, Duarte & Marques (2014), Posio (2017) and Martins (2019) have discussed the use of *pessoa* also for first-person singular as well as impersonal reference in EP, and Amaral and Mihatsch (2019) for impersonal reference in BP.

In spontaneous, colloquial, spoken EP, examples of *a pessoa* or *uma pessoa* are not difficult to come across. In addition to example (1), two further examples from my data (see Section 2) are given in (8) and (9).

- (8) *acho interessante uma pessoa estudar o nosso passado*
‘I find it interesting when a person studies our past.’ (male, 30)
- (9) *eu acho que fico a viver na mesma casa a não ser que ganhe o euromilhões. (0.4)*
se não a pessoa não tem dinheiro para comprar outra casa.
‘I think I’ll stay in the same house if I don’t win the EuroMillions. (0.4)
If not, the person doesn’t have money to buy another house.’ (female, 51)

In example (8), *uma pessoa* can be interpreted as non-referential, not necessarily or primarily pointing at the speaker, although he is included in its referential range because the reading is generic. Example (9) is rather different: here, the speaker uses *a pessoa* when speaking of a situation clearly involving only herself, as suggested by the use of a first-person singular in the same sentence. *A pessoa* and *uma pessoa* thus resembles human impersonal pronouns like *one* or the impersonal use of the second person singular: they can express a generalization potentially concerning the speaker, as in (8), or a reference to speaker potentially generalizable to other persons as well, as in (9). Semantically and pragmatically, the *pessoa* constructions resemble the English pronoun *one*, which has been described by Moltmann (2010) as a first-person oriented generic pronoun: “it does not stand for the speaker’s actual person, but rather for a range of individuals that the speaker identifies with or simulates” (Moltmann 2010: 440). Similarly to *one*, the link between *pessoa* and the first-person singular can be established in two different ways: through a generalization from the speaker’s experience to others, or through the inclusion of the speaker in a range of individuals being referred to (cf. Moltmann 2010: 441).

The impersonal use of *a pessoa* and *uma pessoa* in BP has recently been examined by Amaral & Mihatsch (2019), who study it together with the nouns *pessoal*, literally ‘personnel’, and *povo*, literally ‘people, nation’. Using the typology of human impersonal pronouns and the diagnostic sentences proposed by Gast & van der Auwera (2013), they show that both *a pessoa* and *uma pessoa* occur in veridical and generic clauses with both speaker-inclusive (‘One only lives once’) and speaker-exclusive (‘They eat dragonflies in Bali’) uses, as well as in non-veridical clauses that can be either modal (‘One should never give up’) or non-modal (‘What happens if one drinks sour milk?’). However, they cannot be used in veridical and episodic clauses (e.g. ‘They’re knocking on the door.’, ‘They’ve surrounded us.’, ‘They’ve raised the taxes again.’). Amaral and Mihatsch (2019) also discuss the eventual grammaticalization of the different nouns into impersonal pronouns. While *pessoal* and *povo* show more indices of grammaticalization in BP, like the omission of articles, this is not the case of *pessoa*; however, a relevant piece of evidence for an ongoing grammaticalization of *pessoa* is that it strongly favors the syntactic subject position (Amaral & Mihatsch 2019: 168; see Section 3.1 for a comparison with my data).

Although Amaral and Mihatsch (2019) claim that the emergence of impersonal uses of the nouns they examine, including *pessoa*, is a notably Brazilian phenomenon dating back to the 20th century, an examination of EP data and historical sources shows that impersonal uses of *pessoa* can be found as early as in the 16th century. One of the earliest mentions of impersonal uses of *pessoa* in the literature is from Nunes (1919):

Com sentido idêntico ao vocábulo mencionado, *ome* ou *homem*, a antiga língua, seguindo uma prática já existente no latim vulgar, usava empregar também o substantivo *pessoa*; hoje persiste ainda o mesmo uso com a diferença apenas, que ascende já ao século XVI, de fazer preceder esta palavra do numeral feminino *uma*. (Nunes 1919: 265)

[With identical meaning to the aforementioned vocable, *ome* or *homem*, the ancient language, following a practice already existing in Vulgar Latin, used to employ the noun *pessoa*; today the same use still persists, with the only difference, dating back to 16th century, to have it preceded by the feminine numeral *uma*.]

Unfortunately, Nunes (1919) does not provide any examples of these constructions. However, impersonal uses expressing generalizations comparable to example (1) can be found in 16th and 17th century literary sources, with both indefinite (10) and definite (11) articles.

- (10) *a fruta é de maravilhoso gosto, tão leve e sadia que, por mais que **uma pessoa** coma, não há fartar-se*
 ‘the fruit has a wonderful taste, so light and healthy that, no matter how much a person eats, s/he will not get tired’ (Fernão Cardim, *Carta de relação da viagem e missão a Província do Brasil*, 1590)
- (11) *e tenho observado que o chocolate é alimento dominante que, em se habituando a ele, não se toma quando **a pessoa** quer, senão quando quer ele*
 ‘and I have noticed that chocolate is an addictive foodstuff which, once being used to, is not eaten when the person wants [to eat it], but when it wants [to be eaten]’ (Manuel Bernardes, *Nova Floresta*, 1688)

Although I haven’t come across speaker-oriented uses comparable to (9) in historical data, it is difficult to say whether this means that such uses are a more recent development, or whether their absence is due to genre and medium included in the oldest literary corpora. In general, references to local persons (the speaker and the addressee) are more common in spoken, colloquial discourse than in most written genres.

Previous research provides little information on how grammaticalized the uses of *pessoa* are in Portuguese. Stolz (1991: 12) gives the phrase *A pessoa não deve preocupar-se* translated as “one (literally: the person) should not worry” as an example of the grammaticalization of ‘person’ into an indefinite pronoun. However, considering the semantic and syntactic differences between impersonal pronouns and indefinite pronouns (Cabredo Hofherr 2008), if *a pessoa* is to be considered a pronoun, it falls into the first category. One of the differences between impersonal and indefinite pronouns discussed by Cabredo Hofherr (2008) is how coreference to the pronoun is established: while indefinite pronouns like *alguém* ‘someone’ cannot refer back to themselves within a phrase, human impersonal pronouns can. Consider the two modified versions of example (1), presented here as (12a) where the second occurrence of *uma pessoa* is coreferential with the first occurrence, just like the null pronoun in the original, while in (12b) the second occurrence of the pronoun *alguém* ‘someone’ cannot establish coreference with the first occurrence. Another case of two coreferential occurrences of *uma pessoa* within a sentence is provided in the attested example (20).

- (12) a. *quando **uma pessoa**_i vai para um país estrangeiro, **uma pessoa**_i / \emptyset _i também vai à capital*
 ‘When you_i go ~ one_i goes to a foreign country, you_i also go ~ one_i also goes to the capital.’
- b. *quando **alguém**_i vai para um país estrangeiro, **alguém**_{*i/k} / \emptyset _i também vai à capital*
 ‘When someone_i goes to a foreign country, someone_{*i/k} also goes to the capital.’

As remarked by Cabredo Hofherr (2008), human impersonal pronouns and personal pronouns share the possibility of referring back to themselves, due to the fact that they have unique reference in discourse, while indefinite pronouns establish a new reference every time they are used.

While Stolz (1991) does not provide further data on the purported grammaticalization process of *a pessoa*, it should be noted that grammaticalization of noun phrases into referential devices is not uncommon in Portuguese. In addition to *pessoa* and the better-studied noun phrase *a gente* ‘the people’, there are other collective nouns used to refer

to speaker-including groups such as *a malta* ‘the gang’ and *o pessoal* ‘the personnel’ (cf. Amaral & Mithatsch 2019), as well as address pronouns deriving from nouns such as *você* ‘thou’ lexicalized from *vossa mercê* ‘your mercy’, *o senhor* ‘you’, literally ‘the sir’ and so on (Raposo 2014: 900). Thus, the eventual grammaticalization of *a/uma pessoa* is not a unique development but, rather, can be facilitated by the existence of various models in the language. I will return to this characteristic of Portuguese and its relation with the preference for expressed subjects (as opposed to null subjects) in the discussion (Section 4).

2 Data

The data analyzed in the present paper come from *Português Falado no Porto*, a sociolinguistic interview corpus collected by the author in the town of Porto in Northern Portugal in summer 2014 (Posio, in preparation). The corpus consists of 16 interviews of speakers with a relatively high socioeconomic status and contains ca. 77,000 words. All of the informants were born and were currently living in the region of Porto at the time of the interviews and they had either completed or were carrying out university studies. The interviews were carried out by four female linguistics students who are native speakers of the same language variety and were approximately 20 years old at the time of the data collection. The informants were aged between 22 and 69 years. The main topics of the interviews evolve around the studies and work of the informants. Other topics discussed are the speakers’ attitudes towards their home town and the variety of Portuguese spoken there. The interviews can be characterized as relatively familiar: in most cases, the informants were friends, acquaintances or relatives of the interviewers, and all of them use the familiar second person singular when addressing the interviewers. The interviewers use the same address form with most informants, except in three interviews where the third person singular with no overt subject is used to address the informant.

Due to the relative informality and the use of the second-person singular address, the corpus differs from most available resources of spoken EP, like the *CORDIAL-SIN Syntax-Oriented Corpus of Portuguese Dialects* (Martins 2000–). This corpus consists of data from fieldwork carried out for the construction of dialectal atlases of Portugal (including mainland, Madeira and Azores) between 1974 and 2000 and contains 600,000 words. The informants are elderly people with little formal instruction, living in rural areas, and the topics of the interviews are related with old times, lifestyle and vocabulary related with these topics. Some examples from the CORDIAL-SIN corpus are used in the present paper to illustrate phenomena only scarcely covered by my data.

3 Analysis

3.1 Referential range and definiteness the *pessoa* constructions

The NP *uma pessoa* occurs 52 times and *a pessoa* 33 times in the data (see Table 1). The majority of these occurrences fall under the category of potentially grammaticalizing referential devices that are the focus of this paper: they are found in all but one of the 16 interviews. Only two occurrences (one with indefinite and one with definite article) are referential and specific, i.e. refer to a particular person identifiable in the discourse universe, as in example (13).

- (13) *fui admitido para ser entrevistado ah surgiu uma outra vaga, (0.3) de uma pessoa que tinha seguido tinha (.) conseguido ah uma:, (0.3) um trabalho, (0.5) ah:, (0.4) em Angola*
 ‘I was admitted to be interviewed ah another position came up (0.3) of a person who had continued had (.) achieved ah a (0.3) a job, (0.5) ah, (0.4) in Angola’
 (male, 30)

Table 1: Occurrences of *uma/a pessoa* in the *Português Falado no Porto* corpus.

<i>uma pessoa</i> ‘a person’	subject	object	other	predicative	total
Potentially grammaticalizing use	28	1	0	0	29
Non-referential lexical use, modified with an adjective or relative clause	7	1	5	5	19
Referential lexical use	1	0	0	0	1
Unclear use	0	0	3	0	3
Total	36	2	8	5	51
<i>a pessoa</i> ‘the person’	subject	object	other	predicative	total
Potentially grammaticalizing use	27	2	1	0	30
Non-referential lexical use, modified with an adjective or relative clause	0	0	0	1	1
Referential lexical use	0	1	0	0	1
Unclear	0	0	1	0	1
Total	27	3	2	1	33

There are also 20 cases where the noun *pessoa* is modified by a restrictive adjective or relative clause, as in example (14). In these uses, the indefinite form *uma pessoa* is more frequent than the definite one, as is to be expected given the restrictive nature of the modifiers.

- (14) *uma pessoa que seja mesmo um bom profissional*, (0.2) *arranja em qualquer emprego*.
 ‘a person who really is a good professional, (0.2) gets along in any job.’ (male, 25)

In addition, there are 16 instances of the noun *pessoa* without an article that occur mostly in quantifier phrases (e.g. *cada pessoa* ‘each person, everyone’) or as predicate nominals (e.g. *como pessoa que vive na cidade* ‘as a person who lives in the city’). When the cases where *pessoa* is referential, modified by a relative clause, adjective or quantifier, or that were impossible to analyze (e.g., occurring in false starts or autocorrections) are discarded, we are left with 29 occurrences of *uma pessoa* and 30 occurrences of *a pessoa* that represent the potentially grammaticalizing use as a referential device (see **Table 1**). Similarly to the BP data examined by Amaral and Mihatsch (2019), all these uses are either generic or non-veridical and not episodic, following Gast and van der Auwera’s (2013) typology of impersonal pronouns (see Section 1.3).

As **Table 1** shows, both *uma pessoa* and *a pessoa* have a very strong preference for the syntactic subject position. In Amaral and Mihatsch’ (2019) data from BP, 75.6% of the occurrences of *a pessoa* were syntactic subjects, in contrast to the lexical NPs *o menino* ‘the boy’ (53.7%) and *a cidade* ‘the city’ (24.4%). In my EP data, the percentages of subject uses are even higher: 90% for *a pessoa* and 96.5% for *uma pessoa*. This is not surprising in the light of the grammaticalization hypothesis, given that the key function of human impersonal expressions is to defocus the agent of an action, and subjects typically represent the semantic agent. Canonical *man*-impersonal pronouns (French *on* and Germanic *man*) only occur in subject positions, and other human impersonals like Spanish *uno* or English *one* also tend to occur predominantly as syntactic subjects (Cabredo Hofherr 2008; Siewierska 2011).

Given that the indefinite *uma pessoa* and the definite *a pessoa* between have a similar distribution with regard to syntactic roles, the obvious question is whether these two NPs are functionally differentiated from each other in some other way. Although it would seem

logical that the definite variant would receive a speaker-oriented reading more often than the indefinite one, as the speaker is always definite in the discourse context, native speaker intuition suggests, rather, that the indefinite variant (*uma pessoa*) is always inclusive and the definite variant (*a pessoa*) can be both inclusive and exclusive (Martins 2019).

In order to tackle the question of whether the definiteness of the noun phrase correlates with the type of referential scope of *pessoa*, I classified the potentially grammaticalized occurrences into three categories: **speaker-exclusive** (oriented towards a third person, not including the speaker), **referentially generic** (not oriented towards the speaker, although the speaker may be included in the referential range) and **speaker-oriented** (occurring in a context where the intended referent is mainly the speaker, although other referents could also be involved; cf. Moltmann 2010). The classification is presented in **Table 2**. It should be stressed that the categories are not clear-cut, but rather represent different points on a continuum between speaker-inclusive and speaker-exclusive. In most cases, there is no independent evidence in the context in order to establish which type of referential scope the speaker has in mind, and all such cases fall into the “referentially generic” category.

In total, there were two occurrences of *a pessoa* and one occurrence of *uma pessoa* where the referential scope of *a pessoa* clearly excludes the speaker, as evidenced by contextual clues such as a contrast established between *pessoa* and a first-person singular or plural. Example (15), where the referent of *uma pessoa* can be construed as ‘any people not including us’ is clearly at odds with the generalization that only *a pessoa* (but not *uma pessoa*) can receive speaker-exclusive readings (Martins 2019). In this example, the speaker explains how the dialect spoken in Porto may sound bad to others, and is herself clearly not included in the scope of possible referents.

- (15) *fica mal eu admito h: fica muito mal uma pessoa ouvir-nos (.) mas lá, para Lisboa também não gosto.*
 ‘it sounds bad, I admit h: it sounds bad when a person hears us (.) but there, around Lisbon I don’t like it either.’ (female, 32)

Interestingly, however, the distinction between the indefinite and the definite article may be related to the type of potential third-person referents intended by the speaker and the particular semantic frame of the event being referred to. In (14), the referent may be understood as ‘any person who listens to us talking’, which explains the choice of the indefinite article. In contrast, in example (16) *a pessoa* can be understood to refer to a generic third person, i.e., neither the speaker nor the addressee, but the person is construed as definite due to the particular situation being described. Although the speaker does not refer to any concrete event anchored in the past but rather describes her work in general, the definiteness arises from the semantic frame of the interpretation scenario described by the speaker.²

Table 2: Referential scope of *a/uma pessoa*.

Referential range	3 rd person oriented, excludes the speaker	Referentially generic	1 st person oriented, includes the speaker	Unclear
<i>uma pessoa</i>	1	15	11	2
<i>a pessoa</i>	2	12	15	1

² The English subject pronoun (s/he) is given in brackets in the translation when there is a null subject in the original language of the example.

- (16) *então a pessoa (es)tá a falar, (0.4) e eu tinha que:, apanhar o que a pessoa (es) tava a dizer, percebes,*
 ‘so **the person** is talking (0.4) and I had to:, catch what the person was saying, you understand,’ (female, 40)

In those cases where the referential range concerns mainly or only the speaker, both *a pessoa* (15 cases) and *uma pessoa* (11 cases) are found. Both the definite and the indefinite NP receive their speaker-oriented interpretation by inference from the immediate discourse context, rather than by direct reference. Thus, in example (17) *a pessoa* occurs in a context where the general discourse topic is the speaker’s life and work. In the first line, the speaker uses the first-person plural to refer to herself and others in her situation; the following lines contain several clauses with *a pessoa* as the subject, while in the last line the speaker switches to a first-person singular.

- (17) *penso que o trabalho absorve-nos muito, (1.3) e acho que nos ocupa mesmo muito atualmente*
 ‘I think that the work absorbs us a lot, (1.3) and I think that it occupies us really a lot these days’

acho que a pessoa vive muito para o trabalho. (.) e já sai do trabalho muito cansada. (0.6) eh: (0.4)

‘I think that **the person** lives a lot for the work. (.) and already leaves the job really tired. (0.6) eh: (0.4)

[2 lines omitted]

ah: e portanto (.) esta é a vida atual e penso que enquanto a pessoa está a trabalhar que se- será da da mesma forma. (0.5) o trabalho é absorvente. (0.3)

‘oh, and therefore (.) this is the current life and I think that while **the person** is working that it wi- will be th the same way. (0.5) the work is absorbing. (0.3).’

se a pessoa um dia que se reforme se ainda estiver viva, (.) @ não sei em que estado estarei, @ se tiver saúde acho que vou passear, acho que hei de passear muito

‘if **the person** one day when (s/he) will retire is still alive, (.) @ I don’t know in what state I will be, @ if I’m healthy I think I will travel, I think I should travel a lot.’

(female, 51)

The remainder of the occurrences, where neither inclusive nor exclusive reading is suggested by contextual clues, are classified as “generic” in **Table 2**. In these cases (12 occurrences of *a pessoa* and 15 of *uma pessoa*), the speaker is not excluded from the referential range, but the reading is not directed towards the speaker, i.e. (s)he is not the primary referent. However, even in such cases the speaker may express a generic statement based on his or her personal experience. Thus, in example (18) the speaker utters a generic statement concerning ‘anyone’, but the context suggests that this generalization is drawn from his own experience from working abroad. In example (19), the generalization concerns ‘anybody’, but the context suggests that the speaker is, in fact, thinking chiefly of her daughter.

- (18) *se os outros fazem bem uma coisa uma pessoa tem mas é que, (0.4) ser humilde e: e copiar. (male, 30)*
 ‘if the others do something well, what **a person** has to do is, (0.4) be humble and copy.’

- (19) *tenho que pôr regras (.) ela tem que ter regras logo de pequenita tem que tem que ser (.) é é básico uma pessoa nasce já tem regras (.) tem que seguir as regras.*
(female, 32)

‘I have to establish rules (.) she has to have rules already as a young girl it has it has to be (.) it’s it’s basic a **person** is born (s/he) already has rules (.) (s/he) has to follow the rules.’

In conclusion, the qualitative analysis of the occurrences of *a/uma pessoa* suggests that the referential scope of both *uma pessoa* and *a pessoa* can be oriented towards the speaker or a third person (arguably also towards the addressee, although this is not manifested in our data), and the scope of reference is only deducible from contextual clues like the presence of a first-person singular in the immediate context (as in example 16) or a contrast established between *pessoa* and first-person singular (example 16) or plural (example 15). In other words, *a/uma pessoa* can be characterized as referentially underspecified or having “open” or “arbitrary” reference that can be specified locally based on inferences made by the interlocutors.

3.2 Evidence of grammaticalization of *a/uma pessoa*

When judging the degree of grammaticalization of *a/uma pessoa*, notions such as semantic bleaching or loss of phonological materials, considered as key symptoms of grammaticalization, are difficult to apply directly. On the one hand, since the word *pessoa* is already semantically very broad – inasmuch it only specifies that the referent is a human being – the change of meaning associated with a category shift from noun to pronoun is rather subtle (cf. Amaral & Mihatsch 2019: 165). Even though the pragmatic extension of the NP is evident in some examples, it is difficult to distinguish between a noun with a wide pragmatic usage scope and a referential device resembling a pronoun based on semantics and pragmatics alone. On the other hand, there do not seem to be any readily perceptible phonetic difference between the noun in contexts susceptible of grammaticalization and contexts where it functions as a lexical noun.

The potential evidence of grammaticalization of the *pessoa* constructions in EP is briefly touched upon by Posio (2017: 220), who suggests that the high frequency of the lemma *pessoa* in Portuguese, as opposed to its cognate *persona* in Spanish, could be evidence of grammaticalization. Ongoing grammaticalization typically correlates with a high usage frequency of the lexical item. Not only does grammaticalization most often affect items that already are highly frequent (e.g. *go* is more likely to grammaticalize into an auxiliary verb than *proceed*), but the pragmatic extension leading into new usage contexts and loss of semantic content can also boost usage frequency of lexical items. Thus, it is interesting to note that the normalized frequency of the lemma *pessoa* in my data is 62.07 times per 10,000 words, while in a similar sociolinguistic interview corpus from Peninsular Spanish, *Habla Culta de Salamanca*, the frequency of the cognate noun lemma *persona* is only 9.83 times per 10,000 words (Posio 2017: 220).

Posio (2017: 220) and Amaral and Mihatsch (2019: 163) also mention the repetition of the noun phrase to refer to itself, instead of using a personal pronoun. In my data there are no cases where the personal pronoun *ela* ‘she’ would be used to refer back to the NP; Amaral and Mihatsch (2019: 164) have one such example, but there *a pessoa* is left-dislocated and occurs right before the subject pronoun. On the contrary, there are several examples in the CORDIAL-SIN corpus, as well as in Amaral & Mihatsch’ (2019) data from BP and EP, where the whole NP is repeated within a short passage (see example 20). *A/uma pessoa* thus behaves like any personal pronoun in that it can only be referred to by itself or by a null subject.

- (20) *Mas esse peixe, já **uma pessoa** às vezes não o conhece. Não sabe de que peixe é, não é? Se **uma pessoa** visse a figura do peixe, já **uma pessoa** dizia: “Olha, pode ser a sardinha, pode ser carapau”.* (CORDIAL-SIN, VPA-30)
 ‘But that fish, sometimes a person does not even recognize it. (S/he)³ doesn’t know what fish it is, right? If a person saw the form of the fish, a person would say: “Look, it can be a sardine, it can be a mackerel.”’

A third eventual piece of evidence of semantic bleaching mentioned by Posio (2017: 220) is non-standard gender agreement. This is illustrated in (21), where the predicative *enterrado* ‘buried’ occurs in the masculine although the noun *pessoa* is feminine, probably reflecting the speaker-oriented reference. Since the speaker is male and refers to his own experience, this can be considered a case of semantic agreement (also referred to as, e.g., “notional” or “logical” agreement; Corbett 2006: 155).

- (21) [...] *eu digo muitas vezes isso, fui enterrado: vivo. (0.6) fui enterrado mas, pronto por breves instantes **uma pessoa** não sabe, olha, quando é enterrado não sabe se é por breves instantes ou seja fui enterrado cobriram-me com a capa meteram-me terra* [...] (male, 30)
 ‘I often say that, I was buried alive (0.6) I was buried but, for a short while a person doesn’t know, look, when (he) is buried (MASC.SG) (he) doesn’t know if it is for a short while, I mean, I was buried they covered me with the cape and put soil on me’

More examples of both gender and number mismatches can be found in the CORDIAL-SIN dialect corpus. In (22), the predicative *perdido* ‘lost’ is in the masculine singular and in the predicative *moços* in the masculine plural.

- (22) *Isso dá cabo, tira o viço ao sobreiro. **Uma pessoa vê-se perdido** quando é para arrancar a cortiça. Vem casca e vem tudo.*
 ‘It [a pest insect] destroys, it sucks the strength out of the cork oak. A person is lost (MASC.SG.) when it’s time to separate the cork. The bark falls out and everything falls out.’ (CORDIAL-SINMTV63)
- (23) *Quando **uma pessoa é moços**... Noutra tempo também havia uns moços que também brincavam com isso.*
 ‘When a person is young (MASC.PL.)... Another time there were also some youngsters who would also play with it.’ (CORDIAL-SIN: AJT23)

While in the examples (21)–(23) the masculine agreement probably reflects the gender of the speaker (all three speakers are male), at present I have no explanation to the observation that masculine agreement is found only with the indefinite form *uma pessoa* but not with the definite *a pessoa* in the CORDIAL-SIN corpus (Martins 2019).

3.3 Phonetic realization of a/uma pessoa

In addition to the three phenomena discussed in Section 3.2 (high frequency, referring to itself, non-standard agreement patterns) suggesting at least some degree of pragmatic extension and semantic bleaching, more robust evidence for grammaticalization could be found by looking at the phonetic level, if the impersonal uses of *a/uma pessoa* showed

³ While such uses are not episodic (i.e. anchored to a specific point in time), they are not fully generic either but could rather be considered an intermediary category between generic and episodic uses, as they are loosely anchored to a period of time in the past and typically occur in narrative sequences referring to past events. See Posio (2017) for a discussion of similar occurrences of impersonal second person singulars in Spanish.

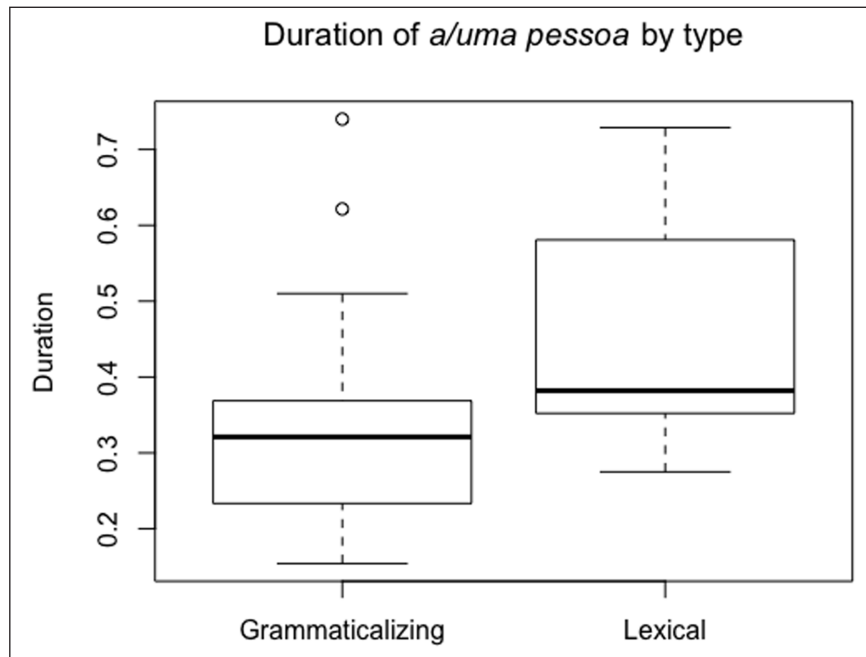


Figure 1: Duration of the sequence *a/uma pessoa* in potentially grammaticalized vs. lexical uses.

signs of phonetic attrition. Thus, I measured the duration of the sequences *a/uma pessoa* occurring without adjectival or relative clause modifiers using *Praat* and calculated their mean length, as shown in **Figure 1**. The Wilcoxon rank sum test with continuity correction applied to the data shows that the difference in the duration of the potentially grammaticalizing occurrences and the lexical occurrences of *a/uma pessoa* is statistically significant ($W = 449, p = 0.001854$).

As can be observed in **Figure 1**, the occurrences of *a pessoa* and *uma pessoa* classified as ‘Grammaticalizing’ are realized over 0.1 seconds shorter than in the ‘Lexical’ uses of these noun phrases, i.e. when they occur with modifiers or refer to an identifiable person in the discourse context. In the data, there was no significant difference in the duration of the definite (*a pessoa*) and the indefinite variant (*uma pessoa*), which were therefore analyzed together. Although speech rates obviously vary between individual informants and between stretches of conversation, the speech rate of the sequence where the *pessoa* constructions occurred was not taken into account in the analysis. Given that both potentially grammaticalized and lexical uses of *a/uma pessoa* are produced by all but one speaker in the data, it is unlikely that speech rate variations would affect the results, in particular because the observed differences are also found within the speech of individual informants. To what extent the difference in length reflects an eventual grammaticalization or rather relates to other semantic or pragmatic properties of the construction is discussed in Section 4.

4 Discussion

In this section, I discuss the findings presented in Section 3 in the light of previous research. The discussion is divided into three topics: evidence for an ongoing grammaticalization of *a/uma pessoa* (Section 4.1), pragmatic preferences and the emergence of human impersonals (Section 4.2) and the relation between human impersonal pronouns and the null subject properties of a language (Section 4.3).

4.1 Formal and functional evidence of an ongoing grammaticalization

As the analysis presented in Section 3 shows, the noun phrases *uma peessoa* and *a peessoa* are used in EP spoken discourse as a referential device with variable referential range: depending on contextual clues, they can receive generic and speaker-oriented, but also speaker-exclusive, third-person oriented uses. Such referential flexibility is not uncommon in the early stages of grammaticalization of referential devices such as (im)personal pronouns: for example, in Old French the pronoun *on* deriving from ‘man’ had several referential values including first and second person singular as well as plural (King, Martineau & Mougeon 2011: 473) and the pronoun *still* has a very wide range of uses, in spite of the main referential value being first-person plural (see examples 2–4). All occurrences of *a/uma peessoa* analyzed are either generic (i.e. not anchored to a specific point in time) or non-veridical, but not episodic; this is also typical of weakly grammaticalized impersonal pronouns, as the episodic uses tend to represent a later development (Siewierska 2011).

As for the results of the phonetic analysis, it is unclear whether the difference in the duration between the grammaticalizing uses and the noun uses of *peessoa* could be attributed to a phonological change accompanying the grammaticalization process. In more canonical cases of grammaticalization, the lexical items undergo a process of phonological attrition involving a loss of morphological content (e.g. *going to* developing into the auxiliary *gonna*). Since unstressed vowels tend to be highly reduced in EP in general, there is little phonological material that could undergo attrition in the sequence *uma/a peessoa* to begin with. Rather than serving as evidence for grammaticalization, the observed difference in duration between different uses of *a/uma peessoa* can be related with the fact that impersonal pronouns cannot receive contrastive stress (cf. an example like **One~you only live once* where the pronouns *one* or *you* cannot be interpreted impersonally if they are stressed). In addition, since the use of *peessoa* in the impersonal sense is more frequent than its use as a lexical noun, the different durations could be caused by a frequency effect affecting the pronunciation of homonyms. In general, the more frequent members of homonym pairs (like English *time* as opposed to *thyme*) tend to be realized shorter than the less frequent member (see Bybee & Napoleão de Souza 2019; Drager 2011; Gahl 2008, 2009; Lohmann 2018).

Although the constructions with the noun *peessoa* present some signs of grammaticalization like semantic bleaching, accompanied by a high usage frequency in comparison with the cognate lexeme *persona* in Spanish (Amaral & Mihatsch 2019; Posio 2017), there is no evidence of an ongoing change eventually leading to a “full” category shift from noun phrase to pronoun. The grammatical status of these constructions can be compared to that of *a gente*, literally ‘the people’, used for impersonal or first-person plural reference. Although there is ample evidence of the use of this referential device as the primary means of referring to first-person plural in BP, its combinatory properties remain somewhat different from those of “old” personal pronouns. For example, according to Taylor (2009), in BP *a gente* cannot be modified by numerals (**a gente os dois* vs. *nós os dois* ‘we two’⁴) or appositive nouns (**a gente os brasileiros* vs. *nós os brasileiros* ‘we Brazilians’). The same restrictions apply for *a/uma peessoa* in Brazilian and European Portuguese: modifying the NP by an adjective or a quantifier is only possible when *peessoa* is used as a lexical noun (e.g. *uma peessoa só* ‘just one person’, *uma peessoa portuguesa* ‘a Portuguese person’).

⁴ However, sporadic examples of *a gente* modified by numerals, at least in apposition if not in subject position, are found in EP corpora, e.g. Nascimento (1989: 482): *a gente os dois*, *a gente governava-se* ‘we two, we took care of ourselves’.

As for the diachronic evidence of an eventual grammaticalization process, there is little to say about the history of the *pessoa* constructions due to lack of studies based on historical data. Attested uses of *pessoa* from the 16th and 17th centuries (see examples 10 and 11) are not fundamentally different from the generic present-day uses. What remains unclear is whether the speaker-oriented uses that constitute almost half of the occurrences attested in my data are a more recent development, or whether they are just more frequent in spoken, colloquial language as opposed to literary texts. Interestingly enough, orientation towards the speaker seems to be more characteristic of *a/uma pessoa* in EP as opposed to BP (cf. Amaral & Mihatsch 2019).

Looking at the development of the specific readings of other human impersonal pronouns (e.g. the use of ‘you’ when the intended referent is the speaker, the use of a *man*-impersonal pronoun like French *on* for the first-person plural), Siewierska (2011: 80) suggests a grammaticalization cline where the specific singular readings are the last stage of the development, following generic readings. Although the distribution of the different readings of third person plural impersonals supports the existence of such a cline (as there are languages permitting all other readings except the singular, specific ones, but not the other way around), it is unclear whether such a cline can be assumed for human impersonals grammaticalizing from other sources. For instance, Auer & Stukenbrock (2018: 281) argue that the speaker-oriented use of the second-person singular pronoun *du* in German cannot be considered as a subsequent development based on the generic use, given that both the generic and the speaker-oriented use are found in the oldest sources where the construction is attested from the 18th century. More research on historical data, examining the eventual changes in the use and frequency of *pessoa* constructions, would be needed to shed light on this question.

4.2 Pragmatics of *a/uma pessoa*

The development of speaker-referring expressions from noun phrases has also been conceptualized in the literature without assuming an underlying grammaticalization process. Collins and Postal (2012) use the term “imposter” to refer to 3rd person NPs (or DPs) that refer to the speaker or the addressee (e.g. *the author, yours truly, my lady*), while also admitting that there are cases where “grammatically 1st or 2nd person forms are notionally some distinct person” (Collins & Postal 2012: 5). Working in a generative framework, these authors postulate that the deictic pronominal features of imposters are due to the presence of a null indexical pronoun in them. In the functional, usage-based approach adopted in the present paper, there is no need to assume an empty pronoun in order to explain the referential flexibility of these noun phrases, as the meaning of pronouns and other referential devices is negotiable and constructed by the speakers in interaction. Hence, in many languages second person singulars can be used to refer to the speaker, first-person plurals can be used to refer to the addressee, and a wide range of NPs formally in the third person can be used to refer to the addressee and the speaker.

In the case of Portuguese, the use of imposter noun phrases in different syntactic functions is particularly common, as evidenced by the multitude of address terms ranging from professional titles (*o doutor* ‘doctor’, *a professora* ‘teacher’) to honorifics (*o senhor* ‘mister’, *a senhora* ‘ma’am’) and proper names (*a Maria, o João*). The tendency to use third person noun phrases to address the interlocutor, instead of the second person singular, may explain why the use of the second person singular for impersonal or generic reference, or reference to the speaker, is less common in EP than in other related languages like Peninsular Spanish. The use of a second person form may constitute a face-threatening act in EP conversation, in particular when the speakers are not in very familiar terms with each other (Carreira 2005; Posio 2017), and arguably this may also reduce the use of the

impersonal second-person singular (see Section 2.1). Similarly, the use of noun phrases like *a/uma pessoa*, *a gente*, *a malta*, *o pessoal* (Raposo 2014: 900) to refer to the speaker or to groups including the speaker may be encouraged by the tendency of avoiding direct personal reference in discourse (Carreira 2005).

4.3 Emergence of impersonal pronouns and null subject properties

Generative studies on impersonal or non-referential subjects, most notably Holmberg (2005), have argued that languages that do not permit null subjects, i.e. so-called non-pro-drop languages, have overt non-referential subject pronouns, while so-called pro-drop languages rely on verbal morphology alone for the expression of non-referential subjects. Siewierska (2011) refines this claim by examining data from European languages, concluding that there is no strict one-to-one relationship but rather a correlation between subject expression and the existence of overt non-referential subject pronouns like *man*: non-pro-drop languages make greater use of *man*-impersonals, while pro-drop languages make greater use of other impersonalization strategies, like third person plurals. Amaral and Mihatsch (2019) refer to this correlation in order to explain the emergence of incipient human impersonal pronouns in BP, given that there is ample evidence to the effect that BP is losing its pro-drop properties. However, they consider the existence of the same or similar constructions in EP an open question, given that EP is “firmly considered a pro-drop language” (Amaral & Mihatsch 2019: 181). They suggest two possible explanations: either EP is also presenting a tendency towards overt subjects or the occurrence of these constructions may be a lexical influence from BP.

Given that *a/uma pessoa* constructions are found in EP historical data as well as in the CORDIAL-SIN dialect corpus, whose informants had not been widely exposed to Brazilian influences, the contact hypothesis suggested by Amaral and Mihatsch (2019) does not seem plausible. There is, however, evidence to support their first hypothesis, suggesting that the assumption of EP being radically different from BP with regard to its null subject tendencies needs to be revisited. A comparison of the expression of subject pronouns in variable contexts (i.e. permitting both expression and omission) between EP and Peninsular Spanish reveals that speakers of EP prefer overt subject pronouns to a much greater extent than speakers of the latter language. Posio (2012: 346) examines data from a corpus of spoken EP where as much as 48.9% of the first-person singular subject pronouns and 32.2% of the first-person plural subject pronouns are expressed, while in a comparable corpus of Peninsular Spanish only 34.6% of first-person singular and 4.5% of first-person plural pronouns are expressed in variable contexts (Posio 2012: 346). Another difference between these languages is that EP permits the use of third person personal pronouns for inanimate entities (Duarte 2000: 15) much more readily than Peninsular Spanish, where it is extremely rare (Enríquez 1984: 177). Thus, while EP may stand out as a consistent null subject language when compared with Brazilian Portuguese, a comparison with Peninsular Spanish shows that EP displays relatively high rates of expressed subjects in spoken discourse.

Whether or not the expression of non-obligatory subject pronouns in EP is increasing over time remains an open question, due to lack of diachronic studies – or even diachronic data, as the phenomenon is most readily observable in spoken discourse. Duarte (2000) presents some diachronic evidence from theater dialogue showing that subject pronoun usage has indeed increased considerably in BP over past centuries. She also connects this development with the emergence of “new” personal pronouns like *a gente* and *você* that do not participate in the variable subject pronoun use, as they can only be omitted in elliptical or coreferential contexts. However, the same pronouns have also emerged in EP, although it is true that their use remains less frequent than in the Brazilian variety due to the more persistent use of the older pronouns *tu* ‘thou’ and *nós* ‘we’ permitting variable

subject expression. More studies comparing authentic spoken data from the Brazilian and European varieties would be needed in order to confirm how similar or dissimilar these varieties actually are with regard to the usage frequency of subject pronouns or human impersonal constructions. However, considering the correlation between *man*-impersonals and expressed pronominal subjects (Siewierska 2011), is not surprising that a construction type resembling the *man*-impersonals has emerged in Portuguese (both European and Brazilian), but not in closely related Spanish, which is a more canonical null subject language in terms of discourse tendencies.

5 Conclusions

In conclusion, the findings discussed in the present paper support the view of a special status of the *pessoa* constructions in EP. The *pessoa* constructions appear to be semi-grammaticalized noun phrases that are used as referential devices, similarly to many other referential devices in Portuguese that share properties with both pronouns and noun phrases. There are nevertheless no strong signs of a full grammaticalization towards canonical *man*-impersonal pronouns (or indefinite pronouns, as suggested by Stotz 1991). While the present paper has examined the use of the *pessoa* constructions in a corpus of spoken EP, in particular the variety spoken in and around the town of Porto, more research is needed to shed light to the use of *pessoa* in other varieties of Portuguese as well as the diachrony of these constructions.

Acknowledgements

Part of the research presented in this paper was discussed in the workshop *Human impersonals in the outskirts: typological and functional perspectives to human impersonals* organized at Stockholm University, Sweden, in October 11–12, 2018. I wish to thank Max Wahlström for his insightful comments on different versions of this paper and for collaboration in the organization of the workshop. Special thanks are due to Ana Maria Martins, Isabel Margarida Duarte, João Veloso, Andrea Pešková, and the students of Linguistics from the University of Porto who collected and transcribed the data. All eventual errors and inaccuracies remain my own responsibility.

Competing Interests

The author has 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. Wolfgruber (Eds.), *Semantic and syntactic aspects of impersonality* (pp. 149–185). Hamburg: Buske.
- Auer, P., & Stukenbrock, A.** (2018). When ‘you’ means ‘I’: The German 2nd ps. sg. pronoun *du* between genericity and subjectivity. *Open Linguistics*, 4(1), 280–309. DOI: <https://doi.org/10.1515/opli-2018-0015>
- Barrett Brown, C.** (1931). The disappearance of the indefinite *hombre* from Spanish. *Language*, 7, 265–277. DOI: <https://doi.org/10.2307/409230>
- Blanco Canales, A.** (2004). *Estudio sociolingüístico de Alcalá de Henares* [Sociolinguistic study of Alcalá de Henares]. Alcalá de Henares: Universidad de Alcalá.
- Bybee, J., & Napoleão de Souza, R.** (2019). Vowel duration in English adjectives in attributive and predicative constructions. *Language and Cognition*, 11(4), 555–581. DOI: <https://doi.org/10.1017/langcog.2019.32>
- Cabredo Hofherr, P.** (2008). Les pronoms impersonnels humains: Syntaxe et interpretation [Human impersonal pronouns: Syntax and interpretation]. *Modèles linguistiques tome XXIX-1*, 57, 35–56. DOI: <https://doi.org/10.4000/ml.321>

- Cameron, R.** (1993). Ambiguous Agreement, Functional Compensation and Nonspecific tú in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change*, 5, 305–334. DOI: <https://doi.org/10.1017/S0954394500001526>
- Carreira, M. H. A.** (2005). Politeness in Portugal: How to Address Others? In L. Hickey & M. Stewart (Eds.), *Politeness in Europe* (pp. 306–316). Clevedon/Buffalo/Toronto: Multilingual Matters. DOI: <https://doi.org/10.21832/9781853597398-023>
- Collins, C., & Postal, P.** (2012). *Imposters. A Study of Pronominal Agreement*. Cambridge, MA: MIT Press. DOI: <https://doi.org/10.7551/mitpress/9780262016889.001.0001>
- Corbett, G. G.** (2006). *Agreement*. Cambridge, England: Cambridge University Press.
- Creissels, D.** (To appear). Impersonal pronouns and coreference: The case of French on. In S. Manninen, K. Hietaam, E. Keiser & V. Vihman (Eds.), *Passives and Impersonals in European Languages*. Amsterdam: Benjamins. Available online at <http://www.deniscreissels.fr/public/Creissels-ON.pdf>.
- Dias, A. E. da Silva.** (1918). *Syntaxe histórica Portuguesa* [Portuguese Historical Syntax]. Lisboa: Livraria Clássica Editora.
- Drager, Katie K.** (2011). Sociophonetic variation and the lemma. *Journal of Phonetics*, 39(4), 694–707. DOI: <https://doi.org/10.1016/j.wocn.2011.08.005>
- Duarte, I. M., & Marques, M. A.** (2014). As formas pronominais EU/TU – valor genérico e distância [Pronominal forms EU/TU – broad value and distancing]. *Revista Galega de Filoloxía*, 15, 69–85.
- Duarte, M. E. Lamoglia.** (2000). The loss of the Avoid Pronoun Principle in Brazilian Portuguese. In M. A. Kato & E. V. Negrão (Eds.), *Brazilian Portuguese and the Null Subject Parameter* (pp. 17–36). Frankfurt am Main: Vervuert Verlag. DOI: <https://doi.org/10.31819/9783964561497-002>
- Enríquez, E. V.** (1984). *El pronombre personal sujeto en la lengua española hablada en Madrid* [Subject personal pronoun in Spanish spoken in Madrid]. Madrid: Instituto Miguel de Cervantes.
- Fonseca-Greber, B., & Waugh, L. R.** (2003). On the Radical Difference between the Subject Personal Pronouns in Written and Spoken European French. In P. Leistyna, C. F. Meyer (Eds.), *Corpus Analysis: Language Structure and Language Use, Rodopi, Amsterdam*. (pp. 225–240). DOI: https://doi.org/10.1163/9789004334410_013
- Gahl, S.** (2008). *Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech*. *Language*, 84, 474–96. DOI: <https://doi.org/10.1353/lan.0.0035>
- Gahl, S.** (2009). Homophone duration in spontaneous speech: A mixed-effects model. *UC Berkeley Phonology Lab Annual Report (2009)*, 279–98. Available online at: http://linguistics.berkeley.edu/phonlab/annual_report.html#2009.
- 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.
- 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>
- Guirado, K.** (2011). Uso impersonal de tú y uno en el habla de Caracas y otras ciudades [Impersonal use of tú and uno in speech varieties of Caracas and other cities]. *Círculo de Lingüística Aplicada a la Comunicación*, 47, 3–27. DOI: https://doi.org/10.5209/rev_CLAC.2011.v47.39017

- Heine, B., & Kuteva, T.** (2002). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511613463>
- Holmberg, A.** (2005). Is there a little pro? Evidence from Finnish. *Linguistic Inquiry*, 36(4), 533–564. DOI: <https://doi.org/10.1162/002438905774464322>
- Kibrik, A. A.** (2011). *Reference in Discourse*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199215805.001.0001>
- King, R., Martineau, F., & Mougeon, R.** (2011). The interplay of internal and external factors in grammatical change: First-person plural pronouns in French. *Language*, 87, 470–509. DOI: <https://doi.org/10.1353/lan.2011.0072>
- Koenig, J.-P., & Mauner, G.** (1999). A-definites and the Discourse Status of Implicit Arguments. *Journal of Semantics*, 16(3), 207–236. DOI: <https://doi.org/10.1093/jos/16.3.207>
- Lohmann, A.** (2018). *Time and thyme* are NOT homophones: A closer look at Gahl’s work on the lemma-frequency effect, including a reanalysis. *Language*, 94(2), e180–e190. DOI: <https://doi.org/10.1353/lan.2018.0032>
- Lopes, C. R. dos Santos.** (2003). *A inserção de a gente no quadro pronominal do português* [The insertion of a gente in the pronominal framework of Portuguese]. Madrid: Iberoamericana, Frankfurt am Main: Vervuert Verlag. DOI: <https://doi.org/10.31819/9783865278494>
- Martins, A. M.** (2019). *Os indefinidos homem e pessoa na história do português* [The indefinites homem and pessoa in the History of Portuguese]. Colloquium presentation, University of Zürich.
- Martins, A. M.** (coord.) [2000-]. *CORDIAL-SIN: Corpus Dialectal para o Estudo da Sintaxe/Syntax-oriented Corpus of Portuguese Dialects*. Lisbon: Centro de Linguística da Universidade de Lisboa. Available online at: <http://www.clul.ulisboa.pt/en/10-research/314-cordial-s>.
- Meyer-Lübke, W.** (1900). *Grammaire des langues romanes. Tome troisième: Syntaxe* [Grammar of Romance Languages. Tome III: Syntax]. Paris: Welter. [French translation of Meyer Lübke, W. (1899). *Grammatik der Romanischen Sprachen, Dritter Band: Syntax*, Leipzig, Reisland.]
- 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>
- Nascimento, M. F. B.** (1989). *A gente, um pronome da 4a pessoa* [A gente, a pronoun of the 4th person]. In [s.n.], *Actas do Congresso sobre a Investigação e Ensino do Português*. Lisboa: Diálogo Compilação. 480–490.
- Nunes, J. J.** (1919). *Compêndio de Gramática Histórica Portuguesa (Fonética e Morfologia)* [Compendium of the Historical Grammar of Portuguese (Phonetics and Morphology)]. Lisboa. Livraria Clássica Editora. (8.^a ed.)
- Posio, P.** (2012). Who are ‘we’ in spoken Peninsular Spanish and European Portuguese? Expression and reference of first person plural subject pronouns. *Language Sciences*, 34(3), 339–360. DOI: <https://doi.org/10.1016/j.langsci.2012.02.001>
- 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.** (In preparation). *Português Falado no Porto* [Portuguese spoken in Porto]. *Sociolinguistic interview corpus*.

- 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>
- Raposo, E. B. P.** (2014). Pronomes [Pronouns]. In E. Raposo, M. F. B Nascimento, M. A. Coelho da Mota, L. Segura & A. Mendes (Eds.), *Gramática do Português* (pp. 883–920). Volume I, Lisbon: Fundação Calouste Gulbenkian.
- Siewierska, A.** (2011). Overlap and Complementarity in Reference Impersonals. Man-Constructions vs. Third Person Plural Impersonals in the Languages of Europe. In A. Malchukov & A. Siewierska (Eds.), *Impersonal Constructions. A Cross-Linguistic Perspective* (pp. 57–89). Amsterdam/Philadelphia: John Benjamins. DOI: <https://doi.org/10.1075/slcs.124.03sie>
- Siewierska, A., & Papastathi, M.** (2011). Third person plurals in the languages of Europe: Typological and methodological issues. *Linguistics*, 43(2), 575–610.
- Stolz, T.** (1991). *Forschungen zu den Interrelationen von Grammatikalisierung und Metaphorisierung: Von der Grammatikalisierbarkeit des Körpers, vol. 1: Vorbereitung (=ProPrinS 2)* [Studies on the interrelations between grammaticalization and metaphorization: On the grammaticalizability of the body, vol. 1: Preliminaries]. Essen: University of Essen.
- Taylor, M.** (2009). On the pronominal status of Brazilian Portuguese *a gente*. In P. Irwin & V. Vázquez Rojas Maldonado (Eds.), *New York University Working Papers in Linguistics, vol. 2, Spring 2009: Papers in Syntax*. Available online at: https://as.nyu.edu/content/dam/nyu-as/linguistics/documents/nyuwpl/taylor_09_a_gente_nyuwpl2.pdf.
- van der Auwera, J., Gast, V., & Vanderbiesen, J.** (2012). Human impersonal pronouns in English, Dutch and German. *Leuvense Bijdragen*, 98(1), 27–64.


How to cite this article: 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: 2, pp. 1–21. DOI: <https://doi.org/10.5334/jpl.254>

Submitted: 05 June 2020

Accepted: 28 October 2020

Published: 07 April 2021

Copyright: © 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Journal of Portuguese Linguistics* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 