

***Uno*: A corpus linguistic investigation of intersubjectivity and gender**

Federica Formato, University of Brighton; Vittorio Tantucci, Lancaster University

Abstract

Generic masculines – masculine forms used for women – are employed in many languages, e.g. English (Mills 2008), French (Coady 2018), Spanish (Bengoechea 2011) and German (Motschenbacher 2016), providing accounts of how gender is made visible in the language through morphological, lexical and syntactic units. These accounts are also linked with how gender is seen in societies and culture, reproducing an imbalance between women and men. Specifically, language discrimination against women is based on the idea that speakers orient themselves toward androcentric language, recognising ‘men’ as a metonym for the group ‘human being’ (Alvanoudi 2014), causing a linguistic invisibility of women.

Similarly, studies in Italian have also discussed the use of masculine forms to refer to, talk about and describe women (Cavagnoli 2013), or have shown how these are used in specialised (Nardone 2016, 2018) or media corpora (Formato 2014, 2016, 2019). This paper investigates the use of a specific (and underexamined) generic masculine in Italian – i.e. the indefinite pronoun *uno*.M.SG (in comparison with *una*.F.SG) labelled ‘impersonal masculine’ (Formato 2019: 69) – in three subcorpora of the Perugia Corpus (TV, Web, and Spoken; Spina 2014). *Uno*.M.SG is seen as constructing ‘extended intersubjectivity’, that is the awareness of a general third party (3rdP) acting as the social bearer of the utterance (Tantucci 2013, 2016, 2017a). The results show that the masculine impersonal *uno*.M.SG is widely used in the three subcorpora and in several functions, confirming that grammatically gendered language is still employed within a ‘masculine as a norm’ order.

Keywords: gendered language; grammatical gender; Italian; extended intersubjectivity; masculine as a norm

1 Introduction

Generic masculines have often been connected to a ‘masculine as a norm’ perspective that tends to hide women by rendering them invisible in the language. While speakers of some languages (e.g., English) have so far managed to adopt more neutral language (e.g., replacing the generic *he* with *he/she*, *s/he*, singular *they* and, in some cases, generic *she*; Earp 2012; Mills 2008), other languages are still governed by grammatical masculine terms that are used in their singular form to refer to women and in their plural forms to refer to undefined gender referents and mixed-gender groups (as well as groups of women only). In this paper we investigate a specific linguistic phenomenon, that of *uno* (one.M.SG) (and *una*, one.F.SG) as a marker of extended intersubjectivity, that is the awareness of a general third party (3rdP) acting as the social bearer of the utterance (Tantucci 2013, 2016, 2017a). As explained in detail in the methodology section, to investigate the indefinite pronoun *una* and *uno* we interrogated the Perugia corpus (Spina 2014) and, more specifically, the subcorpora TV, Web, and Spoken.

We start with an outline of the grammatical debate on generic masculines, providing an overview of previous studies on the topic. This is followed by the methodology for the study and the results of our analyses. We conclude with a discussion of the implications of the findings in light of the way gendered language is used through a cognitive perspective in the corpus by Italian speakers. The research question that this paper addresses is as follows: *How is the generic masculine uno used as an intersubjective marker in spoken and written subcorpora of the Perugia corpus?*

The aim of this research question is to examine how the preferred use of the generic masculine *uno*.M.SG in the subcorpora of the Perugia corpus demonstrates the relation between the linguistic choice and the undergoing cognitive processes of the speakers, who sees the masculine form as a neutral one. The novelty of this investigation lies in examining impersonal forms (see section 3 for an overview of the functions) rather than pair terms – i.e. masculine *vs* feminine – used to refer to job titles (these being the most controversial and the most investigated, see Formato 2016, Nardone 2018).

2 Italian as a grammatical gender language

In this section, we review gender in the Italian language and discuss the notions of availability (i.e. possible options to describe, talk about and address female and male referents), and use (what options are used by the speakers).

Italian is among those that are defined as grammatical gender languages, having a complex morphological, lexical and syntactic system, in which gender is ‘overt’ (Corbett 1991), i.e. gender is attributed to female or male referents through, mostly, morphological inflections which are visible in the language, e.g. *ragazza* (girl), *ragazzo* (boy) (for a full account of grammatical gender in Italian see Marcato and Thüne 2002 and Formato 2019). Grammatical gender is attributed to human beings as well as to objects, though to these last arbitrarily; some scholars (Boroditsky, Schmidt and Phillips 2003) show that gender stereotyping can also be found in describing objects according to their grammatical gender, e.g. using traditionally feminine traits to describe objects with grammatical feminine inflections (e.g. *bridge* in Spanish, *la*.F.SG *puente*.F.SG) or traditionally masculine ones to describe those with grammatical masculine inflections (e.g. *bridge* in German, *der*.M.SG *Brücke*.M.SG). In this paper, we are interested in explaining and discussing (mis)uses of grammatical gender in relation to human referents. Previous studies in Italian (Cavagnoli 2013; Formato 2016, 2019; Fusco 2012; Nardone, 2016, 2018) and other languages (for English, see Mills 2008, for Spanish see Bengoenecha 2011, for French see Coady 2019) demonstrate that widely used generic masculines (masculine terms used to talk about and address female referents) are discriminatory as they tend to hide women, reproducing historical gender imbalance. This is mostly the case for terms which are used to describe women in male-oriented spaces, e.g. politics. Language intertwines with (imbalanced) social systems, contributing to change how masculine and feminine grammatical units are perceived and used. In other words, as argued by Motschenbacher (2016: 152), ‘fixed semantic ascriptions are replaced by complexities of contextual meaning potential’. Not only context, speakers are deemed to have a paramount role in producing this potential and, in doing so, reproducing unequal structures in society. On this topic, Motschenbacher (2016) convincingly explains the active role of speakers in shaping discursive structures. As an example of this, Formato (2019) investigates a social media post that advocates the use of feminine inflections for women, and finds that speakers justify their *opposing* linguistic choices – that is preferring generic masculines to feminine forms for female referents – according to four different reasons (adapted from previous work by Mills 2008): a) opposition to traditions of language, e.g. women not wanting to use the feminine to refer to themselves notwithstanding the grammatical option; b) *there are more important things to deal with*, describing this topic as irrelevant to other ones, c) un-aesthetic nature of the forms, attributing to feminine inflections the status of having been recently created (accentuating that the masculine form is the only existing one for specific terminology related to traditionally male workplaces) or sounding *awful*; and d) (mis)understanding of grammatical gender, that is

creating feminine forms which do not follow grammatical rules or are semantically incorrect (e.g. **sindachessa*¹ F.SG instead of the correct *sindaca* F.SG).

Generic masculines in cognitive-based literature fall within the understanding that men are the prototype of the category ‘human beings’ (Alvanoudi 2014; Martin and Papadelos 2017). This metonymy is connected with the ‘[s]ocial stereotypes [...] in which one member of the category is used as a vehicle for understanding the category as a whole’ (Alvanoudi 2014: 54). Starting from this, the phenomenon under investigation in this paper, that is *uno*.M.SG, is labelled ‘impersonal masculine’ (Formato 2019), i.e. indefinite pronouns used in the masculine forms to refer to people regardless of their gender, as exemplified below:

- (1) *Uno*.M.SG *dovrebbe saper curare il pianeta.*
‘*One*.M.SG should know how to look after the planet.’
(Formato 2019: 69)

Uno.M.SG is studied (as explained in 3) in comparison with the feminine form *una*.F.SG with the aim to investigate which form is used more and how in the chosen corpus. As for other generic masculines and the literature discussed above, it is evident that masculine forms cannot not only be seen through the lens of grammar. They are used to express a wider linguistic and social androcentric experience. These *unstable* understandings and *motivated* use (Abbou 2011) of grammatically masculine and feminine terms in Italian, advantaging men as more linguistically prominent with respect to women, has been discussed in several contexts, e.g. parliament (Formato 2014), in the news when referring to female ministers (Formato 2016), in job adverts (Nardone 2018), in lexicography (Fusco 2012), and in legal language (Cavagnoli 2013). This accounts for discrimination against women in male-oriented contexts, e.g. politics where women are still facing obstacles and are subject to double standards (for an overview of the gender gap in Italy, see Formato 2019).

Generic masculines and feminine forms are here seen through the lens of liberal feminism (Bucholtz 2014), that is the efforts in establishing (linguistic) parity between men and women, a perspective that is proper to second wave feminism. Feminization as a strategy to defy generic masculines, is problematic for some: for instance, queer linguists argue that the binary system (masculine-feminine) is here reproduced, casting doubt on how linguistic equality can be achieved. As discussed by Motschenbacher (2014), neutralization is the only strategy that would allow for a sort of *justice*, yet admitting that this is not always possible. Similarly, Abbou (2011) agrees that other strategies other than feminization (in French) could be used (e.g. adopting a gender *neutral* vowels, i.e. *E*). While it is not our intention to reproduce the gender (essentialist) binary, we recognize that in Italian, feminine forms are still to achieve the same status as the masculine ones, both in how language is used and how speakers perceive it. Guidelines and language recommendations – mostly originating from grass root initiatives in sporadic (work)places – have recently appeared; these follow the seminal ones by Sabatini (1987, 1993), commissioned by the government in the late 80s. One example being the guidelines published on the website of the Ministry of Education (MIUR), titled *Linee guida per l’uso del genere nel linguaggio amministrativo* (Guidelines for the use of gender in the administrative language, 2018). These, as well as some others, focus on recommending a non-sexist use of Italian as far as (feminisation of) job-titles are concerned disregarding other problematic and discriminatory grammatical units, e.g. pronouns.

Neutralization could, potentially, be possible in Italian in many occasions (as well as for other languages, see Motschenbacher 2014 for German and English). In general terms, this can

¹ *Sindachessa* is considered to be incorrect because of the suffix –essa (etymologically related to *wife of*, as for the English –ess, *Earl/Countess*). Sabatini (1987) explains that the feminine form of the root *sindac-* is *sindaca*.

be achieved: i) through the use of the split form, i.e. masculine plus feminine (or vice versa), e.g. *uno*.M.SG or *una*.F.SG (one or one), ii) through a disembodiment of the human referents, e.g. *il corpo docenti* (the teaching body) instead of split forms such as *professoressa*.F.PL e *professori*¹.M.PL (female and male teachers), iii) through the use of * for the written genres, e.g. *un** *dovrebbe saper curare il pianeta*.. In relation to the phenomenon under investigation, there seems to be also another option aimed at replacing impersonal masculines: iv) using *si* instead of *uno*.M.SG (or *una*.F.SG), e.g. *si dovrebbe saper curare il pianeta* (one should know how to look after the planet).

However, some of these choices – especially i) and iv) - would not convey the speakers' involvement in the same way *uno* M.SG or *una* F.SG do, selecting a somewhat *more* impersonal form than the indefinite pronoun, while others – ii) and iii) might either sound unnatural and not economically viable or, also, limited, e.g. the * cannot be used in the spoken register and can cause problems for the other agreeing elements that needs neutralization in the sentence.

It is for this reason that we believe that investigating *uno* (and *una*) through the perspective of extended intersubjectivity, explained below, can shed light on how gendered language is used across time.

3 Extended intersubjectivity: a (novel) cognitive perspective

This section discusses the role of *uno* and *una* not only in relation to its gender properties (grammatical and society-bound, as discussed in Section 2), but also as markers of extended intersubjectivity (henceforth, E-I). In cognitive psychology, intersubjectivity is generally discussed with reference to the so-called theory of mind (ToM) or mind reading (i.a. Apperly 2010; Goldman 2006). More specifically, ToM hinges on the ability to attribute mental states – beliefs, intents, desires, pretending, knowledge, etc. – to oneself and others and to understand that others have beliefs, desires, and intentions that are different from one's own (Premack and Woodruff 1978). In linguistics, Benveniste ([1958] 1971) was the first to point out that “discourse [...] is language in so far as it is taken over by the man [sic] who is speaking and within the condition of intersubjectivity, which alone makes linguistic communication possible” (p. 230). Intersubjectivity is also discussed by Schrifin (1990:142), who defines it as the interaction between an actor's actions – those intended to be perceived and designed as such, as well as those not so intended – and an audience's interpretation of all the information stated and implied. Verhagen (2005) distinctively grounds his analysis of intersubjective constructions in argumentation theory (Anscombe and Ducrot 1983), with a special focus on the cognitive construals activated by the speaker/writer (Sp/w) and shared by the addressee/reader (Ad/r) when an intersubjective construction/strategy comes into play. From this angle, every construing process is motivated by the presence – actual or virtual – of a Sp/w and an Ad/r sharing the conceptual representation of what is said. Different from Verhagen, Traugott (2012) suggests that intersubjective reanalysis of a construction involves the semantic shift from a more Sp/w-centered meaning (comparatively more subjective) to one focusing mostly on the Ad/r (comparatively more intersubjective). She provides evidence for discourse markers in English being used as hedges, constructions of (im)politeness such as the Ad/r-oriented use of *a bit of*, occurring in contexts where Sp/w tries to mitigate potential face-threats. Intersubjectivity has also been discussed in contexts where joint attention and deixis are at play. Langacker (1987, 1990, 1991) suggests that pronouns such as *I*, *we*, and *you* have the function of foregrounding the speaker's communicative setting (what he defines as the “ground”) to identify the referent. In his approach, the conceptualizers' awareness of the *here-and-now* of the speech event can be semantically encoded by deictics that implicitly establish spatial relationships among speakers and objects. Similarly, demonstratives and determining elements like *such* (in English) and *zulk* (in Dutch) are elsewhere also considered intersubjective, serving

to create a “joint focus of attention” (Diessel 2006:465) by which the speaker negotiates discourse referent tracking for the hearer (Ghesquière 2009; Ghesquière and Van de Velde 2011).

Whether from a synchronic or a diachronic perspective, accounts of intersubjectivity are traditionally centered on the *here-and-now* of the conversation, hinging on the awareness that the speakers have of one another through interaction. In this respect, Nuyts (2001a, 2001b, 2012) holds a different view, as he mainly focuses on intersubjective construals of modal meanings “presented as being shared between the assessor and a wider group of people, possibly (but not necessarily) including the hearer” (Nuyts 2012: 58), such as the constructions *it is likely* and *unfortunately* (in English). Drawing on that, Tantucci (2013, 2017a, 2017b) distinguishes between meanings that are specifically aimed at addressing the Ad/r’s potential reactions to what is said and meanings that include a more-or-less general third party, who conceptually functions as the social bearer of the utterance. The former is defined as immediate intersubjective (I-I) and diachronically precedes further reanalysed functions encoding E-I. An illustrative case of I-I construction is the chunk *you don’t want X* when it specifically encodes the Sp/w’s awareness of a specific interlocutor:

(2) **You don’t want to be married.** You are too young – you are.
(COHA – May Agnes Fleming, *A Terrible Secret*, 1874 reported in Tantucci 2017a:105)

With the employment of *you don’t want X*, Sp/w idiomatically informs Ad/r of his/her wants, pre-emptively addressing what Ad/r might be feeling/thinking about getting married. It could be paraphrased as *Aware of what you might think, I am telling you that you don’t want X*. Here the intersubjective awareness conveyed by *you don’t want X* in (2) does not exceed the here-and-now of the conversation.



Figure 1 Immediate intersubjectivity (Tantucci 2013: 217, 2017a: 92)

The line sketched in Figure 1 symbolically connects the interlocutors’ minds during the speech event. The main point of Figure 1 is to represent I-I as a mirrored form of mutual awareness limited to the speech participants (Sp/w and Ad/r). Consider the case of the usage of *you don’t want X* in (3):

(3) But, after fasting so long **you don’t want** to eat too much at first.
(COHA – H. Irving Hancock, *Dick & Co. Start Things Moving*, 1911 reported in Tantucci 2017a:105)

This usage of *you don’t want X* is an instance of a more extended form of intersubjectivity (see Figure 2). Formally, it still includes phoric reference to a second person, yet, while Sp/w’s proposition also profiles the intent to prevent some addressee’s virtual intention, Ad/r is nonetheless no longer a specific interlocutor (s/he could be anyone).

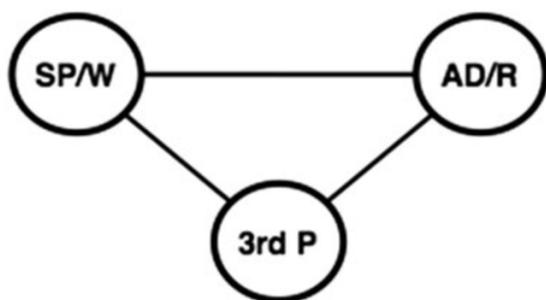


Figure 2 Extended intersubjectivity (Tantucci 2013: 218, 2017a: 94)

Extended intersubjectivity is an important dimension when phoric functions acquire a social meaning. A corpus-based analysis of the COHA (Corpus of Historical American English, Davies 2010) shows that the *you don't want X* construction shifts diachronically from an exclusively I-I usage to a new generic E-I one (Tantucci 2017a: 108). When E-I functions are at play, a general social persona is construed as the social bearer of the utterance, supporting the good sense of what is said.

While *you don't want X* is employed assertively in both (2) and (3), the latter could be directed to anyone, viz. an assumed 3rdP, who is expected to react exactly as the Ad/r would: *after fasting so long you [as anyone else] don't want to eat too much at first*. The shift from I-I to E-I thus corresponds to a reanalysis from a personal meaning (oriented towards the Ad/r) to a social one (extended to a generic 3rdP). The mismatch between I-I and E-I is easily tested by substituting you with (no-)one:

- (4) (a) ***No one** wants to be married. You are too young – you are. [I-I]
 (b) But, after fasting so long **no one** wants to eat too much at first. [E-I]

The E-I dimension of illocutionary acts is especially relevant when ideological construals of social personas are at play. The construals of *you* in (3) and *one* in (4) both express a social meaning, as they refer to what should be expected by anyone to feel, act or say. Whilst extended intersubjectivity is inherently characterized by generic reference, on the other hand not all generic referential expressions are intersubjective. What defines intersubjectivity is a process of 'thinking about thought' (cf. Apperly 2010), which is indeed at stake when a social persona's will is at play (see again (3) above), yet not when a form of generic reference is realised as such, e.g. *a car is a vehicle*, or *clouds are white*. In this sense, the dimension of extended intersubjectivity becomes a crucial one when ideology needs to be disentangled by conventionalised language use. This is because it provides the tools to identify feelings, narratives or ideologies that a speaker considers to be 'socially normal', or in other words what anyone in society would by default agree upon, sympathise with or relate to.

A crucial aim of this paper is thus to unveil whether the generic social persona that is construed in phoric functions of extended intersubjectivity is grammatically marked for gender. Our hypothesis is that the masculine *uno* is the prototypical marker for E-I reference, that is, when generic reference intersects with the emotions, feelings and beliefs of the 'general social persona', the masculine version of the marker is significantly at play. This will be tested in the following corpus-based analysis.

4 Methodology

Starting from what has been discussed in the literature, in this section we outline the methodology used to investigate the occurrences of *uno* (and *una*) in the Perugia corpus (Spina 2014, PEC henceforth). Before delving into a detailed explanation of the corpus, we argue that choosing to investigate a corpus (and using techniques pertaining to the field of corpus

linguistics) aims at producing results that demonstrate ‘the discursive formation, the cultural relativity, heterogeneity, potential incoherence and ideology-relatedness of gender language structures’ (Motschenbacher 2016: 157). Examinations of gender in corpora have extensively shown discriminatory and imbalanced language that was detrimental to women (see Formato 2016, Nardone 2016 for Italian, Baker 2010, 2014 for English). Moreover, providing a diachronic view allows for stronger generalizations about the use and the spread of gendered patterns (Motschenbacher 2016).

The selected corpus – PEC – contains 26 million words (1990-2010) distributed in ten genres, three of which are taken into consideration here, that is, TV, Web, and Spoken. Each of these genres contains text from several subgenres; more specifically, the Web subcorpus contains blog, chat, forum, social network (these, although written, are arguably similarly spontaneous as spoken data), and Wikipedia; the TV subcorpus includes drama, advertising, shows, sport commentary, sport reports, panel shows, and the news; and the spoken subcorpus consists of songs, conference papers, face-to-face conversation, institutional talk, legal talk, religious talk, interviews, and material delivered in class. We choose PEC, hosted by CQPweb, for its richness of data and its functions, for example, XML annotation and part-of speech tagging (POS henceforth).

In the following paragraphs, we explain the step-by-step procedure conducted to investigate the data. We believe that examining the impersonal masculine *uno*.M.SG (together with the feminine *una*.F.SG) can shed light on the quantitative and qualitative differences of, specifically the mostly used *uno*.M.SG. In order to optimize the research, we conducted a syntax query using POS, an example of which is: “[word=“uno”][pos=“VER:fin”][word=“uno”][word=“non”]”, restricted to “genere: 10_WEB”. This syntax query allowed us to retrieve all occurrences of *uno*.M.SG followed by a finite verb or *non* ‘not’ (as this would be likely followed by a verb); three similar queries were run to collect data from the subcorpora chosen for this investigation. The same queries were used to collect the occurrences of *una*.F.SG. The searches of *uno*.M.SG and *una*.F.SG in the three genres were imported in an Excel file in an extended linguistic context of 50 words, and a first selection of occurrences was performed. In other words, we cleaned the imported corpus, excluding occurrences that were not suitable to investigate *uno*.M.SG and *una*.F.SG as intersubjective markers. By using a syntax query (as explained above), occurrences of *uno*.M.SG and *una*.F.SG employed as the indefinite article (e.g., *una*.F.SG *casa sull’albero* ‘a tree-house’) and functioning as a numeral for people (e.g., *uno*.M.SG *studente ha disertato la lezione* ‘a student has not attended the class’²) were automatically excluded. In addition to these cases, we manually excluded those occurrences in which *uno*.M.SG and *una*.F.SG were employed to list or differentiate between two people (e.g., *Ma sono i fratelli di Saverio? Sì, uno*.M.SG *è lo zio Gino*. ‘Are these Saverio’s brothers? Yes, **one** is uncle Gino.’). More interestingly, we also removed those instances where *unique* gendered experiences were embedded, that is, where the impersonal masculine and the feminine seemed to be used in relation to known gendered referents rather than to construct a universal, yet impersonal, E-I experience, as in the following extract:

- (5) *E questa è anche una cosa che a volte passa in secondo piano, che è bello arrivare a Parigi avendo vinto come ha vinto Pantani ed è bello avere il classico seguito, cioè dei corridori che hanno diviso la strada, non esiste che uno*.M.SG *vada via a quattro giorni dalla fine.*

‘And sometimes this is backgrounded, it is great to reach Paris having won as Pantani [an Italian cyclist] did and it is great to have the rest following, that is the cyclist that have divided the road, it is impossible to think that **one** leaves four days before the end.’

From the extended context, we understand that this occurrence refers to the Tour de France, a cycling race in which only male cyclists can participate.³ Further exclusions were taken into consideration, for example, when the negative *non* was followed by an adjective, when there were imprecisions in transcriptions and/or tagging, and also when *una.F.SG* or *uno.M.SG* referred to objects.

Following these considerations, we do not exclusively focus on the frequencies of *uno.M.SG* and *una.F.SG* in the three genres; we also take into consideration specific linguistic variables, namely, polarity, tense, phoricity, illocutionary force, and sentence type. In Table 1, we offer an overview of those that we have considered as tangent linguistic phenomena to examine our data:

Category	Subcategories	Extra information
Polarity	Negative, positive	
Tense	E.g., present tense, past tense	
Phoricity	Non-direct reference, anaphoric and cataphoric	Depends on whether, in the 50-word linguistic context, <i>uno</i> and <i>una</i> refer to a non-gender specific human referent mentioned before (anaphoric) or after (cataphoric) the instance, or for which there does not seem to be a referent on either side
Illocution	assertive, directive, evaluative ⁴	
Sentence type	hypothetical, declarative, interrogative, exclamative	

Table 1 Categories and subcategories investigated for all occurrences of *uno* and *una*

In the analysis section, we highlight the significant interplay between the E-I markers *uno.M.SG* and *una.F.SG* in relation to the above categories. To cross-investigate statistical patterns, we have interrogated statistical tools, whose functions and applications are explained in the following sections.

5 Analysis of the subcorpora

In this section, we present the results of the investigation and we discuss what they mean in relation to gender and intersubjectivity.

Table 2 introduces the absolute frequency (AF) and the percentage (%) of occurrences of the impersonal masculine *uno.M.SG* and the feminine *una.F.SG* as E-I markers.

	TV		Web		Spoken		Total	
	AF	%	AF	%	AF	%	AF	%
<i>Uno</i>	110	30.64	119	33.14	127	35.37	356	99.16
<i>Una</i>	1	0.27	1	0.27	1	0.27	3	0.83
Total	111	30.91	120	33.42	128	35.65	359	100

Table 2 Absolute frequencies and percentages of *uno.M.SG* and *una.F.SG*, divided into subcorpora

Unsurprisingly, Table 2 shows that the impersonal masculine *uno.M.SG* is used markedly more often than *una.M.SG* across the three subcorpora (TV, Web, Spoken), with 99.16% of the overall occurrences. This is already an obvious indicator of the remarkable unbalance between generic masculines and feminine forms when generalized instantiations of a social persona's behavior, feelings, or beliefs are at play.

The limited use of *una.F.SG* and the corresponding wide use of *uno.M.SG* demonstrate how (female and male) speakers tend to conceptualize E-I markers in relation to men as the

prototypical category of human being (also referred to as *people=male, male=people*, Hamilton 1991), and support the results of previous (corpus) studies on masculine forms within the ‘masculine as a norm’ framework (Formato 2016, 2019; Nardone 2016, 2018). Because of the interplay between grammatical and social gender, some occurrences of *una.F.SG* are difficult to rate with respect to a common and universal experience, unlike instances of the impersonal masculine *uno.M.SG*. The instance of *una.F.SG* being used as impersonal, uttered by a female speaker, is presented in (6):

(6) *Una.F.SG chiama e videochiama e mi si addebita anche una chiamata su un numero per cui non ho chiesto assistenza.*

‘**One** calls and videocalls [them – a telephone company] and they have charged me for these calls too therefore I did not ask for assistance.’

It should be noted that the possibility cannot be excluded that the gender of the speaker is here relevant (however it is not always in our corpus) to the generalized experience of the impersonal *una.F.SG*, referring to herself yet through an impersonal form.

Having discussed the only occurrence of *una.F.SG* as E-I, we now move to illustrate in detail how *uno.M.SG* functions in relation to the linguistic phenomena described above.

5.1 Characteristics of *uno.M.SG* in use

In this section we provide a usage-based account aimed at unveiling formal, pragmatic, and contextual dimensions that significantly concur with speakers’ construing of the generalized social referent of *uno.M.SG*. The notion of usage-based linguistics is becoming more and more central in cognitively-inspired linguistics (Gries 2011). The study of cognition based on corpus data is arguably indirect, despite fulfilling desirable qualities such as being natural, representative, and plentiful (cf. Arppe et al. 2010). Still, the domain of cognitively orientated corpus linguistics is growing (i.a. Gries and Stefanowitsch 2006; and Stefanowitsch and Gries, 2006; Tantucci 2018; Tantucci & Di Cristofaro 2019). Meta-discussions on cognitive corpus linguistics often underpin the importance of converging evidence, and, hence, the interaction of corpora with other sources of data. In our case we discuss overt linguistic forms that express a process of ‘thinking about thought’, which is the core issue of experimental psychological on Theory of Mind. The present usage-based analysis canters on illocutionary and formal characteristics of extended-intersubjective usages of *uno.M.SG*.

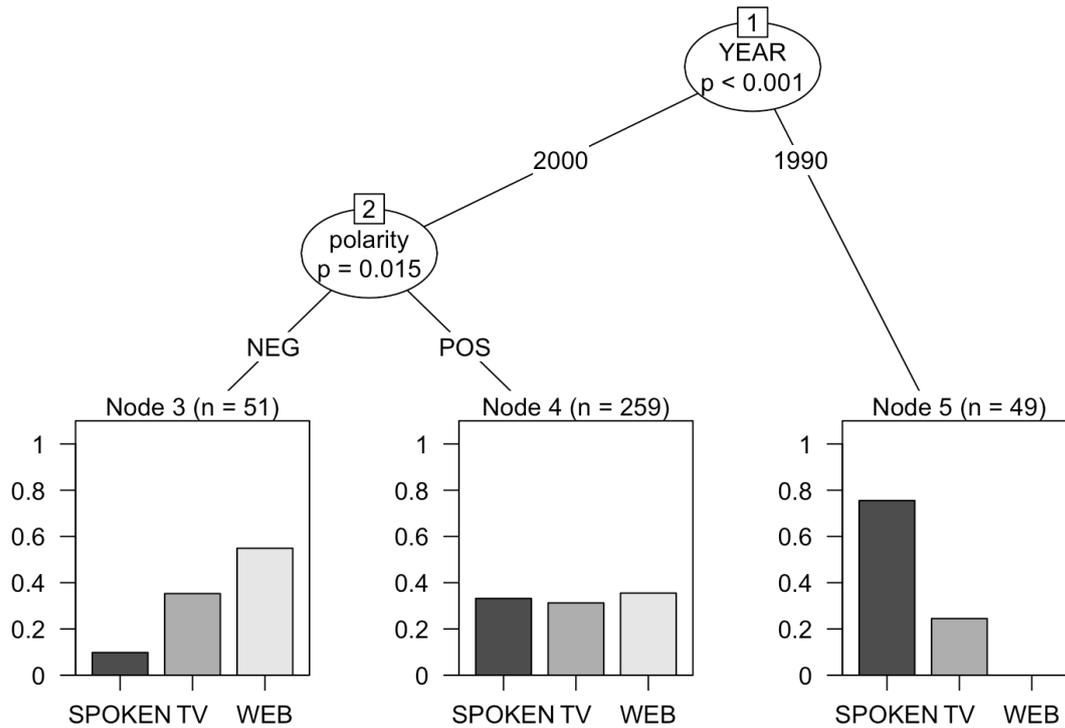


Figure 3 Conditional inference tree of situated usages of *uno.M.SG*

The plot in Figure 3 is obtained with the ‘ctree’ function of the R package ‘party’ (see Levshina 2015: 291) and refers to illocutionary concurrences (ICs) of *uno.M.SG* being spontaneously employed in the respective Spoken, Web, and TV sections of the Perugia Corpus. It is important to stress that the tree above has nothing to do with a generative one. More specifically, conditional dependencies among variables in Figure 3 depend exclusively on statistical significance (the higher the node, the more significant the ‘conditional decision’). They provide context-bound convergences among polarity, tense, and the other variables that we introduced in the Methodology section. The descending order of each split computationally simulates a conditional *decision* made by interlocutors based on degrees of significance of each covariant that comes into play when extended intersubjective usages of *uno.M.SG* are realized.

One important argument of this paper is that phoric functions of *uno* are based on the extended intersubjective awareness of how a general social persona is expected to act, feel, or think. They all correspond to generalized instantiations, viz. abstractions ‘involving instances of a given type’ (Langacker 2009: 9), allowing users to establish mental contact ‘through the mediation of fictive or virtual entities conjured up for that purpose’ (Langacker 2005: 170).

From Figure 3 we can notice the first interesting illocutional concurrence of *uno.M.SG*, namely, a significant tendency of speakers towards negative polarity in Web contexts (i.e. blogs, comments on websites, and so on) after the year 2000. This seems to reflect a conventionalized IC of online speech exemplified by cases where *uno.M.SG* hinges on what the general social persona cannot or should not do, think, or say. In the case below, a blogger interprets Berlusconi’s (former Italian Prime minister) words as ironically referring to the behavior of ones ‘of the kind of Berlusconi’, viz. what anyone would consider as general dishonest people:

- (7) *Berlusconi dice che con questi giudici non si può governare in santa pace. Un lapsus. Forse intendeva dire che, con questa storia della legalità, uno.M.SG non può fare più affari alla cara vecchia maniera.*

‘Berlusconi says that these judges do not allow for governing *in peace*. A slip. Maybe what he meant is that with all this talk about legality, **one** cannot conduct business in the same way as before.’
(Web subcorpus, 2010)

We refer to context-bound intersections of formal and illocutionary dimensions of this kind as illocutional concurrences (ICs) (Tantucci and Wang 2018). ICs encompass converging factors at various levels of verbal experience that contribute both locally (i.e., at the morphosyntactic level) and peripherally (i.e., at the illocutionary level) to the encoding of contextually and temporally situated speech acts (i.a. Tantucci 2016).⁵

In the following extract, there is a similar construction of what can or cannot be done through the employment of the E-I marker *uno*:

- (8) [...] *i settori possono andar bene e possono andar male e quindi uno.M.SG non può dire tutti bene o tutti male automatismi come diceva La Russa e possono andar male no automatismi su questo non ci sono dubbi.*
‘[...] some sectors can be successful or not and **one** cannot generalize by saying that all are successful or unsuccessful as La Russa suggested and all of them could be unsuccessful, there is no doubt about this.’
(TV subcorpus, 2005)

The web subcorpus also includes a number of cases where *uno.M.SG* appears in the protasis of a hypothetical construction, thus contributing to express generalized instantiations of what a *normal* social persona is prevented from doing, saying, or thinking under some circumstances:

- (9) *Queste ultime leggi impediscono ai giovani di qualificarsi se non pagando molti soldi. tanto che io mi chiedo ma se uno.M.SG non è figlio d’avvocato, come farà a diventare avvocato?*
‘These last laws prevent young people to get the qualification unless they spend a lot of money. To the extent to which I ask myself but if **one** is not a barrister’s son, how will they become a barrister?’
- (10) *Se uno.M.S. non partecipa alle votazioni viene sottratta una parte della retribuzione, potrebbe anche essere visto questo per le commissioni.*
‘If **one** does not take part in the vote, a part of the wage is not going to be paid, and this could become valid for some committee too.’
(TV subcorpus, 2009)

One more interesting IC is the distinctive tendency of *uno.M.SG* to occur in spoken interaction before the 2000s (that is in the 90s), viz. before the rapid growth of social networks and online speech. This IC emerges from the right-hand side of the plot.

Figure 4 presents a multiple correspondence analysis (Nenadic and Greenacre 2007) of our annotation. This allows the modelling of associations among variables by calculating the chi-square distance between different categories of the variables and between observations. These associations are then represented graphically as a map, which eases the interpretation of the structures in the data; the closer the distance between variables, the stronger the statistical correspondence.

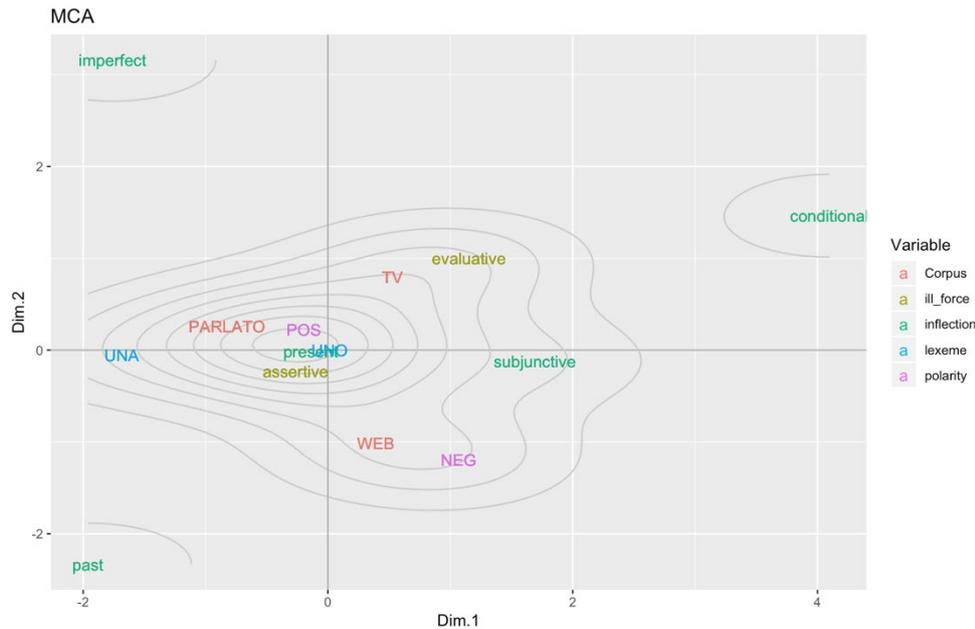


Figure 4 Multiple correspondence analysis of situated usages of *uno.M.SG*

From the plot we can immediately notice a very strong convergence (IC) among the usage of *uno.M.SG*, present tense, assertive speech acts, positive polarity, and spoken interaction (that is, in the subcorpus Parlato). This suggests that the extended intersubjective usages of *uno.M.SG* underpin spontaneous spoken interaction where committed assertions state how a general social persona is expected to act, feel, or believe. These tend to be encoded with the present tense (sometimes associated with the hypothetical *se* (if)), referring to some currently relevant state of affair or recurrent habits, as in the following extracts:

- (11) *Non sto dicendo che è il modo migliore né che dovremmo imitarlo, tutto sommato preferisco il sistema italiano in cui se uno.M.SG non ha i soldi la macchina non se la compra, però comunque vanno riconosciuti i pro insieme ai contro.*
 ‘I am not suggesting that this is the best way and we should all follow this, I somewhat prefer the Italian way that is if **one** does not have money, one does not go and buy a car. However, both pros and cons should be taken into account.’
 (Web subcorpus, 2010)
- (12) *Paura giustificata. Certo, c’è stato il fatto di Chernobyl che uno.M.SG dice ma le centrali nucleari possono essere davvero pericolose.*
 ‘A justified fear. Surely, what happened in Chernobyl and **one** says that nuclear stations can really be dangerous.’
 (TV corpus, 2010)

In connection to this, we can see how past tenses (top and bottom left corners) are extremely distant from the main area of convergence where *uno.M.SG* is situated.

One more point that deserves attention is the correspondence of language of TV and evaluative speech acts. The latter are cases where Sp/w expresses his/her own personal opinion (e.g., *I think that p*), rather than asserting how things are as such (e.g., *p*) (see Tantucci 2016; Tantucci and Wang 2018). This may also indicate that generalized reference to the social

persona's behavior tends to be more cautious in the TV discourse, as the awareness of a large audience might inhibit assertiveness of an interlocutor's statement.

A last point that seems to be interesting in relation to how *uno.M.SG* is employed in this corpus is the hypothetical construction. The hypothetical *se* 'if' precedes *uno.M.SG* in 122 (33.89%) of the 359 occurrences. This opens the debate about the extent to which the impersonal masculine is seen as embodying possible scenarios, as in:

(13) *Se uno.M.SG seguisse le idee di Mari avrebbe poche speranza di pubblicare un libro in Italia*

'If **one** would follow Mari's ideas, they would not be hopeful in publishing a book in Italy.'

(Web subcorpus, 2010)

(14) *Se uno.M.S. prende le cose con un certo spirito, con un occhio ironico penso proprio che ci sia da ridere.*

'If **one** takes what happens with humor, with irony, I think we would be right to laugh.'

(Spoken subcorpus, 2007)

This section has demonstrated that the masculine impersonal *uno* is still the preferred form and is widely employed in several ways according to the linguistic variables investigated.

6 Conclusions

When we started working on this project, we had a suspicion that the impersonal masculine *uno.M.SG* was used more frequently than the grammatical feminine *una.F.SG*, based on previous research and our observations with regard to generic masculine language used to recount a general experience, as that embedded through extended intersubjectivity. Our rigorous investigation has taken into consideration three genres – language on the internet, on TV, and spoken – from the Perugia corpus, which spans three decades (from the 1990s to the 2010s). *Uno* (and *una*) are also examined in relation to specific linguistic variables – polarity, tense, illocutionary force, and phoricity – with the aim of providing a comprehensive picture of how they operate in the language. The initial results – those that deal with the difference in frequencies – demonstrate that Italian, not only for the pair *uno.M.SG* and *una.F.SG* but also for other terms, falls firmly within an androcentric view of language, where men, and therefore grammatically masculine forms, are the prototypical category of reference. This relates to not only the *use* of the masculine form – as seen in the Analysis section – but also the *availability* – that is, the range of options from which the speakers choose their preferred form (Formato 2019).

To summarize the results in relation to the linguistic variables investigated, our analysis found that the E-I marker *uno.M.SG* is used with positive polarity (with the exception of the Web subcorpus after 2000, where negative polarity is noted) in its present tense (as it indicates the action happening during the present or habitual and recurrent actions). A more in-depth investigation also suggests that more than 33% of the instances of *uno.M.SG* are preceded by the hypothetical *if*. These results identify the E-I marker impersonal masculine as a way for speakers to construct generalized ideas about who is prototypically seen as operating in the world and whose experience predominates. One relevant notion is that of the ontological experience: for instance, Bengoënchea (2015) evaluates the *YO* that is the ontological 'I', how we experience and express ourselves. She asks how speakers could use a feminine (generic) term (which she labels 'feminine universal absolute') with the aim of embodying the universal experience of both women and men and replacing the *known* experience of seeing the world

through masculine forms. This highlights one of the main concerns with regard to masculine forms that have seemingly undergone neutralization: how do we move this neutralization to feminine terms (whether generics or for female referents)? And (how) can speakers re-evaluate their experience in terms of feminine or masculine terms?

The efforts to promote feminisation, for instance through guidelines, have not been fruitful, failing to provide a robust platform for discussion on discriminatory practices; the reasons being lack of enforcement, and un-sistematicity as well as un-institutionalization of the debate around gender in/and language. In reviewing the lack of success of these various initiatives and guidelines, Formato (2019: 123) suggests that general audience ‘finds shelter in known misconception about gendered language’. Similarly, initiatives to promote gender fairer language (e.g. see Bengochea 2011 for Spanish) or neutralization (see Motschenbacher for German) encounter many obstacles from the linguistic point of view as well as from speakers. The main implication being that generic masculines are still widely used – as demonstrated in the investigation of the indefinite pronoun *uno* – and plausibly perceived by the speakers as neutral forms.

References

- Abbou, J. (2011) Double gender marking in French: a linguistic practice of antisexism. *Current Issues in Language Planning* 12(1): 55–75.
- Alvanoudi, A. (2014) *Grammatical Gender in Interaction: Cultural and Cognitive Aspects*. Leiden: Brill.
- Anscombre, J.-C. and Ducrot, O. (1983) *L'Argumentation dans la Langue* [The Argumentation within the Language]. Bruxelles: Mardaga.
- Apperly, I. A. (2010) *Mindreaders: The Cognitive Basis of 'Theory of Mind'*. Hove, UK: Psychology Press.
- Arppe, A., Gilquin, G., Glynn, D., Hilpert, M. and Zeschel, A. (2010) Cognitive corpus linguistics: five points of debate on current theory and methodology. *Corpora* 5(1): 1-27.
- Benveniste, É. (1958) Catégories de Pensée et Catégories de Langue [Categories of Thoughts and Categories of Language]. *Les Études Philosophiques* 13(4): 419–429.
- Bengochea, M. (2008) Lo femenino en la lengua: sociedad, cambio y resistencia normativa. Estado de la cuestión [The feminine in the language: Society, change and normative resistance]. *Lenguaje y Textos* 27: 37–68.
- Bengochea, M. (2011) Non-sexist language Spanish policies: an attempt bound to fail? *Current Issues in Language Planning* 12(1): 35–53.
- Bengochea, M. (2015) *Lengua y Género* [Language and Gender]. Madrid: Síntesis.
- Boroditsky, L., Schmidt, L. A. and Phillips, W. (2003) Sex, syntax, and semantics. In D. Gentner and S. Goldin-Meadow (eds), *Language in Mind: Advances in the Study of Language and Thought* 61–79. Cambridge, MA: MIT Press.
- Bucholtz, M. (2014) The feminist foundations of language, gender, and sexuality research. In S. Ehrlich, M. Meyerhoff and J. Holmes (eds), *The Handbook of Language, Gender, and Sexuality* 23–47. Malden, MA: WileyBlackwell.
- Cavagnoli, S. (2013) *Linguaggio Giuridico e Lingua di Genere: Una Simbiosi Possibile* [Language of Law and Gendered Language: A Possible Symbiosis]. Torino: Edizioni dell'Orso.
- Coady, A. (2018) The origin of sexism in language. *Gender and Language* 12(3): 271–293.
- Corbett, G. (1991) *Gender*. Cambridge: Cambridge University Press.
- Davies, M. (2010–) *The Corpus of Historical American English (COHA): 400+ million words, 1810–2009*. <http://corpus.byu.edu/coha> (accessed July 2011).
- Diessel, H. (2006) Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics* 17(4): 463–489.

- Earp, B. D. (2012) The extinction of masculine generics. *Journal for Communication and Culture* 2(1): 4–19.
- Formato, F. (2014) *Language Use and Gender in the Italian Parliament*. PhD Thesis. Lancaster University, UK. <http://www.research.lancs.ac.uk/portal/en/people/federica-formato3c1ec1f2-ac4a-49da-a1fc-148468a26e67.html>.
- Formato, F. (2016) Linguistic markers of sexism in the Italian media: a case study of *ministra* and *ministro*. *Corpora* 11(3): 371–399.
- Formato, F. (2019) *Gender, Discourse and Ideology in Italian*. Basingstoke: Palgrave.
- Fusco, F. (2012) *La Lingua e il Femminile Nella Lessicografia Italiana tra Stereotipi e (In)visibilità* [Language and the Feminine in the Italian Lexicography: Between Stereotyped Representation and Invisibility]. Torino: Edizioni dell’Orso.
- Ghesquière, L. (2009) From determining to emphasizing meanings: the adjectives of specificity. *Folia Linguistica* 43(2): 311–343.
- Ghesquière, L. and Van de Velde, F. (2011) A corpus-based account of the development of English *such* and Dutch *zulk*: identification, intensification and (inter)subjectification. *Cognitive Linguistics* 22(4): 765–797.
- Goldman, A. I. (2006) *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.
- Gries, St. Th. and Stefanowitsch, A. (eds) (2006) *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*. Berlin: Mouton de Gruyter.
- Gries, St. Th. (2011) Corpus data in usage-based linguistics: what’s the right degree of granularity for the analysis of argument structure constructions? In M. Brdar, St. Th. Gries and M. Žic Fuchs (eds), *Cognitive Linguistics: Convergence and Expansion* 237–256. Amsterdam: John Benjamins.
- Hamilton, M. C. (1991) Masculine bias in the attribution of personhood: people = male, male = people. *Psychology of Women Quarterly* 15(3): 393–402.
- Langacker, R. W. (1987) *Foundations of Cognitive Grammar: Theoretical Prerequisites, Volume 1*. Stanford: Stanford University Press.
- Langacker, R. W. (1990) Subjectification. *Cognitive Linguistics* (1): 5–38.
- Langacker, R. W. (1991) Cognitive grammar. In F. G. Droste and J. E. Joseph (eds), *Linguistic Theory and Grammatical Description* 275–306. Amsterdam: John Benjamins.
- Levshina, N. (2015) *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins.
- Marcato, G. and Thüne, E-M. (2002) Gender and female visibility in Italian. In M. Hellinger and H. Bussmann (eds), *Gender across Languages: The Linguistic Representation of Women and Men, Volume 2* 186–217. Amsterdam: John Benjamins.
- Martin, P. and Papadelos, P. (2017) Who stands for the norm? The place of metonymy in androcentric language. *Social Semiotics* 27(1): 39–58.
- Mills, S. (2008) *Language and Sexism*. Cambridge: Cambridge University Press.
- Motschenbacher, H. (2010) Female-as-norm (FAN): a typology of female and feminine generics. In M. Bieswanger, H. Motschenbacher and S. Mühleisen (eds) *Language in its Socio-Cultural Context: New Explorations in Gendered, Global and Media Uses* 35–67. Frankfurt am Main: Peter Lang.
- Motschenbacher, H. (2013) Gentlemen before ladies? A corpus-based study of conjunct order in personal binomials. *Journal of English Linguistics* 41(3): 212–242.
- Motschenbacher, H. (2014) Grammatical gender as a challenge for language policy: the (im)possibility of non-heteronormative language use in German versus English. *Language Policy* 13(3): 243–261.
- Motschenbacher, H. (2015) Some new perspectives on gendered language structures. In M. Hellinger and H. Motschenbacher (eds), *Gender Across Languages: The Linguistic*

- Representation of Women and Men, Volume 4* 27–48. Amsterdam: John Benjamins.
- Motschenbacher, H. (2016) A discursive approach to structural gender linguistics: theoretical and methodological considerations. *Gender and Language* 10(2): 149–169.
- Nardone, C. (2016) *Asimmetrie Semantiche Di Genere: Un'Analisi Sull'Italiano Del Corpus Itwac* [Gendered Semantic Asymmetries: An Investigation into the Italian Corpus ItWac]. *G/S/I* (3).
- Nardone, C. (2018) *Lingua, Genere e Lavoro in Italia e in Germania: Un'Analisi Comparativa su Annunci di Lavoro, sui Corpora ItWaC e DeWac e sulla Stampa* [Language, Gender and Work in Italy and Germany: A Comparative Investigation into Job Adverts, in the ItWaC and DeWac Corpora and the Media]. PhD thesis, Università di Bologna and Heinrich-Heine-Universität Düsseldorf.
- Nuyts, J. (2001a) *Epistemic Modality, Language, and Conceptualization: A Cognitive-Pragmatic Perspective*. Amsterdam: John Benjamins.
- Nuyts, J. (2001b) Subjectivity as an evidential dimension in epistemic modal expressions. *Journal of Pragmatics* 33(3): 383–400.
- Nuyts, J. (2012) Notions of (inter)subjectivity. *English Text Construction* 5(1): 53–76.
- Premack, D. and Woodruff, G. (1978) Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 1(4): 515–526.
- Sabatini, A. (1987). *Raccomandazioni per un Uso Non Sessista Della Lingua Italiana, Commissione nazionale per la realizzazione della parità tra uomo e donna, Presidenza del Consiglio dei Ministri*. Retrieved from http://www.innovazionepa.it/dipartimento/documentazione/documentazione_par_i_opportunita.htm
- Sabatini, A. (1993). *Il Sessismo nella Lingua Italiana. Commissione Nazionale Per La Parità E Le Pari Opportunità Tra Uomo E Donna*. Retrieved from <http://www.funzionepubblica.gov.it/media/962032/il%20sessismo%20nella%20lingua%20italiana.pdf>.
- Schiffrin, D. (1990) The management of a co-operative self during argument: the role of opinions and stories. In A. D. Grimshaw (ed.), *Conflict Talk: Sociolinguistic Investigations of Arguments in Conversations* 241–259. Cambridge: Cambridge University Press.
- Spina, S. (2014) Il Perugia Corpus: una risorsa di riferimento per l'italiano. Composizione, annotazione e valutazione [The Perugia corpus: a resource for the Italian language. Composition, annotation and evaluation]. In R. Basili, A. Lenc and B. Magnini (eds), *Proceedings of the First Italian Conference on Computational Linguistics* 354–359. CLiC-it. Pisa: Pisa University Press.
- Stefanowitsch, A. and Gries, St. Th. (eds) (2006) *Corpus-Based Approaches to Metaphor and Metonymy*. Berlin: Mouton de Gruyter.
- Tantucci, V. (2013) Interpersonal evidentiality: the Mandarin V-^过 *guo* construction and other evidential systems beyond the 'source of information'. *Journal of Pragmatics* 57: 210–230.
- Tantucci, V. (2016) Toward a typology of constative speech acts: actions beyond evidentiality, epistemic modality, and factuality. *Intercultural Pragmatics* 132: 181–209.
- Tantucci, V. (2017a) From immediate to extended intersubjectification: a gradient approach to intersubjective awareness and semasiological change. *Language and Cognition* 9(1): 88–120.
- Tantucci, V. (2017b) An evolutionary approach to semasiological change: overt influence attempts through the development of the Mandarin 吧-*ba* particle. *Journal of Pragmatics* 12: 35–53.
- Tantucci, V. (2018) From co-actionality to extended intersubjectivity: drawing on language change and ontogenetic development. *Applied Linguistics*, amy050, <https://doi.org/10.1093/applin/amy050>

- Tantucci, V. and Wang, A. (2018) Illocutional concurrences: the case of evaluative speech acts and face-work in spoken Mandarin and American English. *Journal of Pragmatics* 138: 60–76.
- Tantucci, V. and Di Cristofaro, M. (2019) Entrenchment inhibition: constructional change and repetitive behaviour can be in competition with large-scale “recompositional” creativity. *Corpus Linguistics and Linguistic Theory*.
- Traugott, E. C. (2012) Intersubjectification and clause periphery. *English Text Construction* 5(1): 7–28.
- Verhagen, A. (2005) *Constructions of Intersubjectivity: Discourse, Syntax, and Cognition*. Oxford: Oxford University Press.

¹ *Professori*.M.SG is also used as a ‘versatile masculine’ (Formato 2019), that is a masculine employed to refer to a gender-mixed group.

² This case would pose the question of whether the generic is the noun (*studente*) rather than the numeral (*uno*).

³ Another race – La Course – takes place in France for women, yet questions about why a Tour de France for women does not exist remain, e.g. <https://www.bbc.co.uk/sport/cycling/44831758> (accessed October 30, 2018).

⁴ Imperative is not included as there are no options that would include the indefinite pronoun.

⁵ The distinction between locally and peripherally has to do with variables that can be identified within the boundaries of the utterance or a grammatical form (local) and co-variants underpinning illocutionary dimensions such as the nature of the speech act that is realised, the turn-taking sequence that leads to a specific usage and so on.