



UNIVERSITÀ
DI PAVIA

DIPARTIMENTO DI STUDI UMANISTICI

CORSO DI LAUREA MAGISTRALE IN LINGUISTICA
TEORICA, APPLICATA E DELLE LINGUE MODERNE

***A Comparative Analysis of General Extenders in
Italian and Spanish Discourse***

RELATRICE

Ilaria Fiorentini

CORRELATRICE

Chiara Zanchi

Tesi di laurea magistrale di

Chaima Badi

Matricola n. 555316

Anno accademico 2025/2026

To my family, with gratitude.

ABSTRACT

This thesis presents a comparative pragmatic study of general extenders (GEs), markers like ‘and stuff like that’ or ‘etcetera’ in Italian and Spanish discourse. Comparative research on general extenders across these two romance languages remains limited, and studies combining elicited interview data with corpora analysis are even rarer. This study addresses that gap through a mixed-methods design combining a self-collected corpus of ten semi-structured interviews (five Italian L1 speakers studying Spanish, five Spanish L1 speakers studying Italian) with a corpus-assisted analysis of the same GE forms emerged from the interviews in two reference corpora of spontaneous speech, KIParla for Italian and CORPES XXI for Spanish. All GE forms identified in the interviews were manually annotated for construction, exemplar type, and pragmatic function, following the taxonomy proposed by Overstreet and Yule (2021). The findings reveal an important asymmetry between the two languages: while Italian informants relied on a small set of highly frequent, stable constructions, Spanish informants produced markedly more ad hoc and idiosyncratic GE variants, a pattern corroborated by the corpus-assisted analysis. The results support a view of general extenders as a multifunctional, cognitively grounded pragmatic category, and demonstrate the value of a mixed-methods, corpus-assisted design for comparative pragmatic research on vague language.

Keywords: general extenders, vague language, comparative pragmatics, semi-structured interview, corpus-assisted analysis.

CONTENTS

1. INTRODUCTION	5
2. VAGUE LANGUAGE IN INTERACTION: SOCIAL AND COGNITIVE PERSPECTIVES	8
2.1 The concept of vagueness in language.....	8
2.2 Politeness theory	12
2.3 Ad hoc categorization	18
2.3.1 Graded structures and typicality.....	23
3. GENERAL EXTENDERS	26
3.1 Defining general extenders	26
3.2 Sociolinguistic variation in general extenders use.....	29
3.3 Classification of general extenders	32
3.4 The grammaticalization of general extenders	33
3.5 The multifunctionality of general extenders	39
3.5.1 General extenders as markers of intersubjectivity	42
3.5.2 Interactive and textual functions of general extenders...	43
3.5.3 Linguistic economy, mitigation and intensification in general extenders usage.....	45
3.6 The case of Italian <i>eccetera</i>	47
3.7 General extenders in non-native speakers	49
4. METHODOLOGY AND INTERVIEW CORPUS	52
4.1 Participants and corpus description	52
4.2 Informant profiles	53
4.3 The interview structure	54
4.4 Transcription phase	55
4.5 Annotation phase.....	57
4.6 Quantitative distributions of GEs	59
5. ANALYSIS OF THE INTERVIEW CORPUS.....	61
5.1 A comparative qualitative analysis	61

5.1.1 Italian sub-corpus	61
5.1.2 Spanish sub-corpus.....	67
5.2 Translation task: cross-linguistic rendering of GEs.....	78
6. CORPUS-ASSISTED ANALYSIS OF GENERAL EXTENDERS: CORPORA OF SPOKEN ITALIAN AND SPANISH	83
6.1 Aims, corpora and sampling procedure	83
6.2 Quantitative and functional distribution in reference corpora	85
6.3 Qualitative analysis.....	90
6.3.1 Qualitative analysis of GEs from the KIParla corpus	90
6.3.2 Qualitative analysis of GEs from the CORPES corpus .	97
7. CONCLUSIONS	109
REFERENCES	113
ONLINE RESOURCES	117
APPENDIX A – Interview questions.....	118
APPENDIX B – Annotation of GEs in the interview data.....	121
APPENDIX C – Translation task	124

1. INTRODUCTION

Vagueness is a defining feature of everyday spoken language. Far from being a sign of imprecision, vague expressions allow speakers to categorize experience economically, to manage interpersonal relationships, to hedge uncertain claims, and to organize discourse in real time (cf. Channell, 1994; Overstreet & Yule, 2021). Among the many linguistic devices used to encode vagueness, general extenders GEs, phrase or clause final expressions such as English ‘and stuff like that’ or ‘or something’, Italian *eccetera* or *e cose così*, and Spanish *y cosas de esas* occupy an interesting position at the intersection of syntax, semantics, and pragmatics. Formed by a conjunction, a pro-form and a similarity marker, GEs extend an otherwise grammatically complete utterance, signaling that the item(s) mentioned belong to a broader category (cf. Overstreet & Yule, 2021; Fiorentini & Magni, 2021; Fiorentini & Zanchi, 2024). Although general extenders have been extensively investigated in English (cf. Cheshire, 2007; Pichler & Levey, 2011; Overstreet & Yule, 2021), recently also in Italian (cf. Fiorentini, 2018; Fiorentini & Magni, 2021; Fiorentini & Zanchi, 2024; Barotto & Mauri, 2018) and Spanish (cf. Fernández, 2015; Kern, 2022). However, comparative studies contrasting the two romance languages remain limited. This gap is surprising given the proximity of Italian and Spanish and the degree of contact between their speaker communities, a context in which general extenders, as markers particularly sensitive to sociolinguistic variation and language contact, might be expected to behave differently. Moreover, much of the existing literature relies either on corpus data alone, which affords limited control over the communicative context in which GEs

are produced, or on elicited data alone, which cannot be independently verified against broader patterns of language use. The present study adopts a mixed-methods design that combines semi-structured interviews with a corpus-assisted analysis to complement and validate the findings emerged from the interview data.

This thesis investigates the pragmatic behavior of general extenders in the spontaneous speech of two groups of young adult bilingual speakers: Italian L1 speakers currently studying or having studied Spanish as a second/foreign language, and Spanish L1 speakers currently studying or having studied Italian. Specifically, the study pursues three interrelated research questions. First, regarding quantitative distribution, it investigates which of the two language groups exhibits greater lexical diversification and frequency in their use of GEs, and how this distribution compares with evidence from reference corpora of spontaneous speech. Second, concerning functional distribution, it examines what pragmatic functions: referential, interpersonal, personal, or textual, following Overstreet and Yule's (2021) taxonomy Italian and Spanish general extenders predominantly fulfil. Third, in terms of cross-linguistic equivalence, it explores whether bilingual speakers preserve the status of a general extender when translating it between their two languages. To address these questions, the study combines two complementary data sources. The first is a corpus of ten semi-structured interviews (five per language group), designed to elicit spontaneous general extender production across a range of communicative tasks, from discussing language-learning difficulties to describing everyday routines and past events, and concluding with a bidirectional translation task in which target GEs were embedded in two short texts. The second is a corpus-assisted analysis in which each of the 24 GE forms identified in the interviews was queried in

KIParla, a reference corpus of spoken Italian, and CORPES XXI, filtered for oral registers, for Spanish. This second stage allows the qualitative patterns observed in the interview data to be assessed against a broader dataset of spontaneous speech, following a corpus-assisted methodology, given the limited size of the interview data.

The thesis is organized into seven chapters. Chapter 2 presents the theoretical basis by situating general extenders within the study of vague language, reviewing politeness theory as a framework for understanding their interpersonal function, and introducing the notion of ad hoc categorization as a cognitive explanation of how GEs are interpreted in context. Chapter 3 narrows the focus to general extenders specifically, offering a detailed definition, an overview of sociolinguistic variation in their use, a classification of GE structures, an explication of their grammaticalization process, and a discussion of their multifunctionality before focusing on two case studies: the Italian form *eccetera* and the use of general extenders by non-native speakers. Chapter 4 details the methodology of the interview, describing the informant sample, the interview structure, the transcription conventions, and the annotation scheme adopted for this study. Chapter 5 presents the qualitative analysis of the interview corpus, organized by sub-corpus (Italian, Spanish) and followed by a discussion of the translation task. Chapter 6 extends this analysis through the corpus-assisted analysis, reporting the quantitative distribution of the target forms in KIParla and CORPES XXI before presenting a qualitative discussion of selected corpus occurrences. Finally, chapter 7 draws together the main findings, discusses their theoretical and methodological implications, and acknowledges the limitations of the present study.

2. VAGUE LANGUAGE IN INTERACTION: SOCIAL AND COGNITIVE PERSPECTIVES

The present chapter outlines the theoretical foundations for the study of general extenders by providing an overview of vague language. Specifically, section 2.1 defines the concept of linguistic vagueness, while section 2.2 examines politeness theory and its functional relationship with general extenders. Finally, section 2.3 explores the cognitive dimension of these markers through the lens of ad hoc categorization.

2.1 The concept of vagueness in language

In the broader literature, two predominant approaches to vagueness are distinguished. The first, rooted in a philosophical tradition dating back to Aristotelian logic, treats linguistic vagueness as a pervasive phenomenon concerning truth-value and semantic indeterminacy. The second, pragmalinguistic vagueness, shifts the focus toward specific categories of vague language employed for communicative purposes. This study on general extenders adopts the latter approach, as it allows for an analysis of how speakers use vagueness as a purposeful strategy to manage interpersonal interaction (Ruzaitė, 2010: 33). General extenders (GEs) represent a class of expressions occurring in phrase or clause-final position, formed by a conjunction and a generic nominal or pro-form (Fiorentini & Magni, 2021: 295). Specifically, Overstreet and Yule (2021: 1) define general extenders as pragmatic markers by stating that they are considered ‘general’ because they are non-specific in reference, and function as ‘extenders’ in that they extend utterances that are otherwise grammatically complete. A more detailed theoretical framework will be presented in chapter 3.

In contrast to traditional views that associate good language usage with absolute precision, Channell (1994) argues that vagueness is a fundamental component of linguistic competence. She claims that linguistic vagueness is not a defect, but a necessary tool for managing the lack of clear boundaries in the non-linguistic world and the unfamiliarity that often occurs in natural discourse. The appropriateness of vagueness is highly context dependent, while formal registers (such as those used in medical reports, legal contracts, or academic writing) demand rigorous precision, the same degree of exactness may be counterproductive in informal settings. Consequently, the ability to adapt linguistic expression according to the communicative setting is a component of a speaker's communicative competence. Understanding how vagueness functions is, therefore, essential for grasping the core nature of human language (Jucker, Smith & Lüdge, 2003: 1738). Channell (1994) emphasizes that vagueness is not a product of ignorance or laziness, but a deliberate communicative strategy. In this view, a speaker employs a vague term, such as a general extender, because it is the most efficient means of achieving specific pragmatic goals, whether hedging on certainty or signaling shared knowledge. Indeed, as Drave (2001: 26) notes, there is a significant semantic overlap between vagueness and hedging,¹ as both are concerned with conveying the speaker's subjective orientation toward their own message. This process relies on a mutual agreement between interlocutors: the speaker assumes the listener will successfully interpret the intended meaning through context and pragmatic inference. Thus, the construction of vagueness is a collaborative process, requiring active participation from both the speaker and the

¹ For a wider view of the overlap between vagueness, ambiguity, and hedging, cf. Fiorentini & Zanchi (2024).

addressee. As noted in the literature, interlocutors who share a high degree of familiarity can interpret vague markers with significant accuracy, relying on shared knowledge regarding each other's lives and habits (McGee, 2018: 43).

Following Channell (1994: 20), vagueness is defined as a communicative strategy that modifies an utterance to make its meaning less precise (cf. Drave, 2001: 25). However, linguistic vagueness is not always a deliberate strategy. Although speakers are often aware of their choice to be vague and the reasons behind it, they may also lapse into vagueness unintentionally, lacking any conscious intent (McGee, 2018: 41).

Lim and Wu (2023) propose the concept of 'literal vagueness' as a unified notion that integrates conventional linguistic imprecision with inherent context dependence. Under this view, language is fundamentally underspecified. They argue that a linguistic system characterized by literal vagueness is functionally superior to a strictly precise one, as it maximizes communicative efficiency. By utilizing less explicit language, speakers reduce their own production effort while inviting the listener to employ pragmatic inference to reconstruct the intended meaning based on the context. As Aijmer (2002: 12) observes, unlike ordinary words, discourse particles are characterized by their indexical nature, which implies that their semantic meaning is not fixed but is instead determined by the specific context in which they occur. Rather than carrying a literal meaning, these elements function as pragmatic pro-forms that derive their interpretation from the surrounding discourse and the shared knowledge of the interlocutors. Consequently, discourse particles function as indexical markers of the speaker's social reality. Those particles can be employed to signal the level of intimacy or distance between interlocutors. Furthermore, their usage often reflects

broader social affiliations, including class, gender, and age. These markers can index the speaker's position within a hierarchy of authority, revealing perceived status or rank during the interaction. Speakers belonging to a specific social group or sharing a common background frequently employ vague expressions. Rather than being a sign of linguistic laziness, this lack of specificity acts as a marker of intimacy; it signals that the interlocutors are sufficiently aligned that explicit details are unnecessary for achieving mutual understanding (Fernandez & Yuldashev, 2011: 2623). For instance, in the case of adolescent language, a defining characteristic of discourse is the elevated frequency of vague expressions. According to Palacios Martínez (2011b: 115), teenagers employ a distinct repertoire of vagueness markers that are comparatively rare in the linguistic patterns of older generations.

When a speaker utilizes a vague marker like a general extender, the category of objects or concepts being referred is not pre-defined. Instead, it is dynamically co-constructed by the interlocutors through the integration of top-down and bottom-up cognitive mechanisms. This interactional process involves an initial formulation, the provision of explicit exemplars, and the eventual indexical abstraction required to anchor the reference within the specific context (Mauri & Sansò, 2018: 26). Sociolinguistic variables such as gender also play a significant role in the distribution of these markers. Certain studies suggest that the higher incidence of vague language among female speakers may be a byproduct of traditional gender roles and societal expectations. Historically, these social norms created a clear split in how men and women were expected to communicate: men were generally encouraged to speak in a bold and direct manner, while women were often taught to be more cautious or indirect in their speech (McGee, 2018: 46). This gender

divergence is reflected in the frequency and function of general extenders, as studies consistently identify a correlation between gender and the employment of these markers. While their usage appears relatively balanced during the preadolescent stage, a divergence emerges in early adolescence, where females exhibit higher usage. This gap seems to close during the mid-teens before reversing in the late teens and early twenties, at which point males tend to surpass females in their use of these markers. However, according to Overstreet & Yule (2021: 132), whether this developmental trajectory remains consistent across different sociolinguistic contexts remains a critical empirical question that requires further investigation.

The role of vagueness is particularly evident in the context of second-language acquisition. For a second-language learner, the use of vague markers like general extenders is therefore not merely related to lexical gaps, but a strategy employed to maintain conversational flow in a linguistic system that is often fuzzy. Consequently, vagueness emerges as a functional and conventionalized component of interaction rather than a failure strategy to be avoided. Beyond simply filling lexical or knowledge gaps, vague language serves several pragmatic purposes. It allows speakers to emphasize specific information, deliberately withhold details when necessary, and convey a sense of tentativeness. Furthermore, it enables the expression of evaluations or expectations while simultaneously fostering an atmosphere of friendliness, informality, or deference within the interaction (Drave, 2001: 27).

2.2 Politeness theory

In a social context, politeness serves as the strategic management of ‘face’, which is defined as the public self-image that individuals

expect their interlocutors to acknowledge. By employing politeness strategies, speakers demonstrate an awareness of this self-image, either by maintaining deference or by fostering solidarity through more intimate interactions. Consequently, linguistic choices are frequently dictated by the need to protect or enhance the interlocutor's social sense of self (Yule, 1996: 60). In this light, general extenders also possess a crucial role in the social dimension of spoken interaction, often described in terms of their interpersonal function. Specifically, their use can fulfil a politeness function by acting as a pragmatic mechanism to soften impositions on the interlocutor. In everyday communicative interaction, individuals generally operate under the assumption that their public self-image will be acknowledged and respected by others. When a speaker makes an utterance that directly challenges or undermines these expectations, it is categorized as a face-threatening act (FTA). Indeed, many speech acts are intrinsically face-threatening: those targeting negative face, such as direct requests or advice, impose on an individual's autonomy, while those targeting positive face, such as disagreements or disapproval, undermine an individual's desire for social acceptance (Brown & Levinson, 2006: 313-314). To maintain social equilibrium and mitigate potential conflict, speakers often employ face-saving acts. These are linguistic strategies designed to lessen the impact of a perceived threat to another's face. For instance, rather than using a direct command that might impose on someone's autonomy, speakers may opt for an indirect request to preserve social harmony (Yule, 1996: 61). As noted by Searle (1975: 49), politeness serves as the primary driver for employing indirect speech acts, particularly in the context of requests. This aligns with Brown and Levinson's (2006: 316) figure of FTA strategies, which

suggests that when a speaker anticipates a threat to face, they may choose to go ‘off record,’ as shown in Figure 1.

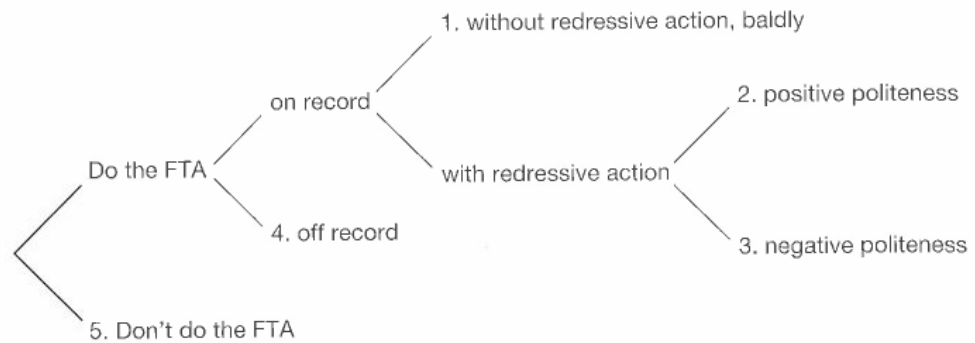


Figure 1. Possible strategies for doing FTAs (taken from Brown & Levinson 2006: 316).

By using a general extender, the speaker may avoid an ‘on record’ imposition, instead opting for a vague, off-record approach that relies on the interlocutor’s ability to infer the intended meaning, thereby minimizing the risk of appearing overly demanding or critical and preserving the interlocutor’s face. This indirection often involves a violation of Grice’s Cooperative Principle, which assumes that participants normally follow specific maxims, such as Quantity, Quality, Relation, and Manner, to ensure an effective exchange of information. For instance, by using a vague general extender to conclude a sentence, a speaker intentionally flouts the Maxim of Quantity, which affirms that one should be as informative as required, as well as the Maxim of Manner, which requires the avoidance of obscurity (Grice, 1975: 45-47). In line with the Cooperative Principle, Overstreet and Yule (2021: 47) argue that a fundamental component of intersubjectivity within interaction is the assumption of mutual collaboration, where interlocutors rely on the shared understanding that participants are providing sufficient information to be understood, without intentionally obscuring meaning or withholding relevant details.

Beyond the flouting of maxims, it is essential to distinguish between conversational and conventional implicatures (Grice, 1975). While conventional implicatures are inherent to the linguistic form itself, conversational implicatures are highly context-dependent and require the listener to perform a pragmatic calculation. In the case of general extenders, speakers may rely on these implicatures to convey meaning for which they lack specific lexical items.

In addition to conversational maxims, the use of general extenders can be understood through Austin's (1962: 98-118) Speech Act Theory, which proposes that utterances possess an illocutionary force beyond their literal meaning. While the locutionary act involves the production of a vague sentence, the illocutionary act represents the speaker's intent to perform a specific social function, such as mitigating a request or affirming shared knowledge. Austin distinguishes these from the perlocutionary act, which is the actual effect achieved on the interlocutor. A typical feature of pragmatic competence involves the use of indirect speech acts, which occur when a speaker performs one illocutionary act indirectly by way of performing another. Drawing upon Austin's (1962) distinction between the locutionary act and the illocutionary force, it becomes clear that not all communicative acts are literal. As noted by Searle (1979: 30-32), speakers frequently produce utterances where the sentence meaning and the intended speaker meaning diverge. For example, a statement like 'Can you reach the salt?' functions primarily not as an inquiry into the listener's physical ability to perform the action (sentence meaning: "Are you able to reach the salt?"), but rather as an indirect request for action (intended speaker meaning: "Pass me the salt"). This same mechanism applies to general extenders, whereby speakers may utilize these vague forms to perform indirect speech acts, relying on the listener's ability to

infer the intended illocutionary force through shared conversational routines.

According to Yule (1996: 65-66) politeness strategies are divided into solidarity strategies, which employ positive politeness to foster a sense of social belonging, and deference strategies, which prioritize negative politeness by emphasizing the hearer's autonomy. While negative politeness involves linguistic strategies that respect an interlocutor's need for independence and lack of imposition, positive politeness is oriented toward the participant's need for in-group membership and acceptance (Brown & Levinson 2006: 317). In the context of the present study, general extenders fulfil a dual pragmatic role: they act as a hedge to protect the listener's negative face or serve as a marker of solidarity to reinforce positive face wants (Yule, 1996: 62-66). For instance, the expression 'nice and all' functions as a pragmatic politeness marker, designed to save face or mitigate the impact of the utterance (Fernandez & Yuldashev, 2011: 2623). In particular, adjunctive general extenders are used to implement a positive politeness strategy. This signals that the speakers share common ground, thereby appealing to intersubjectivity (Overstreet & Yule, 2021: 49). Conversely, the speaker also employs a disjunctive general extender as a hedge to mitigate or reduce the threat to their interlocutor's independence and freedom of action (Overstreet & Yule, 2021: 7). Traugott (2010) distinguishes between two key pragmatic dimensions: subjectivity, which concerns the expression of the speaker's internal beliefs, attitudes, and personal perspective, and intersubjectivity, which involves the speaker's orientation toward the addressee. Within this framework, intersubjectivity is primarily concerned with managing the social relationship and preserving the hearer's social face. However, because individual life experiences produce distinct and

non-identical mental worlds, the assumption of exact reference in discourse is effectively precluded. This inherent subjectivity means that an explicit appeal to shared knowledge does not imply that interlocutors possess identical information (Overstreet & Yule, 2021: 46). General extenders have also been defined as subjective markers. As Cortés (2006: 123) points out, these markers can convey a speaker's affective rejection or indifference toward the topic. In such instances, the speaker may intentionally employ these forms to indicate a desire to distance themselves from the subject, particularly when used as hedges on the informativeness or accuracy of an utterance. This subjective quality is manifested differently depending on the form of the extender: adjunctive general extenders function as hedges on informativeness by signaling that further details have been deliberately omitted. Whereas disjunctive general extenders act as hedges on the accuracy of a description or report (Overstreet, 2014: 123). Parallel to this, Cortés (2006: 115) argues that disjunctive GE forms serve as a mechanism of politeness, offering the interlocutor a range of alternatives rather than imposing a single, rigid one. Drawing on Overstreet, Cortés also links this to negative politeness, where the speaker employs these GEs to provide the listener with options. This function is particularly prominent in questions and requests, where the inclusion of a disjunctive general extender acts as a strategy to preserve the addressee's autonomy. Moreover, they serve to soften the illocutionary force of an utterance; by introducing a range of alternatives, the speaker avoids imposition on the listener. It is also worth observing that the use of disjunctive general extenders is not only for the benefit of the addressee; it also functions as a self-protective mechanism for the speaker. Gille and Häggkvist (2010: 134) notes that presenting alternatives through general extenders also serves as a defensive strategy for the speaker's own

positive face. When making an invitation or an offer, the speaker faces a social risk if the listener refuses. By adding a disjunctive extender, the speaker softens the request, so that the listener can decline without directly rejecting the proposal, thereby preserving the speaker's image while simultaneously respecting the listener's freedom of action.

2.3 Ad hoc categorization

From a cognitive perspective, speakers construct what Barsalou (1983: 211) terms 'ad hoc categories' to achieve their communicative goals (cf. also Mauri, 2021). Unlike established categories, these classifications, such as 'things to take on a camping trip,' are often generated spontaneously within specific situational contexts. At first glance, such categories may appear to be nothing more than idiosyncrasies of human thoughts (Barsalou, 2021: 42). However, because they are driven by immediate contextual needs, these categories are neither stored for long-term use nor maintained once the communicative goal is achieved. On one hand, lexical items belonging to established categories are processed with greater efficiency, as pre-existing memory structures seamlessly assimilate such information. In contrast, the online construction of ad hoc categories presents significant challenges in both organization and retrieval. Nevertheless, the specialized nature of these categories suggests that they are inherently context-dependent; consequently, our brains work efficiently by selectively activating only the specific category required for the immediate context. Over time, through repetition, an ad hoc taxonomy can lead to its consolidation as a well-defined structure within long-term memory (Barsalou, 1983: 223). Overstreet and Yule (2021: 29) applies Barsalou's (1983) concept of ad hoc categories by differentiating between 'lexicalized forms,'

which represent categories with a specific term, and ‘non-lexicalized forms’, which refer to concepts for which no individual word exists. The identification of these ‘non-lexicalized forms’ is fundamentally pragmatic, as it relies on the presupposed background knowledge shared between interlocutors within a specific context. When a speaker employs vague expressions such as a general extender, they are typically signaling this type of non-lexicalized categorization. Traditionally, literature on categorization has viewed language primarily as a mirror of cognition, serving as a tool to reflect pre-existing categories. However, contemporary data suggests a more dynamic relationship: language does not merely reflect categories but actively contributes to their construction through various strategies (Mauri & Sansò, 2018: 2). Mauri and Sansò (2018: 8) identify three semantic layers that characterize the construction of an ad hoc categorization. They refer to explicit exemplars as the specific items named by the speaker. The implicit members are the additional elements associated with these exemplars that share a common property relevant to the current situation. Finally, a superordinate category represents the higher-level class that encompasses both the explicit and implicit members.

Through the lens of Relevance Theory, language is understood as a dynamic process where word meanings are adjusted in context to produce ad hoc concepts. This approach suggests that every lexical item is interpreted according to its specific context of use (Wilson and Carston, 2007, cited in Mauri & Sansò, 2018: 4). The use of general extenders can be viewed as an application of lexical pragmatics within Relevance Theory, as these markers allow speakers to adjust meanings and build ad hoc categories in real time (cf. Carston, 2004: 641-643). According to the Relevance Theory, the concept of relevance is determined by a balance between cognitive

effects,² and processing effort. Essentially, input is deemed less relevant if it requires excessive mental effort to decode the message (Wilson & Sperber, 2004: 608-609). For instance, the hearer reconstructs the speaker's meaning based on the accessibility of information. Rather than evaluating all possible interpretations, the addressees follow a path of least cognitive effort, testing interpretive hypotheses (such as disambiguation or implicatures) in order of accessibility. This process concludes as soon as the hearer reaches an interpretation that satisfies their expectations of relevance, effectively balancing the processing effort with the communicative effect achieved (Wilson & Sperber, 2002: 605). In this context, communication is viewed not merely as a code model, where messages are simply encoded and decoded but according to the inferential model. Within this approach, a speaker provides an utterance that serves as evidence of their intention, which the listener then uses to infer the actual meaning. The core claim is that every utterance creates a specific expectation of relevance, guiding the hearer toward the speaker's intention through a non-demonstrative inference process. Therefore, by following the inferential model of communication, general extenders serve as linguistic evidence that trigger an inferential process in the listener. This allows the hearer to move beyond the literal meaning of the words to identify the specific, context-dependent category intended by the speaker, thereby aligning with the Gricean principles that utterances are designed to be cooperative and meaningful (Wilson & Sperber, 2002: 600, 2004). Just as human cognition is shaped by our physical surroundings, it is equally rooted in the social environment (Barsalou, 2021: 40).

² This refers to the interaction between new information and contextual and existing knowledge.

According to Wilson and Sperber (2004: 610), communication is successful because the search for meaning is a core feature of human cognition. Our inferential and retrieval mechanisms are naturally designed to prioritize assumptions that yield the most relevant interpretation of a given input within its specific context. Speakers exploit this cognitive mechanism by providing linguistic cues, such as general extenders, knowing that the hearer's cognitive processes will automatically treat these inputs as relevant. The hearer then integrates these cues with existing memories or thoughts to reach an optimal inference. This process is driven by the fact that human interests are products of a constant search for relevance, where individuals prioritize information that is highly accessible within their memory. Certain topics produce a greater communicative effect for less processing effort. Consequently, communication is viewed as an act of claiming the interlocutor's attention. Listeners, therefore, will only engage if they presume the utterance is relevant enough to justify the effort. Unlike other pragmatic maxims or conventions that speakers may or may not observe, the Principle of Relevance is a generalization of human behavior; it represents an automatic presumption that every utterance is worth the audience's attention (Sperber and Wilson, 1985: 160).

In accordance with the inferential model, Levinson's (1987) I-Principle suggests that the less a speaker says, the more the listener reads between the lines to fill in the missing information based on what is typical. Unlike Q-implicatures, which are based on what the speaker did not say, I-implicatures are based on what the speaker did say and allow interlocutors to move from a semantically general expression to a specific interpretation. The theoretical foundation of the I-Principle is rooted in Grice's Cooperative Principle, specifically related to the second Maxim of Quantity, which instructs speakers

not to make their contribution more informative than is required. While Grice first framed this principle as a requirement to avoid saying more than is necessary, Levinson (1987: 91) develops this into a model for conversational inference, where the speaker's minimal linguistic effort triggers a reciprocal process of inferential elaboration. Speakers often prefer general expressions to reduce conversational effort. By employing a minimal string, the speaker provides just enough linguistic information to trigger the listener's inferential process, enabling the successful reconstruction of a complex category without the need for exhaustive or precise lexical enumeration, relying instead on the I-Principle to ensure the listener reaches the intended specific interpretation (Levinson, 1987: 84). Sperber and Wilson (1985: 158), pointed out that metaphors and vague expressions (collectively termed 'loose talk') violate the Gricean maxim of truthfulness. In their view, metaphors are interpreted through the same cognitive mechanisms as any other form of vague or literal speech. According to this view, while a listener is entitled to expect that an utterance represents the speaker's thoughts, they should not always expect literal accuracy. Sperber and Wilson suggest that literalness is not a prerequisite for successful understanding; instead, vague talk is often good enough, and sometimes even more suitable for conveying the intended message efficiently. This cognitive efficiency, whereby successful communication does not require literal precision, is what makes the construction of ad hoc categories possible in the first place: if listeners can reconstruct a speaker's intended meaning without exhaustive specification, a general extender needs only provide a minimal cue for the listener to infer, in context, which elements belong to the intended category.

2.3.1 Graded structures and typicality

The theory of Grounded Cognition proposed by Barsalou (2021: 36-39) suggests that cognition emerges from the interaction between diverse brain domains such as perception, emotion, and motor systems which become coupled together during situated action. Central to this framework is the concept of simulation, the process of reenacting perceptual states to represent information conceptually. When the brain perceives an entity or event, it establishes a record of that state across various systems (visual, auditory, tactile, etc.). For example, interacting with a category (like an apple) creates an entrenched network of sensory patterns in the brain. Once triggered by a cue, such as hearing the word “apple” or being in a specific context, this network reactivates to simulate the experience, even if the object is not present. These simulations are not always complete or conscious; they are highly task-dependent and context-specific. Building on this cognitive foundation, the internal organization of these categories can be understood through the lens of Graded Structure, a paradigm whose empirical foundations were laid by Eleanor Rosch’s work on natural categories and prototype effects (Rosch, 1973). According to this theory, categories, such as ‘fruit’, do not possess rigid boundaries but are instead characterized by a continuum of membership. Rosch and Mervis (1975) argue that the degree of typicality attributed to a category member is dictated by its similarity to other instances within that same group. Consequently, a category concept is understood as the aggregate average of its members, where those possessing the highest degree of resemblance are treated as the most salient prototypes. This structure manifests through three core dimensions:

- Typicality: Members are not perceived as equally representative; rather, certain ‘better examples’ serve as prototypes. In the category of ‘fruit’, an apple or an orange is typically identified as a highly representative member.
- Unclear Cases: Categorization is often non-binary, leading to unclear cases where speakers may be uncertain about inclusion. These fuzzy boundaries are an intrinsic property of natural language categories.
- Non-members of a category: The principle of similarity extends even to non-members, where certain external items are rejected more slowly if they share conceptual proximity with the category (e.g., a bat is more difficult to reject as a bird than a chair).

A fundamental distinction must be drawn between established lexical categories and the creation of ad hoc categories. While established categories are organized around a prototype, the categories generated in response to a general extender (e.g., things to buy for dinner) function as open-ended lists of contextually relevant examples. By utilizing general extenders, speakers effectively navigate the graded structure of a category, providing a prototypical core through explicit exemplars while allowing the listener to mentally construct the remaining ad hoc category. However, unlike fixed categories, it is difficult to identify a singular prototype for the set implied by a general extender; in these instances, all members hold equal communicative relevance without a fixed hierarchy of typicality (Barsalou, 1983: 213-218). The construction of ad hoc categorization relies on exemplification. By citing one or more representative members of a category, the speaker provides a concrete starting point that points toward a context-specific category. This indexical abstraction allows the hearer to move from the specific named items

to the intended concept (Mauri & Sansò, 2018: 11). According to Mauri et al. (2021: 30), the process of categorization involves a collaborative effort where both speaker and hearer are simultaneously engaged in defining members and boundaries. This identification is achieved through a ‘progressive zooming-in movement,’ where participants utilize specific exemplification to refine the category in real time. This mechanism is clearly visible in the construction of a list; for instance, whenever multiple lexical items are structured together, their meaning is derived by identifying the semantic commonalities that link them. In this view, the listener recognizes a shared conceptual category that justifies the grouping of these specific elements (Barotto & Mauri, 2018: 100). Barotto and Mauri (2018: 101) suggest that list structures formatted as X, Y, Z trigger an automatic assumption that the listed items possess a shared attribute, identified as property P. Within this framework, these items serve as specific illustrations of the broader category defined by that property. Significantly, the discovery of P is not explicitly stated; instead, it is an inferential process reliant on the shared background knowledge of the interlocutors. Consequently, the property remains outside the realm of ‘what is said,’³ requiring the listener to reconstruct the intended category.

³ The notion of ‘what-is-said’ was originally established by Grice (1989, as cited in Barotto & Mauri, 2018: 101) as a mediating layer between ‘sentence meaning’, the literal, semantic content of an expression, and ‘speaker’s meaning,’ which encompasses the full interpretation once all pragmatic implicatures have been realized. For Grice, what-is-said is aligned with the literal sentence meaning.

3. GENERAL EXTENDERS

This chapter investigates general extenders, characterizing them as a specialized and highly multifunctional category of pragmatic markers. Section 3.1 examines the existing literature regarding the definition of general extenders, followed by an analysis of their sociolinguistic variation in section 3.2 and a formal classification in section 3.3. Section 3.4 then explores the grammaticalization processes that shape these markers. A significant portion of the chapter is dedicated to the multifunctionality of general extenders; specifically, section 3.5 examines how they operate as tools for intersubjectivity, interactive and textual organization, linguistic economy, and the management of mitigation and intensification. Finally, the chapter focuses on two areas: the specific case of the Italian *eccetera* in section 3.6, and the general extender usage among non-native speakers discussed in section 3.7.

3.1 Defining general extenders

In contemporary linguistics, the status of general extenders remains a subject of ongoing academic debate. Early descriptions of general extenders emphasized two main features: their placement at the end of an utterance and their function in signaling a set or a list based on the preceding element (Overstreet & Yule, 2021: 22). While some scholars categorize them broadly as pragmatic markers, others prefer the term ‘set-marking tags’. Within the Italian linguistic tradition, research have categorized these forms as devices for building ad hoc categories or as constructions for expressing non-exhaustiveness. Fiorentini and Sansò (2016) classify general extenders as a subcategory of the broader category of discourse markers. According

to the authors, these forms function as pragmatic tools within the domains of discourse organization and interactional management. Similarly, in Spanish studies, several researchers consider general extenders to be a subcategory of discourse markers. Within this tendency, they employ specialized labels such as *marcadores de inconclusión* or *apéndices generalizadores* to describe their function in discourse (Borreguero Zuloaga, 2021: 91). A significant contribution to the study of general extenders in Spanish comes from Gille and Häggkvist (2010), who characterize these elements as *apéndices conversacionales*. Conversely, Aijmer (2013), Beeching (2016), and Brinton (2017) (as cited in Overstreet & Yule, 2021: 5) categorize general extenders as pragmatic markers rather than discourse markers. The key distinction lies in their functional domains: pragmatic markers are fundamentally concerned with establishing and maintaining social cohesion by indexing shared background knowledge and managing interpersonal relationships. Whereas, discourse markers are primarily focused to establish and maintain textual cohesion, indicating the sequential connections between different segments of the text. Notwithstanding this functional distinction, Aijmer and Simon-Vandenberg (2011: 229) point out that pragmatic markers: “can overlap with other markers in some of their meanings”.

In the literature, the concept of non-exhaustivity extends beyond the presence of additional category members; it depends upon the interlocutor’s ability to reconstruct the implied category. In this context, the items referenced by a general extender are unified by an underlying property P.⁴ This property functions as a pragmatic variable, the specific value of which is determined by the

⁴ A characteristic shared by all elements of the set.

communicative context. Consequently, the use of a general extender signals to the listener that they must rely on the context to retrieve the intended category members that share this specific property (Mauri, Goria, & Fiorentini, 2019: 294). Non-exhaustivity is defined by Barotto and Mauri (2018: 103) as encompassing both open lists and lists of exemplars. Open lists encode non-exhaustivity directly, through a dedicated marker occurring at the end of the list, typically a general extender itself (e.g. ‘etcetera’). Lists of exemplars, by contrast, convey non-exhaustivity indirectly: the very act of exemplifying (e.g. via ‘for example’) implies that the listed items were selected from a larger set, even though no explicit GE is used. Existing research into non-exhaustive lists primarily emphasizes their relationship with linguistic vagueness and indeterminacy. Within this field, general extenders are identified as the specific linguistic marker where non-exhaustivity is encoded. This is further supported by Mauri, Goria, and Fiorentini’s (2019: 296) study of spoken Italian, which validates that general extender function as the primary indicators of non-exhaustive lists in discourse. Moreover, these expressions demonstrate a conventionalized relationship between structure and semantics, in which the final meaning cannot be derived from the sum of its individual components. Instead, the construction possesses a specific conceptual frame, allowing the listener to infer an extensive set of referents than those explicitly articulated in the discourse (Mauri, Goria, & Fiorentini, 2019: 304). Therefore, successful interpretation of a non-exhaustive list depends on the listener’s ability to achieve ‘saturation’ of the property P. In other words, the interlocutor does not need to identify every member of the set but must instead detect the specific context that defines the current category (Barotto & Mauri, 2018: 106).

3.2 Sociolinguistic variation in general extenders use

Early research into general extenders, especially within the tradition of Anglophone sociolinguistics (Dines, 1980; Norrby & Winter, 2002, as cited in Borreguero Zuloaga, 2025: 190), associated these pragmatic markers with speakers perceived to have limited discursive competence. From a diachronic perspective, tracing the evolution of general extenders presents significant challenges, as they are spoken forms often absent from conventional historical records. For instance, the Oxford English Dictionary traditionally relegated forms like ‘and that’ to substandard speech, effectively marking them as indicators of social class rather than neutral pragmatic markers (Cheshire, 2007: 165). It emerged that the use of general extenders is not uniform; rather, it manifests in different ways across a variety of social dimensions. Their frequency and form are conditioned by factors such as cultural background, register, and age, among other social variables (Ruzaitė, 2010: 34). Research indicates a strong correlation between age and the deployment of these markers, frequently characterizing them as prototypical features of youth discourse (Overstreet & Yule, 2021: 126). According to the findings of Palacios Martínez (2011a: 2460), there is a significant sociolinguistic divergence in the use of general extenders correlated to the speaker’s age. His data reveals that certain forms, such as ‘and stuff,’ ‘and that,’ ‘and everything’ are more prevalent in adolescent language than in adult speech. This higher frequency suggests that these items serve as key markers of youth identity and informal peer-group interaction (cf. Pichler & Levey, 2011: 453-456). Consequently, specific general extenders enable adolescents to negotiate their personal and collective identities; the integration of these markers into spontaneous discourse allows

younger speakers to delineate a unique personal stance while concurrently affirming their alignment with peers (Palacios Martínez, 2011b: 119). Furthermore, Palacios Martínez (2011b: 106) pointed out that adolescents are primary agents of linguistic innovation and change; they introduce variations that may eventually be integrated into the broader grammatical and lexical framework of a language. This creative capacity is particularly evident at the lexical level, as younger speakers frequently adopt and adapt terms from foreign languages or jargons. Palacios Martínez (2011b) define youth language as a specific variety emerging from the linguistic features prevalent in the informal written and oral communications of adolescents. While this variety shares some features with other registers used in colloquial contexts, it is uniquely shaped by the intersection of age and various sociolinguistic variables, including gender, socioeconomic status, and cultural background. Consequently, teenage discourse should be viewed not as a static or uniform variety, but as a heterogeneous phenomenon that evolves according to geographical and situational contexts, with the speaker's age serving as the primary defining characteristic. Similarly, Cheshire (2007: 187) observes that adolescents utilize adjunctive general extenders more frequently than disjunctive general extenders. Consequently, Cheshire argues that for a rigorous analysis, it is essential to treat adjunctive and disjunctive general extenders as separate categories rather than a single phenomenon.

In her study, Dines (1980: 18-19) analyzes general extenders, referred to as 'clause terminal tags,'⁵ as a sociolinguistic variable. She observes that the frequency of the marker 'and stuff like that'

⁵ These markers were later defined by Dines (1980: 23) as 'set marking tags.' She argues that their primary discourse role is to signal that the prior lexical item should be understood as a representative member of a larger category or set (as cited in Jucker, Smith & Lüdge, 2003: 1748).

varies across social strata, noting a higher prevalence in working-class speech compared to that of the middle class. Conversely, variants such as ‘and this sort of thing’ or ‘or whatever’ occurred with low frequency and lacked the clear social distribution found in more common tags. She noted that these variants did not achieve the high level of usage observed for markers like ‘and that’ within the working-class sample (1980: 23). Ultimately, Dines argues that while speakers across different groups may face similar discourse purposes, their choice of linguistic variants differs. Furthermore, she highlights a methodological challenge in the study of these markers, namely the difficulty of identifying a complete taxonomy of variants. This emphasizes the importance of identifying alternatives utilized by speakers who avoid using such terminal tags. In addition, Dines (1980: 19) explores the sociolinguistic salience of general extenders by examining how they influence listener perception. Through an experiment, she demonstrates that clause-terminal tags function as powerful social markers; middle-class evaluators in her study associate the use of these forms with working-class speakers. Crucially, Dines suggests that these markers may be subject to social stigma among certain groups, as they are often perceived as indicators of vague or inexplicit communication rather than as multifunctional tools used to manage the discourse.

These studies suggested that general extenders were characteristic of specific social groups, including the working class and youth. Under this early paradigm, the use of general extenders was interpreted as a compensatory mechanism for a lack of linguistic fluency, serving primarily to bridge gaps during the process of discourse planning. Consequently, general extenders were often defined as semantically vague elements or approximators. Current data, however, indicates that general extenders are employed by

speakers regardless of their socio-cultural background, appearing independently of gender, age, or level of education. Nevertheless, according to Cheshire's (2007: 188) study, while the distribution of general extenders showed no significant correlation with social class or gender, such findings do not signify a complete lack of sociolinguistic variation.

3.3 Classification of general extenders

According to the classification proposed by Mauri (2014; cf. Fiorentini, 2018: 22), general extenders can be categorized into two primary macro-categories: analytic and synthetic. Analytic general extenders consist of a conjunction, a pro-form, and an optional similarity marker, which primary function is to extend a category by implicitly referencing other potential members, as exemplified by Italian *e cose del genere* 'and stuff like that'. In contrast, synthetic general extenders result from the univerbation of previously analytic structures; a classic example is *eccetera* 'etcetera', which derives from the Latin locution *et cētera*. At the same time, general extenders are distinguished by their functional roles. Adjunctive forms are introduced by the conjunction 'and' (or the Italian *e* and Spanish *y*) and indicate that 'there is more' that could be listed within the category. According to Borreguero Zuloaga (2021: 79), syntactic variations characterize the use of adjunctive general extender. While many forms require a conjunction, certain forms allow general extenders to appear independently. For instance, Spanish speakers frequently employ non-conjunctive forms such as *tal* 'such' or *esas cosas* 'things like that'. A parallel phenomenon is observed in Italian, where the construction *cose così* 'things like that' serves as a prevalent example of a general extender appearing without a preceding conjunction. In contrast, disjunctive forms, introduced by

the conjunction ‘or’ (Italian/Spanish *o*), present alternative possibilities or express the speaker’s uncertainty regarding the content of an utterance. Crucially, both types predominantly occur in informal, everyday spoken interaction (Overstreet & Yule, 2021: 1). However, some forms, such as ‘etcetera’ are more frequently encountered in formal speech and writing, contrasting with the general extenders that are typical of spoken language (Overstreet & Yule, 2021: 3).

The typical placement of a general extender occurs at the utterance-final boundary, serving to signal that a specific topic list is concluded. This functional role is described as list-completion and turn-completion, as the general extender effectively marks the end of the speaker’s contribution within the ongoing interaction (Overstreet & Yule, 2021: 6). In this regard, Cortés (2006: 104) argues that the list-completer, such as *y todo eso* ‘and all that’ in Spanish, does not actually function as a placeholder for missing vocabulary. Instead, it serves as a delimiting marker, informing the listener that the speaker has provided a representative and complete set of examples. Because the preceding terms have already defined the boundaries of the topic, the general extender here acts primarily as a textual marker of closure, indicating that the list is complete and no further specific items are necessary.

3.4 The grammaticalization of general extenders

Grammaticalization is defined as a process consisting of a set of linguistic shifts occurring within a specific construction (Bybee, 2003). Hopper and Traugott (2003: 1) define the process of grammaticalization as:

That part of the study of language change that is concerned with such questions as how lexical items and constructions come in

certain linguistic contexts to serve grammatical functions or how grammatical items develop new grammatical functions.

According to Bybee (2003), these changes typically include phonetic reduction, decategorization and pragmatic shift. This phenomenon develops through the mechanism of generalization or bleaching. Throughout this process, linguistic forms lose their specific semantic features, allowing them to be employed across a broader range of contexts. This transition is rooted in habituation, where the high frequency of use of a linguistic expression diminishes its original impact (Meillet, 1919; Lehmann, 1982, as cited in Bybee, 2003: 605). According to Bybee (2003: 602), grammaticalization does not target words in isolation; rather, the process affects an entire construction, including its specific lexical components. This structural evolution is driven by the cognitive process of online categorization, which Mauri and Sansò (2020: 5) describe as a pervasive force within linguistic systems. They argue that ad hoc categorization is not limited to lexical choice but is central to grammatical domains such as number, plurality, and other construction like general extenders.

The relationship between frequency of use and reduction in form is a well-established principle of language change, and general extenders provide a clear illustration of this phenomenon. The short forms (e.g., *e cose* ‘and things’) are typically subject to phonological reduction in spontaneous speech, resulting in their frequent integration inside the tone unit, creating a tighter bond with the preceding constituent. This high integration correlates with their higher overall frequency. The employment of these reduced forms may primarily function as a conversational maintenance strategy, contributing to the prosodic rhythm of the utterance rather than

adding informational content (Overstreet, 2014: 112). For instance, variants such as the Italian *e cose così* ‘and things like that’, frequently incorporate an indexical element, typically featuring the pro-form ‘that’ in English or its equivalent *così* in Italian, which serves to connect anaphorically to a preceding constituent in the utterance. The indexical element signals that the vague category *cose* ‘things’ must be interpreted as similar to or analogous with the elements previously mentioned (Overstreet & Yule, 2021: 6). This extended form moved from its longer form *e cose così* ‘and things like that’ to *cose così* ‘things like that’ and ultimately to the short form *così* ‘that’, exemplifying the ongoing process of grammaticalization and decategorialization. Conversely, longer forms are less reduced and are far more likely to occupy a separate tone unit, signaling a greater pause, which reflects their distinct usage and lower frequency (Aijmer, 2002; Channell, 1994; cf. Overstreet & Yule, 2021: 3). As these reduced forms become routinized through frequent use, their pragmatic functions become increasingly fixed. As Levinson (1987: 99) argues, when an expression is repeatedly used to convey a specific meaning, over time, this repeated use causes the implied meaning to become a permanent part of the expression itself.

Traugott (2010: 30) argues that subjectification and intersubjectivity are distinct from the process of grammaticalization. Nevertheless, a significant overlap exists between these phenomena; because grammaticalization aligns with markers that signal a speaker’s attitude or establish discourse connectivity, it is inherently linked to the development of subjective and intersubjective meanings. In this sense, while their processes are independent, they function in close interaction.

General extenders should be distinguished from discourse markers for several reasons; they constitute a homogeneous category of words formed by a conjunction followed by a noun phrase, whereas discourse markers are formed by a wide variety of forms, such as conjunctions, adverbs, or idiomatic expressions. However, they do share some common characteristics with discourse markers, as they have a core meaning that is not conceptual and varies according to the context. Finally, both categories serve to express interpersonal functions and convey the speaker's attitude towards the message (Aijmer, 2002; cf. Palacios Martínez, 2011a: 2454). Nevertheless, speakers often pair general extenders with discourse markers as a compensatory mechanism, ensuring that the communicative adequacy of the utterance is maintained despite the vague nature of the general extender (Cheshire, 2007: 187).

The evolution of both general extenders and discourse markers is rooted in the process of grammaticalization. While the grammaticalization of discourse markers generally involves a higher degree of flexibility, as they can be placed in various positions within a sentence. Traditionally, general extenders have been characterized by a fixed position, consistently appearing at the end of a sentence or clause. However, recent studies have documented a shift in this paradigm; newer observations indicate that general extenders are beginning to occupy more internal positions within the utterance. Another defining feature of the process of grammaticalization is semantic transparency and interchangeability; once grammaticalized, the meaning of most general extenders remains invariant across contexts, allowing varied forms to fulfil pragmatic functions. This is often accompanied by decategorialization, where the pro-form loses its original referential meaning as a noun phrase to function as a non-referential pragmatic marker (Borreguero

Zuloaga, 2021: 94-95). According to Borreguero Zuloaga (2025: 193-195), the degree of grammaticalization within general extenders can be measured through several specific linguistic shifts. Grammaticalization often leads to the phonetic reduction of the conjunction, which may disappear entirely, or the univerbation of the conjunction and the core element (e.g., *eccetera* ‘etcetera’). Another tendency of this process is the loss of syntactic complexity, as the general extender structure moves from complex to simple core elements. Research indicates that general extenders with simple core elements represent a more advanced stage of grammaticalization than their complex counterparts, a fact supported by their higher frequency in speech. Overstreet and Yule (2021: 12) further refine this evolution by identifying three interrelated processes: decategorialization, morphosyntactic reanalysis, and phonological attrition. A primary indication that decategorialization has occurred is the loss of semantic and syntactic constraints on coordination.

Extract 6 (Reading, m.cl.)

Michael: no well they own this sort of minibus company. they just have a I think it's two or three minibuses and they take children to school and they take disabled children **and things** around (Cheshire, 2007: 169).

Here, ‘things’, which semantically refers to non-human entities, is attached to human referents (disabled children). In this instance, ‘things’ has lost its specific semantic feature, thus becoming integrated into a fixed pragmatic phrase (cf. Overstreet and Yule, 2021: 12). The second process is associated with morphological reanalysis or fusion which involves the reduction of multi-word phrases. This reduction results in fusing the whole elements into a single, unified form, such ‘some’ + ‘thing’ which becomes

‘something’. Finally, according to Overstreet and Yule (2021: 120), as these expressions experience decategorialization and reanalysis, they typically undergo phonological reduction, where highly frequent markers are subjected to phonetic erosion in speech. General extenders exemplify this phenomenon by illustrating the gradual loss of core elements for the sake of efficiency. This process leads to the sequential loss of initial parts, such as the conjunction or the noun, ultimately leaving only the minimal indexical element to carry the full pragmatic function of vagueness and completion. On the other hand, according to Cheshire (2007: 166-167), the most problematic area is the analysis of pragmatic shift. This difficulty arises because new interactional functions often emerge in discourse contexts before the internal structure of the construction has become fully fixed, meaning that the literal meaning of individual words may still be transparent while the expression already performs a new pragmatic role.

In instances where a general extender has not yet undergone grammaticalization, the head noun, such as ‘stuff’ or ‘things’, typically maintains a strict anaphoric relationship with a preceding referent, reflecting its original syntactic and semantic properties. The choice of a general extender is traditionally governed by the grammatical number and category of its antecedent: a singular mass noun would typically trigger the use of ‘and stuff like that,’ whereas a plural count noun would require the corresponding plural form ‘and things like that’. However, as these constructions undergo decategorialization, these rigid grammatical constraints begin to weaken. This structural weakening is exemplified by forms such as ‘and that,’ which often lack a clear, singular referent and may instead function as a marker for the entire preceding utterance (Cheshire, 2007: 168). Similarly, Dines (1980: 27) notes that while certain tags,

such as ‘and things’, typically follow plural count nouns, others, most notably ‘and stuff’, demonstrate greater flexibility appearing regardless of the grammatical agreement of the preceding element. A similar pattern is observed in Spanish, in which general extenders show a notable flexibility; despite their singular form, they can be appended to plural noun phrases. However, these characteristics observed primarily in Peninsular Spanish may not be universal across all Spanish varieties (Overstreet & Yule, 2021: 159).

3.5 The multifunctionality of general extenders

Concurrently with the standardization of the term ‘general extender,’ the assumption that these forms serve a primarily referential or categorizing function has been subjected to re-evaluation. Recent scholars have moved beyond a strictly set-marking analysis, underlining the multifunctionality inherent in these markers. It is now widely recognized that a single general extender can perform several communicative functions (Overstreet & Yule, 2021: 41). Supporting this idea, Cheshire (2007) pointed out that general extenders are not just used to extend a category, possessing instead a multifunctional role in language. A central aspect of her analysis is that the structure of a general extender is frequently tied to its specific role in interaction. This suggests that different general extender forms are not interchangeable; rather, they are selected by speakers to perform different communicative tasks within the discourse. In her analysis, Cheshire (2007: 183) observes that short-form general extenders are predominantly multifunctional, often operating across several linguistic layers simultaneously. Even when a general extender performs a distinct textual function, it frequently serves a parallel interpersonal one. For instance, in the context of reported speech, the general extender acts as a signal to the listener, framing

the utterance not in terms of its literal meaning, but as a representative or typical example of what was said. On the other hand, Aijmer's central thesis outlines that the multifunctionality observed in discourse particles and in this case, the general extenders, is a consequence of the grammaticalization process. This evolution begins when an expression loses its literal meaning through frequent use, thereby shifting into pragmatic and interpersonal functions, such as managing the discourse, signaling vagueness, or implementing politeness strategies. Consequently, because these elements are no longer anchored to a rigid literal definition, they gain the flexibility to perform multiple communicative roles depending on the context. Ultimately, the capacity for general extenders to simultaneously serve as list completers, attenuators, and stance markers confirms their status as pragmatic markers currently undergoing "ongoing grammaticalization" (Aijmer, 2002: 5). In this regard, Overstreet and Yule (2021: 116) argue that general extenders should be analyzed as constructions integrated into the grammar, performing specific grammatical functions, notwithstanding the fact that these forms are frequently categorized as pragmatic markers. To further understand the function of general extenders, it is useful to distinguish between the two core components of an utterance: the propositional and the expressive. The propositional component is tied to truth-conditional relations and uses referential forms to link language to the world of facts and objects. In this sense, a general extender serves to signal the remaining members of an ad hoc category. The expressive component is non-referential and focuses on the speaker's personal attitude toward the topic, the interlocutor, or the communicative event itself (Overstreet, 2014: 114). Cortés (2006: 110-111) categorizes the functions of general extenders into two primary

groups based on the speaker's communicative intent. The first group is interpersonally driven, focusing on the relationship with the listener, aiming for politeness, and fluid interaction. The second group pertains to the speaker's presentation of the discourse, conveying attitudes such as doubt, uncertainty, or, conversely, the intensification of certainty. Regarding this latter function, Cortés observes that a general extender can act as an intensifier of the speaker's assertion. When accompanied by specific prosodic features, such as higher pitch or increased emphasis, the GE does not merely close a list; it reinforces the validity of the preceding segments. This creates a dual effect: it reflects the speaker's total conviction in their statement while simultaneously increasing the persuasive force exerted on the interlocutor. Conversely, Cortés (2006: 120) identifies a specific class of general extenders used as informational mitigators. These forms signal a speaker's doubt or insecurity regarding the accuracy of their preceding statements. Cortés divides these mitigators into two distinct sub-groups: doubt and insecurity. In case of doubt, the GE serves as a corrective tool, helping the speaker satisfy the Gricean Maxim of Quality by signaling that the information provided may not be entirely precise or true. Similarly, markers of insecurity convey a sense of informative insufficiency, where the speaker acknowledges that their contribution lacks the necessary certainty or detail.

General extenders frequently cluster with other pragmatic devices, such as approximators ('sort of,' 'like') or epistemic stance adverbials ('probably,' 'maybe'). This synergy of markers in spontaneous speech shows that speakers often employ multiple expressions to manage both conversation and interpersonal functions simultaneously. The pairing of markers like 'you know' and 'and stuff' represents, according to Overstreet and Yule (2021: 86), a

highly intensified manifestation of a common textual organization pattern.

3.5.1 General extenders as markers of intersubjectivity

The utility of a general extender extends beyond the mere indication of additional members within a clearly delineated category. Frequently, speakers employ these expressions without specific items in mind, utilizing vagueness as a strategic tool to preserve conversational fluidity. Because general extenders lack fixed referential meaning, they function as pragmatic pro-forms whose interpretation is entirely contingent upon the immediate discourse context (Guil Povedano, 2000, as cited in Borreguero Zuloaga 2025: 187). Consequently, these expressions necessitate a semantic interpretation based on the mutual knowledge shared by the interlocutors of the interaction, where the listener relies on common ground to complete the implied list. This inferential process and the cognitive demand placed on a listener to decode a general extender varies according to the specificity of the implied information. The process relies on shared knowledge, ranging from universal cultural concepts to highly private information accessible only to the participants of a specific interaction (Overstreet, 1999: 69, as cited in Fernandez & Yuldashev, 2011). From this perspective, linguistic interaction is characterized by the use of conventionalized markers that explicitly signal intersubjectivity (Overstreet & Yule, 2021: 46). For instance, the Spanish marker *y eso* ‘and that’ appears to perform an intersubjective function closely aligned with that of ‘and that’ in British English (Fernández, 2015: 9). Whereas the Spanish marker *y tal* ‘and so on’ operates similarly to English adjunctive general extenders, such as ‘and so on’ or ‘and that kind of thing’ (Fernández, 2015: 10).

3.5.2 Interactive and textual functions of general extenders

In the context of discourse organization, general extenders predominantly fulfil interactive and textual (or meta-discursive) functions. Cortés (2006: 111) emphasizes the interactive dimension of general extenders, noting that they function as ‘initiatives’ designed to elicit a reaction from the interlocutor. In this view, the topic of conversation becomes secondary, and by using these markers, the speaker signals a sense of complicity, suggesting that further details are unnecessary because both parties share sufficient common ground to grasp the unspoken implications. Beyond their interpersonal role, general extenders perform textual functions related to discourse boundaries. Specifically, they can be utilized to foreground the introduction of a new topic or, conversely, to signal the segmentation of talk by marking the conclusion of a conversational turn (Overstreet & Yule, 2021: 92). In this capacity, general extenders play a crucial role in turn-taking by marking transition relevance places (TRPs), which function as mechanisms for yielding the turn or as pre-closing elements. Alternatively, they can act as floor-holding markers, enabling the speakers to structure their discourse while maintaining their conversational turn. While often found in turn-final positions to assist in topic management, their use is not restricted to the end of a turn; when a general extender appears within the turn, it typically signals that the speaker intends to continue discussing the current topic (Cheshire, 2007: 181). For instance, in spontaneous speech, Spanish forms such as *y tal* ‘and so on’ act as signals to divide discourse into smaller units; this strategy allows speakers to both organize their discourse and hold the floor (Borreguero Zuloaga, 2021: 89-90, 2022: 177). Furthermore, a general extender can trigger cooperative turn-taking when paired

with discourse markers like *claro* ‘sure’ or *pues sí* ‘so yes’, which signal mutual agreement and shared understanding between interlocutors (Borreguero Zuloaga, 2021: 84).

Beyond managing the speaker’s own turn, general extenders are also employed in reporting the speech of others. Spanish discourse features what Borreguero Zuloaga (2021: 80) calls ‘reported discourse extender’. This type occurs when a speaker reports the speech of others, serving to indicate an extension of the discourse itself rather than an expansion of a lexical set or list. This function is manifested through three primary structures: adjunctive extenders, which often incorporate onomatopoeic nouns, such as in *y patatín y patatán* ‘and so on and so forth’ or *y bla bla bla* ‘and blah blah blah’. The repetition of phonemes in these cases is strategically employed to evoke the impression of a lengthy and ongoing discourse. Within subordinate clauses, the construction involves a verb of saying followed by the complementizer *que* ‘that’ and the conditional conjunction *si* ‘if’. Prevalent in indirect reported speech, these structures typically conclude with a repetitive noun or demonstrative (e.g., *que si esto y aquello* ‘that it’s this and that’, *que si tal y cual* ‘that such-and-such’). Finally, epistemic interrogative clauses utilize an epistemic verb to introduce an interrogative pronoun, often appearing in parallel pairs such as *no sé qué* ‘I don’t know what’, *no sé cuántos* ‘I don’t know how many’ (Borreguero Zuloaga 2022: 162-163). As noted by Gille and Håggvist (2010: 136), in the context of reporting the speech of others, general extenders act as specialized reported discourse extenders that evoke the sensation of an exchange without the need for lexical exactness. In this way, the speaker’s primary goal may shift from literal repetition to conveying a general impression of the original interlocutor’s perspective, emotional tone, or stance. This

communicative effect is managed by articulating a single representative term or utterance, followed by a general extender.

3.5.3 Linguistic economy, mitigation and intensification in general extenders usage

General extenders function as a specific type of list marker that typically appear when a speaker is providing a list. As Masini, Mauri and Pietrandrea (2018: 50) argue, a list is a concatenation of units filling the same function, generally requiring at least two members (2018: 56). In spoken interaction, these structures allow speakers to manage turns and signal the boundaries of a topic. In this regard, general extenders serve as a mechanism for linguistic economy, facilitating timesaving in two distinct ways. First, they allow the speaker to omit information that is already part of the common ground or shared knowledge between interlocutors. Second, they function as a politeness strategy; by avoiding a detailed explanation of facts the listener is presumed to already know, the speaker prevents the social offense of appearing patronizing (Borreguero Zuloaga, 2022: 171). Indeed, as pointed out by Fernández (2015: 7), providing an exhaustive list is often deemed “interactionally awkward”; by using a general extender, the speaker signals that the remaining items are easily inferable. These economy-driven functions are deeply rooted in the Gricean Maxim of Quantity. By signaling that ‘there is more,’ adjunctive general extenders act as a hedging strategy, allowing speakers to avoid being more informative than is strictly required by the context. In doing so, general extenders mark the boundaries of a message’s informativeness while simultaneously indicating the amount of information the speaker assumes is shared knowledge (Overstreet, 2014: 119-122). However, the pragmatic effect of this hedging varies significantly depending

on the form chosen. For instance, the degree of mitigation provided by disjunctive general extenders depends on their specific form and level of vagueness. Forms like *o algo* ‘or something’ or *o lo que sea* ‘whatever’ indicate high uncertainty or a lack of commitment to accuracy, which is often regarded as unimportant by the speaker. In contrast, the use of *o algo así* ‘or something like that’ effectively delimits the field of possibilities; consequently, it suggests a lower degree of mitigation than that provided by broader expressions (Borreguero Zuloaga, 2021: 86). Another significant function of these disjunctive markers is to signal that the speaker’s preceding statement is an approximation of the speaker’s actual thoughts (Gille and Häggkvist, 2010: 135). Crucially, this implies that general extenders are not pragmatically interchangeable; for instance, while ‘or something’ (It. *o qualcosa del genere*) typically signals approximation, the choice of ‘or whatever’ (It. *o quel che è*) may convey indifference or even irritation on the part of the speaker (2021: 8).

General extenders can alternatively function as intensifiers to highlight the importance or surprising nature of an idea. This aligns with what Overstreet and Yule (2021: 8) categorize as a ‘personal function,’ reflecting the capacity of these markers to signal a speaker’s stance or feelings toward the ongoing utterance. Consequently, in addition to their referential role of signaling ‘there is more,’ general extenders function as intensifiers that reinforce the preceding utterance. In such instances, speakers appear to flout the Maxim of Quantity; by using a general extender to maximize the utterance, they indicate that the intensity of the experience exceeds the constraints of a conventional list (Overstreet & Yule, 2021: 63). As noted by Gille and Häggkvist (2010: 138), these markers can signal intentional exaggeration or frame the preceding information

as particularly noteworthy or surprising. The intensification is often achieved through repetition (e.g., *y etcétera etcetera* ‘and etcetera etcetera’) or by the combination of different general extenders, suggesting to the hearer that the implicit discourse is significantly longer, and that additional information could be integrated into the current topic (Borreguero Zuloaga, 2021).

3.6 The case of Italian *eccetera*

In the current Italian linguistic repertoire, *eccetera* holds the highest frequency among general extenders (Guil, 1999, as cited in Fiorentini & Magni, 2021: 306). While these markers are typically associated with spontaneous oral discourse, *eccetera* also appears with significant frequency in formal written language. The Latin expression *et cētēra* literally means ‘and the remaining things’ or ‘and the other things’ (Fiorentini & Magni, 2021: 296). During the Old Italian period, *et cetera* experienced phonological reduction, resulting in the univerbation of the modern form *eccetera*. A similar process occurred in other romance languages, yielding forms such as Spanish *etcétera*, Portuguese *etcétera*, and French *etcétéra*. Historically, this marker, frequently appearing in abbreviated forms such as ‘etc.’ or ‘ecc.’, was placed at the end of a list or a citation. Its primary function was to serve as a substitute for information or omitting details already familiar to the interlocutor (Fiorentini & Magni, 2021: 303; cf. also Fiorentini, 2018). The form was originally categorized as an adjunctive general extender composed of a conjunction and an indefinite. However, its usage has evolved through the incorporation of adjectival modifiers, a notable example of which is the Spanish phrase *y un largo etcetera* (‘and a long etcetera’), in which the adjective *largo* (‘long’) is added to the base form *etcetera* to reinforce it. In this context, the modifier serves as

an intensification mechanism, allowing the speaker to highlight the surprising nature of the category while maintaining linguistic economy (Borreguero Zuloaga, 2022: 161). Like other general extenders, *eccetera* signals the existence of alternative set-members that share a common property with the explicitly mentioned items. In terms of its internal structure, *eccetera* stands apart from the prototypical Italian general extender. While most Italian general extenders are analytic, *eccetera* represents a synthetic form (Fiorentini & Magni, 2021: 310). In the case where the general extender *eccetera* functions as a pro-form, analogous to *cosa* ('thing'), it requires an initial conjunction to specify whether it performs an adjunctive or disjunctive function within the discourse (Fiorentini, 2018: 35). When preceded by the conjunction 'and' (*e* or *y*), it functions as an adjunctive general extender, implying a continuous list of equivalent elements. Conversely, when introduced by the disjunction 'or' (*o*) it operates as a disjunctive general extender; in this capacity, the speaker provides two alternatives as representative examples to define the category (Fiorentini & Magni, 2021: 309). According to Overstreet and Yule (2021: 33) a specific function of *etcetera* is to signal a continuation of the preceding discourse, suggesting that while a sequence remains incomplete, it has reached a point of communicative sufficiency. In the context of list construction, the marker indicates that further items could be added, though they are not strictly necessary for the listener's understanding. However, speakers and writers often employ *etcetera* to invite the interlocutor to mentally complete a broader line of thought or narrative, relying on the audience's ability to infer the implied extension of the list.

Fiorentini (2018) identifies four distinct pragmatic functions for the Italian general extender *eccetera*:

- Completion: Invites the listener to mentally fill in missing parts of a specific, known text, such as a proverb or fixed expression.
- Completion with Abstraction: Indicates additional, unspoken information that shares a defining property with those already mentioned.
- Alternative and Abstraction: Indicates the existence of other elements sharing a property with the explicit list, though these members function as alternatives to one another, similar to the disjunctive general extenders.
- Abstraction: Suggests that *eccetera* does not refer to additional items but rather highlights a specific contextual property necessary to interpret the intended meaning based on shared knowledge.

3.7 General extenders in non-native speakers

In the field of second language acquisition, learners frequently overuse or underuse a specific linguistic marker compared to native speakers. This is particularly evident in spontaneous interaction, which is characterized by its unplanned nature. Since such conversation is generated under significant cognitive and processing constraints, these pressures manifest through various features such as disfluencies, including both filled and unfilled pauses, repetition of words or phrases and the appearance of incomplete grammatical structures (Aijmer, 2004: 174). In this context, learners may employ general extenders as fillers or extended constructions, a practice that can be interpreted as a discourse strategy aimed at managing cognitive load. By utilizing these markers, learners gain the necessary temporal gap to organize their thoughts and select the most appropriate linguistic expression for their next utterance (Buysse,

2014: 13). On the other hand, competent native speakers employ discourse particles with high precision to facilitate the listener's comprehension. Consequently, the incorrect application or total omission of these particles by non-native speakers can obscure the intended logic of the discourse, potentially resulting in communicative breakdowns or misunderstandings (Aijmer, 2002: 3). As Ruzaitė (2010: 34) points out, the omission of general extenders can have significant negative implications for the informative quality of discourse. These units serve as essential cues, guiding the listener to interpret the preceding lexical item not as an isolated entity, but as an illustrative example of a broader category. Without these markers, the listener may fail to recognize the speaker's intended categorization, leading to a less accurate interpretation of the message. Similarly, Dines (1980) defines the primary function of these markers as a signal to the interlocutor; specifically, the tag serves to: "cue the listener to interpret the preceding element as an illustrative example of some more general case".⁶ According to Ruzaitė's (2021) study, an unexpected finding emerges regarding the frequency of vague language. Contrary to the previous research (cf. Aijmer, 2004; Drave, 2001), which typically indicates a lower frequency of vague language in non-native speaker (NNS) discourse, Ruzaitė's results demonstrate that certain categories of vagueness are more prevalent among non-native speakers than native speakers. Ruzaitė suggests that a possible interpretation for this phenomenon is that: "learners lack awareness of register variation"⁷ leading them to use these markers in contexts where a native speaker might opt for

⁶ Dines, E. R. (1980). Variation in discourse: "And stuff like that". *Language in Society*, 9(1), p. 22. <https://www.jstor.org/stable/4167109>

⁷ Ruzaitė, J. (2021). Vague language in argumentative essays written by advanced Lithuanian learners of English. *Eesti Rakenduslingvistika Ühingu aastaraamat Estonian Papers in Applied Linguistics*, 17, p. 240.

more precise or formal alternatives. However, beyond that lack of register awareness, vagueness also serves as a crucial pragmatic function in the learner's interlanguage, particularly regarding politeness. Because requests are face-threatening acts, learners often struggle to employ them with the necessary sociopragmatic nuance to avoid appearing unintentionally imposing or intrusive. The strategic use of a general extender within a request can help learners to preserve the listener's negative face by softening the force of the imposition and making the utterance seem less definitive (Fernández, 2015: 11). Conversely, within native speaker interaction, pragmatic markers are primarily associated with interpersonal function rather than merely signaling a lack of clarity or uncertainty. According to the framework established by Brown and Levinson (1987, as cited in Aijmer, 2004: 182), these elements are used for face-management, allowing interlocutors to navigate the complexities of social rapport through politeness strategies and communicative indirectness. Rather than indexing imprecision or uncertainty, their presence often reflects a deliberate effort to soften the impact of an utterance or maintain social harmony.

The theoretical foundation outlined in these chapters provides a comprehensive framework for understanding the nature, functions, and cross-linguistic behavior of general extenders. Having established these foundational concepts, the following chapters shift the focus to the empirical dimension of the research. Chapter 4 outlines the methodology, dataset, transcription and annotation criteria employed to investigate the interview corpus, while chapter 5 presents a comparative qualitative analysis of the findings.

4. METHODOLOGY AND INTERVIEW CORPUS

This chapter outlines the methodological framework and the characteristics of the corpus compiled for this study. Section 4.1 provides an overview of the participants and the general features of the dataset. Section 4.2 presents the sociodemographic and linguistic profiles of the informants, while section 4.3 details the structure of the interviews. The methodological steps are then described in sections 4.4 and 4.5, which cover the transcription and annotation phases. Finally, section 4.6 introduces the quantitative distribution of the identified GE forms.

4.1 Participants and corpus description

The corpus for this study consists of semi-structured interviews conducted with a sample of 10 informants aged under 35. To ensure a balanced comparative analysis between Italian and Spanish, the sample was divided into two groups based on the participants' native language; five Italian native speakers and five Spanish native speakers who are currently studying or have previously studied Spanish or Italian as a foreign or second language. Convenience sampling was adopted, with informants recruited from the researcher's personal and academic spheres.

The primary aim of this study is to investigate and compare the pragmatic behavior of general extenders in the spontaneous speech of two distinct groups of bilingual/L2 speakers: Italian native speakers studying Spanish, and Spanish native speakers studying Italian. Specifically, this research aims to address two core questions. The first one is related to the quantitative distribution of GEs, namely, to determine which of the two language groups exhibits a

higher frequency of GE production during the interaction. Finally, to classify and analyze the specific function behind the use of GEs by both groups. The study investigates whether the informants employ these markers primarily with a referential, personal, interpersonal, or textual function.

4.2 Informant profiles

In alignment with established research and ethical guidelines, all participants took part in the study on a voluntary basis. Each informant read and signed an informed consent form, which explicitly outlined the academic purposes of the research and the voluntary nature of their participation. Crucially, to guarantee absolute confidentiality, all personal data were anonymized. The real identities of the informants were replaced with alphanumeric codes (e.g., INF_1, INF_2) throughout the data processing, analysis, and final transcription phases, ensuring that no recognizable personal background could be traced back to any individual participant. Following standard data management protocols, each speaker's turn is captured as an isolated textual segment and marked with an informant-specific label (e.g., INF_NO). As highlighted in qualitative data literature, this systematic labelling protocol is crucial because it connects the textual data directly to the interviewee demographic profile (McLellan et al., 2003: 67-68). This structured tracking ultimately facilitates a rigorous comparison of linguistic patterns across different informants' linguistic, cultural, and pragmatic backgrounds. Table 1 provides an overview of the sociodemographic and linguistic profiles of the informants included in this sample. For each participant, key variables are detailed, including language background, age, gender, country of origin, level

of education attained, and reported target language proficiency level according to the CEFR framework.⁸

Informant	Group (L1/L2)	Age/ Gender	Country of origin	Educational level	LS level
INF_1	Italian L1/ Spanish LS	23/Female	Italy	MA	B2
INF_2	Italian L1/ Spanish LS	24/Female	Italy	BA	C1
INF_3	Italian L1/ Spanish LS	23/Female	Italy	MA	C1
INF_4	Italian L1/ Spanish LS	24/Female	Italy	MA	C1
INF_5	Spanish L1/ Italian LS	19/Female	Venezuela	High School	B2
INF_6	Spanish L1/ Italian LS	25/Male	Peru	High School	B1
INF_7	Spanish L1/ Italian LS	33/Female	Chile	MA	C1
INF_8	Spanish L1/ Italian LS	27/Male	Ecuador	MA	C1
INF_9	Spanish L1/ Italian LS	31/Female	Chile	Postgrad	B1/B2
INF_10	Italian L1/ Spanish LS	24/Female	Italy	MA	C1

Table 1. Sociodemographic and linguistic profile of informants.

4.3 The interview structure

The interview was structured into three distinct phases.⁹ The first phase comprised the collection of sociodemographic and linguistic data from the informants. The second phase consisted of a semi-

⁸ <https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions> (Accessed: 10 July 2026).

⁹ The interview questions are provided in Appendix A.

structured elicitation task, featuring targeted questions specifically tailored to trigger the spontaneous production of general extenders. The questions were designed to address different communicative dimensions, starting with an initial set that explored the informants' personal perceptions regarding their linguistic difficulties in acquiring the target language, as well as their study habits and the specific pedagogical materials they regularly utilize. Subsequent questions focused on daily routines and personal experiences, requiring the informants to detail their supermarket shopping habits, outline the organization of a hypothetical travel or event, describe a recent celebration, and recommend tourist attractions within their hometown. Finally, the third phase introduced a bidirectional translation. This task involved two fragments: an Italian source text to be translated into Spanish, and a Spanish source text to be translated into Italian. Crucially, specific general extender forms had been artificially embedded within both texts (*e roba del genere* 'and stuff like that' and *o qualcosa del genere* 'or something like that' in the Italian text; *y esas cosas* 'and things like that' in the Spanish text). This experimental design aimed to observe how bilingual informants translated these vague expressions.

4.4 Transcription phase

The empirical phase was executed in sequential phases. First, the semi-structured interviews were audio-recorded with the participants' permission and subsequently transcribed manually using Microsoft Word. As noted in qualitative research literature, the scope and depth of a transcript must strictly complement the intended depth of the analysis, meaning that decisions regarding text inclusion should be fundamentally driven by the core research questions (McLellan et al., 2003: 67). To balance overall text relevance with

the specific needs of this study, the entirety of each interview was first fully transcribed by hand. From this comprehensive textual dataset, individual utterances containing general extenders were isolated and extracted for further linguistic analysis. The transcription process adhered to the established criteria of the KIParla corpus,¹⁰ incorporating a simplified version of the Jefferson transcription system, which is widely utilized in Conversation Analysis (CA) (Mauri et al., 2019); the specific transcription conventions applied throughout this process are systematically outlined in Table 2.

Symbol	Description
,	Rising intonation
.	Falling intonation
:	Prolonged sound
(.)	Brief pause
> hello <	Bracketed speech is delivered more rapidly
< hello >	Bracketed speech is delivered more slowly
[hello]	Overlap between participants
(hello)	Hardly intelligible speech (transcriber's best guess)
xxx	Unintelligible speech
((laughs))	Non-verbal behavior
=	Prosodically attached units

Table 2. Simplified Jefferson transcription system adaptation. Adapted from the official transcription guidelines of the KIParla corpus.

¹⁰ <https://kiparla.it/il-corpus/> (Accessed: 24 May 2026).

4.5 Annotation phase

Following the completion of the text transcripts, the study proceeded with the manual data annotation phase.¹¹ All identified general extender forms, and their corresponding pragmatic functions were systematically compiled within an Excel spreadsheet for subsequent analysis. Structurally, the corpus was mapped to distinguish the language of production, the specific informant code, and the context in which the GE occurs. The analysis involved a detailed examination of the immediate co-text for each occurrence, focusing on the following dimensions: the number of exemplars provided, and the type of conjunction used (and/or). Each GE was coded for its pragmatic function as well as for the presence of cues or explicit exemplification markers (e.g., It. *diciamo* ‘let’s say’, Sp. *por ejemplo* ‘for instance’).

This study adopts a pragmatic annotation scheme based on the theoretical framework proposed by Overstreet and Yule (2021). The manual annotation process employs a tagset designed to capture the multifunctional nature of GEs across four main pragmatic dimensions:

- Referential Function (FUN_REF): Annotated under the dimension of categorization, this tag marks instances where mentioned items serve as exemplars of an open-ended set or category.
- Interpersonal Function (FUN_INT): Operating within the dimension of intersubjectivity, this tag identifies appeals to shared background knowledge, mutual understanding, or common ground between interlocutors.

¹¹ The full annotation of each GE is provided in Appendix B.

- Personal Function (FUN_PER): Tied to the dimension of subjectivity, this tag is applied to express speaker-centered attitudes, such as uncertainty, mitigation, or hesitation.
- Textual Function (FUN_TEX): Related to turn construction, this tag denotes discourse-structuring devices used to indicate list closure or to manage conversational turns.

Finally, the underlying cognitive operations triggered by the GEs were categorized into distinct sets, classes, or frame. This systematic matrix ensured a rigorous, multidimensional taxonomy capable of capturing how bilingual speakers strategically deploy vague language across both Italian and Spanish. To categorize the cognitive operations triggered by general extenders, the present study adopts the theoretical framework articulated by Mauri (2017: 302). This approach suggests that languages use specific contextual clues to create abstract categories. In this process, the speaker provides one or more examples, which the listener uses as a starting point to infer a broader category. Depending on the conceptual relationship tying these specific exemplars together, the resulting inferential process yields three distinct mental representations:

- Set: Formed when the mentioned exemplars are co-occurring elements that combine to achieve a specific goal (e.g., listing ingredients required for a recipe).
- Class: Constructed when the exemplars function as equivalent alternatives belonging to a shared taxonomic group (e.g., listing distinct yet interchangeable quiet activities suitable for a child).
- Frame: Evoked when the exemplars represent a sequential chain of actions that fit into a narrative scheme (e.g., the chronological steps involved in dining at a restaurant).

By applying this tripartite taxonomy, each GE token in the corpus was coded based on the specific type of mental abstraction it requires the listener to perform.

4.6 Quantitative distributions of GEs

To provide an initial overview of the quantitative findings, Table 3 outlines the distribution and frequencies of GEs identified across the Italian and Spanish sub-corpora. Although the dataset is relatively small and most individual GE forms feature a single occurrence, these distributional patterns nonetheless allow for meaningful observations. The data reveal a clear disparity in the distribution of GEs between the two languages, with Spanish accounts constituting two-thirds of the total occurrences (66.7%), compared to the Italian data which account for the remaining third (33.3%), as detailed in Table 3 below.

GEs form	Frequency	%
ITALIAN DATA		
eccetera	1	4.2%
e tutto	1	4.2%
queste cose qua	2	8.3%
tutte cose buone	1	4.2%
queste cose	1	4.2%
le solite cose	1	4.2%
un po' di tutto	1	4.2%
Total Italian GEs	8	33.3%
SPANISH DATA		
y cosas de esas	1	4.2%

y cosas que se hacen gas rápido	1	4.2%
o cualquier cosa	1	4.2%
o algo tecnológico	1	4.2%
lo que sea	1	4.2%
y otras cosas	1	4.2%
un poco de todo	1	4.2%
cosas de la vida cotidiana	1	4.2%
y bueno eso	1	4.2%
los clásicos	1	4.2%
eso así	1	4.2%
como cosas didácticas	1	4.2%
esas clásicas diferencias	1	4.2%
algo más dulce	1	4.2%
o algo así	1	4.2%
y luego todo lo que viene	1	4.2%
Total Spanish GEs	16	66.7%
Global Total GEs	24	100.0%

Table 3. General extenders forms and frequency distribution across the Italian and Spanish sub-corpora.

Chapter 4 has outlined the methodological steps adopted for the study, detailing the sample selection, data collection procedures, and the analytical criteria used for data transcription and annotation. The following chapter will present the analysis of general extenders in the corpus, followed by a discussion of the main findings considering the research questions.

5. ANALYSIS OF THE INTERVIEW CORPUS

This chapter presents the qualitative analysis and discussion of empirical findings. Section 5.1 offers a comparative qualitative analysis of the GEs identified in the dataset, examining the Italian sub-corpus (Subsection 5.1.1) and the Spanish sub-corpus (Subsection 5.1.2). Section 5.2 expands on these findings by examining the translation task, focusing on the cross-linguistic rendering of the target general extenders markers and their pragmatic equivalence across the two languages.

5.1 A comparative qualitative analysis

This section presents a qualitative analysis of the 24 general extenders identified in the interview data, organized by sub-corpus (Italian, subsection 5.1.1; Spanish, subsection 5.1.2). For each extract, the general extender is discussed with reference to its immediate co-text, its syntactic construction, and its predominant pragmatic function, following the annotation scheme based on the taxonomy proposed by Overstreet and Yule (2021) and presented in chapter 4. Where relevant occurrences displaying construction-specific features (such as the presence of an explicit exemplification cue, a non-canonical syntactic position, or evidence of multifunctionality) are discussed in additional detail. All extracts are reproduced in their original transcription, followed by an English translation.

5.1.1 Italian sub-corpus

This subsection presents the qualitative analysis of the general extenders identified in the Italian sub-corpus. The analysis opens

with Example (1), which illustrates the use of *eccetera*. Although traditionally regarded as a canonical form of the written domain; its presence in spoken interview data is well attested and aligns with previous studies highlighting *eccetera* as a frequent general extender in contemporary spoken Italian (cf. Fiorentini, 2018; Fiorentini & Magni, 2021).

- (1) *quindi prima cosa sicuramente scelgo la meta poi quando ho scelto la meta e: e ho appunto delle date l'alloggio poi successivamente: l'hotel eh il volo o comunque i mezzi di trasporto e: e poi dopo tutto il resto quindi i luoghi di interesse itinerari eccetera* (INF_1)
 ‘so first thing for sure I choose the destination then when I have chosen the destination a: and I have precisely the dates the accommodation then subsequently: the hotel eh the flight or anyway the means of transport and: and then after all the rest so the places of interest itineraries **etcetera**’

In (1), the informant utilizes *eccetera* to signal that the list of holiday-related components (places of interest, itineraries) could be further extended. From the pragmatic perspective, the function is classified as referential (FUN_REF): the general extender signals that further, unspecified elements belong to the same category of preparatory actions. In line with Fiorentini’s (2018) taxonomy, this instance exemplifies completion with abstraction, as *eccetera* invites the listener to mentally infer additional, unspoken items that share a defining property. On a cognitive level, this utterance triggers a narrative frame according to Mauri’s (2017) classification. The speaker does not present a categorization of interchangeable alternatives (class) nor a static bundle of co-occurring situational entities (set). Instead, the discourse is highly structured through explicit temporal and sequential anchors (‘first thing’, ‘then’, ‘then subsequently’, ‘and then’). The exemplars provided by the informant represent consecutive, logical steps embedded in a chronological

script, specifically, the process of travel planning. Additionally, the vocalic prolongations captured in the transcript (*e*: ‘and:’) serve as hesitation cues and turn-holding devices. This allows the speaker to balance the cognitive load of real-time planning while maintaining conversational control over the ongoing listing. As illustrated in (1), this represents the only instance in the Italian sub-corpus where the general extender fulfils a referential function. The analysis now shifts to cases where these expressions serve an interpersonal, textual, or personal function as shown in Example (2).

- (2) *vabbè cose per la casa ehm (.) sì queste cose qua* (INF_2)
 ‘oh well things for the house ehm (.) yes **these things here**’

In (2), the GE follows a single, already vague exemplar (*cose per la casa* ‘things for the house’). The preceding pause (.) suggests that the speaker struggles to retrieve further specific items; consequently, she resorts to the GE *queste cose qua* ‘these things here’ as a closing device on a topic she treats as self-evident and not worth elaborating further. The classification as FUN_PER reflects this: the GE does not extend a taxonomic set so much as signal the speaker’s own limited engagement with specifying the category, a subjective stance arising from this processing difficulty rather than a referential or interpersonal function. A similar pragmatic function can be observed in Example (3) below.

- (3) *abbiamo mangiato cibi tipici calabresi un po’ troppo piccanti per i miei gusti però tutto molto buono c’era vabbè classica salsiccia piccante un po’ di (.) taralli adesso non mi ricordo più i nomi però. tutte cose buone* (INF_2)
 ‘we ate typical calabrian food a bit too spicy for my taste but everything was really good there was like oh well the classic spicy sausage some (.) taralli now I don’t really remember the names anymore but. **good stuff like that**’

In Excerpt (3), the occurrence departs from the canonical general extender pattern in that it does not straightforwardly imply additional

unnamed members of a taxonomic category. Although a referential function was initially intended to complete the list, the speaker explicitly admits an inability to recall further items (*adesso non mi ricordo più i nomi* ‘now I don’t really remember the names anymore’). Consequently, rather than expanding the category, the GE functions as an evaluative closure, leading to its classification as FUN_PER. The speaker employs this position to convey personal stance. This case illustrates how the same syntactic position can be exploited for stance-taking beyond category extension, thereby demonstrating the inherent multifunctionality of GEs. The shift toward the interpersonal function is exemplified in occurrence (4).

- (4) *c'è stata la proclamazione e tutto abbiamo un po' festeggiato lì in università e la sera siamo andati a mangiare in un ristorante* (INF_1)
 ‘there was the graduation ceremony **and everything** we celebrated a bit there at the university and in the evening we went out to eat at a restaurant’

In (4), the informant utilizes the adjunctive marker *e tutto* ‘and everything’ immediately following a single exemplar (the graduation ceremony). By attaching *e tutto* to the graduation ceremony, the speaker signals that the event entailed a series of conventional activities (such as taking photos or receiving congratulations) that belong to the interlocutor’s shared common ground, thereby fulfilling an interpersonal function (FUN_INT). The shared cultural background between the speaker and the interviewer enables these unmentioned details to be presupposed and taken for granted within the interaction. On a cognitive level, following Mauri’s (2017) framework, this occurrence is codified as a narrative frame. Unlike (1), where the frame was constructed dynamically through sequential steps, here the frame is evoked instantaneously through a socio-cultural script. The noun ‘graduation ceremony’ acts as a cognitive

trigger for the “university graduation script.” Consequently, the GE serves as a cue for the listener to infer and reconstruct the scenario. This same interpersonal function is exemplified in (5).

- (5) *cioè ti faccio un esempio sono andata a malta eh l'anno scorso e quindi vabbè ho controllato ovviamente lì si guida a (.) all'opposto rispetto a qua e ehm anche il fatto ad esempio le prese della corrente sono diverse quindi informarsi anche. queste cose* (INF_3)
'I mean let me give you an example I went to malta uh last year and so oh well I checked obviously there they drive on (.) the opposite side compared to here and um also the fact for instance the the power outlets are different so like getting information. **these things**'

As shown in (5), the GE is preceded by an explicit exemplification cue (*ti faccio un esempio... ad esempio* ‘let me give you an example... for instance’) and follows a clausal exemplar (differences in electrical sockets) within a broader discussion of practical precautions when travelling abroad. Forms involving *cose* ‘things’ represent one of the most frequent GE strategies in contemporary spoken Italian, after *eccetera* (cf. Fiorentini, 2018; Fiorentini & Zanchi, 2024). Moreover, the use of a demonstrative/deictic element within this structure (e.g., It. *queste cose* ‘these things’) enhances the strong (inter)subjectivity characteristic of spoken dialogic interaction (cf. Voghera, 2017, as cited in Fiorentini & Zanchi, 2024: 165). The presence of the exemplification marker, combined with the general extender’s role in generalizing from a single example to a wider category of “things one should check before travelling,” strongly supports an interpersonal reading (FUN_INT): the speaker is guiding the interlocutor toward a shared inferential frame (“you know the kind of thing I mean”), rather than merely listing further unspecified referents. A similar interpersonal function emerges in

Example (6), with the slight difference that the GE precedes rather than follows the exemplars.

- (6) *allora sì recentemente c'è stata una laurea di una mia amica e: abbiamo fatto un aperitivo e: e nulla c'erano **le solite cose** pizzette focacce e: panini e: era: siamo state qua a pavia in un locale qua a pavia era ma sì un contesto abbastanza informale.* (INF_3)
 'so yeah recently there was a graduation of my friend and: we did an aperitivo and: and basically there were **the usual things** mini pizzas *focaccia* and: sandwiches and: it was: we stayed here in pavia at a place here in pavia it was like yeah a pretty informal context.'

In (6), the GE precedes rather than follows its exemplars, introducing the list of finger-food items (mini pizzas, *focaccia*, sandwiches) associated with an aperitivo. The adjective *solite* ('usual') itself carries a categorizing function, framing the subsequent items as instantiations of a familiar, socially shared script. This prepositioning, together with the evident appeal to shared cultural knowledge of what an Italian aperitivo typically includes, motivates the FUN_INT classification. The same function is observed in Example (7) below.

- (7) *abbiamo fatto soltanto questo aperitivo in cui: mhm (.) vabbè si chiacchierava facevano le foto, **queste cose qua** insomma* (INF_3)
 'we just did this aperitivo where: mhm (.) oh well we were chatting they were taking photos, **these things here** basically'

As illustrated in (7), two exemplars (*si chiacchierava, facevano le foto* 'chatting, taking photographs') are extended by *queste cose qua* 'these things here', further reinforced by the closing discourse marker *insomma* 'basically'. The speaker treats the described activities as unremarkable and mutually understood, inviting the interlocutor's tacit agreement and serving an interpersonal function

rather than adding referential specificity. The textual function, which was identified only once in the Italian sub-corpus, is exemplified in (8).

- (8) *mangiare quindi diciamo diciamo cibo di per la cena o per il pranzo la pasta piuttosto che un secondo per la cena o magari non lo so prodotti make-up, quindi per truccarmi o: vestiti se in supermercato ci sono alcuni vestiti magari di qualche centro commerciale o anche sai prodotti per l'igiene insomma prima di prima necessità di tutti i giorni ecco (.) **un po' di tutto*** (INF_4)
 ‘food so like let's say food for for dinner or for lunch pasta or a main course for dinner or maybe I don't know make-up products, so to put on make-up or: clothes if in the supermarket there are some clothes maybe from some shopping center or even you know care products anyway everyday basic necessities well (.) **a bit of everything**’

Example (8) is the most structurally elaborate occurrence in the Italian sub-corpus, closing a five-member list of purchasable everyday items (food, make-up, clothing, care products) explicitly introduced by the marker *diciamo* ‘let’s say’. The speaker combines other pragmatic markers with the GE. As discussed in section 3.5, this alignment of pragmatic markers (*you know*) alongside the general extender reflects what Overstreet and Yule (2021: 86) term a highly intensified manifestation of a common textual organization pattern. The FUN_TEX classification directly accounts for this structural complexity, reflecting the general extender’s role in organizing and closing a syntactically and cognitively demanding list-building sequence, marking the boundary of the enumerative unit within the ongoing turn.

5.1.2 Spanish sub-corpus

This subsection presents the qualitative analysis of general extenders in the Spanish sub-corpus, examining their pragmatic, syntactic, and

structural features in conversation. The first excerpt, Example (9), provides a clear instance of a referential GE.

- (9) *eh a veces la: la dobla cuando hay doble consonante un poco confuso pero creo que leyendo escuchando libros y cosas de esas eh lo aprendo poco a poco* (INF_5)
'eh sometimes they: they double it when there's a double consonant a bit confusing but I think that by reading listening to books **and things like that** eh I learn it little by little'

Focusing on the referential dimension, Excerpt (9) provides a clear instance where two exemplars (*leyendo, escuchando* 'reading, listening') referring to language learning strategies are extended by the Spanish adjunctive GE *y cosas de esas* 'and things like that'. This occurrence is categorized as 'class' within the annotation scheme, as the GE signals that further, unspecified elements belong to the same category (language learning activities). The FUN_REF classification is thus well supported, with no additional interpersonal or textual function being evident in the co-text. A similar function is illustrated in Example (10) below.

- (10) *depende de las palabras porque hay algunas palabras ahorita mismo no tengo un ejemplo pero hay algunas palabras que bueno hoy vimos una que era volátil que eso en español es eh (.) una persona que se que que se deja llevar por sus emociones pero en italiano volatile es para los para los pájaros y para las los gases y cosas que se hacen gas rápido* (INF_5)
'it depends on the words because there are some words right now I don't have an example but there are some words that well today we saw one that was *volatile* which in spanish is uh (.) a person who lets who lets who lets themselves get carried away by their emotions but in italian *volatile* is for for birds and for for gases **and things that turn into gas quickly**'

As shown in (10), this occurrence exemplifies how the speaker constructs an *ad hoc* general extender while explaining the strategies

she employs to avoid confusing the meanings of false friends between Spanish and Italian. The two exemplars (*pájaros* ‘birds’, *gases* ‘gases’) are extended by an idiosyncratic structure rather than a canonical GE, illustrating a distinguishing feature of the Spanish data. Given its primary categorizing function, this instance is classified as FUN_REF. Conversely, Example (11) exemplifies an interpersonal function.

- (11) *si es un evento eh (.) me enfocaría en las personas por ejemplo si es que es un evento como un cumpleaños o una boda o un evento de fiesta o cualquier cosa* (INF_5)
 ‘if it’s an event uh (.) I would focus on the people for example if it’s an event like a birthday or a wedding or a party event **or whatever**’

Excerpt (11) illustrates three exemplars (*un cumpleaños o una boda o un evento de fiesta* ‘a birthday, or a wedding, or a party’) explicitly introduced by *por ejemplo* ‘for example’ and closed by the disjunctive GE *o cualquier cosa* ‘or whatever’. The presence of the exemplification marker and the disjunctive coordination suggest that the speaker is inviting the interlocutor to complete the category inferentially. This motivates the FUN_INT classification: the GE operates at an interpersonal level by relying on a shared understanding of “the kind of event” being discussed, rather than simply signaling referential vagueness, a function that is instead illustrated in Example (12) below.

- (12) *¿si van a estar sentadas? ¿si van a estar bailando? ¿si van ¿si va a haber música? ¿si va a haber comida? más o menos las decoraciones si hace falta si se necesita una presentación o algo tecnológico las luces los detalles técnicos* (INF_5)
 ‘if they’re going to be sitting down? if they’re going to be dancing? if they’re if there’s going to be music? if there’s going to be food? more or less the decorations if needed if it is required a presentation **or something technological** the lights the technical details’

As shown in (12), a single exemplar (*una presentación* ‘a presentation’) is extended by the Spanish disjunctive GE *o algo tecnológico* ‘or something technological’ generalizing to a broader category of technical event requirements later specified in the co-text (lights, technical details). Labelled as FUN_REF, the general extender performs straightforward categorial extension, preceding the detailed enumeration that immediately follows in the turn. A similar function is observed in Example (13).

- (13) *okay pero para empezar si voy a algun lugar lo primero que busco es si hay un lu- un hospedaje en general si hay un b&b hotel lo que sea cerca del lugar o al concierto o al evento que iría para no estar tan perdido* (INF_6)
 ‘okay but to start with if I go somewhere the first thing I look for is if there’s a lo- accommodation in general if there’s a b&b hotel **whatever** near the place or the concert or the event I’d go to so I’m not that lost’

In (13), two exemplars (B&B, hotel) are extended by the vague expression *lo que sea* ‘whatever’. The FUN_REF classification captures the general extender’s categorizing function (any type of accommodation), its inherent semantic vagueness highlights it as a particularly prominent example of referential indeterminacy in the Spanish sub-corpus. As discussed in subsection 3.5.3, this instance aligns directly with Borreguero Zuloaga’s (2021: 86) observation that forms like *o lo que sea* ‘or whatever’ signal a high degree of uncertainty or a low commitment to accuracy. Here, the exact type of accommodation is treated as secondary or non-essential by the speaker, whose primary communicative goal is simply to establish a broad, functional category (“accommodation in general”). A shift in pragmatic function is observable in Example (14) below.

- (14) *si puedo recordar algo era que a mí me gustó mucho un plato que era como una paella pero era con arroz (.) de*

color negro con tomates y otras cosas que era muy rica
 (INF_6)
 ‘if I can remember anything it was that I really liked a dish
 that was like a paella but it was made with rice (.) black in
 colour with tomatoes **and other stuff** that was very tasty’

In (14), the speaker is attempting to reconstruct the ingredients of a specific culinary experience recalled from memory. The cognitive difficulty of this retrieval process is signaled right at the start of the utterance by the disclaimer *si puedo recordar algo* ‘if I can remember anything’. This memory limitation is further marked, mid-utterance, by a pause immediately preceding the enumeration of ingredients (black rice, tomatoes). This occurrence clearly meets the criteria for the personal function (FUN_PER): the GE *y otras cosas* ‘and other stuff’ enables the speaker to close the list precisely when recollection fails, thereby preserving narrative continuity and interactional face despite the underlying memory lapse. The subsequent evaluative closure *que era muy rica* ‘that was very tasty’ reinforces this personal, experience-centered framing. By contrast, Example (15) exemplifies a textual function.

(15) *eh (.) wow los describiría como únicos variados eh existe una palabra en latín que es sui generis porque es un poco de todo*
 (INF_6)
 ‘uh (.) wow I would describe them as unique varied uh there is a word in latin which is sui generis because it's **like a bit of everything**’

In (15), three adjectival exemplars (unique, varied, *sui generis*) are extended by *un poco de todo* ‘a bit of everything’, which brings to a close a metalinguistic reflection on selecting the most appropriate words to describe places in his city. The FUN_TEX classification reflects how the GE is used by the speaker to signal the end of a detailed sequence where he was fully aware of expanding the discourse, thereby marking the end of the turn rather than performing

a referential or interpersonal function. The latter function is illustrated in Example (16) below.

- (16) *bueno la comida para mi gata ((ride)) la comida para mi gata pero: bueno queso mucho queso me gusta el queso ehm sí en el supermercado compro fundamentalmente comida comida cosas de de la vida cotidiana ehm* (INF_7)
'well food for my cat ((laughs)) food for my cat but: well cheese lots of cheese I like cheese ehm yeah at the supermarket I basically buy food food **everyday life stuff** ehm'

Excerpt (16) exemplifies how a single exemplar (*comida* 'food') is extended by *cosas de la vida cotidiana* 'everyday life stuff'. The FUN_INT classification reflects the general extender's reliance on shared, culturally grounded understanding of what "everyday life" typically entails. Furthermore, following the speaker's initial laughter and hesitations, the GE draws upon a common experiential frame, inviting the interlocutor to complete the list based on mutual knowledge of routine supermarket purchases. The same pragmatic function is illustrated in Example (17) below.

- (17) *eh ah y en lo que se refiere a la: digamos a la a la a la gramática las preposiciones para mí son una pesadilla no no nunca he entendido bien cómo ((ride)) cómo se usan quizás también un poco el contraste entre ir y venir los clásicos* (INF_7)
'uh ah and when it comes to the: let's say to the to the to the grammar prepositions are a nightmare for me I've never really understood how ((laughs)) how they are used maybe also a bit the contrast between *ir* and *venir* **that typical stuff**'

In (17), the GE is preceded by the exemplification marker *digamos* 'let's say'; therefore, the GE *los clásicos* 'that typical stuff' extends a single exemplar (the *ir/venir* contrast) to a broader, presupposed category of classic grammatical difficulties for language learners. The FUN_INT classification is motivated by the general extender's

appeal to mutual background knowledge regarding widely recognized learning obstacles. A similar interpersonal function is manifested in (18), where shared background knowledge is again invoked to construct a broad, mutually recognizable category.

- (18) *cuando aprendí obviamente diferencias como: mantequilla burro que para nosotros burro es asino entonces esa esa es como esas clásicas diferencias* (INF_9)
 ‘when I learned obviously differences like: butter butter in italian donkey in spanish which for us *burro* means donkey so that that is like **those typical differences**’

In (18), a single clausal exemplar (the Italian-Spanish false-friend pair ‘*mantequilla/burro*’) is extended by *esas clásicas diferencias* ‘those typical differences’. The FUN_INT classification reflects the general extender’s appeal to a shared category (classic cross-linguistic differences) that language learners are assumed to recognize collectively. Conversely, Example (19) illustrates a shift in the pragmatic function of GEs.

- (19) *ehm de cajón huevos leche verduras frutas eso así.*
 (INF_8)
 ‘uhm obviously eggs milk veggies fruits **stuff like that.**’

As illustrated in (19), a four-member list of grocery items (*huevos leche verduras frutas* ‘eggs, milk, veggies, fruits’) is closed by the canonical GE form *eso así* ‘stuff like that’. According to Overstreet and Yule (2021), this occurrence exemplifies decategorialization. Here, the demonstrative element *eso* loses its specific deictic referent and undergoes semantic bleaching, functioning instead as a fixed pragmatic marker of non-exhaustivity. Furthermore, while *eso así* typically co-occurs with the coordinating conjunction *y* ‘and’, its use here without the conjunction confirms the decategorized status of the general extender in dynamic spoken discourse. The FUN_REF classification is clear: the GE signals that further, similar grocery

items belong to the same category. This function is likewise present in the subsequent occurrence:

- (20) *yo diría que entre libros y: videos eh dígase o aplicaciones como cosas didácticas <quizás>* (INF_9)
'I would say like between books and: videos eh let's say or apps **like educational things** <maybe>'

In (20), three exemplars (*libros* 'books', *videos* 'videos', *aplicaciones* 'applications') are extended by *como cosas didácticas* 'like educational things', preceded by the marker *dígase* 'let's say', which itself explicitly signals imprecise or approximate denotation. The FUN_REF classification reflects the general extender's categorizing function, as *cosas didácticas* include educational resources in general. Furthermore, the combination of *dígase* 'let's say' and the final hedge *quizás* 'maybe' highlights the speaker's real-time cognitive planning and mitigation. Notably, this mitigating use of 'quizás' is observed in Excerpt (21) below.

- (21) *pastas arroz ehm algún tipo de proteína que normalmente compro pollo eh (.) o algo más dulce quizás alguna mermelada* (INF_9)
'pasta rice ehm some kind of protein like usually I buy chicken eh (.) **or something sweeter** maybe some jam'

As illustrated in (21), the occurrence closes a list of grocery items (*pastas arroz* 'pasta, rice', protein exemplified by *pollo* 'chicken') with a shift toward a distinct subcategory, introduced by the disjunctive GE *o algo más dulce* 'or something sweeter'. Unlike most GE in the sample, which extends a set of same-type exemplars, here the GE marks a boundary between two different grocery subcategories (savory items vs. sweet items) while simultaneously performing vague categorial extension within the second subcategory itself. The GE is immediately followed by the hedging marker *quizás* 'maybe' and a single further exemplar *mermelada*

‘jam’. The FUN_REF classification is well supported: *o algo más dulce* signals that jam is only one possible instance of a broader, unspecified category of sweet grocery items.

A disjunctive general extender is likewise utilized to construct a category of alternative possibilities in Example (22) below.

- (22) *eh (.) ¿qué más? (.) verduras si es que no las compré en en algún mercado o algo así frutas también ehm (.) cereales avena ehm yogur.* (INF_9)
 ‘uh (.) what else? (.) veggies if I didn’t buy them at at some market **or something like that** fruits too ehm (.) cereals oats ehm yogurt.’

In (22), a single exemplar (*verduras* ‘veggies’) is extended by the Spanish GE *o algo así* ‘or something like that’, coordinated with the disjunctive *o* ‘or’. This utterance is categorized as a ‘class’ according to Mauri’s (2017) framework, as the disjunctive marker *o* ‘or’ constructs a category of interchangeable alternative options. The sequence opens with the self-directed question *¿qué más?* ‘What else?’, which serves as an explicit cognitive planning marker. The FUN_REF classification captures the general extender’s function in signaling alternative, unspecified purchasing contexts belonging to the same category within an enumerative grocery-listing sequence, which the speaker subsequently continues by providing further exemplars (fruits, cereals, oats, yogurt). Turning to the textual function, Example (23) illustrates this pattern.

- (23) *ah bueno hicimos ehm (.) hicimos sándwiches así como pancitos de miga ehm (.) comimos pasteles porque <en chile nos gusta mucho comer pastel> bebimos cerveza eh (.) y: y bueno eso, sí bueno papas fritas que son son muy ricas <las papas fritas>* (INF_7)
 ‘ah well we made ehm (.) we made sandwiches like crustless sandwiches ehm (.) we ate cakes because <in chile we really like eating cake> we drank beer eh (.) and: **and well that**, yeah well french fries which are they’re very tasty <french fries>’

In (23), three exemplars (*sándwiches* ‘sandwiches’, *pasteles* ‘cakes’, *cerveza* ‘beer’) are extended by *y bueno eso* ‘and well that’, which itself embeds the discourse marker *bueno* ‘well’. Interestingly, the first exemplar contains a further internal exemplification (*sándwiches así como pancitos de miga* ‘sandwiches like crustless sandwiches’), functioning as an “example within an example” that specifies the broader subcategory before the speaker resumes the list. This pattern of embedded exemplification aligns with what Barotto and Mauri (2018: 130) describe as “two levels of listing, one embedded into the other”. In the present case, the speaker achieves a comparable effect by using the simulative marker *como* ‘like’ to introduce a more specific exemplification nested within the first exemplar of the main list. Moreover, as noted by Nogueira da Silva (2011: 164-171), the combination *y bueno* ‘and well’ in Spanish functions simultaneously as a discourse boundary marker serving to clarify, qualify, or elaborate upon the preceding utterance, a mechanism that often appears in closing sequences that reformulate or summarize narrated events. However, the utterance construction and the subsequent continuation of the list (*papas fritas* ‘French fries’) after the GE indicate that *y bueno eso* here functions less as a definitive closure and more as a textual pivot within an ongoing, loosely planned enumeration. This is further evidenced by the fact that the speaker recalls *papas fritas* only afterwards. The GE thus serves to hold the floor and restructure the turn when additional items come to mind. This justifies the FUN_TEX classification, as the GE marks a structural boundary within the turn.

Example (24) illustrates the most syntactically elaborate general extender in the entire sample.

- (24) *primero buscar vuelos eh luego el alojamiento y luego todo lo que lo que lo que viene a lo mejor programar los días eh a dónde ir* (INF_9)
'first looking for flights uh then accommodation and then everything that everything that **everything that comes with it** like maybe planning the days uh where to go'

In (24), *y luego todo lo que viene* 'and then everything that comes with it' closes a two-item list of trip-planning steps (*vuelos* 'flights', *alojamiento* 'accommodation') while simultaneously projecting forward to further, sequentially ordered steps still to come. Followed by the exemplification marker *a lo mejor* 'like maybe', the construction is classified as FUN_TEX: the general extender's primary function is organizational, structuring an ongoing procedural narrative and signaling that further steps will follow in due course. This utterance is categorized as a 'frame' according to Mauri's (2017) classification, where the informant evokes a temporal, event-based script (specifically the travel planning). The explicit sequential markers ('first', 'then', 'and then') structurally reinforce this chronological progression. Furthermore, the triplication of the pronoun (*lo que lo que lo que* 'everything that') alongside the vocalic hesitations (*eh*) reflects real-time planning and turn-holding effort. The subsequent insertion of *a lo mejor* 'like maybe' combined with exemplars ('planning the days', 'where to go') function as a mitigating strategy, framing the upcoming steps not as strict requirements, but as illustrative, hypothetical components of the travel script.

These qualitative findings highlight that categorization can also be achieved through what Overstreet and Yule (2021: 40) identify as 'specific extenders'. These are essentially long-form general extenders that have been enriched by the insertion of lexical material, such as adjectives, prepositional phrases, or relative

clauses, which specifies or restricts the domain of the category being created. Structurally, this inserted material typically modifies the core constituent of the GE (e.g., Eng. ‘thing’, ‘stuff’, It. *cose*, Sp. *cosas*). This phenomenon is clearly visible in the present interview data, where informants frequently modified standard GE structures to delimit ad hoc categories. In the Italian sub-corpus, this is illustrated by forms such as *tutte cose buone* ‘good stuff like that’ and *le solite cose*, ‘the usual thing’. Similarly, the Spanish data reveal various degrees of lexical enrichment, ranging from adjective-modified disjunctives like *o algo tecnológico* ‘or something technological’, *como cosas didácticas* ‘like educational things’, and *algo más dulce* ‘something sweeter’, to prepositional and clausal constructions such as *cosas de la vida cotidiana* ‘everyday life stuff’ and *y cosas que se hacen gas rápido* ‘and things that turn into gas quickly’. In all these cases, the added lexical content explicitly defines the shared property that binds the category members together in context.

5.2 Translation task: cross-linguistic rendering of GEs

As a final stage of the interview, informants were asked to translate two short texts, one from Italian into Spanish and one from Spanish into Italian, each containing general extenders embedded in the source text (*e roba del genere* ‘and stuff like that’ or *o qualcosa del genere* ‘or something like that’ in the Italian-to-Spanish text; *y esas cosas* ‘and things like that’ in the Spanish-to-Italian text). Unlike the spontaneous speech data discussed in section 5.1, this task constitutes a form of elicited, controlled production: rather than observing how general extenders emerge spontaneously in discourse, it captures how speakers actively render an already-present GE when translating between their two languages. As such, it offers a

complementary, though methodologically distinct source of evidence.

All ten informants' GE translations were classified as instances of equivalent translation. The source-language GE was rendered with a GE in the target language, rather than through a paraphrase, an omission, or a literal translation. This uniformity suggests that general extenders are not perceived by bilingual speakers as idiomatic or language-specific expressions resistant to translation, but rather as instances of a shared, readily accessible functional category. This finding is consistent with Overstreet and Yule's (2021) characterization of general extenders as a cross-linguistically robust pragmatic category, and it lends further support to the broader comparative claim of this thesis, namely that Italian and Spanish, despite their structural and lexical differences, converge in the pragmatic function fulfilled by GEs even when the specific lexical realization varies. Table 4 summarizes the lexical variants produced for each item, together with their frequency across the ten informants.¹²

Source GE	Direction	Target-language variants produced	Frequency
e roba del genere	ITA to ESP	y algo así	2
		y cosas por el estilo	2
		y algo similar	1
		y algo de este tipo	1
		y cosas así	1
		y cosas como relacionadas	1
		eh <algo así>	1
		y cosas similares	1

¹² The translations together with the corresponding informant codes are provided in Appendix C.

		oTRAS cosas de este género	1
o qualcosa del genere	ITA to ESP	o algo así	4
		o algo similar	2
		o algo por el estilo	1
		algo de este género	1
		una cosa por el estilo	1
		o cosas similares	1
y esas cosas	ESP to ITA	e cose del genere	4
		e cose così	2
		e queste cose	2
		e cose così e altre cose	1
		e queste cose e cose così	1

Table 4. GEs variants produced for the three general extenders in the translation task.

A second observation concerns the degree of lexical variation within each translation direction. For the Italian GE *e roba del genere* ‘and stuff like that’, informants produced a comparatively wide range of forms: *y algo similar* ‘and something similar’, *y algo así* ‘and something like that’, *y algo de este tipo* ‘and something of this kind’, *y cosas por el estilo* ‘and things of this sort’, *y cosas así* ‘and things like that’, *y cosas similares* ‘and similar things’, *oTRAS cosas de este género* ‘other things of this kind’. For the Spanish GE *y esas cosas*, ‘and things like that’ by contrast, the Italian renderings converge more narrowly around a small set of variants: *e cose così* ‘and things like that’, *e queste cose* ‘and these things’, *e cose del genere* ‘and things of this sort’. This asymmetry mirrors the pattern already observed in the spontaneous speech data (Section 5.1), where the Spanish sub-corpus displayed markedly greater lexical diversification than the Italian one. The fact that this asymmetry re-emerges in a controlled translation task strengthens the claim that it reflects a cross-linguistic tendency.

A further point of interest concerns the informant's rendering of *e roba del genere* ('and things like that') as *y cosas como relacionadas eh <algo así>* ('and things like related uh <something like that>'), which displays visible hesitation and lexical search (marked by the filled pause 'uh' and the false start *cosas como relacionadas* 'and things like related' before the proper GE, *algo así* 'something like that', is produced). This suggests that, even where the translation task is ultimately completed successfully with a canonical GE, the process of retrieving an appropriate equivalent may not always be immediate, a point that is worth noting as evidence of the online cognitive effort involved in producing pragmatically appropriate vague language under translation conditions. Furthermore, a notable pattern observed in the translation data is the occurrence of GE doubling within the exact same utterance. Specifically, two informants produced a sequence containing two distinct GEs to render a single target GE form: one informant generated *e cose così e altre cose* 'and things like that and other things', while the other produced *e queste cose e cose così* 'and these things and things like that' to render the Spanish GE *y esas cosas* 'and things like that'. This redundancy highlights how speakers, uncertain about which specific target form is most appropriate, may combine multiple GEs to compensate for uncertainty during lexical retrieval.

These results indicate that the translation task corroborates two of the central findings of this thesis: (i) the pragmatic and cross-linguistic transferability of general extenders, and (ii) the greater lexical variability of Spanish general extenders relative to their Italian counterparts, in line with the literature outlined in section 3.2, which highlights that the use and frequency of GEs are influenced by

sociolinguistic factors, including the cultural background of speakers.

To sum up, this chapter was dedicated to the analysis of the GEs that emerged from the interviews. The next chapter will focus on a corpus-assisted analysis of the forms that arose spontaneously during the interviews and were discussed in section 5.1, drawing on the KIParla corpus for Italian and the CORPES XXI for Spanish.

6. CORPUS-ASSISTED ANALYSIS OF GENERAL EXTENDERS: CORPORA OF SPOKEN ITALIAN AND SPANISH

This chapter presents an empirical, corpus-assisted analysis of the use of general extenders in spoken Italian and Spanish. Building upon the theoretical framework outlined in previous chapters, this chapter is structured as follows. Specifically, section 6.1 outlines the main research aims, describes the two spoken corpora (KIParla for Italian and CORPES XXI for Spanish), and details the sampling procedure used to extract the target general extenders: 24 forms identified in the interview corpus (Section 4.6), corresponding to 23 lexical forms queried in the reference corpora (one Italian form, *queste cose qua* ‘these things here’, was produced by two different informants). Section 6.2 reports the quantitative findings, illustrating both the overall absolute frequency and the distribution of pragmatic functions across the selected GE forms. Finally, section 6.3 carries out a qualitative analysis, starting with the Italian GE forms from KIParla in subsection 6.3.1, and concluding with the Spanish GE forms from CORPES in subsection 6.3.2.

6.1 Aims, corpora and sampling procedure

While the analysis presented in section 5.1 relies exclusively on the interview corpus, the present section extends the investigation by examining how the same general extender forms behave in two corpora of spontaneous speech: KIParla¹³ for Italian, and CORPES XXI,¹⁴ filtered for oral registers, for Spanish. KIParla is a corpus of spontaneous spoken Italian, collected through the collaboration of

¹³ <https://kiparla.it/il-corpus/> (Accessed: 24 May 2026).

¹⁴ <https://www.rac.es/corpes/> (Accessed: 24 May 2026).

the Universities of Bologna and Turin and accessible via the NoSketchEngine interface. The monitor corpus KIParla is continuously expanded; at the time of data collection for this thesis, the corpus comprises 3 million words (specifically 3,021,069 words across 493 recordings/documents). Conversations are drawn from a range of communicative situations including office-hour interactions between professors and students, guided group interactions, spontaneous conversations, oral exams, university lectures, and student-collected semi-structured interviews (Mauri, Ballarè, Gorla, Cerruti, & Suriano, 2019). By contrast, CORPES XXI (Corpus del Español del Siglo XXI), developed by the Real Academia Española (RAE) in collaboration with the Association of Academies of the Spanish Language (ASALE), is a large-scale reference corpus comprising over 455 million orthographic forms across more than 420,000 documents drawn from both written and oral texts produced across the Spanish-speaking world, including Spain, Latin America, the Philippines, and Equatorial Guinea. For the purposes of this study, queries were restricted to the corpus's oral component by applying the "oral" register filter available in the search interface. It should be noted that this filter is not restricted to informal conversational speech: the oral section of CORPES XXI comprises a broad range of registers, encompassing both informal spontaneous conversation and considerably more formal spoken genres, such as media interviews, radio and television broadcasts, public addresses, and press conferences. Some of the corpus occurrences discussed in subsection 6.3.2 were in fact drawn from this more formal end of the register spectrum (e.g. a government press conference, a public presidential address). Regarding geographical variety, CORPES XXI maintains a deliberate balance between Peninsular and Latin American Spanish, with texts from Latin America constituting most

of the corpus (approximately 70%, against 30% for Spain; RAE, n.d.). Including Latin American varieties alongside Peninsular Spanish was relevant for this study, as the Hispanophone informants who took part in the interviews are speakers of Latin American varieties of Spanish. The aim of this corpus-assisted methodology is not to establish statistically robust frequency generalizations, which would require larger per-form samples, but rather to verify whether the GE forms identified in the interview data follow similar patterning. For each of the general extender forms identified in the interview (Section 5.1),¹⁵ a query was submitted to the corresponding reference corpus. Where a form yielded more than 20 occurrences, a random sample of up to 20 instances was extracted and manually examined; where fewer than 20 occurrences were attested, all available occurrences functioning as GEs were analyzed. Each examined occurrence was annotated using the same coding scheme applied to the interview data (construction, exemplar type, pragmatic function, following Overstreet and Yule’s 2021 taxonomy, among other categories), allowing for a comparison across the two datasets. Occurrences in which the queried string did not function as a general extender, for instance, cases of purely anaphoric use were excluded from the analysis.

6.2 Quantitative and functional distribution in reference corpora

To provide an overview of the empirical data, the analysis begins by examining the distribution of the selected GE forms retrieved from the two corpora, KIParla and CORPES. Figure 2 displays the

¹⁵ Italian GEs: *eccetera, e tutto, queste cose qua, tutte cose buone, queste cose, le solite cose, un po' di tutto*. Spanish GEs: *y cosas de esas, y cosas que se hacen gas rápido, o cualquier cosa, o algo tecnológico, lo que sea, y otras cosas, un poco de todo, cosas de la vida cotidiana, y bueno eso, los clásicos, eso así, como cosas didácticas, esas clásicas diferencias, algo más dulce, o algo así, y luego todo lo que viene*.

absolute frequency of all general extender forms queried in the two reference corpora (7 Italian forms in KIParla, 16 Spanish forms in CORPES). The graph reveals a marked asymmetry between the two corpora. Among the Italian forms, *eccetera* ‘etcetera’ (984 occurrences), *queste cose* ‘these things’ (558), *e tutto* ‘and everything’ (417), and *queste cose qua* ‘these things here’ (111) stand out as highly frequent general extenders, while *le solite cose* ‘the usual things’ (10) is comparatively rare, and *un po' di tutto* ‘a bit of everything’ (1) and *tutte cose buone* ‘good stuff like that’ (0) are marginal or absent in the corpus. Among the Spanish forms, only *lo que sea* ‘whatever’ (275) and *o algo así* ‘or something like that’ (155) reach frequencies comparable to the most widespread Italian GE forms; the remaining fourteen forms cluster at considerably lower frequencies, and seven of them (*o algo tecnológico* ‘or something technological’, *y cosas que se hacen gas rápido* ‘and things that turn into gas quickly’, *cosas de la vida cotidiana* ‘everyday life stuff’, *cosas didácticas* ‘like educational things’, *esas clásicas diferencias* ‘those typical differences’, *algo más dulce* ‘or something sweeter’, and *y luego todo lo que viene* ‘and then everything that comes with it’) are entirely unattested in the CORPES corpus. In contrast to previous literature, which emphasizes that most general extenders consist of fewer than three words in both English and Spanish (cf. Aijmer, 2002; Fernández, 2015; Kern, 2022), the present interview data reveal the presence of longer and highly descriptive structures. This distribution corroborates the qualitative observation already advanced in section 5.1.2: Spanish general extenders emerged from the interview display a higher degree of lexical diversification, often featuring idiosyncratic constructions created ad hoc by speakers. Consequently, several of these improvised forms are rarely attested

in Spanish corpora, whereas the Italian forms examined cluster more tightly around a small set of highly frequent, stable constructions.

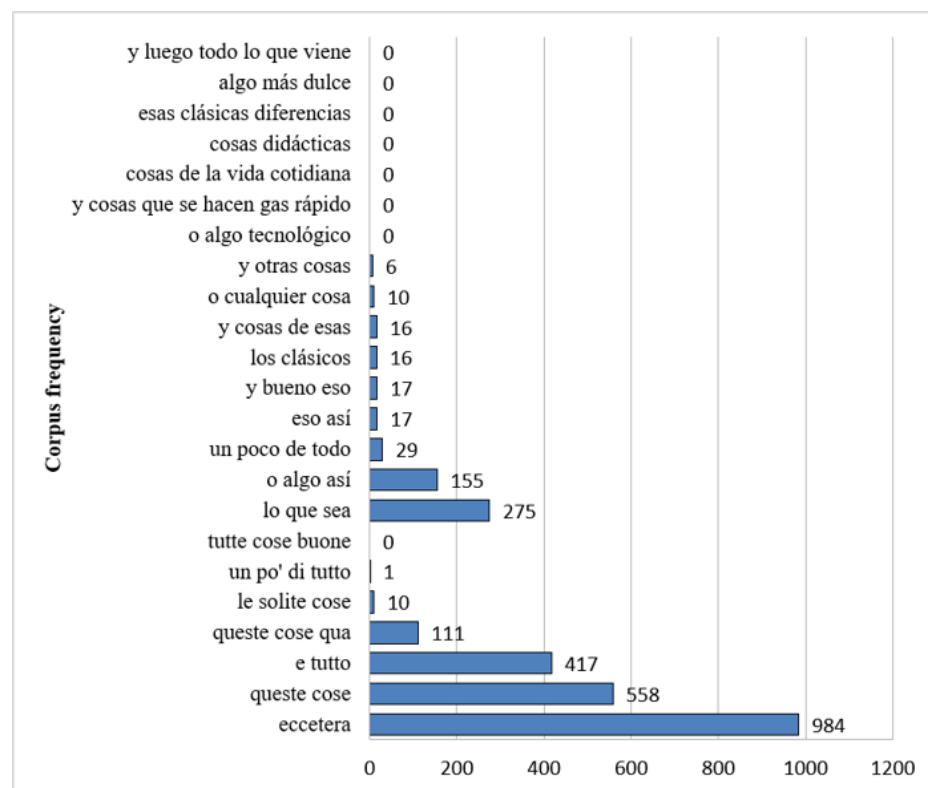


Figure 2. Absolute frequency of each GE form across the KIParla and CORPES corpora.

Notably, the empirical data gathered from the interviews reveals a predominance of adjunctive general extenders over disjunctive variants. This pattern strongly mirrors the quantitative findings of Kern (2022: 221), who reported that adjunctive forms introduced by *y* ‘and’ accounted for the majority (71%) of all occurrences in Spanish conversational data, whereas disjunctive forms introduced by *o* ‘or’ were considerably less frequent (20%), a distribution that aligns with broader tendencies identified in Peninsular Spanish discourse (cf. Fernández, 2015).

Figure 3 summarizes, by language, how many of the general extender forms identified in the interview corpus were subsequently attested (i.e. yielded at least one occurrence) versus not attested in the corresponding KIParla and CORPES corpora. Of the eight Italian

GE forms queried in KIParla, seven were attested and only one, *tutte cose buone* ‘good stuff like that’ returned no results. Of the sixteen Spanish GE forms queried in CORPES, by contrast, only nine were attested, while seven yielded no occurrences at all. This asymmetry offers a corpus-attested confirmation of the productivity contrast already visible in Figure 2: several of the general extender forms produced by Spanish informants in the interviews appear to be non-conventionalized expressions rather than recurrent constructions attested in spontaneous Spanish speech, whereas the Italian forms in the sample are, with one exception, all attested constructions.

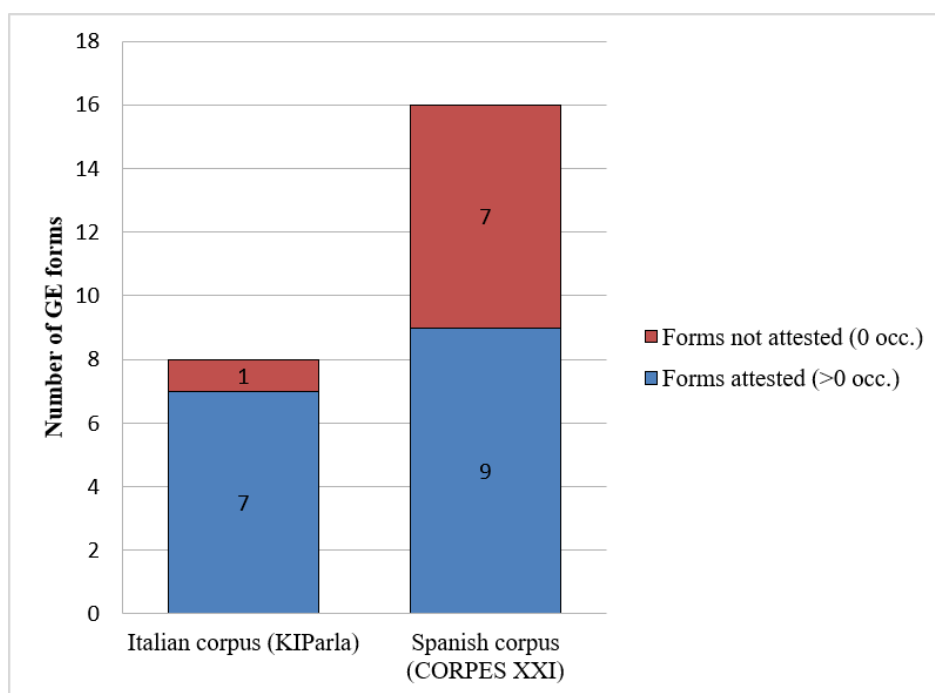


Figure 3. Number of attested and unattested GE forms across the corpora.

Figure 4 below presents the pragmatic functions of the GE forms attested in the two corpora following the taxonomy of Overstreet and Yule (2021): referential (FUN_REF), interpersonal (FUN_INT), personal (FUN_PER), and textual (FUN_TEX). The graph shows that most forms, both Italian and Spanish, are strongly REF-dominant, consistent with the canonical, list-closing use of general

extenders: *queste cose* ‘these things’ (19/20 REF), *eccetera* ‘etcetera’ (18/20), *queste cose qua* ‘these things here’ (18/20), *lo que sea* ‘whatever’ (17/20) and *o algo así* ‘or something like that’ (15/20) all show referential extension. Three forms, however, depart from this pattern in ways that merit closer attention. *Y bueno eso* ‘and well that’ is attested exclusively with a textual function across all eight examined occurrences (8/8 FUN_TEX), never once closing a referential list; *eso así* ‘stuff like that’, in its single occurrence, performs a personal function (1/1 FUN_PER); and *e tutto* ‘and everything’, while still REF-dominant (17/20), performs a personal function (3/20 FUN_PER), similar to the one attested in the interview data in Example (4), not observed to the same extent in the other Italian forms. *Un poco de todo* ‘a bit of everything’ stands out as the most multifunctional GE form in the entire sample, distributing across all four categories (10 REF, 2 INT, 2 PER, 5 TEX).

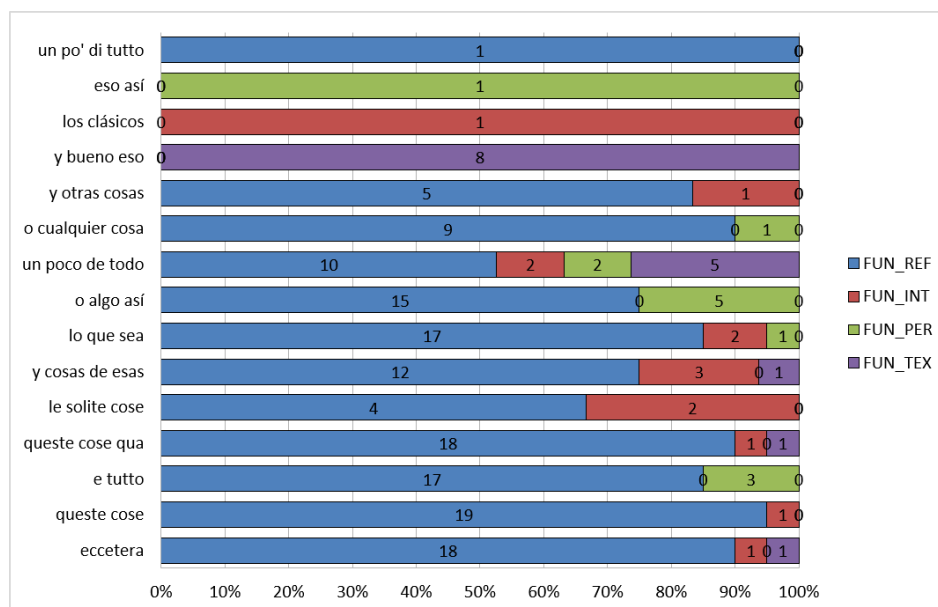


Figure 4. Pragmatic functions per GE form (adapted from Overstreet & Yule, 2021).

Moreover, while Cheshire’s (2007) hypothesis suggests that longer general extenders tend to perform a referential function more

frequently than shorter forms (cf. Kern, 2022: 228), the present data reveals a different distribution. In fact, the referential function is overwhelmingly dominant across almost all GE forms, including short forms such as *eccetera* ‘etcetera’ (90%) and *queste cose* ‘these things’ (95%).

Having outlined the quantitative distribution, the next section turns to a qualitative analysis by selecting at least two attested occurrences for each GE form across the KIParla and CORPES corpora.

6.3 Qualitative analysis

Expanding upon the quantitative overview presented in section 6.2, this section examines a selection of representative corpus occurrences in qualitative depth, following the same analytical format adopted in section 5.1. For each general extender form, two illustrative examples are discussed with reference to their construction, pragmatic function, and any notable points of convergence or divergence between them. Where relevant, each form’s behavior in the reference corpus is further compared with its use in the interview data (Section 5.1), in order to assess the extent to which the functional patterns identified in spontaneous, elicited speech are corroborated by, or diverge from, the corpus evidence. The discussion is organized by language, beginning with Italian (Subsection 6.3.1) and followed by Spanish (Subsection 6.3.2).

6.3.1 Qualitative analysis of GEs from the KIParla corpus

To begin the qualitative analysis, occurrences (25) and (26) illustrate the referential function already identified as dominant for *eccetera* ‘etcetera’ in the interview data (Subsection 5.1.1), extending an open category of tasks (25) or cultural and religious practices (26) beyond

the explicitly named exemplar(s). The two examples nonetheless diverge with regard to its structural aspects: in (25), *eccetera* occupies a medial, non-final position, it interrupts the enumeration of the brother's household tasks, followed rather than preceded by a further exemplar (*accendere il fuoco* 'lighting the fire'). Whereas in (26) it occupies the canonical, list-final position, additionally doubled (*eccetera eccetera*) to reinforce the sense of an open, unspecified category. In both cases, the use of *eccetera* introduces a certain degree of vagueness to the utterance by deliberately omitting specific elements within the given discourse context (cf. Fiorenza, 2024: 456). Furthermore, in (26), *quindi* 'so' introduces the exemplar as a consequential connective (*che è ebreo quindi la cucina kosher* 'since he's Jewish so kosher cuisine'). Example (25), by contrast, is worth noting as a distinct constructional variant within the data, useful for discussing the positional flexibility of the GE in Italian. The medial use in (25) is not an atypical pattern; *eccetera* in Italian spontaneous speech shows a recurrent tendency to occur mid-list, a flexibility of construction not typically associated with general extenders in the theoretical literature, where list-final position is generally treated as canonical (cf. Overstreet & Yule 2021). The reduplicated GE form observed in (26), meanwhile, reflects a similarly frequent pattern previously noted in the LIP corpus (cf. Fiorentini, 2018). Furthermore, in line with Fiorentini's (2018) findings, *eccetera* consistently performed a referential function, as well as in the present analysis of the KIParla corpus, accounting for 90% of occurrences.

(25) *ma era proprio un lavoro di famiglia perché c'era mio
fratello che è quello che pensa alla legna **eccetera** a
accedere il fuoco* (KIParla_BOR032)

‘well it was really a family job because my brother was there he’s the one who takes care of the wood **etcetera** lighting the fire’

- (26) *la sua famiglia di lui che è ebreo quindi la cucina kosher eccetera eccetera* (KIParla_PKP063)
‘his family since he’s jews so kosher cuisine **etcetera etcetera**’

Turning to *e tutto* ‘and everything’ the second most frequent Italian GE form analyzed in the KIParla corpus (417 occurrences; section 6.2). The two corpus occurrences examined below provide concrete exemplars of its primary pragmatic functions:

- (27) *devi vedere se c’è la persona di turno se ci sta il bus disponibile // se c’hai il pieno e tutto* (KIParla_PKP141)
‘you have to see if the person on duty is there if there’s a bus available // if you have a full tank **and everything**’

- (28) *cioè non ti dice neanche buongiorno e tutto* (KIParla_SBCA001)
‘I mean she doesn’t even say good morning **and everything**’

Both occurrences illustrate the referential function (FUN_REF), yet they differ in their construction. Example (27) closes a formed list of three parallel clausal exemplars, all introduced by the same conditional frame (*se...* ‘if..’), describing the organizational conditions that must be checked before a bus trip can proceed; *e tutto* ‘and everything’ here extends the category to further, unspecified logistical conditions of the same kind. Example (28), by contrast, extends a single exemplar (*non ti dice neanche buongiorno* ‘she doesn’t even say good morning’), generalizing it to a broader category of missing politeness behaviors. These two occurrences diverge from the interpersonal function (FUN_INT) attested for the same GE form in the interview (4) (Subsection 5.1.1), where the GE relied on the interlocutor’s shared cultural knowledge of what a

graduation celebration typically involves, rather than just extending a list. The contrast suggests that GE are capable of multiple pragmatic functions depending on the context in which they are employed.

Regarding *queste cose qua* ('these things here'), the third Italian GE form under analysis, the interview data revealed two distinct functions, namely personal classification in (2) and an interpersonal function in (7). Conversely, the two corpus occurrences examined below both perform a referential function.

- (29) *abbiamo in tradizione a livello dell'emilia romagna sono la esce bolognese esse molto // la zeta bolognese // queste cose qua* (KIParla_BOI119)
 'we have in our tradition in emilia romagna there's the bolognese s the s a lot // the bolognese z // **these things here**'

- (30) *non lo so perchè poi alla fine i i lavoretti sono più o meno gli stessi cioè fare la promoter nel supermercato come andare a fare la promoter alle fiere queste cose qua mh* (KIParla_BOI002)
 'I don't know because in the end the jobs are more or less the same I mean being a promoter in the supermarket or working as a promoter at trade fairs **these things here** mh'

Both occurrences close a list of parallel exemplars, two regional phonetic features in (29), two variants of the same type of job in (30), and extending each category to further, unspecified members. The two examples differ in their degree of syntactic structure. In (30), the exemplars are explicitly linked by the similarity marker *come* ('like'), functioning as an exemplification cue that frames the two jobs as interchangeable instances of the same category. In (29), by contrast, no such marker is present; the two exemplars are simply juxtaposed. These latter occurrences diverge from the interview data in an interesting way. In (2), the use of the GE form was motivated by processing difficulty, thereby signaling the speaker's subjective

detachment from the topic (FUN_PER); in (7), by contrast, it invoked the interlocutor's tacit agreement on a shared topic. Within the KIParla occurrences analyzed, with the exception of a single GE performing a textual function, the GEs serve a categorial function, closing lists of two clearly identifiable, parallel exemplars. This three-way contrast suggests that *queste cose qua* 'these things here' is a highly flexible construction, capable of referential, personal, interpersonal and textual readings depending on contextual factors, exemplar availability, speaker fluency, and the presence or absence of similarity markers inviting interlocutor engagement.

Turning to *queste cose* 'these things', the two KIParla corpus occurrences analyzed are exemplified below.

- (31) *la mia mamma è terrorizzata dalle cose viscide tipo serpente rane e topi // e dall' altro lato però // e è un' amante degli animali li difende non uccide neanche gli insetti // tipo è vegetariana* **queste cose** (KIParla_TO081)
 'my mom is terrified of slimy things like snakes frogs // and mice and on the other hand though // and she's an animal lover she defends them she doesn't even kill insects // like she's vegetarian **these things**'

- (32) *cioè quando io dicevo delle cose siccome molte volte ero a casa pomeriggio con lei // e parlavo a voce alta* **queste cose qui** (KIParla_BOI032)
 'I mean when I said things since a lot of times I was at home in the afternoon with her // and I spoke out loud **these things here**'

Occurrences (31) and (32) perform a referential function, though they diverge considerably in their construction. Example (31) closes an elaborate list of multiple heterogeneous exemplars animals feared as slimy things (snakes, frogs, mice) contrasted with the mother's animal-loving behaviors (not killing insects, being vegetarian), all introduced by the exemplification marker *tipo* ('like'), used twice within the same turn; *queste cose* 'these things' here extends the

category to further behaviors of the mother's relationship with animals. Example (32), by contrast, extends a vaguer expression (*delle cose... parlavo a voce alta* 'things... I spoke out loud'), and is realized with a distinct morphological variant, *queste cose qui* 'these things here' rather than *queste cose* 'these things'. This deictic pattern (deictic + *cose* 'things' + deictic) is documented as the single most frequent construction involving *cose* 'things' as a GE in contemporary spoken Italian, accounting for 34.5% of occurrences in LIP and 38.8% in KIParla (cf. Fiorentini & Zanchi 2024: 165). This finding further corroborates the internal variability of this construction (e.g., *queste cose qua* 'these things here') observed in (29) and (30). Fiorentini and Zanchi (2024) attribute the deictic pattern to the strong intersubjectivity characteristic of spoken dialogue, whereby spatial deictic elements contribute to creating a shared focus of attention among interlocutors (Voghera 2017: 87). This alignment is directly consistent with the interpersonal and personal readings identified in occurrences of this form in the interview data (Subsection 5.1.1). Additionally, the majority of occurrences in both the LIP and KIParla corpora (cf. Fiorentini & Zanchi, 2024: 166-167) are preceded by a single exemplar (62% and 64.5% respectively) as seen in (31) and (32). Moreover, explicit exemplification markers are frequently absent (present in only 17.2% of LIP and 27.6% of KIParla corpora), a pattern illustrated in (32), which lacks such a marker, in contrast to (31), where the exemplification marker *tipo* 'like' occurs twice.

Analyzing *le solite cose* 'the usual things' the interview data documented this GE in an initial position: in occurrence (7), *le solite cose* 'the usual things' precedes rather than follows its exemplars (*pizzette, focacce, panini* 'finger food typically served at an aperitivo'). This initial positioning, together with the evident appeal

to shared cultural knowledge of what an Italian aperitivo typically includes, motivated an interpersonal reading. The two KIParla corpus occurrences below, by contrast, both perform a referential function.

- (33) **BOR031:** *la pizza di natale?*
BOI117: *no vabbè eh c'è il barbarossa a medicina*
BOR031: *ah sì*
BOI118: *sì dai questi eventi il barbarossa a natale qualche evento non so un compleanno questo*
BOI117: *eh questi eventi*
BOR031: *okay*
BOI117: *poi ci troviamo qualche volta sì mh le solite cose*
(KIParla_PBB032)
‘BOR031: christmas pizza?
BOI117: no come on uh there’s the barbarossa in medicine
BOR031: ah right
BOI118: yeah come on these events the barbarossa at christmas any event I don’t know a birthday this
BOI117: yeah these events
BOR031: okay
BOI117: then we meet up sometimes yeah mh **the usual things’**
- (34) **TOI066:** *sì ma in realtà non ho mai fatto sport di squadra*
TOR009: *mh*
TOI066: *mh in gen~ cioè ho fatto danza le solite cose*
(KIParla_PTD011)
‘TOI066: yeah but actually I’ve never done team sports
TOR009: mh
TOI066: mh gen~ I mean I did dance **the usual things’**

Example (33) is notable for its multi-turn co-construction involving three speakers, where the list of social events (the Barbarossa gathering, a birthday) is jointly built across turns. This distributed construction of the category across three interlocutors parallels the dialogic, co-constructed patterns documented in the KIParla corpus by Fiorentini and Zanchi (2024: 172), showing how interlocutors may incorporate, reuse, and re-elaborate part of a preceding turn to demonstrate understanding and cooperation in jointly building a non-

exhaustive list. Example (34), by contrast, is co-constructed within a conversational sequence following a single speaker's utterance, where the exemplar *danza* ('dance') is supplied to clarify a previously vague statement. Here, the deployment of vagueness serves to economize cognitive effort; since the main communicative goal is to establish the absence of team sport experience (cf. Fiorenza, 2024: 452). Before providing this exemplar, the primary speaker produces an open-ended assertion (*non ho mai fatto sport di squadra*, 'I've never done team sports'). Once the interlocutor responds with a backchannel (mh), the main speaker elaborates on their intended meaning by providing the specific exemplar (*danza* 'dance') and appending the GE *le solite cose* 'the usual things' at the end of the turn.

In Example (35) below, the GE *un po' di tutto* 'a bit of everything' performs a referential function, preceding the enumeration itself. Rather than closing a list of exemplars, the GE precedes the list of jobs and skills, anticipating the overall category before the first specific exemplar (*idraulico* 'plumber') is provided. Notably, this is the only occurrence of this specific form identified and analyzed within the KIParla corpus.

- (35) *mah mio padre ha fatto un po' di tutto ha fatto l'idraulico ha fatto* (KIParla_TOI077)
 'well my dad did **a bit of everything** he worked as a plumber he did'

6.3.2 Qualitative analysis of GEs from the CORPES corpus

The following subsection examines the occurrences identified for the Spanish GE forms retrieved from the CORPES corpus. The interview data documented *y cosas de esas* 'and things like that' in a referential use. The two corpus occurrences below illustrate distinct applications.

(36) *te presto tres pero / me las debes jeh! bueno y cosas de esas* (CORPES_SCOM_H21_053)
 ‘I’ll lend you three but / you owe me alright! well **and things like that**’

(37) *el último año que / tuvo yo matemáticas // yo creo que en otros institutos a lo mejor sí que daban derivadas e integrales y cosas de esas* (CORPES_SCOM_M32_022)
 ‘the last year that / I had math // I think that in other high schools maybe they did do derivatives and integrals **and things like that**’

Occurrences (36) and (37) diverge sharply in function. Example (37) performs a referential function (FUN_REF), directly convergent with the interview occurrence: two nominal exemplars (*derivadas e integrales* ‘derivate and integrals’), preceded by the epistemic stance adverbial *a lo mejor* ‘maybe’, are extended to other unspecified math topics. Example (36), by contrast, does not close a list of parallel exemplars at all; it terminates an extended reported speech sequence, a reconstructed childhood exchange about lending marbles, with the discourse-closing marker *bueno* ‘well’ immediately preceding the GE. Here *y cosas de esas* ‘and things like that’ functions not to extend a taxonomic category but to mark a structural boundary, returning from quoted speech to the main narrative, thereby performing a textual function (FUN_TEX). This usage mirrors the textual behavior documented by Cheshire (2007), who notes that general extenders, including longer variants such as ‘and things like that’, recurrently serve to mark a section or signal the end of reported speech. Extending this view, Borreguero Zuloaga (2022) highlights that GEs in spontaneous speech allow speakers to convey the general impression of an utterance without literal precision, while structurally marking the end of the reported speech fragment and signaling the return to the primary narrative text. This is consistent with the broader functional profile that emerged for this form across

the wider corpus sample examined for this study (Subsection 6.2): while *y cosas de esas* ‘and things like that’ performs a FUN_REF as in (37), and as in the single interview occurrence, it also recurrently patterns toward non-referential functions in context, whether interpersonal, through its marked tendency to co-occur with interlocutor directed tag questions (*¿sabes?* ‘you know’), or textual, as in (36), where the discourse marker of a reported speech episode reflects the discourse-organizing function. Taken together, these two occurrences suggest that *y cosas de esas* ‘and things like that’, far from being restricted to the canonical referential function observed in the interview data, is in fact a multifunctional construction.

Examples (38) and (39) illustrate the referential function of the disjunctive GEs *o cualquier cosa* ‘or whatever’. Although both occurrences perform a primary FUN_REF, they exhibit different mechanisms of referential extension.

- (38) *este fin de semana lo organizo yo // yo cualquier cena o cualquier cosa de mis amigas las organizo yo siempre*
 (CORPES_SCOM_M13_008)
 ‘this weekend I’m organizing it // any dinner **or whatever** with my friends I always organize it’

- (39) *no te preocupes / puedes dejarlo encendido no sea que te llamen de la guardería o cualquier cosa*
 (CORPES_SCOM_M11_040)
 ‘don’t worry / you can leave it on just in case they call you from kindergarten **or whatever**’

In (38), *o cualquier cosa* ‘or whatever’ occurs within a left-dislocated, topicalized nominal structure. The GE functions referentially by closing an open-ended list following a specific nominal exemplar (*cualquier cena* ‘any dinner’), extending the category of social gatherings organized by the speaker beyond mere dinners to include other equivalent activities with her friends. Notably, the exemplar itself already carries the quantifier *cualquier*

‘any’ in *cualquier cena* ‘any dinner’, creating a redundant reinforcement of vagueness. The co-occurrence of the quantifier with the matching GE form *o cualquier cosa* ‘or whatever’ doubles down on the non-exhaustive nature of the category openness. In (39), *o cualquier cosa* ‘or whatever’ likewise performs a FUN_REF but operates at a clausal level rather than extending a set of nominal hyponyms. Here, the GE attaches to a single clausal exemplar within a hypothetical frame (*no sea que...* ‘in case...’), namely receiving a call from the kindergarten (*te llamen de la guardería*). The GE allows the speaker to construct the specific kindergarten scenario as merely one instance of a broader category of possible contingencies, justifying keeping the phone on without having to enumerate every emergency.

Examples (40) and (41) illustrate the pragmatic multifunctionality of the GE *lo que sea* ‘whatever’: (40) performs a referential function, whereas (41) serves an interpersonal function.

(40) *eeh comprar que comprar un par de pisos o quizá una casa chalé lo que sea somos gallegos*

(CORPES_SCOM_H23_003)

‘uh buying what buying a couple of apartments or maybe a chalet **whatever** we’re galicians’

(41) *al final comunicamos / ¿qué comunicamos? pues cultura información lo que sea ¿no?*

(CORPES_CORALES_Herrera_y_punto)

‘at the end of the day we communicate / what do we communicate? well culture information **whatever** right’

In (40), *lo que sea* ‘whatever’ performs a primary referential function. The GE closes a list following two explicit nominal exemplars (*un par de pisos* ‘a couple of apartments’, *una casa chalé* ‘a chalet’), where the second exemplar is pre-modified by the epistemic adverbial *quizá* (‘maybe’). The GE here serves to extend the category of real-estate purchases to any other equivalent property.

The interaction between *quizá* ‘maybe’ and *lo que sea* ‘whatever’ signals the speaker’s indifference toward the specific property type, projecting vagueness over a set of hypothetical architectural choices. In (41), *lo que sea* ‘whatever’ operates primarily in the interpersonal domain. The list is introduced via a self-interrogative question (*¿qué comunicamos? pues...* ‘what do we communicate? well...’), followed by two abstract nominals (*cultura, información* ‘culture, information’). Crucially, the GE is immediately followed by the tag question *¿no?* (‘right?’). *Lo que sea* ‘whatever’ functions intersubjectively to downplay the precision of the offered examples signaling that the speaker does not wish to sound overly authoritative. The immediate co-occurrence with the interlocutor directed tag *¿no?* ‘right’ actively invites the hearer’s agreement and shared alignment regarding the open-ended nature of the category.

Occurrences (42) and (43) are drawn from public political discourse, a government press conference in (42) and a presidential public address in (43). Despite belonging to formal and institutional contexts, both examples employ the adjunctive GE *y otras cosas* ‘and other stuff’ within reported speech, illustrating how speaker-driven reported speech allows for informal, open-ended category extension even in a more formal situation; thereby performing a referential function.

(42) *no exageramos a pesar de que se nos dijo que habíamos exagerado y otras cosas*
 (CORPES_Ruedas_de_prensa_del_Consejo_de_Ministros
 _del_Gobierno_español_enero_2004)
 ‘we’re not exaggerating even though we were told that we had exaggerated **and other stuff**’

(43) *que tenían las tiendas pues decían sí te vendo el azúcar pero me tienes que comprar el jabón el aceite y otras cosas*
 (CORPES_Conferencia_presidente_AMLO)

‘that the stores had well they were like yes I’ll sell you sugar but you got to buy soap oil **and other stuff**’

In (42), *y otras cosas* ‘and other stuff’ is embedded within a reported speech sequence introduced by the passive framing *se nos dijo que...* ‘we were told that...’. The GE extends a single clausal exemplar *habíamos exagerado* ‘we had exaggerated’. Crucially, this instance represents a reported speech case that mirrors the structural boundary-marking highlighted earlier in Cheshire (2007) and Borreguero Zuloaga (2022). Rather than citing the full list of accusations, the speaker uses the GE to condense a series of past criticisms into a single representative example. In (43), *y otras cosas* ‘and other stuff’ occurs within a direct quoted speech (*pues decían sí...* ‘they were like yes...’) embedded in a public address. Structurally, (43) represents a canonical instance of list extension, where the GE closes an open-ended list of nominal hyponyms within the same grocery/shopping domain. By attaching the GE to these concrete nominal exemplars, the speaker presents the cited items as typical exemplars of a broader commercial practice. The reported speech format lends the narrative the conversational authenticity of a previously lived situation.

In (44) and (45), *un poco de todo* ‘like a bit of everything’ perform a referential function.

(44) *sí sí que bailan sí / porque bueno tengo amigos / argentinos // de Brasil // de México / o sea que / tengo una variedad ahí **un poco de todo***

(CORPES_SCOM_M11_050)

‘yeah yeah they dance yeah / cause well I’ve got friends / Argentinian // from Brazil // from Mexico / so / I have a bit of a variety there **like a bit of everything**’

(45) *o sea el perío pues hojeo normalmente hojeo El País / El Correo La Voz El Mundo bueno mmm **un poco de todo***

(CORPES_SCOM_M23_018)

‘I mean the paper well I usually browse through *El País* / *El Correo* *La Voz* *El Mundo* well mmm **like a bit of everything**’

Both occurrences close a list of nominal exemplars, friends’ nationalities in (44), named newspaper titles in (45), and are structurally very close to one another. In (44), the three exemplars (*Argentinos, de Brasil, de México* ‘Argentinian, from Brazil, from Mexico’) are preceded by an explicit category label, *una variedad* ‘a variety’, functioning as a property clue¹⁶ (cf. Barotto & Mauri 2018) that names the abstract category before the GE itself supplies it a second time, a degree of redundancy between label and GE not found in (45), where the four named newspapers are simply listed in sequence, with the discourse marker *bueno* ‘well’ as the only element separating the list from the GE. Both examples are also marked by online planning: the filled pause ‘mmm’ in (45), and the self-correcting reformulation *o sea que tengo una variedad ahí* ‘so I have a bit of a variety there’ in (44).

Moving on to *y bueno eso* ‘and well that’, the interview data documented this GE in a textual use. This textual reading (FUN_TEX) is strongly corroborated by the wider corpus sample for this form (Section 6.2), in which *y bueno eso* ‘and well that’ was attested exclusively with a textual function across every examined occurrence, never once performing a referential, interpersonal, or personal function. The two occurrences below illustrate two distinct realizations of this same discourse organizing use.

(46) *la semana eso no es trabajar / eso lo hice por / bueno por otro favor personal / que me pidieron ir allí y bueno eso //*
(CORPES_PRESEGAL:SCOM_M23_018)

¹⁶ According to Barotto and Mauri (2018: 109), “property clues” are co-textual elements that facilitate category building by acting as interpretive guides for the hearer and as real-time formulation devices, such as exemplification and elaboration, for the speaker.

‘the week that’s not working / I did that for / well for another personal favor / they asked me to go there **and well that**’ //

- (47) *la pasión / sobre todas las cosas // la pasión eso es // eso es / y bueno eso*
(CORPES_Canal_Oficial_Jesús_Quintero_Alejandro_Sanz_en_Ratones_Coloraos_22/02/2010)
‘passion / above everything else // / passion that’s what it is // that’s it / **and well that**’

Neither occurrence closes a list of exemplars in the canonical sense; in both cases, *y bueno eso* ‘and well that’ instead marks the end of a sequence, a personal favor being justified as not “real work” in (46), the value of passion being emphatically restated twice in (47) rather than extending a taxonomic category to further unnamed members. The two examples nonetheless diverge in tone: (46) closes the topic with a note of justification; (47), by contrast, closes with an assertive, almost declarative finality, the repeated affirmations (*eso es / eso es* ‘that’s what it is / that’s it’) building toward the GE as a rhetorical full stop rather than a hedge. These two occurrences reinforce the picture that emerged from the interview data: *y bueno eso* ‘and well that’ shows a markedly narrower functional range than any other Spanish form examined in this study, patterning almost categorically toward discourse organization and turn/topic closure rather than referential categorization, a striking contrast with the GE form *y cosas de esas* ‘and things like that’, which, as shown in (36) and (37), shows greater pragmatic functional flexibility.

Regarding *los clásicos* ‘that typical stuff’, the interview data documented this GE appealing to a shared understanding among interlocutors, thereby fulfilling an interpersonal function. Similarly, the single corpus occurrence available for this form likewise performs an interpersonal function, as illustrated in (48).

(48) *eso por supuesto / el estudiar // nunca está de / que nunca te digan // bueno eso no sirve para nada eso que tal // los clásicos*

(CORPES_PRESEGAL:SCOM_H21_053)

‘that for sure / studying // it’s never / never let them tell you // well that’s good for nothing that’s this or that // **that typical stuff**’

Occurrence (48) closes an exemplar (*estudiar* ‘studying’) realized in reported speech, typical discouraging remarks about the value of studying (*eso no sirve para nada, eso qué tal* ‘that’s good for nothing, that’s this or that’), rather than a directly asserted list, as in the interview (17). Despite this difference, the underlying function converges closely with the interview data: the GE here invokes a shared, community-recognizable category of common discouraging comments, the kind of thing “people always say”, consistent with the interpersonal reading. The convergence between the two datasets suggests that the interpersonal function of *los clásicos* ‘that typical stuff’ is closely tied to the semantics of the adjective *clásico* ‘typical’ itself, invoking a stereotypical, collectively recognized category.

Turning to (*o*) *eso así* ‘(or) stuff like that’, the interview data documented this GE performing a referential function (19), appearing strictly in its bare form *eso así* without the initial disjunctive conjunction *o* ‘or’. In contrast, the single corpus occurrence retrieved from CORPES for this GE includes the disjunctive conjunction.

(49) *si yo tengo veintisiete ellos tienen cuarenta y nueve cuarenta y ocho o eso así*

(CORPES_PRESEGAL:SCOM_H12_028)

‘if I’m twenty-seven they’re forty-nine forty-eight **or stuff like that**’

Here, *o eso así* ‘or stuff like that’ closes a pair of estimated ages, functioning as a numerical approximator rather than a categorical

extender in the strict sense. This aligns with Fiorenza's (2024: 453) definition of "approximation", conceived as one of the key manifestations of intentional vagueness, where the speaker deploys imprecise or vague terms to fulfill a specific communicative goal. Unlike the interview occurrence, however, this instance was classified as personal function rather than referential: the speaker's hesitant, self-correcting production of the two age estimates signals uncertainty about her own recollection, and *o eso así* 'or stuff like that' closes this sequence not by pointing to further comparable items of an open category, but by hedging the precision of the speaker's claim. This diverges from the interview occurrence (19), where *eso así* 'stuff like that' closed a list of grocery items with no comparable epistemic hedging.

The behavior of the disjunctive GE *o algo así* 'or something like that', is illustrated by the two examples below:

(50) *desde fuera se puede ver como algo extraño o como una secta o algo así*

(CORPES_PRESEGAL:SCOM_H13_013)

'from the outside it can look like something strange or like a sect **or something like that**'

(51) *bueno te pongo ahí un vasito de agua / unas galletas / o algo así*

(CORPES_PRESEGAL:SCOM_M33_005)

'well let me set out a glass of water / for you some cookies / **or something like that**'

Both occurrences perform a referential function consistent with the interview data, though they differ considerably in construction. Example (50) closes two nominal exemplars (*algo extraño* 'something strange', *una secta* 'a sect'), each individually introduced by the simulative marker *como* ('like'), a clear instance of what Barotto and Mauri (2018: 117) term an exemplifying simulative

construction,¹⁷ here combined with a general extender in final position, extending the category to further unspecified negative connotations under the same category. Example (51), by contrast, closes two juxtaposed nominal exemplars (*un vasito de agua* ‘a glass of water’, *unas galletas* ‘some cookies’) with no similitive or exemplification marker at all, offering a guest something to drink and eat; the GE here extends to other similar snack or refreshment items. Both examples therefore converge with the interview occurrence in preserving a purely referential function, while illustrating a wider range of constructional possibilities for this GE form than the single interview form alone could show.

Overall, the corpus-assisted analysis presented in this chapter offers confirmation of the two central patterns already identified in the interview data (Chapter 5). First, the marked asymmetry in the GEs production between the two languages, a small, stable set of highly frequent forms in Italian against a considerably more diversified, and in several cases entirely unattested, set of forms in Spanish, recurs consistently across both data sources, reinforcing the conclusion that this asymmetry reflects a property of each language’s general extender system. Second, the functional analysis of the corpus occurrences confirms that referential extension remains the predominant function in both languages, while also revealing the same degree of functional pragmatic flexibility already observed in the interview data, with some forms, such as *y bueno eso* ‘and well that’ displaying a narrower functional range than others, such as *as un poco de todo* ‘a bit of everything’, which distributed across all four functions depending on context.

¹⁷ Exemplifying similitive constructions (cf. Barotto & Mauri, 2018: 117) refer to similitive expressions linking multiple hyponyms to a hypernym, signaling to the hearer that the listed members serve as illustrative instances of a larger category.

The corpus-assisted analysis was not without its limitations, most notably the difference in size, composition, and register between KIParla and CORPES XXI. A comparison of the two corpora revealed a structural difference: KIParla is a corpus of spoken Italian, entirely composed of transcribed spontaneous speech, whereas CORPES XXI is a general reference corpus in which the oral component constitutes a subsection of a much larger, predominantly written resource.

Having analyzed the empirical findings, the next chapter synthesizes the main conclusions, as well as the theoretical and methodological implications of this study.

7. CONCLUSIONS

This thesis aimed to investigate the pragmatic behavior of general extenders in the spontaneous speech of Italian and Spanish bilingual/L2 speakers, combining a self-collected interview corpus with a corpus-assisted analysis of the KIParla and CORPES corpora. This concluding chapter revisits the research questions outlined in chapter 1, discusses the theoretical and methodological implications of the findings, and acknowledges the limitations of the present study.

Regarding the quantitative distribution of GEs across the two languages, the study found an important asymmetry between the Italian and Spanish interview sub-corpora. While the Italian informants relied on a comparatively small set of highly frequent, stable constructions, most notably *eccetera* ‘etcetera’, *queste cose* ‘these things’, and *e tutto* ‘and everything’, the Spanish informants produced ad hoc and highly idiosyncratic GE variants (e.g. *cosas de la vida cotidiana* ‘everyday life stuff’, *o algo tecnológico* ‘or something technological’). Moreover, this quantitative observation was corroborated by the corpus-assisted analysis: seven of the eight Italian forms queried were attested in KIParla, against only nine of the sixteen Spanish forms in CORPES XXI, with the remaining seven entirely unattested. This convergence between the two data sources, the interview corpus and the corpora analysis, strengthens the reliability of the findings and suggests that Spanish, at least within this sample, affords speakers greater flexibility to construct ad hoc general extender constructions, whereas Italian general extenders display a highly frequent set of forms. Regarding the functional distribution of GEs in the corpus-assisted analysis, the

analysis confirmed that referential extension (FUN_REF) is the predominant function in both languages, consistent with the canonical characterization of general extenders as markers of open-category (Overstreet & Yule, 2021). However, the qualitative analysis also revealed functional flexibility both across and within individual forms. Some forms showed a stable, narrow functional pattern: *y bueno eso* ‘and well that’, for instance, was attested exclusively with a textual, discourse-organizing function across the entire corpus sample, never once closing a referential list. Others, such as *un poco de todo* ‘a bit of everything’, distributed across all four functions depending on context. This multifunctionality is consistent with the theoretical claim, advanced throughout the general extender literature (cf. Overstreet & Yule, 2021; Cheshire, 2007), that the interpretation of a GE must be reconstructed from its immediate co-text. The data collected in this study also corroborates two observations already advanced in the literature on general extenders (section 3.4). First, contrary to the assumption that GEs occupy a clause-final position, several occurrences in both the interview and corpus-assisted samples showed the general extender in a medial, non-final position, as in the Italian *eccetera* occurrences discussed in section 5.1.1 and 6.3.1. Second, the data confirms that GEs frequently co-occur with other pragmatic and discourse markers. Markers such as It. *cioè* ‘I mean’, It. *insomma* ‘well’, and Sp. *bueno* ‘well’ were repeatedly found immediately preceding or following the GE, often signaling reformulation, summary, or a shift in discourse structure. Similarly, interpersonal tag questions (Sp. *¿no?* ‘right’, *¿sabes?* ‘you know’) were found to co-occur systematically with several Spanish GEs. These patterns support the view that general extenders are frequently embedded alongside other pragmatic markers that jointly shape their interpretation in context.

Concerning cross-linguistic equivalence, the translation task, within the interview design, produced a striking result: across all thirty translations, every single rendering employed an equivalent general extender in the target language, with no instances of omission, paraphrase, or literal translation. This uniformity suggests that, for bilingual speakers of Italian and Spanish, general extenders are not perceived as idiomatic, language-specific expressions resistant to translation, but rather as instances of a shared, accessible functional category, for which both languages offer an equivalent form. Interestingly, the same asymmetry in lexical variability observed in the spontaneous interview data reemerged in the translation task, with the Spanish renderings showing a comparatively wider range of variants than the corresponding Italian ones, indicating that this cross-linguistic asymmetry is not a result of spontaneous production alone, but a more stable property of how each language's speakers approach the general extender category, even under the more controlled conditions of a translation task.

These findings offer two main contributions. Theoretically, the study lends further cross-linguistic support to the view of general extenders as a multifunctional, cognitively grounded pragmatic category, whose interpretation depends on the online construction of ad hoc categories (cf. Barsalou, 1983; Mauri, 2017; Mauri & Sansò, 2018). From a methodological perspective, the study demonstrates the clear value of a mixed-methods research design. Because the interview setting yielded a relatively small number of general extender forms, relying solely on qualitative elicitation would have provided an incomplete picture. Supplementing the interview sample with a corpus-assisted analysis of up to 20 occurrences per GE form allowed for a more comprehensive empirical examination.

Some limitations of the present study should be acknowledged. First, the interview sample is small (10) and was recruited through convenience sampling, which limits the generalizability of the quantitative findings; the frequencies reported throughout this thesis should therefore be read as indicative of qualitative tendencies. Second, the two reference corpora used in the corpus-assisted analysis, KIParla and CORPES, differ considerably in size, composition, and register, which limits the comparability of raw and normalized frequencies across the two languages. Finally, the annotation of pragmatic function, while carried out systematically and with explicit criteria, remains an interpretive task. Building on these findings and limitations, several directions for future research emerge. Future work could also usefully extend the corpus-assisted methodology adopted here to fully comparable corpora, matched for size and register to allow more reliable normalized-frequency comparisons than in the present study. In conclusion, this thesis has shown that general extenders are richly multifunctional pragmatic devices whose behavior in Italian and Spanish converges in some respects, most notably, the predominance of referential extension and the systematic applicability of the same functional taxonomy, while diverging markedly in others, particularly in the interview corpus; the Italian general extenders feature lower frequency and less idiosyncratic GEs variants than their Spanish counterparts.

By combining elicited interview data and corpora analysis within a single comparative design, this study has provided a solid framework for further comparative research on vague language and GEs in Italian, Spanish, and beyond.

REFERENCES

- Aijmer, K. & Simon-Vandenberg, A. M. (2011). Pragmatic markers. In J. Zienkowski, J-O. Östman, & J. Verschueren (Eds), *Discursive Pragmatics*, p. 229. Amsterdam: John Benjamins.
- Aijmer, K. (2004). Pragmatic Markers in Spoken Interlanguage. *Nordic Journal of English Studies*, 3, pp. 173–190.
- Aijmer, K. (2002). English Discourse Particles: Evidence from a corpus. Amsterdam & Philadelphia: *John Benjamins*, pp. 1–14.
- Austin, J. L. (1962). How to do things with words. Oxford University Press. Pp. 98-118.
- Barotto, A., & Mauri, C. (2018). Constructing lists to construct categories. *Italian Journal of Linguistics*, 30(1), pp. 95–134. <https://hdl.handle.net/11585/653550>
- Barsalou, L. W. (2021). Categories at the interface of cognition and action. In C. Mauri, I. Fiorentini, & E. Gorla (Eds.), *Building Categories in Interaction: Linguistic resources at work*, pp. 35–71. John Benjamins Publishing Company.
- Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, 11(3). Pp. 211–227.
- Bybee, J. (2003). Mechanisms of change in grammaticization: The role of frequency. In Brian Joseph and Richard Janda (Eds.) *The Handbook of Historical Linguistics*. Oxford: U.K.: Blackwell. Pp. 602–605.
- Borreguero Zuloaga, M. (2025). Percorsi di grammaticalizzazione e sviluppo delle funzioni pragmatiche dei general extenders. Uno studio diacronico di e tutto il resto e e cose del genere. *Studi e Saggi Linguistici*, 63(2), pp. 184–201.
- Borreguero Zuloaga, M. (2022). General Extenders in Spanish Interactions: Frequent Forms, Pragmatic Functions *y todo eso*. *Anuari de Filologia. Estudis de Lingüística*, 12, pp. 155–187. <https://doi.org/10.1344/AFEL2022.12.8>
- Borreguero Zuloaga, M. (2021). Are general extenders a type of discourse markers? In M. Saiz-Sánchez & S. Gómez-Jordana Ferrary (Eds.), *Études de sémantique et pragmatique en synchronie et diachronie: Hommage à Amalia Rodríguez Somolinos*. Presses Universitaires Savoie Mont Blanc. Pp. 77–97.
- Brown, P., & Levinson, S. C. (2006). Politeness: Some universals in language usage. In A. Jaworski & N. Coupland (Eds.), *The discourse reader* (2nd ed., pp. 311–323). London: Routledge.
- Buyse, L. (2014). ‘We Went to the Restroom or Something’. General Extenders and Stuff in the Speech of Dutch Learners of English, in

Romero-Trillo J. (ed.), *Yearbook of Corpus Linguistics and Pragmatics New Empirical and Theoretical Paradigms*, Springer, Dordrecht, p. 13.
https://doi.org/10.1007/978-3-319-06007-1_10

Carston, R. (2004). *Relevance Theory and the Saying/Implicating Distinction*. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics*, pp. 641–643. Blackwell Publishing.

Channell, J. (1994). *Vague language*. Oxford: Oxford University Press.

Cheshire, J. (2007). Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics*, 11(2), pp. 155–193.

Cortés Rodríguez, L. (2006). Los elementos de final de serie enumerativa del tipo y todo eso, o cosas así, y tal, etc. Perspectiva interactiva. *Boletín de Lingüística*, 18(26), pp 102–129.

Dines, E. R. (1980). Variation in discourse: "And stuff like that." *Language in Society*, 9(1), pp. 13–31. <https://www.jstor.org/stable/4167109>

Drave, N. (2001). Vaguely speaking: A corpus approach to vague language in intercultural conversations. *Language and Computers*, 36(1), pp. 25–40.

Fernandez, J., & Yuldashev, A. (2011). Variation in the use of general extenders and stuff in instant messaging interactions. *Journal of Pragmatics*, 43, pp. 2613–2623.
<https://doi.org/10.1016/j.pragma.2011.03.012>

Fernández, J. (2015). General extender use in spoken Peninsular Spanish: metapragmatic awareness and pedagogical implications. *Journal of Spanish Language Teaching*, 2(1), pp. 1–17.
<https://doi.org/10.1080/23247797.2015.1010846>

Fiorentini, I., & Sansò, A. (2016). Interagire in contesto multilingue e cose così. Il caso dei general extenders. In C. Andorno & R. Grassi (Eds.), *Le dinamiche dell'interazione. Prospettive di analisi e contesti applicativi*, pp. 189–202. Milano: Studi AltLA 5.

Fiorentini, I. (2018). Eccetera eccetera e così via di seguito. I general extenders dell'italiano contemporaneo. In F. Masini & F. Tamburini (Eds.), *CLUB Working Papers in Linguistics*, 2, pp. 20–35.

Fiorentini, I., & Magni, E. (2021). Et cetera, eccetera, etc. The development of a general extender from Latin to Italian. In C. Mauri, I. Fiorentini, & E. Gorla (Eds.), *Building Categories in Interaction. Linguistic resources at work*, pp. 295–256. John Benjamins Publishing Company.

Fiorentini, I., & Zanchi, C. (Eds.). (2024). *Vagueness, Ambiguity, and All the Rest: Linguistic and pragmatic approaches* (Pragmatics & Beyond New Series, Vol. 347). John Benjamins Publishing Company.
<https://doi.org/10.1075/pbns.347>

- Fiorenza, E. (2024). Osservare e insegnare l'espressione dell'approssimazione nell'italiano parlato accademico: L'uso di liste ed eccetera. *Italiano LinguaDue*, 16(2), 448–471. <https://doi.org/10.54103/2037-3597/27859>
- Gille, J. & Häggkvist, C. (2010). Apéndices generalizadores introducidos por o. *Oralia: Análisis Del Discurso Oral*, 13, p. 136. <https://ojs.ual.es/ojs/index.php/ORALIA/article/view/8105>
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts*, pp. 41–58. Academic Press.
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2nd ed.). Cambridge University Press.
- Jucker, A. H., Smith, S. W., & Lüdge, T. (2003). Interactive aspects of vagueness in conversation. *Journal of Pragmatics*, 35(12), pp. 1737–1748. [https://doi.org/10.1016/S0378-2166\(02\)00188-1](https://doi.org/10.1016/S0378-2166(02)00188-1)
- Kern, J. (2022). General Extenders in Bilingual Speech. In E. Peterson, T. Hiltunen, & J. Kern (Eds.), *Discourse-Pragmatic Variation and Change: Theory, Innovations, Contact*, pp. 212–229. chapter, Cambridge: Cambridge University Press.
- Levinson, S. C. (1987). Minimization and conversational inference. In M. Bertuccelli Papi, & J. Verschueren (Eds.), *The pragmatic perspective: Selected papers from the 1985 International Pragmatics Conference*, pp. 61–99. Benjamins.
- Lim, W., & Wu, Q. (2023). Vague language and context dependence. *Frontiers in Behavioral Economics*, 2, pp. 1–14. <https://doi.org/10.3389/frbhe.2023.1014233>
- Mauri, C. (2017). Building and interpreting ad hoc categories: A linguistic analysis. In J. Blochowiak, C. Grisot, S. Durrleman, & C. Laenzlinger (Eds.), *Formal Models in the Study of Language: Applications in Interdisciplinary Contexts*, pp. 297–326. Springer.
- Mauri, C., & Sansò, A. (2018). Linguistic strategies for ad hoc categorization: theoretical assessment and cross-linguistic variation. *Folia Linguistica Historica*, 39, pp. 1–26. <https://doi.org/10.1515/flih-2018-0001>
- Mauri, C., Gorla, E., & Fiorentini, I. (2019). Non-exhaustive lists in spoken language: A construction grammatical perspective. *Constructions and Frames*, 11(2), pp. 290–316. <https://doi.org/10.1075/cf.00032.mau>
- Mauri, C., & Sansò, A. (2020). Ad hoc categorization and languaging: The online construction of categories in discourse. *Language Sciences*, 81, p. 5. <https://doi.org/10.1016/j.langsci.2020.101312>

- Mauri, C. (2021). Ad hoc categorization in linguistic interaction. In C. Mauri, I. Fiorentini, & E. Gorla (Eds.), *Building Categories in Interaction Linguistic resources at work*, p. 30. John Benjamins Publishing Company.
- Mauri, C., Ballarè, S., Gorla, E., Cerruti, M. & Suriano, F. (2019) “KIParla corpus: a new resource for spoken Italian”. In: Bernardi, Raffaella, Roberto Navigli & Giovanni Semeraro (eds.), *Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it*.
- Masini, F., Mauri, C., & Pietrandrea, P. (2018). List constructions: Towards a unified account. *Italian Journal of Linguistics*, 30(1), pp. 50, 56. <https://www.italian-journal-linguistics.com/2018-2/>
- McGee, P. (2018). Vague language as a means of avoiding controversy. *Training, Language and Culture*, 2(2), pp. 40–54. <https://doi.org/10.29366/2018tlc.2.2.3>
- McLellan, E., MacQueen, K. M., & Neidig, J. L. (2003). Beyond the qualitative interview: Data preparation and transcription. *Field Methods*, 15(1), 63–84. <https://doi.org/10.1177/1525822X02239573>
- Nogueira da Silva, A. M. (2011). *Enseñanza de los marcadores del discurso para aprendices brasileños de E/LE: análisis de manuales de E/LE y de narrativas orales de aprendices brasileños* (Tesis doctoral, Universidad de Salamanca, pp. 164–171). Repositorio Documental GREDOS. <https://hdl.handle.net/10366/110547>
- Overstreet, M. (2014). The role of pragmatic function in the grammaticalization of English general extenders. *Pragmatics*, 24(1), pp. 105–129. <https://doi.org/10.1075/prag.24.1.05ove>
- Overstreet, M. & Yule, G. (2021). General extenders: The forms and functions of a new linguistic category. Cambridge University Press. <https://doi.org/10.1017/9781108938655>
- Palacios Martínez, I. M. (2011a). I might, I might go I mean it depends on money things and stuff: A preliminary analysis of general extenders in British teenagers’ discourse. *Journal of Pragmatics*, 43, pp. 2452–2470.
- Palacios Martínez, I. M. (2011b). The language of British teenagers. A preliminary study of its main grammatical features. *Atlantis. Journal of the Spanish Association of Anglo-American Studies*, 33(1), pp. 105–126.
- Pichler, H., & Levey, S. (2011). In search of grammaticalization in synchronic dialect data: general extenders in northeast England. *English Language and Linguistics*, 15(3), 441–471. <https://doi.org/10.1017/S1360674311000128>
- Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, 4(3), pp. 328–350.
- Rosch, E. H., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), pp. 573–605.

- Ruzaitė, J. (2010). Translation equivalents of vague language items: A study of general extenders in a parallel corpus. *Kalby studijos*, (16), pp. 33–38. <https://doi.org/10.5755/j01.sal.1.16.43359>
- Ruzaitė, J. (2021). Vague language in argumentative essays written by advanced Lithuanian learners of English. *Eesti Rakenduslingvistika Ühingu aastaraamat Estonian Papers in Applied Linguistics*, 17, pp. 231–248.
- Searle, J. R. (1979). Indirect speech acts. In *Expression and Meaning: Studies in the Theory of Speech Acts*, pp. 30–57. Cambridge University Press.
- Sperber, D., & Wilson, D. (1985). Loose talk. *Proceedings of the Aristotelian Society*, 86, pp. 153–171.
- Traugott, E. (2010). (Inter)subjectivity and (inter)subjectification: a reassessment. Kristin Davidse, Lieven Vandelanotte, and Hubert Cuyckens, Eds., *Subjectification, Intersubjectification and Grammaticalization*.
- Voghera, M. (2017). Dal parlato alla grammatica. Roma: Carocci.
- Wilson, D., & Sperber, D. (2004). Relevance Theory. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics*, pp. 607–610. Oxford: Blackwell Publishing.
- Wilson, D., & Sperber, D. (2002). Truthfulness and relevance. *Oxford University Press on behalf of the Mind Association*, 111(443), pp. 583–632.
- Yule, G. (1996). *Pragmatics*. Oxford University Press. Pp. 59–66.

ONLINE RESOURCES

- Council of Europe (n.d.) *Level descriptions*, Common European Framework of Reference for Languages (CEFR). Available at: <https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions> (Accessed: 10 July 2026).
- Real Academia Española. (n.d.). Corpus del Español del Siglo XXI (CORPES XXI). <https://www.rae.es/corpes/> (Accessed: 24 May 2026).
- KIParla. (n.d.). Il corpus. KIParla – Corpus di italiano parlato. <https://kiparla.it/il-corpus/> (Accessed: 24 May 2026).

APPENDIX A – Interview questions

Original guide used with the ten informants, followed by an English translation in italics for reference. Section headings have been added for clarity and were not read aloud to informants.

Demographic data

¿Cómo te llamas?

What is your name?

¿Cuántos años tienes?

How old are you?

¿Cuál es tu género?

What is your gender?

¿Cuál es tu fecha de nacimiento?

What is your date of birth?

¿Cuál es tu nivel de educación actual?

What is your current level of education?

¿Cuál es tu país de origen?

What is your country of origin?

¿Cuál es la variedad de español que hablas principalmente?

Which variety of Spanish do you speak primarily?

¿Cuánto tiempo llevas viviendo o estudiando en Italia?

How long have you been living or studying in Italy?

¿Cuál es tu nivel de italiano/spagnolo?

What is your level of Italian/Spanish?

¿Has hecho algún curso de italiano?

Have you taken any Italian/Spanish course?

¿Ves series o películas en italiano?

Do you watch series or films in Italian?

Además del italiano y el español, ¿hablas o estudias otros idiomas?

Besides Italian and Spanish, do you speak or study other languages?

¿Has vivido en algún otro país?

Have you lived in any other country?

Questions on language and language learning

¿Qué aspectos del italiano te resultan más difíciles o confusos? ¿La pronunciación, la gramática, o algo más?

*Which aspects of Italian do you find most difficult or confusing?
Pronunciation, grammar, or something else?*

¿Cómo estudias normalmente? ¿Usas libros, aplicaciones, ves películas?

How do you usually study? Do you use books, apps, do you watch films?

Si te falta una palabra en italiano, ¿qué haces para salir del apuro? ¿Usas gestos, o intentas describir?

If you're missing a word in Italian, what do you do to get by? Do you use gestures, or try to describe it?

¿Hay reglas o palabras del español que te confunden al hablar italiano?

¿Qué mecanismos usas para no mezclar todo?

*Are there Spanish rules or words that confuse you when speaking Italian?
What strategies do you use to avoid mixing everything up?*

Questions on everyday life

Ahora te voy a hacer otras preguntas sobre tu vida cotidiana. / Now I'm going to ask you some other questions about your everyday life.

Cuando vas al supermercado, ¿qué cosas sueles comprar de forma regular?

When you go to the supermarket, what things do you usually buy regularly?

Cuando tienes que organizar un viaje o un evento, ¿qué es lo primero que haces? ¿Reservar el hotel, buscar vuelos?

When you have to organize a trip or an event, what's the first thing you do? Book the hotel, look for flights?

¿Recuerdas una celebración reciente importante? ¿Qué tipo de gente había, qué comieron?

Do you remember a recent important celebration? What kind of people were there, what did they eat?

En tu ciudad, ¿cuáles son los lugares que la gente siempre recomienda visitar?

In your city, what are the places people always recommend visiting?

Translation task

As a final stage of the interview, informants were asked to translate two short texts, each containing embedded general extenders, from Italian into Spanish and from Spanish into Italian (see section 5.2).

Little Red Riding Hood (Italian source text)

Buon giorno Cappuccetto Rosso”, disse il lupo. “Buon giorno a te”, rispose la bambina. “Dove vai Cappuccetto Rosso, così presto di mattina.” “Vado dalla nonna.” “Cosa porti nascosto sotto il grembiolino?” “Vino, torta, e roba del genere. È fresca. L'abbiamo cotta proprio ieri sera, così la nonna che è debole e malata si rinforzerà.” “Dove sta la tua nonna?” “A un quarto d'ora da qui, nel bosco; proprio sotto le tre querce, là c'è la sua casetta, e lì vicino c'è un gran cespuglio di noccioli, o qualcosa del genere, hai capito dove?

Good morning, Little Red Riding Hood,” said the wolf. “Good morning to you,” replied the girl. “Where are you going, Little Red Riding Hood, so early in the morning?” “I’m going to my grandmother’s.” “What are you carrying hidden under your apron?” “Wine, cake, and stuff like that. It’s fresh. We baked it just last night, so that grandmother, who is weak and ill, will regain her strength.” “Where does your grandmother live?” “A quarter of an hour from here, in the woods; right beneath the three oak trees, there’s her little house, and nearby there’s a big hazelnut bush, or something like that, do you understand where?”

The Little Mermaid (Spanish source text)

La más joven de las princesas, la sirenita más pequeña, era un alma curiosa y amable. Mientras que a sus hermanas mayores les encantaba jugar en los jardines profundos y perseguir delfines, y esas cosas, la sirenita más pequeña soñaba con el mundo sobre el mar. Su abuela le contaba cuentos encantadores sobre el mundo de los humanos, y la sirenita deseaba verlo por sí misma.

The youngest of the princesses, the smallest little mermaid, was a curious and gentle soul. While her older sisters loved playing in the deep gardens and chasing dolphins, and things like that, the smallest little mermaid dreamed of the world above the sea. Her grandmother would tell her enchanting stories about the world of humans, and the little mermaid longed to see it for herself.

APPENDIX B – Annotation of GEs in the interview data

A total of 24 occurrences of GEs identified in the interview were annotated and are presented below.

No.	Lang.	Speaker	Context	GE form	Construction	Ex. type	N. ex.	Label	Function	Cue	Ex. marker	Marker form	Co-const.	e/y/o	Set/class/frame
1	ITA	INF_1	primitissima cosa prenoto l'hotel prenoto l'alloggio perché così ho un alloggio e ho delle date e: ah no un'att si se però diciamo se ho già scelto la destinazione perché a volte magari ho l'idea di andare da qualche parte però non so	eccetera	FRAME + EX + GE	N	2	no	FUN_R EF	yes	no		no	e	frame
2	ITA	INF_1	c'è stata la proclamazione e tutto abbiamo un po' festeggiato lì in università e la sera siamo andati a mangiare in un ristorante eh (.) cosa abbiamo mangiato, allora in realtà era una sorta di cena buffet quindi c'era un po' di tutto	e tutto	EX + GE	N	1	no	FUN_I NT	no	no		no	e	frame
3	ITA	INF_2	vabbè cose per la casa ehm (.) sì queste cose qua	queste cose qua	EX + GE	N	1	no	FUN_P ER	no	no		no		set
4	ITA	INF_2	abbiamo mangiato cibi tipici calabresi un po' troppo piccanti per i miei gusti però tutto molto buono c'era vabbè classica salsiccia piccante un po' di (.) taralli adesso non mi ricordo più i nomi però, tutte cose buone.	tutte cose buone	EX + GE	N	2	no	FUN_P ER	yes	no		no		class
5	ITA	INF_3	cioè ti faccio un esempio sono andata a malta eh l'anno scorso e quindi vabbè ho controllato ovviamente lì si guida a (.) all'opposto rispetto a qua e ehm anche il fatto ad esempio le le prese della corrente sono diverse quindi informarsi anche, queste cose	queste cose	SD + EX + GE	Cl	2	no	FUN_I NT	yes	yes	ti faccio un esempio / ad esempio	no		frame
6	ITA	INF_3	allora si recentemente c'è stata una laurea di una mia amica e: abbiamo fatto un aperitivo e: e nulla c'erano le solite cose pizzette focacce e: panini e: era: siamo state qua a pavia in un locale qua a pavia era ma sì un contesto abbastanza informale.	le solite cose	GE + EX	N	3	no	FUN_I NT	yes	no		no		frame
7	ITA	INF_3	abbiamo fatto soltanto questo aperitivo in cui: mhm (.) vabbè si chiacchierava facevano le foto, queste cose qua insomma.	queste cose qua	EX + GE + SD	V	2	no	FUN_I NT	yes	no		no		frame
8	ITA	INF_4	mangiare quindi diciamo diciamo cibo di per la cena o per il pranzo la pasta piuttosto che un secondo per la cena o magari non lo so prodotti make-up, quindi per truccarmi o: vestiti se in supermercato ci sono alcuni vestiti magari di qualche centro commerciale o anche sai prodotti per l'igiene insomma prima di prima necessità di tutti i giorni ecco (.) un po' di tutto	un po' di tutto	EX + SD + GE	Cl	5	no	FUN_T EX	yes	yes	diciamo	no		set
9	ESP	INF_5	eh a veces la: la dobla cuando hay doble consonante un poco confuso pero creo que leyendo escuchando libros y cosas de esas eh lo aprendo poco a poco	y cosas de esas	EX + GE	V	2	no	FUN_R EF	yes	no		no	y	class
10	ESP	INF_5	depende de las palabras porque hay algunas palabras ahorita mismo no tengo un ejemplo pero hay algunas palabras que bueno hoy vimos una que era volátil que eso en español es	y cosas que se	EX + GE	N	2	no	FUN_R EF	yes	no		no	y	set

No.	Lang.	Speaker	Context	GE form	Construction	Ex. type	N. ex.	Label	Function	Cue	Ex. marker	Marker form	Co-const r.	e/y/o	Set/class/frame
			eh (.) una persona que se que se deja llevar por sus emociones pero en italiano volatile es para los para los pájaros y para las los gases y cosas que se hacen gas rápido	hacen gas rápido											
11	ESP	INF_5	si es un evento eh (.) me enfocaría en las personas por ejemplo si es que es un evento como un cumpleaños o una boda o un evento de fiesta o cualquier cosa eh (.) más o menos las personas dónde van a estar	o cualquier cosa	EX + GE	N	3	no	FUN_I NT	yes	yes	por ejemplo	no	o	class
12	ESP	INF_5	¿si van a estar sentadas? ¿si van a estar bailando? ¿si van ¿si va a haber música? ¿si va a haber comida? más o menos las decoraciones si hace falta si se necesita una presentación o algo tecnológico las luces los detalles técnicos	o algo tecnológico	EX + GE	N	1	no	FUN_R EF	yes	no		no	o	frame
13	ESP	INF_6	okay pero para empezar si voy a algun lugar lo primero que busco es si hay un lu- un hospedaje en general si hay un b&b hotel lo que sea cerca del lugar o al concierto o al evento que iría para no estar tan perdido	lo que sea	EX + GE	N	2	no	FUN_R EF	yes	no		no		class
14	ESP	INF_6	si puedo recordar algo era que a mí me gustó mucho un plato que era como una paella pero era con arroz (.) de color negro con tomates y otras cosas que era muy rica	y otras cosas	EX + GE	N	2	no	FUN_P ER	yes	no		no	y	set
15	ESP	INF_6	¿cómo los describiría? eh (.) wow los describiría como únicos variados eh existe una palabra en latín que es sui generis porque es un poco de todo.	un poco de todo	EX + GE	ADJ	3	no	FUN_T EX	yes	no		no		set
16	ESP	INF_7	bueno la comida para mi gata ((ride)) la comida para mi gata pero: bueno queso mucho queso me gusta el queso eh si en el supermercado compro fundamentalmente comida comida cosas de de la vida cotidiana eh	cosas de la vida cotidiana	EX + SD + GE	N	1	yes	FUN_I NT	yes	no		no		frame
17	ESP	INF_7	ah bueno hicimos eh (.) hicimos sándwiches así como pancitos de miga eh (.) comimos pasteles porque <en chile nos gusta mucho comer pastel> bebimos cerveza eh (.) y: y bueno eso si bueno papas fritas que son son muy ricas <las papas fritas>	y bueno eso	SD + EX + GE	N	3	no	FUN_T EX	yes	no		no	y	frame
18	ESP	INF_7	eh ah y en lo que se refiere a la: digamos a la a la a la gramática las preposiciones para mí son una pesadilla no no nunca he entendido bien cómo ((ride)) cómo se usan quizás también un poco el contraste entre ir y venir los clásicos	los clásicos	EX + GE	N	2	yes	FUN_I NT	yes	yes	digamos	no		frame
19	ESP	INF_8	ehm de cajón huevos leche verduras frutas eso así.	eso así	EX + GE	N	4	no	FUN_R EF	yes	no		no		set
20	ESP	INF_9	ah okay eh: yo diría que entre libros y: videos eh digase o aplicaciones como cosas didácticas <quizás>	como cosas didácticas	EX + ADV + GE	N	3	yes	FUN_R EF	yes	yes	digase	no		frame
21	ESP	INF_9	eh hm hm hm hm (.) si no bueno al inicio al inicio cuando aprendí obviamente diferencias como: mantequilla burro que para nosotros burro es asino entonces esa es como esas clásicas diferencias	esas clásicas diferencias	EX + GE	N	2	yes	FUN_I NT	yes	yes	como	no		frame
22	ESP	INF_9	pastas arroz eh algún tipo de proteína que normalmente compro pollo eh (.) o algo más dulce quizás alguna mermelada	algo más dulce	GE + EX	N	1	yes	FUN_R EF	yes	yes	quizás	no		set
23	ESP	INF_9	eh (.) ¿qué más? (.) verduras si es que no las compré en en algún mercado o algo así frutas también eh (.) cereales avena eh yogurt.	o algo así	EX + GE	N	1	no	FUN_R EF	yes	no		no	o	class

No.	Lang.	Speaker	Context	GE form	Construction	Ex. type	N. ex.	Label	Function	Cue	Ex. marker	Marker form	Co-const. r.	e/y/o	Set/classes/frame
24	ESP	INF_9	primero buscar vuelos eh luego el alojamiento y luego todo lo que lo que lo que viene a lo mejor programar los días eh a dónde ir	y luego todo lo que viene	EX + GE	Cl	2	no	FUN_T EX	yes	yes	a lo mejor	no	y	frame

APPENDIX C – Translation task

Complete set of individual translations produced by the 10 informants for the three general extenders embedded in the translation task (Section 5.2). All 30 translations (10 informants x 3 items) were classified as instances of equivalent GE translation.

Original GE	Source lang.	Target lang.	Informant	Translation produced	Strategy
roba del genere	ITA	ESP	INF_1	y algo similar	Equivalent GE
roba del genere	ITA	ESP	INF_2	y algo así	Equivalent GE
roba del genere	ITA	ESP	INF_3	y algo así	Equivalent GE
roba del genere	ITA	ESP	INF_4	y algo de este tipo	Equivalent GE
roba del genere	ITA	ESP	INF_5	y cosas por el estilo	Equivalent GE
roba del genere	ITA	ESP	INF_6	y cosas así	Equivalent GE
roba del genere	ITA	ESP	INF_7	y cosas por el estilo	Equivalent GE
roba del genere	ITA	ESP	INF_8	y cosas similares	Equivalent GE
roba del genere	ITA	ESP	INF_9	y cosas como relacionadas eh <algo así>	Equivalent GE
roba del genere	ITA	ESP	INF_10	otras cosas de este género	Equivalent GE
o qualcosa del genere	ITA	ESP	INF_1	o algo así	Equivalent GE
o qualcosa del genere	ITA	ESP	INF_2	o algo así	Equivalent GE
o qualcosa del genere	ITA	ESP	INF_3	o algo así	Equivalent GE
o qualcosa del genere	ITA	ESP	INF_4	algo de este género	Equivalent GE
o qualcosa del genere	ITA	ESP	INF_5	una cosa por el estilo	Equivalent GE
o qualcosa del genere	ITA	ESP	INF_6	o algo similar	Equivalent GE
o qualcosa del genere	ITA	ESP	INF_7	o algo por el estilo	Equivalent GE
o qualcosa del genere	ITA	ESP	INF_8	o algo similar	Equivalent GE
o qualcosa del genere	ITA	ESP	INF_9	o cosas similares	Equivalent GE
o qualcosa del genere	ITA	ESP	INF_10	o algo así	Equivalent GE
y esas cosas	ESP	ITA	INF_1	e cose così	Equivalent GE
y esas cosas	ESP	ITA	INF_2	e cose così e altre cose	Equivalent GE
y esas cosas	ESP	ITA	INF_3	e queste cose	Equivalent GE
y esas cosas	ESP	ITA	INF_4	e queste cose e cose così	Equivalent GE
y esas cosas	ESP	ITA	INF_5	e cose del genere	Equivalent GE
y esas cosas	ESP	ITA	INF_6	e cose così	Equivalent GE
y esas cosas	ESP	ITA	INF_7	e cose del genere	Equivalent GE
y esas cosas	ESP	ITA	INF_8	e cose del genere	Equivalent GE
y esas cosas	ESP	ITA	INF_9	e cose del genere	Equivalent GE
y esas cosas	ESP	ITA	INF_10	e queste cose	Equivalent GE