Portuguese Language Journal | ISSN 1942-0986 |v. 16 (2022)
American Organization of Teachers of Portuguese

A REFERENCE LIST FROM BP NATIVES' PRODUCTION AND INTERPRETATION DATA FOR THE TEACHING OF NUMERIC IMPRECISION IN L2/L3 PORTUGUESE

DOI: 10.56515/PLJ562476671

Abril Jimenez[1]
Davidson College
Diogo Cosme[2]
Salt Lake Community College

Abstract: In this study, we examined how numeric imprecision is expressed and interpreted in Portuguese. Specifically, we looked at approximators (APs) (e.g., *umas 40 pessoas* 'some 40 people', *quase 10 alunos* 'almost 10 students'), which are lexical items that make the semantic boundaries of expressions containing numeric values vague. APs are generally not part of the language curricula or textbooks for non-native speakers despite them being highly present in everyday communication. With this in mind, our goal is to create a pedagogical reference list of the most commonly used APs in Portuguese, as this information can assist Portuguese language instructors and material creators, especially those serving populations whose native language is Spanish, as it is often the case in the US. Using the data from 57 Brazilian Portuguese native speakers, we analyzed the range and frequency of APs produced during a written elicited production task, as well as their interpretation of the meaning of imprecise quantities in a forced-choice task. Descriptive and inferential statistics showed that (i) APs are employed with high frequency, (ii) several different APs are used, and (iii) APs that exist or are cognates in Spanish do not always share the same semantic meaning in Portuguese.

Keywords: Approximators, Vagueness, Numeric Imprecision, Portuguese, Language Teaching

Resumo: Neste estudo, examinamos como se expressa e se interpreta imprecisão numérica em português. Em particular, analisamos aproximadores (APs) (por exemplo, *umas* 40 pessoas, *quase* 10 alunos), que são itens léxicos que tornam vagos os limites semânticos de expressões contendo valores numéricos. APs geralmente não fazem parte de currículos de idiomas ou livros didáticos para falantes não nativos, apesar de estarem altamente presentes na comunicação cotidiana. Com isso em mente, nosso objetivo é criar uma lista pedagógica de referência com os APs mais utilizados em português, pois esses dados podem auxiliar instrutores e criadores de materiais de língua portuguesa, principalmente aqueles que atendem populações cuja língua nativa é o espanhol, contexto frequente nos Estados Unidos. Usando os dados de 57 falantes nativos do português brasileiro, analisamos o alcance e a frequência de APs produzidos durante uma atividade de

1 Abril Jimenez is an Assistant Professor of Hispanic Studies at Davidson College in North Carolina, and she holds a Ph.D. in Bilingualism and Second Language Acquisition from Rutgers University. Her work focuses on the acquisition of Spanish and Portuguese as heritage and second languages in the areas of lexical pragmatics and different aspects of grammar. ORCID ID: https://orcid.org/0000-0002-5498-7651

2 Diogo Cosme is an Assistant Professor of Spanish and Latin American Studies in the Humanities and Languages Division at Salt Lake Community College. He received his Ph.D. in Romance Languages/Linguistics from the University of Georgia. His research interests comprise experimental approaches to syntax and multilingualism. In particular, tense and aspect in Second and Third Language Acquisition. Additional scholarly work includes educational technology and heritage language in Spanish and Portuguese. ORCID ID: https://orcid.org/0000-0003-0404-7829

produção escrita elicitada, bem como a interpretação do significado de quantidades imprecisas de APs em uma atividade de escolha forçada. Estatísticas descritivas e inferenciais mostraram que (i) APs são empregados com alta frequência, (ii) vários APs diferentes são usados, e (iii) APs que existem ou são cognatos em espanhol nem sempre compartilham o mesmo significado semântico em português.

Palavras-chave: Aproximadores, Vaguidade, Imprecisão Numérica, Português, Ensino de Línguas

1. Introduction

Estimating quantity is part of everyday communication (Kennedy 1987), as speakers are often required to discuss information that is not clear and determinate in their mind. Thus, when quantities and numeric values are not readily available in the speaker's mind, they resort to vagueness for practical reasons (Lasersohn 1999). Approximators (APs) such as *about, around,* or *approximately* (e.g., about $10, around 30 people, approximately 50 feet) are lexical items with properties that give rise to vagueness by introducing imprecision in numeric references. APs give rise to vagueness by designating intervals of possible interpretation (i.e., a range of possible values that are acceptable) that are always in the adjacency of the exemplar number (i.e., the reference number). That is, they modify numeric expressions by inserting a proximal component (Amaral 2010), which signals that the approximated number should not be interpreted as precise, but as being in the vicinity of the actual value.

APs have been studied in a variety of languages: English (Amaral 2010; Channell 1994; Cheng and Warren 2003; Drave 2002; Gibbs and Bryant 2008; Jucker et al. 2003; Ruzaite 2004), Spanish (e.g., García-Medall 1993; González Rodríguez 2008; Kern 2012; Jimenez 2019, 2022; Said-Mohand 2006), French (Champaud and Bassano 1987), Cantonese (Drave 2002), and Swedish (Holmlander 2011). However, with the exception of a brief commentary in Mihatsch (2009) about the scarce use of *como* 'like' with numeric expressions in Portuguese, to our knowledge and to the date of our writing, no study has examined this topic in Brazilian Portuguese. In the context of language instruction in the US, Portuguese is designated as a critical language by the North American government (Critical Language Scholarship n.d.), which suggests more visibility to the language and more funding allocated to training instructors and creating programs to teach Portuguese as a second language. However, some topics in Portuguese remain unexplored in both formal and applied linguistics. Indeed, in the case of APs, several authors have noted the lack of exposure language learners have to vague vocabulary (Channell 1994; Cutting 2007; De Cock et al. 1998); thus, documenting what, how, and when APs are used in Portuguese is an important task, as it would facilitate the integration of these lexical items in the teaching curriculum.

Our goal is not only to contribute to the field, which currently lacks research on numeric vagueness in BP, but we also aim to make this information available to Portuguese language instructors and material creators, in particular, because several Portuguese learners in the US are also speakers of Spanish. The estimation is that almost half (45%) of the Portuguese classes in the US consists of Spanish speakers (Milleret 2012); therefore, using APs in Portuguese may be challenging because of the lack of instructional resources, which may also motivate the reliance on cognates from the previously acquired language(s). In fact, Simões, Carvalho, and Wiedemann (2011) have observed fossilization[3] in the interlanguage of students enrolled in Portuguese for Spanish speakers courses.

[3] Selinker (1972) put forth the term, proposing that learners may transfer non-facilitative properties that can become a permanent part of their production. In other words, learners may exhibit re-occurance of ungrammatical features or infelicitous word choices despite progressing to more advanced levels of proficiency.

In this study, we explore APs in Brazilian Portuguese (BP), and we focus only on scalar APs that combine with numerals (Sauerland and Stateva 2011), such as *quase* 'almost', *umas* 'some', *mais ou menos* 'more or less', and *cerca de* 'close to', with the purpose of using empirical evidence to generate a reference list of the most commonly used APs in BP and to understand the semantic meaning of some frequently used APs in this language. We examined the use and interpretation of APs among 57 native speakers of BP by using a written elicited production task (for production data) and a forced-choice task (for interpretation data). With this study, we aim to establish (i) whether APs are the preferred form or not when expressing numeric imprecision, to uncover (ii) what APs are more commonly used in BP (and to what extent this list compares to Spanish data), as well as to examine (iii) how speakers interpret the meaning of these lexical items, and the extent to which their semantic values are comparable to those APs that are morphologically similar in Spanish. The data collected in the present study is exclusive for Portuguese, but we use data from previous studies about Spanish to compare the findings.

This experimental study on APs in BP contributes to the understanding of how numeric vagueness is conveyed lexically and what speakers understand when APs are used. In addition to that, linguistic approximation is a phenomenon that is widely present across languages (Mkhitaryan and Ayvazyan 2017), but, as mentioned before, it is rarely taught in the language classroom (Peires 1997).

2. Background

2.1 Vagueness and APs

Vagueness refers to the lack of sharp semantic boundaries, and it is a property of a variety of words such as adjectives (e.g., short, big, long) and nouns (e.g., a pile, a mass, a heap) with indeterminacy in their limits. Round numbers (e.g., 100 meters, $50, 5 kilograms) can also be associated with vagueness, as round numbers (100 meters, $55, 5 kilograms) are often preferred over non-rounded numbers (103 meters, $54.67, 5.232 kilograms) for practical reasons. However, when speakers are explicitly vague with regard to a quantity or numeric value, previous research in different languages has shown that speakers have a strong tendency to use APs (Channell 1994; Fuentes Rodríguez 2008; Prince, Frader, and Bosk 1982; Said-Mohand 2006), as shown in the examples (1-3) below:

(1) John got here at *about* 3:30 PM
(2) The shirt costs *like* $15
(3) The athlete raced a distance of *approximately* 40 kilometers

While the literature has approached vagueness and APs in different ways, we adopt Sauerland and Stateva's (2011) semantic analysis of APs. According to this proposal, the APs under examination in the present study are scalar, which, as the name suggests, denote an interval on a scale. These can be clearly visualized with numerals, and can make expressions less precise (4), as opposed to more precise (5).

(4) Jack has *around/about/approximately* 30 pairs of shoes
(5) Jack has *exactly* 30 pairs of shoes

The APs exemplified in (4) have received different names in the literature. For instance, Prince, Frader, and Bosk (1982) use the term *rounders*, while Wierzbicka (1986) called them *approximatives*. In this study, we follow Sauerland and Stateva's (2011) classification, and we use the term *scalar APs* to refer to those that combine with numerals to express imprecision.

2.2 APs in utterance interpretation

We understand APs as words that modify numeric expressions and that are used as a strategy by speakers to indicate imprecision about quantity or numeric values, which is often motivated by indetermination (Peirce 1902), practicality (Lasersohn 1999), or uncertainty at the time of speaking (Channell 1994; Jucker et al. 2003). Furthermore, these lexical items play a very important role in the process of utterance interpretation given that they assist hearers by signaling what type of inferences should be activated in order to reach the appropriate interpretation of a vague quantity. For example, depending on the AP, the direction of the interpretation can be skewed a) upwards: signaling values above the exemplar number (e.g., *40-ish)*; b) downwards: signaling values below the exemplar number (e.g., *almost 40*); or c) in both directions (e.g., *about 40).*

Such interpretations are formalized in Garcia-Medall's (1993) proposal of a three-type categorization, which is based on the inferences that APs activate: negative, neutral, or positive. The first type is defective APs. As the name suggests, they activate negative inferences (i.e., the interpretation is skewed downwards), since they lead interlocutors to interpret a lower value than the exemplar number (i.e., the reference number). Example (6) shows the defective approximator *almost* in English (6a) and *quase* in Portuguese (6b). In both sentences, the exemplar number is 30, which is the threshold number we compare to. With defective APs, our interpretation cannot be higher than the exemplar number; thus, we can only interpret *almost 30* as a value lower but close to 30, viz., 27-29.

(6) a. It costs *almost* 30 dollars
b. Custa *quase* 30 dólares

The second category refers to APs that activate both a positive and negative inference (i.e., skewed symmetrically around the exemplar number), hence, they are called neutral APs. In English, an example of this category is *like* (7a), whereas the Portuguese counterpart is *tipo* (literally 'type') (7b). In these cases, the interval of possible interpretation ranged from values that are slightly above or below, viz. 27-33.

(7) a. It costs *like* 30 dollars
b. Custa t*ipo* 30 dólares

Finally, the third category consists of excessive APs, those that activate positive inferences (i.e., skewed upwards). In other words, the APs that denote a value slightly higher than the exemplar number, which is the case of *-something* in English (8a) and *e tantos/tantas* in Portuguese (8b).

(8) a. There were 30-*something* people
b. Havia 30 *e tantas* pessoas

Table 1 summarizes the three categories of APs with examples in Spanish, Portuguese, and their English correspondence. We ask the reader to note the morphological similarities among many of the Portuguese and Spanish APs. On one hand, such similarities may enhance acquisition among Portuguese learners who speak Spanish since the morphology is easily transferable from one language to another. On the other hand, some semantic values or the frequency in which those cognates are used across languages may not always be directly related, hence the need for comparative data that can serve as a resource for learners. We further discuss the pedagogical gaps and implications in the following section.

	Spanish	**Portuguese**
Defective	*casi* 'almost' *prácticamente* 'practically' *cerca de* 'close to'	*quase* 'almost' *praticamente* 'practically'
Neutral	*como* 'like' *unos/unas* 'some' *aproximadamente* 'approximately' *más o menos* 'more or less' *alrededor de* 'around'	*cerca de 'close to'*[4] *tipo* 'like' *uns/umas* 'some' *aproximadamente* 'approximately' *mais ou menos* 'more or less' *ao redor de* 'around' (also *em torno de*, and *por volta de)*
Excessive	*y tantos* '-something' *largos* 'or so' *y pico* '-ish'	*e tantos/tantas* '-something' *e poucos* 'ish'

Table 1. Categories of APs based on the taxonomy proposed by García-Medall (1993)

2.3 Pedagogical approaches (or the lack thereof) for teaching vague language and Aps

Vague language is acquired in a particular socio-cultural context (Peires 1997), and although it is highly frequent in everyday language use, it is rarely introduced in second language classrooms (Channell 1994; Cutting 2007; De Cock et al. 2014; Peires 1997). This is, perhaps, because vague language has been mistakenly seen as a negative feature of language that deviates from precision and clarity (Jucker, Smith, and Lüdge 2003) or because its absence in language textbooks and curricula keeps perpetuating the cycle. The reality is that vague language is an inherent characteristic of all languages and an element for successful communication (e.g., Channell 1994; Cutting 2007; Jucker et. al. 2003; Metsä-Ketelä 2006; Sabet and Zhang 2015), and, as such, language learners should have the opportunity to learn how to speak vaguely. However, teaching vague language can be a challenging task if there are no materials or resources available, or if the instructors are not aware of the frequency of vagueness in everyday communication.

There are only a handful of studies on how vague quantities are expressed by learners of English, even fewer for learners of Spanish; and, to our knowledge, none for Portuguese learners. In the particular case of APs, research on second language learners has shown that they tend to use APs less frequently and make different lexical choices than native speakers when they discuss, for instance, imprecise quantities or numeric values (e.g., Cheng and Warren 2003; Drave 2002; Jimenez 2022; Yu 2009). These studies also offer information on the most frequent lexical items in both native and non-native repertoires. Comparing the previous literature, especially in Spanish, to Portuguese is useful to determine (i) what lexical items should be prioritized in the teaching, (ii) which ones can successfully be transferred from the first language, and (iii) which forms require more attention as they may pose a challenge to the learners.

Previous research on English has identified the most frequent lexical items that native and second language speakers use to express numeric imprecision by using corpus data (e.g., Drave 2002; Cheng and Warren 2003). The findings show that native English speakers use the APs *like,*

[4] Portuguese *cerca de* was originally categorized as defective. The table was modified after some of our findings, which we will further address in the results and discussion sections.

about, around, more or less, and *almost* frequently. Similarly, in Spanish, studies such as Jimenez (2019, 2022) uncovered that second language learners use APs less frequently and favor lexical items that do not always match native speakers' preferences. These studies uncovered that native Spanish speakers favor the use of *como* 'like', *unos/as* 'some' *alrededor de* 'around', and *más o menos* 'more or less', while non-native speakers use *cerca de* 'close to' and *casi* 'almost'. The frequent use of *como* has also been documented among native Spanish speakers of several dialects such as Peninsular, Mexican, Colombian, and Argentinean, and among other dialects of Spanish in Latin America and in the US (e.g., Camacho 2011; Grasso 2012; Kern 2012; Mihatsch 2009, 2010; Said-Mohand 2006)

A different study by Jimenez (2018) compared how native and second language speakers of Spanish interpreted quantities modified by the APs *casi* 'almost', *prácticamente* 'practically', *como* 'like', and *cerca de* 'close to'. The data indicates that while there seems to be a shared understanding about the meaning of most APs, many second language learners, specially the intermediate-proficiency ones, interpreted *cerca de* and *casi* differently. That is, while native speakers understood these APs as defective (i.e., signaling values that are below the reference number), learners interpreted them as neutral (i.e., signaling values that are both below and above the reference number). This finding is interesting because in more recent studies about AP use, Jimenez (2019, 2022) found that it is precisely *casi* and *cerca de* the APs that intermediate learners prefer to use in oral production.

Collectively, these studies have concluded that non-native speakers sometimes rely on vague forms that also exist in their first language and may experience difficulties mapping the forms in the second language with the appropriate semantic values. This example of negative transfer and/or developmental issues can occur even within a language pair that is extremely closely related. By transfer, we mean the acquisition of properties from the target language that can be promptly acquired, i.e., the English word *information* contains similar morphology and semantics to the Spanish word *información* or Portuguese *informação*. Thus, it can be transferred to the Spanish/Portuguese interlanguage without going through a long process of acquisition. Negative transfer (Faerch and Kasper 1987) occurs when learners misanalyse the correspondence among their previously acquired languages and the target language. For instance, there is a mismatch between the morphology and the semantics of English *embarrassed* and Spanish *embarazada* 'pregnant'. Spanish speakers may quickly add the English word to their interlanguage without analyzing the semantics, negatively affecting acquisition.

The lack of pedagogical resources on numeric imprecision and the absence in language curricula might also be related to the unfamiliarity with the communicative functions of APs. APs take a significant part in the process of utterance interpretation, as they provide information at the procedural level. Procedural information encodes instructions that specify the inferential operations that the hearer needs to perform during interpretation, as well as instructions to contextualize the interpretation (Escandell-Vidal, Leonetti, and Ahern 2011). Unlike conceptual information, which refers to the way in which elements of linguistic structures map onto concepts, procedural information maps linguistic structures onto mental processes (Blakemore 1987). APs give hearers processing instructions to interpret utterances appropriately, rather than telling them something about the real world. APs are a lexical material with meanings that are flexible enough to represent a variety of experiences such as indetermination (Peirce 1902), practicality (Lasersohn 1999), and uncertainty (Channell 1994; Jucker et al. 2003). Thus, they allow speakers to craft messages that carry affective communication (Blakemore 2008) (i.e., impressions with non-propositional effects), in order to express their state of mind.

APs involve not only decoding linguistic meaning but relying on inferential processes to properly activate the interpretation that is appropriate to each lexical item and to the context. Therefore, ignoring APs in language instruction prevents learners from acquiring important semantic and contextual features to successfully craft nuances in their speech to intentionally convey affective communication, indetermination, and uncertainty, and, so, getting in the way from their achieving higher proficiency levels, such as ACTFL's (2012) advanced and superior

levels, which demand the ability to actively participate in the conversation from concrete and abstract perspectives.

3. The study

Considering the existing information about APs in the literature and its gap related to Portuguese, the questions that guided our study are the following:

1. Are APs frequently produced in Brazilian Portuguese as markers of numeric imprecision?
2. What APs do Portuguese speakers favor, and to what extent do the findings compare to those in the Spanish literature?
3. What are the semantic meanings associated with APs in Portuguese, and how do they relate to their Spanish cognates?

For RQ 1, the hypothesis states that APs will be the preferred lexical items for conveying numeric imprecision, while for RQ2, the hypothesis is that neutral APs (e.g., *tipo* 'like', *mais ou menos* 'more or less', *em torno de* 'around') will be favored over defective (e.g., *quase* 'almost', *praticamente* 'practically') or excessive (e.g., *e poucos* '-ish', *e tantos* '-something') APs. The first two hypotheses are supported by previous literature that has identified the pervasive use of APs, and specifically of neutral approximators in languages close to Portuguese, such as Spanish (Grasso 2012; Jimenez 2019, 2022; Kern 2012; Mihatsch 2009, 2010; Said-Mohand 2006). Neutral APs are expected to be preferred as they do not establish a specific semantic limit, and, thus, requiring less precision by the speaker and providing less restricted margins of interpretation without being too far from the exemplar number. Regarding RQ3, we hypothesize that approximators such as *tipo* and *uns/umas* will be interpreted as neutral (e.g., *tipo 10 garrafas* 'like 10 bottles', meaning that 9 or 11 would acceptable readings), while *cerca de* and *quase* will be interpreted as defective (e.g., *quase 10 garrafas* 'almost 10 bottles', 9 would be acceptable but 11 would not). The hypothesis also states that these APs in Portuguese, which also exist in Spanish with minimal morphological changes, will share the semantic meaning with their Spanish counterparts.

4. Methods

4.1 Participants, screening methods, and sampling

57 native speakers of Portuguese participated in the study (mean age = 46.55; sd = 13.99). The participants were born and raised in Brazil and were (a) residing in Brazil or (b) in the United States when the study was conducted. We ran four one-way ANOVA tests to check if there were statistical differences related to their age ($F(1)=1.894$, $p > 0.174$), knowledge on standard Portuguese, i.e., level of education ($F(1)=1.937$, $p > 0.17$), as well as their responses in the production ($F(1)=2.456$, $p > 0.118$) and interpretation tests ($F(1)=1.175$, $p > 0.289$). Results revealed there were no statistically significant differences among groups. Therefore, we merged all the participants into one group.

Self-reported data were gathered by using a language background questionnaire developed by the research group, in which participants responded to questions about their linguistic background, language experience, and use, as well as self-rated proficiency in their language(s). They were given a list of proficiency levels and sublevels with a description of communicative abilities associated with each that are based on ACTFL's (2012) can-do statements and sub-level descriptions. The study employed a purposive sampling technique in order to achieve comparability among the participants. Purposive sampling is a subject selection technique that involves selecting participants based on a specific purpose rather than randomly. Participants were recruited using a recruitment flier that was distributed on a social media platform, as well as using

the snowball technique, which involves research participants identifying other potential subjects. The study was conducted over the internet using Qualtrics - a cloud-based platform for creating and distributing web-based surveys. They received two surveys. The first one was an electronic consent form on Qualtrics that was previously approved by the Institutional Review Board (IRB). Then, we assigned each participant an identifier code and sent them the link to the data collection instrument, where they typed the identifier code at the beginning, keeping the data protected from personal information. Participants received the link to the tasks and consent form via email and completed the study individually.

4.2 Written elicited production task (WEPT)

A WEPT was used to examine how numeric imprecision was expressed in Portuguese. In this task, participants received a general context in which they were asked to help a student that was new on a university campus. The participants read a series of questions that the new student had about the university. They were told that even though they did not know all the information with complete precision, they would try to help the student. The participants were also asked to express that their answers were not precise; however, they were not told which words they had to use. This task only elicited written responses; thus, our data do not reflect the use of APs in spoken language.

The task was administered using the software Qualtrics, which randomized the order in which the items were presented to each participant. Each task item contained one question that served as the preamble and its corresponding answer prompt that contained a blank space where the participants typed a lexical item that conveyed imprecision. A total of 36 question/answer items were included in this task, and the experimental items that elicited the APs of both discrete (i.e., people and objects) and non-discrete items (i.e., time and money), as well as fillers. With the purpose of maintaining homogeneity regarding the type of item under investigation in the production and interpretation tasks, in this study, we are reporting exclusively on the responses provided for the 6 items containing discrete items (i.e., objects and places). Experimental items are listed below in Table 2. Participants only received the Portuguese version of the preamble and prompt. English translations are provided here for the reader's convenience.

Preamble	**Prompt**
Quantas mesas tem na cafeteria? 'How many tables are there at the cafeteria?'	A cafeteria tem _____ 10 mesas. 'There are _____ tables at the cafeteria.'
Quantos prédios tem a faculdade? 'How many buildings are there in the college?'	A faculdade tem _____ 20 prédios. 'There are _____ 20 buildings in the college.'
Quantas cadeiras tem em cada sala de aula? 'How many chairs are there in each classroom?'	Em cada sala aula tem _____ 40 cadeiras. 'In each classroom there are _____ 40 chairs.'
Quantos escritórios tem em cada prédio? 'How many offices are there in each building?'	Em cada prédio tem _____ 50 escritórios. 'In each building there are _____ 50 offices.'
Quantos computadores tem o laboratório? 'How many computers are there in the lab?'	O laboratório tem _____ 60 computadores.

	'In the lab there are _____ 60 computers.'
Quantos livros tem em cada departamento? 'How many books are there in each department?'	Em cada departamento tem _____ 80 livros 'In each department there are _____ 80 books.'

Table 2. Experimental items for the elicited production task.

4.3 Forced-choice task (FCT)

Participants completed an FCT with a total of 108 items, 24 of them were experimental and the rest served as distractors. The purpose of the task was to evaluate the participants' interpretation of six different APs (each one with $k = 6$): *quase* 'almost', *uns/umas* 'some', *tipo* 'like', and *cerca de* 'close to'. The APs for the FCT were selected following Jimenez (2018), who investigated this phenomenon in Spanish. To ensure that the selected APs are actually frequently used in Portuguese, we also searched for occurrences of these APs in the *Corpus do Português* (Davis and Ferreira 2006).

The goal of the task was to test whether these lexical items were interpreted as defectives, neutrals, or excessives, following Garcia-Medall's (1993) taxonomy for APs. Recall that APs can be separated according to the values they signal in relation to the exemplar number (i.e., the number used in the approximation): defective APs only signal values that are slightly lower than the exemplar number, neutral APs signal values that can be either slightly lower or slightly higher, and excessive approximators convey values that are slightly higher.

During the task, the participants read sentences containing approximated quantities, such as the ones provided in Table 3, and had to decide between two possible interpretations (A and B). Based on their choices, we were able to gather the participants' interpretation of the APs' meanings. That is, if there was a tendency to choose the number below the exemplar number for a given approximator, it informed us that it was considered defective, whereas a preference for the higher number signaled that the approximator was interpreted as excessive. However, if there was not a strong tendency to choose either a number below or above (i.e., both interpretations were possible), it indicated that the AP was considered neutral, as neutral APs, theoretically speaking, can activate both interpretations. Table 3 provides some sample items.

Sentence	**Option A**	**Option B**
O pai abriu quase 20 garrafas 'The father opened almost 20 bottles'	21	19
A amiga comprou umas 70 saladas 'The friend bought some 70 salads'	73	67
A avó cozinhou tipo 50 biscoitos 'The grandmother cooked about 50 cookies'	51	49
O tio limpou cerca de 60 cadeiras 'The uncle cleaned close to 60 chairs'	58	62

Table 3. Sample experimental items from the forced-choice task.

4.3 Data analysis

The quantitative analysis used descriptive statistics as well as a Generalized Linear Mixed-Effects Model (GLMM) run in R using the lme4 package (Bates, Mächler, Bolker, and Walker 2015). The descriptive statistics show the distribution of responses both in the written production and the interpretation of APs. The GLMM analyzed the participants' likelihood to interpret the APs *cerca de*, *quase*, *tipo*, and *uns/umas* as signaling values that are below versus values that are above the exemplar number. Only categorical variables were entered in the model. Participant and item were set as random factors.

5. Results

5.1 Production data

The first two RQs inquired about the use of APs as markers of numeric imprecision in Portuguese. RQ1 focused on determining whether APs were the favored form when numeric imprecision had to be expressed, and a written elicited production task (WEPT) was used to gather data. Figure 1 shows the distribution of responses containing APs versus responses containing other markers that convey numeric imprecision. This figure illustrates that the participants exhibited a strong tendency to use APs (e.g., *mais ou menos* 'more or less', *uns* 'some', *cerca de* 'close to') as opposed to other lexical items (i.e., non-approximators) that lessen the degree of certainty or the degree of precision. The non-APs that were found in the data were adverbs of doubt (e.g., *possivelmente* 'possibly', *provavelmente* 'probably', *talvez* 'maybe', *quiçá* 'perhaps') and partial specifiers (*no máximo* 'at most', *no mínimo* 'at least', *mais de* 'more than', *menos de* 'less than').

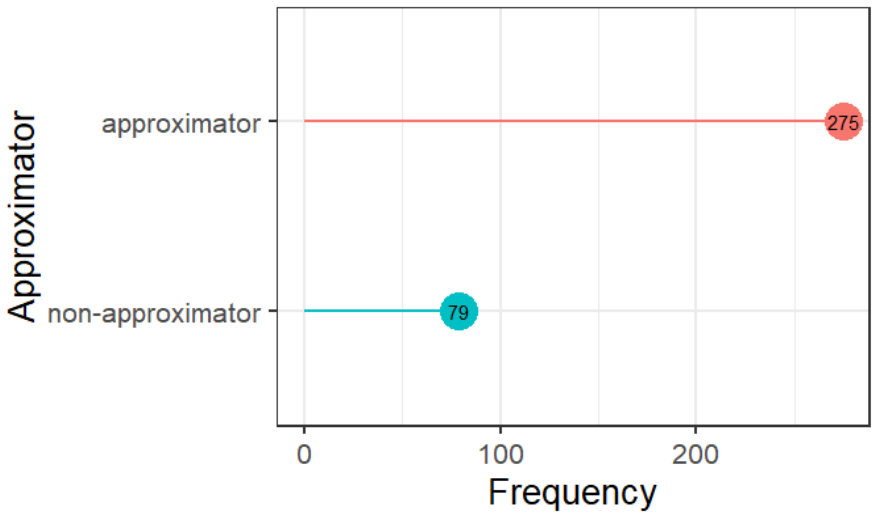

Figure 1. Distribution of responses containing approximators versus other lexical items (n = 354).

RQ2 investigated the range and frequency of APs produced by the participants. Figure 2 shows that several different APs were used, but there were marked preferences for using certain items. Later, in the discussion section, we will address what previous literature about Spanish has found about the use of APs. This information will help us compare the two languages.

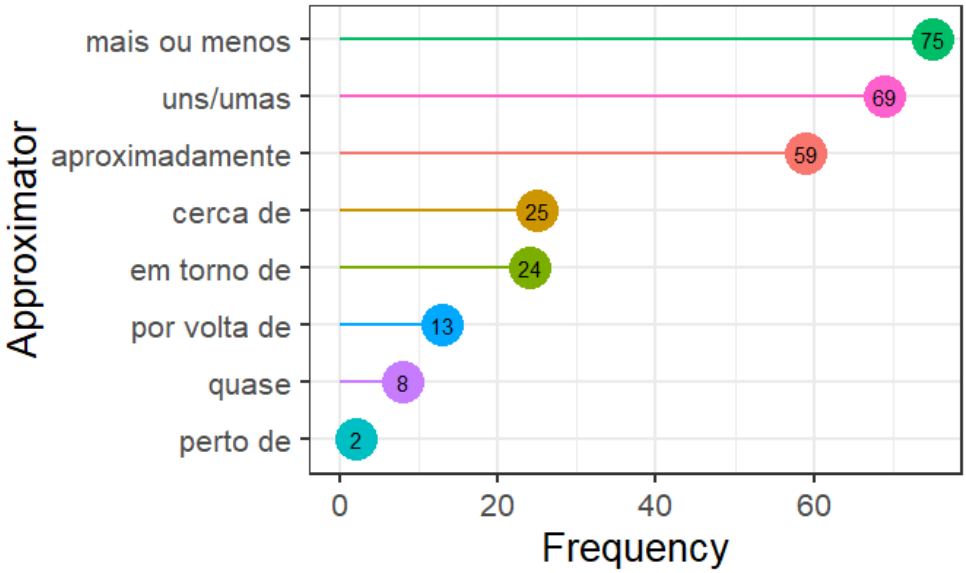

Figure 2. Distribution of approximators produced in responses (*n* = 354).

From Figure 2, we gather that *mais ou menos* 'more or less', uns/umas '*some*', and aproximadamente '*approximately*' were preferred over the other APs. *Cerca de* 'close to' and *em torno de* 'around' were used with similar frequency. Meanwhile, *por volta d*e 'around', *quase* 'almost', and *perto de* 'close to' occurred with the least frequency. No excessive APs (*e poucos* '-ish', *e tantos/as* '-something') were found in the production data.

5.2 Interpretation data

The FCT was administered with the purpose of addressing RQ3 and investigating how Portuguese speakers interpreted the meaning of APs. The distribution of interpretations (i.e., signaling a value that is above or below the exemplar number) of *uns/umas, tipo, quase,* and *cerca de* are presented in Figure 3.

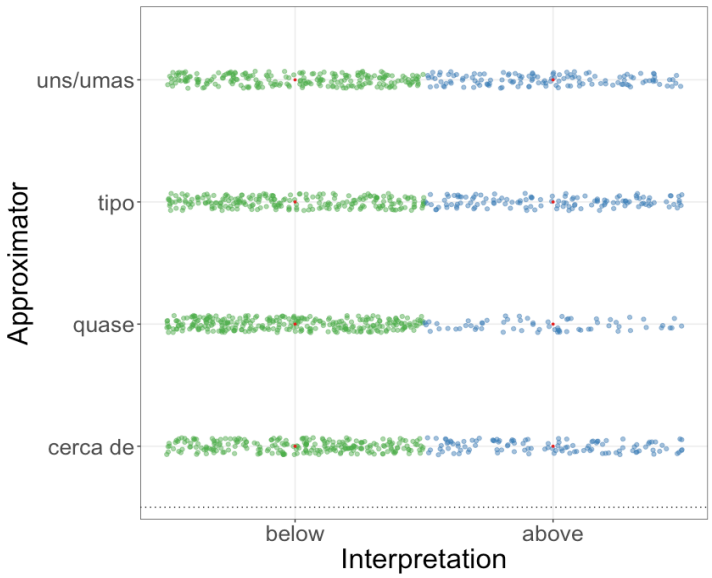

Figure 3. Participants' interpretation of approximators according to whether they signal values that are below or above the exemplar number (*n* = 1,392).

From Figure 3, we gather that the interpretations of *uns/umas, tipo,* and *cerca de*, were symmetrically distributed around the exemplar number. That is, these APs seem to be treated as neutral APs since both interpretations - above and below - were accepted fairly the same number of times. In contrast, *quase* was markedly interpreted as defective, signaling a preference for values below the exemplar number.

In order to determine whether the interpretations of the APs were significantly different, we ran a GLMM, which examined the likelihood of the participants interpreting *quase, cerca de, tipo,* and *uns/umas* as defective APs. The results point out that *quase* was significantly more likely to be interpreted as a defective AP than *cerca de* ($\beta = 1.04$, $SE = 0.21$, $z = 4.89$, $p < .000$), *tipo* ($\beta = 1.30$, $SE = 2.21$, $z = 6.14$, $p < .000$), and *uns/umas* ($\beta = 1.32$, $SE = 0.21$, $z = 6.22$, $p < .000$). No significant differences were found between *cerca de*, *tipo*, and *uns/umas*.

In sum, our data indicated that Portuguese native speakers frequently use APs to express imprecision in numeric expressions, and that the preferred forms were *mais ou menos*, *uns/umas* and *aproximadamente*. When it comes to interpretation, the speakers in this study demonstrated that their understanding of neutral APs is predominantly cohesive among themselves (as expected, given that they are all BP-dominant), by having rated *cerca de*, *tipo*, and *uns/umas* as neutral APs and *quase* as a defective AP.

6. Discussion

This paper focused on the examination of APs in BP. We analyzed which APs were more frequently produced and how they were interpreted. The data we gathered in the elicited written task corroborates our first hypothesis. APs were strongly preferred over other lexical items, such as those that may convey lack of commitment to the content of a proposition: parenthetical verbs (e.g., *acho que* '(I) think that') or adverbs of doubt (e.g., *provavelmente* 'probably'), which are epistemic devices used to express lack of certainty. Indeed, these only came up in 22.31% of the tokens.

We argue that the disfavoring of using adverbs of doubt and parenthetical verbs is not surprising and it has to do with the processing instructions provided via procedural information. In section 2.2, we introduced the terms procedural information and conceptual information (Blakemore 1987; Escandell-Vidal, Leonetti, and Ahern 2011;), and we explained the latter has to do with elements of linguistic structures that map onto concepts. That is the case of adverbs of doubt, which inform the very abstract concept of possibility. On the other hand, APs carry procedural information, which contextualizes and provides instructions for building interpretations. Thereby, saying there were approximately 10 people explicitly instructs the hearer on how to interpret the number: as in something slightly above or below. It automatically also instructs the hearer to exclude other possible meanings, such as "a number largely above or below", "only a number below" and "only a number above". In addition to that, the AP works in a much smaller scope: the noun phrase. By modifying only the number, in lieu of the proposition, APs permit faster and clearer interpretation.

The information provided above accounts for the preference for APs over adverbs of doubt but still leaves us with the partial specifiers (e.g., "at most", "at least", "more than", "less than") found in the data unaccounted for. For those, we posit that partial specifiers also offer procedural information, but they make quantity interpretations even vaguer, since only one point of the scale has a limit. The other point is infinite. Let's take example (9), where the left side of the scale (the minimum number) is set, but the maximum number of people is never limited, providing an infinite number for interpretation. In other words, any number above 20 people would be valid (e.g., 21, 25, or even 50). On the other hand, the AP *around* in (10) limits the range of interpretation. That is, even though it creates fuzzy limits, the values must always be in the vicinity of the exemplar number. For that reason, the number of people that might show up to the party in (10) should

range, for example, between 17 and 23, whereas 50 people might not be an acceptable interpretation for "around 20 people".

(9) At least 20 people will come to the party
(10) Around 20 people will come to the party

Our second RQ inquired about what APs are more commonly used in BP. Given the high presence of Spanish speakers taking Portuguese classes in the US, we also wanted to compare our findings with what has been previously documented in Spanish so we can contribute to educational resources. As predicted by our hypothesis, we found that neutral APs (e.g., *mais ou menos, uns/umas*, and *aproximadamente*) were favored, which is in line with what previous studies found in Spanish (Grasso 2012; Jimenez 2019, 2022; Kern 2012; Mihatsch 2009, 2010; Said-Mohand 2006). The overall preference for neutral APs was expected due to their semantic nature of these lexical items. Inherently - as addressed in the previous paragraph - neutral APs signal values that are in the vicinity of the exemplar number, but they do not set strict semantic boundaries with regard to the direction of the interpretation.

When comparing our findings with the Spanish data in previous studies, we observe that both languages resemble each other in the frequent use of neutral APs, especially Portuguese *uns/umas* and Spanish *unos/unas*. However - as already addressed in the literature review -, *como* 'like' has been documented as the favored form across many Spanish varieties. We did not have any tokens of such lexical item in Portuguese, even though the literature confirms it does scarcely exist (Mihatsch 2009). There is, however, a semantic equivalent of it in Portuguese: *tipo* 'like', which did not appear in our data, despite our knowledge of its frequency in oral BP. We argue that there was a task effect that influenced the register used in our production task. In Portuguese, *tipo* may be restricted to informal registers, similarly to the English AP *like* (e.g., like 20 feet), which usually arises in informal oral conversations (Blythe, Recktenwald, and Wang 1990), mostly among young people (Jørgensen and Stenström 2009). Due to the fact that Brazilians were asked to produce APs in writing, in a linguistics survey, the formality and written register may have activated their own preconceptions of formal written language and skewed their answers to avoid the production of *tipo*.

Let us now consider the third and final RQ, which dealt with the speakers' interpretation of APs and examined whether the BP APs under investigation share the same meaning with their Spanish cognates. In Table 1, we laid out a list of APs in Portuguese divided into excessives, neutral, and defectives. The provided items for Portuguese came from our own intuitions based on the data available from Spanish. Given the lack of existing data for the semantic values of Portuguese APs, we tested the interpretation of such items in the forced-choice task (FCT). Also, since our goal is to contribute to Portuguese learning, and many of these students in the US are also Spanish speakers or learners, we tested the semantic values of the most frequent APs with direct counterparts in Spanish.

For our third RQ, we hypothesized that *tipo* and *uns/umas* would be interpreted as neutrals, while *cerca de* and *quase* would be assigned a defective interpretation. The data from the FCT partially corroborates our hypothesis. As expected, participants did interpret *tipo* and *uns/umas* as neutral APs (i.e., denoting a number slightly below or above the exemplar number), and *quase* as defective (i.e., denoting only values below the exemplar number). Once again, the choice of these particular lexical items was based on our intuition and the data from Spanish. Therefore, we were surprised to find variation among these languages, when comparing our results to Jimenez's (2018) Spanish study. Our Portuguese data showed that *cerca de* was interpreted as neutral while, in Spanish, this AP is considered a defective AP.

We take these findings as evidence that, while most interpretations of the APs under investigation are equivalent in both languages, regardless of morphological similarity, the semantic meaning of APs is not always shared across languages and must not be taken for granted in the

process of language acquisition. This is important because it has been attested that students enrolled in Portuguese for Spanish speakers courses often suffer from fossilization in their interlanguage (Simões, Carvalho, & Wiedemann 2011). Even though their listening and reading skills can achieve high levels very early on, Spanish speakers, whether native or sequential, have difficulties realizing subtle differences between these two closely related languages. Because mutual comprehension takes place in the first semester or even without any formal instruction, learners may not apply linguistic awareness when acquiring Portuguese. In other words, what can be perceived as the same is transferred to their Portuguese interlanguage. This directly affects the acquisition of APs, given that the morphological similarities between the APs in both languages would lead learners to ignore differences between Spanish and Portuguese. Our data has demonstrated that this is not the case; thus, language curricula need to be attentive to these fine divergences.

We suggest that educators incoporate vague language early on. This can be done by simply modifying some activities. For instance, students are often instructed on how to express their routine in their first semester. If they are provided with common APs, they can re-adjust the task to not only say at what time they go to bed or wake up, but also to be able to express at *around* what time they do those things. For task-based curricula, intructors may experiment some role-plays in which they are specifically asked to be vague. For example, students may be asked by their ficticious supervisor to plan the company anniversary party and send them estimates of *approximately* how many people are going to be invited and how many boxes of pizza are going to be ordered. Likewise, when practicing narrating in the past, instructors may ask the class to re-tell a story that happened *some* years ago. Thus, learners would not be able to express details and naturally would need to use approximators and vague language.

7. Conclusion

This study examined the production and interpretation of APs in Brazilian Portuguese (BP). We found that the use of vague language in the form of APs is a very common strategy that speakers use to express numeric imprecision, surpassing the use of epistemic lexical items that convey uncertainty or doubt. We also concluded that APs that are the same lexical item or a cognate in a different language may have different semantic meanings, such is the case of *cerca de* in Spanish and Portuguese. This difference has direct implications for second language learning since some items are not readily available for transfer and need reconfiguration of meaning to form. Furthermore, our findings also have pedagogical implications for language teaching, in particular, Portuguese classes for Spanish speakers. Educational resources of Portuguese as an Additional Language (in Portuguese, PLA - *Português como Língua Adicional*) do not offer explicit instruction on how to express numeric imprecision. The data from the WEPT demonstrate the most common APs in written language and can be used to help learners incorporate these lexical items in their language production.

We would also like to acknowledge the limitations of this study. This paper has not examined how some socioeconomic variables may influence the production and representation of APs. It is important that variationist approaches observe how level of education, age, among other factors, can shed light on generational differences or preferences for distinct levels of formality. Finally, Milleret (2012) estimates that 45% of US Portuguese classes are composed of Spanish speakers, who are also English-Spanish bilinguals learning Portuguese as a third language. Therefore, studies on the acquisition of Portuguese APs would contribute to Portuguese programs in the US and in pedagogies that can enhance acquisition, taking in consideration a multilingual and multicultural audience that requires specific educational approaches.

References

Amaral, Patricia. 2010. "Almost means 'less than': preschoolers' comprehension of scalar adverbs". *On-line Proceedings Supplement of BUCLD*, 34.

American Council on the Teaching of Foreign Languages. "The ACTFL Proficiency Guidelines". 2012. Retrieved June 9, 2022, from https://www.actfl.org/resources/actfl-proficiency-guidelines-2012/english/speaking

Bates, Mächler D., Bolker B., Walker S. 2015. "Fitting linear mixed-effects models using lme4, *Journal of Statistical Software* 67, no. 1: 1-48.

Blakemore, Diane. 1987. *Semantic Constraints on Relevance.* Oxford: Blackwell.

Blakemore, Diane. 2008. "Apposition and affective communication". *Language and Literature* 17, no.1: 37–57. doi: 10.1177/0963947007085054.

Blythe, Carl, Sigrid Recktenwald, and Jenny Wang. 1990. "I'm like," say what?!": A new quotative in American oral narrative". *American Speech* 65, no. 3: 215-227.

Cabrelli-Amaro, J., João Felipe Amaro, and Jason Rothman. 2015. "The relationship between L3 transfer and structural similarity across development: Raising across an experiencer in Brazilian Portuguese". *Transfer effects in multilingual language development* 4: 21-52.

Camacho, José. "Evaluatives, focus and syntactic computation". Unpublished Manuscript, 2011. Department of Spanish and Portuguese, Rutgers University, New Jersey, USA.

Champaud, Christian, and Dominique Bassano. 1987. "Argumentative and Informative Functions of French Intensity Modifiers "presque" (almost), "à peine" (just, barely) and "à peu près" (about): An Experimental Study of Children and Adults". *Cahiers de Psychologie Cognitive/European Bulletin of Cognitive Psychology 7, no. 6*: 605-631.

Channell, Joanne. 1994. *Vague Language.* Oxford: Oxford University Press.

Cheng, Winnie, and Martin Warren. 2003. "Indirectness, inexplicitness and vagueness made clearer." *Pragmatics* 13 no. 3: 381-400. doi: 10.1075/prag.13.3.02che.

Critical Language Scholarship (n.d.). *About the Program.* Retrieved June 9, 2022, from https://clscholarship.org/about

Cutting, Joan. "Introduction to vague language explored". Vague Language Explored, edited by Joan Cutting, 3-20. New York: Macmillan, 2007. https://doi.org/10.1057/9780230627420

Davies, Mark, and Michael Ferreira. 2006. Corpus do Português: 45 million words, 1300s-1900s. *Georgetown: Universidade de Georgetown.*

De Cock, Sylvie, Sylviane Granger, Geoffrey Leech, and Tony McEnery. "An automated approach to the phrasicon of EFL learners". In Learner English on computer, edited by Sylvie De Cock*,* 67-79. London: Longman, 1998.

Drave, Neil. "Vaguely speaking: a corpus approach to vague language in intercultural conversations." In New Frontiers of Corpus Research. Papers from the Twenty First International Conference on English Language Research on Computerized Corpora edited by Pam Peters, Peter Collins, and Adam S. Cohen, 25-40. Amsterdam/New York: Rodopi, 2002.

Escandell-Vidal, Victoria, L Manuel Leonetti, and Aoife Ahern, eds. 2011. *Procedural meaning: Problems and perspectives.* Bingley: Emerald Group Publishing.

Faerch, Claus, and Gabriele Kasper. 1987. "Perspective on language transfer". *Applied Linguistics* 8: 111-136.

Fuentes Rodríguez, Catalina. 2008. "La aproximación enunciativa". *Lingüística Española Actual* 30, no. 2: 223-258.

García-Medall, Joaquín. 1993. "About almost and other approximators." *Dicienda: Cuadernos de Filología Hispánica* 11:153-170.

Gibbs, Raymond W., and Gregory A. Bryant. 2008. "Striving for optimal relevance when answering Questions." *Cognition* 106, no.1: 345-369. doi: 10.1016/j.cognition.2007.02.008.

González Rodríguez, Raquel. 2008. "The Positive Polarity of Spanish." Unpublished

doctoral dissertation. Department of Philology, Universidad Complutense de Madrid. Retrieved from http://eprints.ucm.es/8145/1/T30427.pdf.

Grasso, Marina. 2012. "El lenguaje vago y su potencial comunicativo". *IV Jornadas Internas de Español como Lengua Segunda y Extranjera: Experiencias, Desarrollos, Propuestas.* Facultad de Humanidades y Ciencias de la Educación Universidad Nacional de La Plata, Argentina, 53-63.

Holmlander, Disa. "Strategies of Attenuation in L1 and L2 Spanish: A Contrastive Study in Spanish and Sweedish Speakers." PhD diss., (Lunds University, 2011)

Jimenez, Abril. "What to say when you are not sure? A developmental perspective on the acquisition of numeric approximators among Spanish L2 learners. (Conference presentation, 2018). Hispanic Linguistics Symposium, October 25th, The University of Texas, Austin.

Jimenez, Abril. "Approximating quantities in Spanish: A study on the production of numeric approximators among second language learners". (Conference presentation, 2017) Hispanic Linguistics Symposium, October 28th, Texas Tech University, Lubbock, TX.

Jimenez, Abril. 2022. "The Effects of L2 Spanish Proficiency and Length of Exposure Abroad in the Expression of Imprecise Quantities". *Languages* 7, no.1: 20-42. https://doi.org/10.3390/languages7010020

Jimenez, Abril. "Lexical and Pragmatic Manifestations of Uncertainty: An Examination of Numeric Approximators in L2 Spanish." PhD diss., (Rutgers University, 2019).

Jørgensen, Annette Myre, and Anna-Brita Stenström. 2009. "Dos marcadores pragmáticos contrastados en el lenguaje juvenil: el inglés like y el español como". *Español Actual* 92: 103-122.

Jucker, Andreas H., Sara W. Smith, and Tanja Lüdge. 2003. "Interactive aspects of vagueness inconversation." *Journal of Pragmatics* 25: 1737-1769. doi: 10.1016/S0378-2166(02)00188-1.

Kennedy, Graeme D. 1987. "Quantification and the use of English: A case study of one aspect of the learner's task". *Applied Linguistics* 8, no. 3: 265-286.

Kern, Joseph. "'Como' in Commute: The Travels of a Discourse Marker Across Languages." MA Thesis, (University of Arizona, 2012).

Lasersohn, Peter. 1999. "Pragmatic halos." *Language* 75, no. 3: 522–551. doi:10.2307/417059.

Metsä-Ketelä, Maria. 2006. "Words are more or less superfluous. The case of more or less in academic Lingua Franca English". *Nordic Journal of English Studies* 5, no. 2: 117-143.

Mihatsch, Wiltrud. "The approximators French *comme*, Italian *come*, Portuguese *como* and Spanish *como* from a grammaticalization perspective". In *Grammaticalization and Pragmatics: Facts, Approaches, Theoretical Issues*, edited by Corinne Rossari, Claudia Ricci, and Adriana Spiridon, 65-92. Bingley, UK: Emerald Publications, 2009. https://*doi*.org/10.1163/9789004253193_006

Milleret, Margo. 2012. "Portuguese Study in Higher Education in the US." *Hispania* 95.1: 135–50.

Mkhitaryan, Yelena Saribek, and Zhanna Yegor Ayvazya. "A Study on Expression of Numerical Quantities in English and Armenian". In *The first European conference on the achievements and development of young scientists in the Humanities and Social Sciences*, pp. 59-68. 2017.

Peires, Mary-Louise. 1997. "Come about five-ish: Vagueness in language". *Southern African Journal of Applied Language Studies* 5: 56-69 https://doi.org/10.1080/10189203.1997.9724667

Peirce, Charles S. 1902. "Vague." In *Dictionary of Philosophy and Psychology, Vol. II*, edited by James Mark Baldwin, 321-323. London: Macmillan and Co.

Prince, Ellen F., Joel Frader, and Charles Bosk. 1982. "On hedging in physician-physician discourse." *Linguistics and the Professions* 8, no. 1: 83-97.

Prins, Ronald, and Roelien Bastiaanse. 2004. "Analyzing the spontaneous speech of aphasic speakers." *Aphasiology* 18, no.12: 1075-1091. doi: 10.1080/02687030444000534.

Ruzaitė, Jūratė. 2004. "Academic Precision Reconsidered: A Corpus-Based Account." *Journal of Linguistics* 17: 217-247.

Sabet, Peyman GP, and Grace Q. Zhang, eds. 2015. *Communicating through Vague Language: A comparative study of L1 and L2 speakers*. UK: Palgrave McMillan.

Said-Mohand, Aixa. "Sociolinguistic study of the markers like, so, and you know in the speech of bilingual Americans."PhD Diss, (University of Florida, 2006).

Sauerland, Uli, and Penka Stateva. 2011. "Two Types of Vagueness." In: Vagueness and Language Use. Palgrave Studies in Pragmatics, Language and Cognition, edited by Paul Égré and Nathan Klinedinst, 121-145. Palgrave Macmillan: London. doi: 10.1057/9780230299313_6.

Selinker, Larry. 1972. "Interlanguage". *IRAL* 10: 209-31.

Simões, A. R. M., Carvalho, A. M., and L. Wiedeniann. 2011. "Português para falantes de Espanhol. Portuguese for Spanish speakers". Campinas, Sao Paulo, Brazil: Pontes Editores.

Ullmann, Stephen. 1962. *Semantics: An Introduction to the Science of Meaning*. Oxford: Blackwell.

Wierzbicka, Anna. 1986. "Precision in vagueness: The semantics of English 'approximatives"'. *Journal of Pragmatics* 10: 597-613. https://doi.org/10.1016/0378-2166(86)90016-0

Yu, Shengming. 2009. "The pragmatic development of hedging in EFL learners". PhD Diss. (City University of Hong Kong).