doi:10.5128/ERYa17.13

EESTI RAKENDUSLINGVISTIKA ÜHINGU AASTARAAMAT 17, 231–248

VAGUE LANGUAGE IN ARGUMENTATIVE ESSAYS WRITTEN BY ADVANCED LITHUANIAN LEARNERS OF ENGLISH

Jūratė Ruzaitė

Abstract. This paper examines the use of vague language (VL) in argumentative essays written by advanced Lithuanian learners of English. The study focuses on two main categories of VL: general extenders (GEs), e.g. *and so on, etc./etcetera,* and *or so*, and vague quantifiers, e.g. *some, many, a lot of*, and *a little*. The present research, first of all, aims to assess how extensively Lithuanian learners of English use VL in comparison to native speakers. Second, quantitative corpus data is interpreted in view of the main functions that VL performs in non-native speakers' essays to explain why there are differences in the frequency of VL. Finally, the frequency of VL is interpreted with regard to the formality of the items under investigation. The findings indicate that there are some major differences in the frequency of VL in English L1 and L2, both VL categories being more frequent in English L2.

Keywords: frequency, pragmatic functions, (in)formality, argumentative essays, English, learner language, native speakers

1. Introduction

The present study aims to account for the use of vague language (VL) in argumentative essays written by advanced Lithuanian learners of English. The study accounts for two main categories of VL: general extenders (GEs), e.g. *and so on, etc. /etcetera,* and *or so*, and vague quantifiers, e.g. *some, many, a lot of*, and *a little*, the use of which in Lithuanian learner English (or non-native speakers' language, NNS) is compared to that in written language of students who are native speakers of English (L1, or NS).

Though some authors use alternative terms for quantifiers or their subcategories (e.g. Quirk et al. 1985: 598, Lin 2013, Channell 1994), in this paper the term 'quantifiers' is used consistently to refer to different types of non-numerical

expressions used to refer to quantities without a numeral. Quantifiers are used to refer to 1) a quantity of a noun, which is the most prototypical use of quantifiers, and 2) the intensity of a verb, adverb or adjective (a less prototypical use) (Ruzaitė 2007: 42). That is, quantifiers in this study are analysed when they modify nouns, verbs, adverbs, and adjectives, especially the comparative degree of the latter two categories.

Based on their meaning, two categories of quantifiers are distinguished and considered in this paper: multal quantifiers referring to large quantities, e.g. *many* and *much*, and paucal quantifiers referring to small quantities, e.g. *a few* and *a little* (cf. assertive multal pronouns and assertive paucal pronouns in Quirk et al. (1985)). The distinction between multal and paucal quantifiers is important in the current paper since they differ in their communicative function.

GEs are highly formulaic tags that start with the conjunction *or* or *and*, and follow an item that can be a noun phrase, verb phrase, embedded sentence, or prepositional phrase (Channell 1994). Depending on which conjunction is used, two types of GEs are distinguished: adjunctive GEs are formed with the conjunction *and*, and disjunctive GEs appear with *or*. The conjunction is usually followed by the pronouns *something* or *anything*, or the noun *things*.

So far, research on Lithuanian learner English has mostly focused on linking words, writer positioning, hedging, and multi-word clusters, and VL has not been addressed yet. However, a growing body of research on VL (e.g. Warren 1993, Gao, Zhu 2005, Martínez 2011) shows that it constitutes an important part of pragmatic language competence and thus should be addressed in language teaching in a systematic way. As Gao and Zhu (2005: 41) note, L2 learning is strongly hindered by general intolerance of ambiguity and overestimation of explicitness (cf. Ruzaitė 2007, Overstreet 2012). At the end of the 20th century, researchers started stressing the importance of VL in teaching curricula (e.g. Warren 1993), which had considerable impact on later developments in EFL teaching practices and materials, but more focus is still needed on a larger variety of EFL learners, especially those who use lesser used languages such as Lithuanian.

To study VL in Lithuanian learner English, the present analysis addresses three research questions:

1. How extensively do Lithuanian learners of English use VL in comparison to native speakers?
2. What is the frequency of different VL categories?
3. What are the main functions that VL performs in non-native speakers' essays? How can these functions explain the differences in the frequency of VL? How does the frequency of VL in the learner corpus relate to the formality of the items under investigation?

Thus, the present research, first of all, aims to assess how extensively Lithuanian learners of English use VL and its different categories in comparison to native speakers. Its second goal is to explain why there are differences in the frequency of VL by interpreting quantitative corpus data in view of the main functions that VL performs in non-native speakers' essays. The frequency of VL is also interpreted regarding the formality of the items under investigation.

As can be seen in the RQs above, in this paper there is a strong focus on overused/underused forms. Such a focus is characteristic of learner corpus research.

Since interlanguage (or learner language) includes both correct and deviant forms, learner corpus research aims not only to identify generic L2 errors as opposed to L1-influenced errors or identify lexical errors to help in compilation of learners' dictionaries, but it also examines over- or underuse of some constructions or word(group)s (Granger 1998, 2004, Granger et al. 2002).

As Granger observes, advanced interlanguage (which is under focus in this paper) is the result of a very intricate interrelation of different factors: "developmental, teaching-induced and transfer-related, some shared by several learner populations, others more specific" (2004: 135). In the present study, the learner population is rather homogeneous: it includes undergraduate students of English Philology of a similar age, educational background, and the same L1 background. Therefore, this population is expected to share similar transfer-related uses. This research, however, cannot account for teaching-induced uses since the students' exposure to different types of teaching and reference materials was not considered.

One of the most important aspects of learner corpus research is the possibility to examine learners' language use against the background of correct language (Granger 1998). Hence, in this paper, learner data (a dataset of argumentative student essays totalling 244,746 words) is compared to native speaker language by resorting to the sub-corpus of texts for the discipline "English" in the British Academic Written English corpus (BAWE-English; 458,780 words in total).

To contextualise the current analysis, the following section overviews the main trends in prior research on VL in (learner) English. More information on the data and methods of analysis is provided in Section 3. Section 4 reports the main results of this study by focusing on the frequency and functions of different VL categories in Lithuanian learner English, as contrasted to native speakers' essays. The analysis is followed by a discussion interpreting the current results against the backdrop of previous research, and finally the study offers some concluding remarks.

2. VL in (learner) English

In research on VL, the focus typically lies on spoken language, since VL is considered to be more typical of spoken interaction than written discourse (cf. Biber et al. 1999, Sabet, Zhang 2015, Ruzaitė 2018). As a highly representative study based on the International Corpus of English (ICE-GB) and the British National Corpus (BNC) by Martínez (2011) shows, in English L1 the majority of GEs appear more frequently in spoken data (e.g. *and stuff, and that*, *and everything*, *and things* and *and all*). The frequency of GEs in speech is almost a hundred times higher than in writing (ibid.). When used in writing, they mainly occur in fiction or in informal writing, such as emails (ibid.).

VL, as a typical feature of spoken interaction, is often associated with informality (e.g. Overstreet 1999, Stenström et al. 2002, Cheshire 2007, Ruzaitė 2018). VL, however, does not form a homogeneous category in that respect: there appears to be a continuum of the degree of formality within the category of VL, and some VL items are more associated with formality, whereas others are less formal. For instance, most of the GEs and some quantifiers (e.g. *a lot of, a bit, a little,* and *a little bit*) are predominantly informal (Biber et al. 1999). However, the GE *and so*

on is more frequent in writing and is more closely associated with more formal (academic) style (cf. Biber et al. 1999: 116). In Overstreet and Yule's study (1997), *and so on* and *et cetera* appear mainly in formal discourse, whereas *and everything*, *and stuff*, *or something*, *or anything,* and *or whatever* are used in informal speech. Similarly, Stubbe and Holmes (1995) report that *and so on* and *or so* dominate in formal interviews.

In learner language, VL has also been studied mainly in spoken interaction, e.g. English spoken by Dutch speakers (Buysse 2014), Iranian EFL learners' speech as contrasted to mother-tongue Persian (Parvaresh et al. 2012), and Chinese and Persian speakers of English as contrasted to American native speakers (Sabet, Zhang 2015). Gassner (2012) accounts for the VL item *thing* in English L1 and L2 discourse in Australian job interviews and argues that, in fact, *thing* hardly ever introduces vagueness into discourse (2012: 26).

VL has been studied to some extent in non-native speakers' English, but not all the VL categories have been examined to the same degree. There is too little evidence on the frequency and usage patterns of vague quantifiers, and the VL category that received most attention in learner language studies is that of GEs. GEs have been researched in English spoken by Dutch speakers (Buysse 2014) and Iranian EFL learners' speech as contrasted to mother-tongue Persian (Parvaresh et al. 2012). In spoken communication, as such research shows, L2 learners tend to use fewer GEs than native speakers (Aijmer 2004, de Cock 2004, Drave 2002, Fernández, Yuldashev 2011, Parvaresh et al. 2012, Lin 2013). It has also been observed that L2 learners tend to use GEs more often after a period abroad compared to at-home learners (Grieve 2010). Interestingly, Metsä-Ketelä's (2012) research shows that in academic settings lingua franca speakers use VL almost twice as frequently as native speakers, but the range of VL expressions is narrower in the lingua franca data.

Some of the research on VL in spoken interaction focuses on intercultural communication between native and non-native speakers of English (e.g. Cheng, Warren 2001, Drave 2002, Lin 2013). It is difficult to specify straightforward trends in such studies, since there are some contradictory results, but one trend is that native speakers of English (NSE) use more vague language than non-native speakers (NNSE). For instance, research on intercultural communication among British and Taiwanese adolescents by Lin (2013) points to the underuse of VL among NNS: in spoken interaction all the categories under investigation were more frequent among NS participants.

Drave (2002), who studied VL in intercultural conversations between NSE and native speakers of Cantonese (NSC), also revealed that NSE use more VL than NNSE, but "the range of different types was similar for the two groups of speakers, as were major collocations. However, there were important functional differences, notably the greater NSE exploitation of vague language for affective and interpersonal purposes" (Drave 2002: 25). A very similar investigation carried out by Cheng and Warren (2001) reveals some contrary results: both native English speakers and non-native speakers of English in Hong Kong use vague language to a similar extent and for similar purposes.

A cross-linguistic and cross-cultural perspective to VL is important, since, as the research of Pervaresh et al. (2012: 277) convincingly shows, the use of GEs in EFL depends, at least to some extent, on first language norms. Their study of GEs

in native Persian and non-native English communication shows that GEs in Persian English appear in a slightly different structural pattern, in a different syntactic position, and perform different functions. For example, intensification was not common when soliciting agreement in either Persian NS or Persian EFL, and, differently from native English, GEs were used for expressing outrage and arousing curiosity.

As Lin's study on Taiwanese EFL textbooks used in junior high schools (2012) alerts, some textbooks do fail in exposing language learners to authentic use of conversational English. The categories accounted for in Lin's study include some syntactic, discourse, and lexical features. The last category is mainly represented by VL categories, including vague expressions, approximations, and hedging. His research on spoken grammar points to a gap between textbook conversations and actual intercultural communication, and teachers' role is to bridge this gap to make learners better communicators.

The trends mentioned above have mainly been observed on the basis of spoken data; however, NS and NNS differences in writing still lack researchers' attention. The typical association of VL with spoken grammar naturally has led to more extensive research on spoken interaction. As the research of Martínez (2011) and Biber et al. (1999) shows, VL is not a homogeneous category, and some VL items are more formal and appear more frequently in writing, whereas others are less formal and appear in speech. Thus, it is important to see how much language learners are aware of these differences in formality and to what extent they are capable of adhering to the requirements of the register.

3. Data and methods

In learner language research, learner corpora have become a conventional empirical resource (Granger 2004), and learner corpus research has developed into a well-established branch since its emergence in the late 1980s (for a detailed overview of learner corpus research, see Granger et al. 2015). The present study is also based on corpus data, and as such, it uses a written corpus representing Lithuanian learner English, which was developed using the methodology of the International Corpus of Learner English (ICLE). The essays were written on topics chosen from a list of 14 topics suitable for argumentative essays. The topics were formulated to elicit argumentative style and required the students to discuss some debatable and socially relevant issues, such as crime, pollution, censorship, and feminism.

The dataset consists of 419 argumentative essays (or 244,746 words); the average number of words per essay is 584. The texts were written by Lithuanian undergraduate students majoring in English Philology. They had the experience of 7–11 years of learning English at school and 2–4 years of studying English at university. Since the majority of philology students are female, the sample includes mainly female speakers. This imbalance may have an effect on the results; however, the variable of gender is not considered in the current research and the possible correlations between the use of VL and speaker's gender are to be tested in future research. The essays were mainly written by students in their own time (under untimed conditions), and the length was controlled by requiring the minimum of 500 words per essay.

Learner data is compared to native speaker language by resorting to the sub-corpus of texts for the discipline "English" in the British Academic Written English corpus (BAWE-English, accessed through Sketch Engine)[1]. BAWE-English consists of 106 essays, or 458,780 words. The average number of words per essay is 4,328 tokens, which is considerably higher than that of the Lithuanian learner corpus. The types of NS texts are also more varied than those in the learner corpus as they include not only essays but a large variety of other academic genres, such as case studies, methodology recounts, critiques, design specifications, and proposals. This may have an effect on the comparability of the corpora and will be taken into account when interpreting the quantitative results.

The BAWE-English sub-corpus was used as a reference corpus for comparative NS and NNS analysis aiming to establish which VL usage patterns could be potentially indicative of NNS. The data obtained from BAWE-English was mainly used to examine the frequency (overall number of tokens) of different VL items in NS essays and to identify items overused and underused in NNS. Since the two corpora are of different sizes, all the frequencies were normalised per 10,000 words to make the data comparable.

To process the data in the corpus of Lithuanian learner English, the AntConc software was applied to obtain the frequency of each VL item and the distribution of these items across individual texts. To examine the frequency of VL items in NS, the Sketch Engine tools were applied to search the sub-corpus BAWE-English and to identify word frequencies. The overall frequency is important in showing how extensively Lithuanian learners of English use VL in comparison to NS. The distribution of these items across different learner texts indicates how widely VL items are distributed in the learner corpus across different texts and helps to determine individual differences between speakers.

In addition to the analysis of word frequencies, this study also examined the main functions that VL performs in NNS essays. The functions were determined manually by analysing the concordance lines generated with AntConc and by paying special attention to the collocational patterns of VL items and other contextual information. It is impossible to specify clear-cut functions without knowing the real intentions of the speaker. In addition, VL categories often perform more than one function in the same utterance. Therefore, the functions determined here are tentative, and the ones discussed in the analysis are seen as the primary functions a VL category performs, but some secondary interpretations are also possible. The framework used in the analysis of VL language functions mainly resorts to the prior research of Channell (1994), Ruzaitė (2007), and Sabet and Zhang (2015).

The study thus approaches the data both quantitatively (especially with regard to RQs 1 and 2) and qualitatively (especially with regard to RQ 3). The subsequent sections now turn to the discussion of each of the research questions presented in the introduction.

[1] https://app.sketchengine.eu/#open (16.1.2018).

4. Results

This section reports on the overall frequency of VL in the NS and NNS data, the frequency per VL category in both corpora, and the functions of VL in NNSE. The qualitative analysis of VL functions is expected to explain the frequency trends observed in the data.

4.1. Overall frequency in NS and NNS

The findings indicate that there are some major differences in the frequency of VL in NNSE and NSE, with both VL categories being more frequent in NNSE (see Table 1). The overall frequency of VL in NNSE is 84.25 tokens, as opposed to only 18.93 tokens in NSE (here and elsewhere the frequency is provided per 10,000 words if not stated otherwise).

Table 1. Overall frequency in NNSE and NSE

Feature	NNSE Raw freq.	f/10,000	NS Raw freq.	f/10,000
GEs	262	10.7	24	0.87
Quantifiers	1,820	74.36	498	18.06
Total	2,082	84.25	522	18.93

Thus, the results suggest a strong overuse of VL in NNSE, which could be indicative of learners' inclination to use conversational style in writing due to some lack of awareness of register variation.

The most noticeable differences between NSE and NNSE in the present study appear in the use of vague quantifiers. The overall frequency of quantifiers in NNSE is four times as high as that in NSE (74.36 occurrences vs. 18.06 occurrences respectively). The overall frequency of GEs in both databases is considerably lower than that of vague quantifiers, but in NNSE they are still significantly more frequent than in NSE (10.7 occurrences vs. a bit less than 1 occurrence respectively). This trend again seems to support the earlier-mentioned assumption that even advanced learners lack competence in register variation and awareness of some important differences between conversational style and formal writing.

To further account for the VL frequency per category, a more detailed quantitative analysis of GEs and vague quantifiers follows in the next section.

4.2. Frequency per VL category

The quantitative analysis has revealed that the most frequent VL items are **vague quantifiers**, which are more frequent than GEs in both NSE and NNSE but are overused in NNSE relative to NSE. All the quantifiers under investigation are more frequent in the NNS corpus, and *some* is the most frequent item of all of them (27.54 occurrences; see Table 2).

The data points to a relatively high frequency of informal items in NNSE. Four out of nine quantifiers are associated with informal use in general English (see the overview in Section 2): *a lot of* (7.89 tokens), *a little* (0.61 token), *a bit* (0.41 token), and *a little bit* (0.25 token), which make up 9.16 tokens in total. In NSE, the GEs *a bit* and *a little bit* do not appear at all; *a little* is the only item that appears more frequently than in NNSE; and *a lot of* appears with a considerably lower frequency.

Table 2. Frequency per category (ranked by frequency in NNSE)

Quantifiers	NNS		NS	
	Raw freq.	f/10,000	Raw freq.	f/10,000
some	674	27.54	97	3.52
many	546	22.31	214	7.76
much	269	10.99	150	5.44
a lot of	193	7.89	5	0.18
majority	103	4.21	12	0.44
a little	15	0.61	20	0.73
a bit	10	0.41	0	0
a little bit	6	0.25	0	0
minority	4	0.16	0	0
Total	1,820	74.36	498	18.06

Interestingly, four out of the top five quantifiers in NNSE are those referring to large quantities: *much, many, a lot of,* and *majority*. In contrast, mitigating quantifiers, which refer to small quantities, e.g. *a bit* and *a little (bit)*, are very rare. The only quantifier that refers to a small number and is highly frequent is *some* (27.54 tokens). These two types of quantifiers are associated with different pragmatic functions, which are under focus in the next subsection. The functions of GEs and quantifiers are expected to explain, at least partly, some frequency trends in NNSE.

The analysis of **GEs** has demonstrated that their frequency depends on whether they are disjunctive or adjunctive. The results show that disjunctive GEs are almost never used either in NNSE or NSE, expect for a single occurrence of *or so* in both NSE and NNSE (see Table 3). This finding is not unexpected since earlier studies on GEs also point to a generally higher incidence of adjunctive GEs, especially in more formal communication.

The frequency of GEs presented in Table 3 also shows that not only do GEs in NNSE outweigh those in NSE, but also there is considerable individual variation within this category, some items being far more frequent than others. The most frequent GE in NNSE is *etc.* (6.41 tokens, as opposed to 0.11 tokens in NSE). The second most frequent GE is *and other* (+ noun), which appears 2.49 times. This GE is the most frequent one in NSE, but its frequency is still much lower than that in NNSE.

All the other GEs appear with much lower frequency in both NSE and NNSE, and several of them can be seen as variations of the two most dominant forms. In NNSE, the GE *etc.* has the alternative full form *etcetera*, which does not appear in NSE at all. The GE *and other* (+ noun) has alternative forms with the plural form of *other* and/or an intervening quantifier (*some* or *many*).

Table 3. Frequency per category (ranked by frequency in NNS)

GEs	NNS		NS	
	Raw freq.	f/10,000	Raw freq.	f/10,000
etc.	157	6.41	3	0.11
*etcetera**	1	0.04	0	0
and other (+ noun)	61	2.49	10	0.36
and others	17	0.69	3	0.11
and many other	7	0.29	3	0.11
and many others	4	0.16	1	0.04
and some other	2	0.08	0	0
and so on	12	0.49	3	0.11
or so	1	0.04	1	0.04
Total	262	10.7	24	0.87

* The alternative spelling form *et cetera* did not occur in either corpus

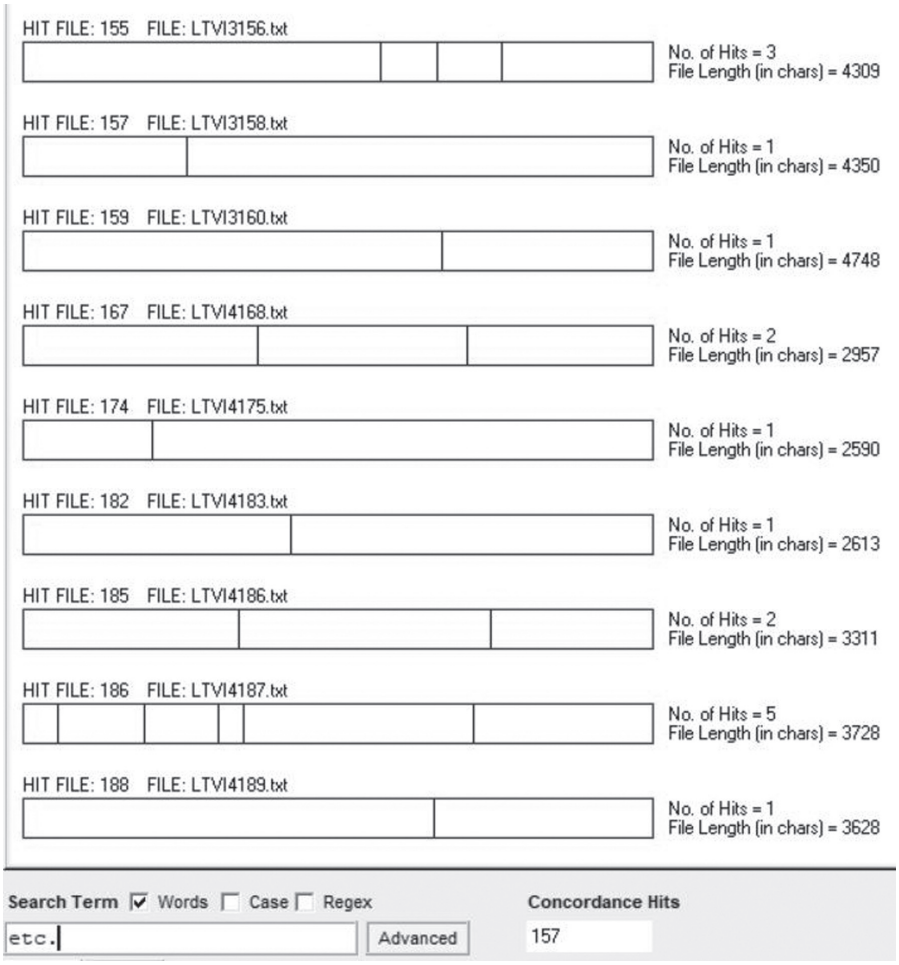

Figure 1. Distribution of etc. across NNS texts in the Lithuanian learner corpus

To find out how widely GEs are distributed in the learner corpus across different texts and to ascertain that they are not a feature of an idiolect, the concordance plot of *etc.* was examined. Only the distribution of *etc.* was analysed since it is the most frequent and thus the most representative GE in NNSE. Figure 1 presents a fragment from the concordance plot, which displays the varied distribution of *etc.* across NNS texts in the Lithuanian learner corpus.

As the data shows, the range of the frequency of *etc.* in different texts is not very large: its raw frequency ranges from 1 to 5 occurrences per text. Such relatively even distribution of *etc.* in student essays indicates that the high frequency of this GE cannot be explained by strong individual preferences for it, but rather it is a common lexical item in the current learner data.

In brief, the quantitative analysis of VL has demonstrated that some VL categories and individual forms vary in their frequency, but both the overall frequency of VL and the frequency of different VL categories are higher in NNSE than NSE. One possible interpretation, as has already been suggested, is that language learners lack awareness of register variation: VL is a dominant feature of conversational English, and its overuse in writing is stylistically marked. In addition, these quantitative results can be interpreted not only in view of register but also some pragmatic aspects. The next section thus turns to the discussion of the main functions of VL in learner data to further explain some learner (mis)uses of VL.

4.3. Functions

As has been observed through manual analysis of the concordance lines, in NNSE **vague quantifiers** serve three main functions: they a) indicate the scope of generalisations (also when they are part of a GE) and thus can perform the function of hedging, b) give contradictory viewpoints, c) present quotes and references to previous studies, and d) are used for emphasis.

When vague quantifiers are used to indicate the scope of generalisations, they restrict the categoricity of a claim to a certain extent depending on the meaning of the quantifier. Multal quantifiers (which indicate large quantities), refer to a broader scope (as in examples (1) and (2)), whereas paucal quantifiers (which indicate small quantities) narrow down the extent of a generalisation (as in example (3)). In such instances, vague quantifiers may appear in GEs, as in example (2), where the GE is underlined.

(1) Myth is a semantic core of **many** modernist writers, consider W. Faulkner's usage of [---]
(2) Women are the presidents of Latvia, Finland and **many** other countries, military advisers for U.S. [---]
(3) However, **some** sceptics may say that language is not so important.

Such explicit indicators of scope help to avoid universal claims, which are often criticised and cautioned against in academic style guides. Vague quantifiers here have a self-distancing effect, are used as a negative politeness strategy, and thus serve as hedges.

To avoid universal claims, quantifiers are sometimes used alongside some other hedging items, such as adverbs indicating the degree of certainty, as in example (4).

(4) Probably **many** people would start arguing on this issue [---]

Here the adverb reduces the relatively high degree of assertiveness of the quantifier *many*.

Quantifiers are used not only to report the scope of a generalisation when giving other people's opinions, but also to provide contradictory viewpoints, as in examples (5) and (6):

(5) Nevertheless, **many** influential scholars hold the view that the language [---]
(6) Postmodernism shares some features with Modernism, but there are **some** important differences as well.

The clues that signal such uses are linkers and conjunctions indicating contrast, e.g. *nevertheless* and *but* in examples (5) and (6).

Quantifiers are also used when introducing quotes and providing references to previous studies. However, these serve as pseudo-references, since they provide no specific references to any specific authors, but make abstract claims about more or less dominant opinions, as in examples (7) and (8).

(7) On the other hand, **some** linguists notice the tendency that English language [---]
(8) As **many** critics say, Postmodernism is a sceptical view to [---]

These examples could be seen as instances of intertextuality, but in none of them actual references are provided, so they just imitate scientific referencing.

Multal quantifiers are used emphatically in such structures as '*very much*, *a lot* + comparative adjective' and '*much* + adjective/noun', where they strengthen the effect of the preceding adjective (or sometimes a noun). This use is clearly demonstrated by the collocates of *much*, the top 25 of which are presented in Figure 2. The most frequent comparatives collocating with *much* are *easier*, *higher*, *faster*, and *better*.

For even more emphasis, the quantifier *much* is sometimes preceded by such intensifying adverbs as *very*, *too*, and *so*. Qualitative analysis has also revealed that the emphatic *much* appears in highly evaluative content, as in examples (9–12).

(9) [---] To sum up, language is **much** more than a mix of various rules of morphology and syntax [---]
(10) Modernist aesthetics was the first to get so **much** disgressed from the tradition.
(11) To my opinion, there was too **much** hurry and too **much** subborness in preparing this reform.
(12) First of all, lots of people spend too **much** time in front of TV.

Multal quantifiers used in such structures as '*very much, a lot* + comparative adjective' and '*much* + adjective' strengthen the effect of the adjective that they precede. Therefore, to highlight their power to increase the assertiveness of speakers' claims, such quantifiers are termed as 'amplifiers' by Hinkel (2003). As such, they are contrasted to downtoners, which include paucal quantifiers and are used in similar patterns with adjectives.

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	266	0	0	-1	much
2	48	0	48	5.30778	more
3	38	32	6	2.64339	is
4	33	20	13	4.00969	as
5	26	26	0	5.26996	very
6	26	26	0	7.34641	too
7	26	26	0	5.97244	so
8	12	12	0	5.10418	how
9	12	12	0	2.18446	are
10	11	3	8	0.25251	of
11	9	2	7	0.19340	to
12	9	0	9	7.23421	easier
13	9	9	0	1.69590	be
14	8	0	8	3.67572	time
15	8	4	4	0.94591	that
16	8	0	8	6.48809	attentior
17	7	6	1	2.38504	with
18	7	0	7	3.31761	money
19	6	6	0	2.13025	have
20	5	0	5	5.22506	violence
21	5	3	2	0.16969	in
22	5	0	5	4.49515	higher
23	4	0	4	8.26068	faster
24	4	0	4	7.75818	effort
25	4	0	4	3.94478	better

Figure 2. Collocates of *much*

As has been shown in Section 4.2, multal quantifiers, which can have the amplifying effect, outweigh downtoning paucal quantifiers in student writing in general, but especially in NNS essays.

Due to the lower frequency of GEs, a smaller array of functions of these items could be distinguished. GEs, which are not typical in NS academic texts, are mainly used in NNS essays to extend a list or an enumeration (usually of examples), as in (13) and (14).

(13) [---] so the first modernists had to rely on European thinkers: Freaud, Nietzsche, **etc.**

(14) [---] the common Lithuanian language examination for both Lithuanian and non-Lithuanian, (i.e. Russian, Polish, Belorussian, **etc.**) schools.

These examples can also be argued to express uncertainty and thus perform the hedging function. However, this form of hedging is not common in written academic English.

To generalise, the analysis of VL functions in NNSE suggests that the overuse of VL and especially the use of quantifiers may result in an unwanted stylistic effect of NNS written texts. The extensive use of emphatic quantifiers and a limited use of downtoning quantifiers may make the style too assertive, and the overuse of GEs may make them too colloquial.

4.4. Discussion of the results

As has already been mentioned, VL in learner language has been investigated to a limited extent and has been examined mainly in spoken interaction, but at least some results of the present study can be assessed and interpreted against the background of some previous research. Though direct and systematic comparisons are not possible due to different designs of the studies under discussion, some tentative implications will be suggested. As will be shown further on, some results of the present study go against the trends observed earlier and thus were unexpected, but most of the findings appear to be in line with prior research.

In view of previous research, an unexpected finding is that the two VL categories under investigation appear to be more frequent in NNSE than NSE. This goes against most of the earlier studies on the use of GEs in spoken (intercultural) communication (e.g. Aijmer 2004, de Cock 2004, Drave 2002, Fernández, Yuldashev 2011, Parvaresh et al. 2012, Lin 2013), which report lower frequency of VL in NNS speech. The current results, though, are in line with Metsä-Ketelä's (2012) study, which, as has already been noted, demonstrates that in academic settings lingua franca speakers use VL almost twice as frequently as NS do. Further exploration of this trend thus is needed with more extensive and more varied data that would allow for systematic contrastive analysis of different L1 backgrounds.

One of the findings that is consistent with some earlier results is that disjunctive GEs are almost never used either in NNSE or NSE. Earlier studies on GEs also point to a generally higher incidence of adjunctive GEs, especially in more formal communication (e.g. Cheshire 2007, Martínez 2011, Ruzaitė 2018). In fact, adjunctive GEs have been reported to be considerably more frequent not only in general English but also other languages (e.g. in Persian (Parvaresh, Tayebi 2014); Slovene (Verdonik 2015); Lithuanian (Ruzaitė 2018)).

It has also been observed in this study that multal quantifiers, which are often emphatic, outweigh downtoning paucal quantifiers in student writing in general, but especially in NNS essays. This is in line with the results of Hinkel (2003), who also reports that in written essays downtoners are infrequent in both NS and NNS texts, whereas emphatics (e.g. verb + *a lot*) and amplifiers (e.g. *very much*, *a lot* + comparative adjective, *much* + adjective) are considerably more frequent in the essays of NNS than in those of NS. The overuse of emphatics and underuse of downtoners make NNS essays sound colloquial, overly assertive, and categorical (Hinkel 2003).

The frequency of individual GEs in NNSE is at least to some extent in line with some previous research. As has already been mentioned, the GE *and so on* is reported to be more frequent in writing and is associated with more formal (academic) style in Biber et al. (1999: 116). Overstreet and Yule's study (1997) shows

that *and so on* and *et cetera* appear mainly in formal discourse, and Stubbe and Holmes' (1995) research demonstrates that *and so on* and *or so* are frequent in formal interviews. All these GEs reported in previous studies as being more formal appear also in the data of this study. Thus, on the one hand, NNS clearly overuse GEs, but the GEs that they use are of an appropriate degree of formality in academic essays.

5. Concluding remarks: General trends and teaching implications

The current study aimed to examine the use of VL in NNS and NS academic texts by considering two main categories of VL: general extenders and vague quantifiers. The research questions raised addressed three main aspects regarding the use of these VL categories: 1) the overall frequency of VL in Lithuanian learner English in comparison to native speakers (RQ1); 2) the frequency of different VL categories (RQ2); and 3) the main functions of VL in non-native speakers' essays and how these functions relate to the formality of VL items (RQ3).

In view of RQ1, the findings indicate that there are some major differences in the frequency of VL in NNSE and NSE, both VL categories being more frequent in NNSE. GEs are ten times more frequent in NNSE (10.7 vs. 0.87; f/10,000), and quantifiers in NNSE are four times as frequent as in English L1 (74.36 occurrences vs. 18.06 occurrences respectively). The overuse of VL in Lithuanian learner English suggests that Lithuanian learners may lack competence in register differentiation. Another possible interpretation of the differences observed in this study is that English L2 speakers perhaps have a different argumentation style, which could be determined by native language interference; however, this remains an open question to be addressed in further research.

Regarding RQ2, the most noticeable differences between NSE and NNSE appear in the use of vague quantifiers. All quantifiers under investigation are more frequent in NNSE, and *some* is the most frequent item of all of them (30 occurrences). Interestingly, four out of the top five quantifiers in NNSE are those referring to large quantities: *much, many, a lot of,* and *majority*. In contrast, mitigating (or downtoning) quantifiers, which refer to small quantities, e.g. *a bit* and *a little*, are very rare. Regarding the formality of quantifiers used in NNSE, the results indicate the overuse of informal items.

The overall frequency of GEs in both databases is considerably lower than that of vague quantifiers, but in NNSE they are significantly more frequent than in NSE (10 occurrences vs. 1 occurrence respectively). GEs with the conjunction *and* clearly outweigh those with *or*. Most importantly, all GEs are more frequent in NNSE, which again suggests the overuse of informal items in NNSE.

Considering RQ3, it has been observed that quantifiers and GEs perform largely the same functions in NNSE and NSE academic texts (cf. Cheng, Warren 2001 for a similar observation about quantifiers in NS and NNS). Quantifiers and GEs can be used as hedges and may reduce the scope or assertiveness of a generalisation, but what is important in the present study is the unexpected dominance of emphatic quantifiers and underuse of downtoning quantifiers. These findings point to the

learners' lack of pragmatic competence of such phenomena as indirectness, inexplicitness, and vagueness.

The results of this study thus have some direct implications for (academic) language teaching. Teachers seem to overlook the importance of vagueness, which is especially register-sensitive. Already very early research on VL and related phenomena points to the importance of mastering the use of vagueness, e.g. Cheng and Warren (2003) stress the importance of indirectness and related phenomena; Stubbs (1996) observes the learners' difficulties to acquire ways of hesitating and saying nothing much; Eisenberg and Phillips (1991) claim that openness and clarity can cause miscommunication; and Tarone and Yule (1987) draw attention to approximation as a non-native speaker strategy. In recent EFL textbooks, it has become customary to include some sections on VL, but there is still a need for even more focus on helping learners to become skilled at differentiating between registers in order to meet genre conventions.

Despite some strong implications of the current findings, the trends observed here should be further tested and verified on the basis of a larger learner corpus. In addition, a systematic contrastive analysis could yield some important results. As Granger observes, learner language, being highly variable, is influenced by a number of "linguistic, situational and psycholinguistic factors" (2004: 125). Further research thus could consider speakers of English from a larger number of native-speaker backgrounds to test the prediction that perhaps some VL categories are overused/underused in learner English in general, but some may be restricted only to some varieties. A more varied sample of English language learners (e.g. a sample including speakers from more diverse age groups, social backgrounds, and with more diverse language learning experience) could also be useful in disclosing how much and in what ways VL variation is determined by social factors. Finally, the present study did not investigate native language interference or the possible influence of teaching materials and methods on the learner output, but both factors are worth considering in further research.

References

Aijmer, Karen 2004. Pragmatic markers in spoken interlanguage. – Nordic Journal of English Studies, 3, 173–190. https://doi.org/10.35360/njes.29

Biber, Douglas; Johansson, Stig; Leech, Geoffrey; Conrad, Susan; Finegan, Edward 1999. Longman Grammar of Spoken and Written English. London–New York: Longman.

Buysse, Lieven 2014. 'We went to the restroom or something': General extenders and stuff in the speech of Dutch learners of English. – J. Romero-Trillo (Ed.), The Yearbook of Corpus Linguistics and Pragmatics: New Empirical and Theoretical Paradigms. Springer, 213–237. https://doi.org/10.1007/978-3-319-06007-1_10

Channell, Joanna 1994. Vague Language. Oxford: Oxford University Press.

Cheng, Winnie; Warren, Martin 2001. The use of vague language in intercultural conversations in Hong Kong. – English World-Wide, 22 (1), 81–104. https://doi.org/10.1075/eww.22.1.05che

Cheng, Winnie; Warren, Martin 2003. Indirectness, inexplicitness and vagueness made clearer. – Pragmatics, 13 (3), 381–400. https://doi.org/10.1075/prag.13.3.02che

Cheshire, Jenny 2007. Discourse variation, grammaticalization, and stuff like that. – Journal of Sociolinguistics, 11, 155–193. https://doi.org/10.1111/j.1467-9841.2007.00317.x

de Cock, Sylvie 2004. Preferred sequences of words in NS and NNS speech. – Belgian Journal of English Language and Literatures (BELL), New Series 2, 225–246.
Drave, Neil 2002. Vaguely speaking: A corpus approach to vague language in intercultural conversations. – Pam Peters, Peter Colins, Adam Smith (Eds.), Language and Computers, New Frontiers of Corpus Research. Papers from the Twenty First International Conference on English Language Research on Computerized Corpora Sydney 2000. Amsterdam–New York: Rodopi, 25–40. https://doi.org/10.1163/9789004334113_003
Eisenberg, Eric M.; Phillips, Steven R. 1991. Miscommunication in organizations. – Nikolas Coupland, Howard Giles, John M. Wiemann (Eds.), "Miscommunication" and Problematic Talk. Newbury Park: Sage Publications, 244–258.
Fernández, Julieta; Yuldashev, Aziz 2011. Variation in the use of general extenders *and stuff* in instant messaging interactions. – Journal of Pragmatics, 43 (10), 2610–2626. https://doi.org/10.1016/j.pragma.2011.03.012
Gao, Hongyun; Zhu, Jue 2005. A research in the influences of vague language on second language learning. – US-China Education Review, 2, 41–45.
Gassner, Denise 2012. Vague language that is rarely vague: A case study of 'thing' in L1 and L2 discourse. – International Review of Pragmatics, 4, 3–28. https://doi.org/10.1163/187731012X632045
Granger, Sylviane (Ed.) 1998. Learner English on Computer. London: Longman
Granger, Sylviane 2004. Computer learner corpus research: Current status and future prospects. – Ulla Connor, Thomas Upton (Eds.), Applied Corpus Linguistics: A Multidimensional Perspective. Amsterdam & Atlanta: Rodopi, 123–145. https://doi.org/10.1163/9789004333772_008
Granger, Sylviane; Hung, Joseph; Petch-Tyson, Stephanie (Eds.) 2002. Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.6
Granger, Sylviane; Gilquin, Gaëtanelle; Meunier, Fanny (Eds.) 2015. The Cambridge Handbook of Learner Corpus Research. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139649414
Grieve, Averil 2010. Adolescent identity and pragmatic marker acquisition in a study abroad context. PhD thesis. Arts, School of Languages, Department of Linguistics and Applied Linguistics, The University of Melbourne.
Hinkel, Eli 2003. Adverbial markers and tone in L1 and L2 students' writing. – Journal of Pragmatics, 35 (7), 1049–1068. https://doi.org/10.1016/S0378-2166(02)00133-9
Lin, Yen Liang 2012. Mind the gap! Textbook conversation vs. authentic intercultural interaction. – Yiu-Nam Leung (Ed.), Selected Papers from the 21st International Symposium on English Teaching. Taipei: Crane Publishing, 42–54.
Lin, Yen Liang 2013. Vague language and interpersonal communication: An analysis of adolescent intercultural conversation. – International Journal of Society, Culture and Language, 1, 69–81.
Martínez, Ignacio M. Palacios 2011. '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 (9), 2452–2470. https://doi.org/10.1016/j.pragma.2011.02.011
Metsä-Ketelä, Maria 2012. Frequencies of vague expressions in English as an academic lingua franca. – Journal of English as a Lingua Franca, 1 (2), 263–285. https://doi.org/10.1515/jelf-2012-0019
Overstreet, Maryann 1999. Whales, Candlelight, and Stuff Like That: General Extenders in English Discourse. Oxford: Oxford University Press.
Overstreet, Maryann 2012. Pragmatic expressions in cross-linguistic perspective. – Applied Research on English Language, 1 (2), 1–14. https://doi.org/10.22108/are.2012.15451

Overstreet, Maryann; Yule, George 1997. On being inexplicit and stuff in contemporary American English. – Journal of English Linguistics, 25 (3), 250–258. https://doi.org/10.1177/007542429702500307

Parvaresh, Vahid; Tavangar, Manoochehr; Eslami Rasekh, Abbas; Izadi, Dariush 2012. About his friend, how good she is, 'and this and that': General extenders in Persian and EFL discourse. – Journal of Pragmatics, 44 (3), 261–279. https://doi.org/10.1016/j.pragma.2011.12.003

Parvaresh, Vahid; Tayebi, Tahmineh 2014. Vaguely speaking in Persian. – Discourse Processes, 51 (7), 565–600. https://doi.org/10.1080/0163853X.2013.874545

Quirk, Randolph; Greenbaum, Sidney; Leech, Geoffrey; Svartvik, Jan 1985. A Comprehensive Grammar of the English Language. London: Longman.

Ruzaitė, Jūratė 2007. Vague Language in Educational Settings: Quantifiers and Approximators in British and American English. Frankfurt am Main: Peter Lang.

Ruzaitė, Jūratė 2018. General extenders and discourse variation: A focus on Lithuanian. – International Journal of Corpus Linguistics, 23 (4), 469–494. https://doi.org/10.1075/ijcl.17019.ruz

Sabet, Peyman G. P.; Zhang, Grace Q. 2015. Communicating through Vague Language: A Comparative Study of L1 and L2 Speakers. New York: Palgrave Macmillan. https://doi.org/10.1057/9781137486387

Stenström, Anna-Brita; Andersen, Gisle; Hasund, Ingrid Kristine 2002. Trends in teenage talk: Corpus compilation, analysis and findings. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.8

Stubbe, Maria; Holmes, Janet 1995. You know, eh and other 'exasperating expressions': An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. – Language & Communication, 15 (1), 63–88. https://doi.org/10.1016/0271-5309(94)00016-6

Tarone, Elaine; Yule, George 1987. Communication strategies in East-West interactions. – Larry Smith (Ed.), Discourse Across Cultures. New York: Prentice Hall, 49–65.

Verdonik, Darinka 2015. Internal variety in the use of Slovene general extenders in different spoken discourse settings. International Journal of Corpus Linguistics, 20 (4), 445–468. https://doi.org/10.1075/ijcl.20.4.02ver

Warren, Martin 1993. Inexplicitness: A feature of naturalness in conversation. – Mona Baker, Gill Francis, Elena Tognini-Bonelli (Eds.), Text and Technology: In Honour of John Sinclair. Philadelphia: John Benjamins Publishing Company, 37–53. https://doi.org/10.1075/z.64.04war

ÄHMANE VÄLJENDUSLAAD LEEDULASTE INGLISE ÕPPIJAKEELES

Jūratė Ruzaitė
Vytautas Magnuse Ülikool

Artiklis analüüsitakse ähmast väljenduslaadi leedu emakeelega inglise keele edasijõudnud õppijate (L2) argumenteerivates tekstides. Uurimus keskendub ähmase väljenduslaadi kahele põhitüübile: üldlaiendid (nt *and so on*, *etc. / etcetera*, *or so*) ning ebamäärased hulgasõnad (nt *some*, *many*, *a lot of*, *a little*). Eesmärk on 1) hinnata, kui ulatuslikult kasutavad ähmast väljenduslaadi leedu emakeelega inglise keele õppijad (L2) võrdluses inglise keelt emakeelena kõnelejatega (L1), 2) määrata ähmase väljenduslaadi peamised funktsioonid argumenteerivas tekstis ja 3) analüüsida uuritavate keeleüksuste formaalsust.

Uurimuse andmestiku moodustavad inglise filoloogia leedu emakeelega üliõpilaste arutlevad kirjutised (kokku 244 746 sõna). Võrdlusandmestik pärineb Briti akadeemilise inglise kirjakeele korpuse (BAWE: British Academic Written English Corpus) inglise keele õppeaine alamkorpusest (kokku 458 780 sõna). Andmestiku analüüsiks kasutati AntConc tarkvara.

Tulemused osutavad suurtele erinevustele L1 ja L2 inglise keeles. Kõige silmatorkavam erinevus on ebamääraste hulgasõnade kasutuses: kõik uuritud hulgasõnad on sagedamad L2 inglise keeles, kusjuures kõige sagedasem keeleüksus on *some* (30 esinemisjuhtu) ning viis sagedamat hulgasõna osutavad suurele hulgale (*much*, *many*, *a lot of*, *majority*). Seevastu vähendavad hulgasõnad (nt *a bit*, *a little*) esinevad väga harva. Üldlaiendite sagedus mõlemas andmestikus on märgatavalt madalam kui ebamäärastel hulgasõnadel, ent L2 inglise keeles esinevad need oluliselt sagedamini kui L1 inglise keeles (vastavalt 10 *vs*. 1 esinemisjuhtu).

Võtmesõnad: sagedus, pragmaatiline funktsioon, (in)formaalsus, argumenteeriv kirjutis, inglise keel, leedu keel, õppijakeel, emakeelekõneleja

Jūratė Ruzaitė's (Vytautas Magnus University, Centre of Intercultural Communication and Multilingualism) research interests include sociolinguistics, pragmatics, discourse analysis, language and ideology, and hate speech.
K. Donelaičio g. 58, 44248 Kaunas, Lithuania
jurate.ruzaite@vdu.lt