Applied Corpus Linguistics 3 (2023) 100040

Contents lists available at ScienceDirect

Applied Corpus Linguistics

journal homepage: www.elsevier.com/locate/acorp

Epistemic stance in written L2 English: The role of task type, L2 proficiency, and authorial style

Maria Pyykönen

University of Turku, Department of English, Koskenniemenkatu 4, 20014, Turku, Finland

ARTICLE INFO

Keywords:
Learner language
Epistemic stance
CEFR
Language assessment
Learner corpus research

ABSTRACT

The present study examines the relationship between the use of epistemic stance expressions (i.e., hedges and boosters) and task type, L2 proficiency, and individual authorial style in 1,773 essays representing three different kinds of tasks (complaint, letter, and opinion) written by 591 Finnish L2 English speakers on four different levels of proficiency (CEFR levels B1-C2). The results of the study show that the frequency of both hedges and boosters is mainly governed by task type, as the opinion tex contained a higher number of both hedges and boosters than the other tasks examined. Proficiency-related patterns were, nevertheless, also observed, as it was shown that in the complaint and opinion tasks, the frequency of both hedges and boosters tends to increase with proficiency, while in the letter task, the frequency of both ED types shows signs of decrease. Individual authorial style was shown to play a very limited role in the frequency of EDs in the data, but the results also suggest that the influence of authorial style may be greater with respect to boosters than it is with hedges.

1. Introduction

The ability to successfully develop and express a position on an argument is one of the key components of academic writing. To this end, the expression of (un)certainty plays an important role, as it allows a speaker to either recognize the existence of alternative points of view, or conversely, completely close them down (Hyland, 2005, pp. 52–53). In the framework of stance, expressions of (un)certainty are referred to as epistemic stance markers (or epistemic devices, henceforth EDs), which are defined as "speaker comments on the status of information in a proposition" (Biber et al., 2021, p. 964). By using EDs, the speaker can express either doubt or certainty towards a proposition, as exemplified by (1) and (2), respectively (both examples retrieved from the learner data used in the present study).

(1) It seemed to be quite all right. (56660-2-C2)
(2) There is always two sides of one coin. (57401-3-B1)

The use of EDs has often also been described as a challenge from the point of view of learning English as a second language (L2) as there is both a wide, grammatically heterogeneous range of lexical items used as EDs, and, subsequently, a wide range of different meanings attributed to them (e.g., Bartley and Hidalgo-Tenorio, 2016, p. 21; Hyland, 2005, p. 133). In addition to this, the use of EDs has also been shown to greatly vary across different contexts of language use (see e.g., Biber, 2006; Larsson, 2019), which presents a further challenge to learners looking to master the use of EDs.

The use of EDs is also included in the functionally-oriented *Common European Framework of Reference for Languages* (CEFR; Council of Europe, 2001). In the CEFR, different levels of proficiency are described using "can-do" statements that relate to multiple facets of language use – with respect to *propositional precision*, for example, the expression of (un)certainty is first mentioned in reference to level C1, where the learner is described as being able to for example "qualify opinions and statements precisely in relation to degrees of, for example, certainty/uncertainty, belief/doubt, likelihood, etc." (Council of Europe, 2001, p. 129). This indicates that while EDs can be used on lower levels of proficiency as well (see e.g., Pérez-Paredes and Díez-Bedmar, 2019), they should become more salient from level C1 onwards; on level C2, the use of EDs is described as becoming increasingly nuanced, as the learners should be able to "convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of qualifying devices" (Council of Europe, 2001, p. 129).

There is, nevertheless, fairly little research on how these descriptors relate to empirical learner language (but see Pérez-Paredes and Díez-Bedmar, 2019; Yoon, 2020). Moreover, previous research on the use of EDs in both native and non-native language suggests that their use is also influenced by register (e.g., Biber et al., 2021; Larsson, 2019), task type (e.g., Hinkel, 2009; Pérez-Paredes and Díez-Bedmar, 2019; Yoon, 2020), and individual speaker styles (e.g., Gablasova et al., 2017), for example. Thus, understanding how these different writer- and text-related variables interact with ED use is an important step in describing how and if the use of stance expressions can be used to distinguish between differ-

E-mail address: mahepy@utu.fi

https://doi.org/10.1016/j.acorp.2022.100040
Received 19 November 2021; Received in revised form 18 November 2022; Accepted 17 December 2022
Available online 18 December 2022

ent levels of proficiency. These descriptions can also be used to inform foreign language teaching, for example, as they can be used to make instructors aware of possible task- and proficiency-related patterns of ED use.

The present study aims to extend previous research by examining a wide range of EDs used in texts representing three different tasks written by Finnish learners of English on four levels of L2 proficiency (CEFR levels B1–C2). In the present study, the focus is on the frequency of EDs in individual texts. Accordingly, the paper discusses the following questions: To what extent is the frequency of i) hedges and ii) boosters influenced by task type, the level of proficiency, and individual authorial style in written L2 English?

The article is structured as follows: in the following section, a brief summary of epistemic stance and previous research related to the variation of ED use across different contexts is provided. Section 3, in turn, describes the material and methods used in the present study, while Section 4 presents the results. The results are summarized and discussed in Section 5. Section 6 presents the concluding remarks and discusses the limitations and implications of the study.

2. Background

2.1. Epistemic stance

Epistemic stance is used for a variety of social functions: in addition to simply displaying certainty or uncertainty, it can be used to signal politeness, construct authority, or regulate the interaction between the interlocutors (Hyland, 2005, p. 53; Kärkkäinen, 2003, p. 26). The present study adopts the framework of stance in which stance expressions consist of a lexical stance item and a claim which that item modifies (Biber et al., 2021, p. 961); here, the focus is on the lexical stance items in particular (i.e., how the claims are modified). The group of lexical items used as EDs is a varied one, and includes modal auxiliaries (e.g., *may*), lexical verbs (e.g., *think*), adverbs (e.g., *perhaps*), nouns (e.g., *opinion*), and adjectives (e.g., *possible*) (e.g., Biber et al., 2021, pp. 961–962). A complete list of the EDs examined in the present study can be found in Appendix A (see section 3.2 for details on the compilation of the list).

In the present study, EDs found in the data are also examined from the perspective of metadiscourse (Hyland, 2005), in which expressions of (un)certainty are further divided into hedges and boosters based on their pragmatic function. Of these, hedges (e.g., *maybe, I guess*) are used to indicate not only a lack of commitment towards the proposition itself, but also as a way of recognizing alternate viewpoints and giving room for possible disagreements (Hyland 2005: 52). Boosters (e.g., *surely, I'm certain*), on the other hand, are used to underline the speaker's confidence and to narrow down the dialogical space by denying the possibility of alternative points of view (Hyland 2005: 52–53). In this sense, the expression of stance and hedging/boosting can be seen as complementary perspectives on the same phenomenon; in the framework of stance, the focus is on the lexical items that are used to express personal opinions, while the terms hedge and booster are used to differentiate between stance expressions that serve different pragmatic purposes in the text.

2.2. Sources of variation in the use of epistemic expressions

Previous studies have made many attempts to uncover the underlying reasons for the variation in the use of epistemic expressions in different contexts. Especially with respect to L1 English, the use of stance markers has been studied from the point of view of different registers. For example, it has been shown that ED use is heavily influenced by the level of formality and personal involvement inherent to different kinds of language-use contexts (e.g., Biber et al., 2021; Biber and Conrad, 2009). These observations have also been echoed in studies examining learner language, where the phenomenon has often been examined with respect to different task types; for example Gablasova et al. (2017, p. 633), who studied spoken L2 English, found that while more formal tasks seemed to elicit fewer EDs in general, tasks involving a higher level of interaction between the interlocutors tended to evoke a higher degree of tentativeness in the use of EDs (ibid.).

Task-related differences were also reported by Crosthwaite and Jiang (2017, p. 13), who examined academic L2 English in a longitudinal setting, and found that essays in particular showed a longitudinal increase in the use of boosters, while reports, in turn, were written in a more tentative manner; they also noted qualitative changes in the kinds of stance expressions used. Similar observations were also made by for example Pérez-Paredes and Díez-Bedmar (2019), who examined the use of three certainty adverbs (i.e., *really, actually*, and *obviously*) in spoken L2 English produced by Spanish-speaking learners and found that the adverbs were used in a different manner in dialogic and monologic tasks, but also noted that some of the expressions examined seemed more task-independent than others. Yoon (2020), who examined argumentative L2 English written by East-Asian learners, also found clear task-related differences in the use of hedges and boosters, as one of the topics seemed to elicit a higher number of boosters than the other; in this case, Yoon suggested that this might be attributable to prompt wording, as the prompt eliciting the most frequent use of boosters also contained three boosters itself.

Another possible source of variation in the use of EDs is the level of L2 proficiency, although studies examining the use of EDs on various levels of L2 proficiency are not entirely conclusive, however: while some have reported proficiency-related patterns in the use of EDs (e.g., Bax et al., 2019, p. 89; Chen and Baker, 2016, p. 873; Hyland and Milton, 1997, pp. 189–196; Pérez-Paredes and Díez-Bedmar, 2019), others have not found any evidence for such patterns (e.g., Staples et al., 2013, p. 222; Yoon, 2020). Of these, for example Hyland and Milton (1997, pp. 191–195) examined the used of EDs in L2 English written by Chinese-speaking learners and reported a general increase in the frequency of EDs across the ability bands examined, and, more recently, Pérez-Paredes and Díez-Bedmar (2019) also found that the frequency of the certainty adverbs *really, actually,* and *obviously* tended to increase with proficiency in spoken L2 English across the levels B1-C1. On the other hand, for example Yoon (2020), who examined the matter in a multifactorial setting, found no statistically significant differences in the overall frequency of neither hedges nor boosters in argumentative essays written by East-Asian learners of English on levels A1-B1.2.

Some of the differences emerging between previous studies examining ED use on different levels of proficiency may of course be explained by for example register differences (i.e., spoken vs written) or the fact that they examine different L1 populations (on L1-related differences in the use of EDs, see e.g., Bartley and Hidalgo-Tenorio, 2016; Pérez-Paredes and Bueno-Alastuey, 2019; Yoon, 2020), but it is noteworthy that many studies also focus on (partially) different proficiency bands. It could thus be that, on the lower levels of proficiency, EDs are used in a similar manner across the different groups (e.g., Yoon, 2020), while on the higher levels of proficiency, the use of EDs gains importance (e.g., Chen and Baker 2016). This would also be in line with in the CEFR descriptors acknowledging the salience of EDs from level C1 onwards.

There is also some evidence that individual speaker styles may influence the use of EDs (e.g., Kirkham, 2011); in the context of learner language, the phenomenon has been examined by Gablasova et al. (2017, p. 631), who found that individual variation seemed to influence multiple levels of ED use, ranging from the overall frequency of EDs to the systematic use of particular stance expressions. Many previous studies have also noted substantial within-group variation related to the use of epistemic stance markers in both native (e.g., Uccelli et al., 2013, p. 49) and non-native English (e.g., Aijmer, 2011, p. 49), which could also be related to individual differences in ED use. There is, nevertheless, little quantifiable evidence of the extent to which individual differences can be used to explain this variation, although the recent insurgence of multifactorial analyses (e.g., mixed effects modeling) in the field of

Table 1
An overview of the corpus used.

	Letter		Complaint		Opinion	
	N	Tokens	N	Tokens	N	Tokens
B1	97	13,017	97	12,737	97	15,459
B2	178	25,485	178	25,318	178	31,468
C1	217	36,667	217	35,994	217	42,396
C2	99	18,313	99	16,902	99	22,831
Total	**591**	**93,482**	**591**	**90,951**	**591**	**112,154**

learner corpus research and corpus linguistics in general has allowed for more fine-grained analyses of individual variation with respect to language use (see e.g., Cvrček et al., 2020; Murakami, 2016; Wulff and Gries, 2021).

In the light of previous research, it thus seems that task type, L2 proficiency, and individual differences are all potential sources of variation in the use and frequency of EDs (i.e., hedges and boosters) in L2 English. Nevertheless, as few studies have examined the matter using a multifactorial approach, little is known as to how these variables interact with each other, making it difficult to draw conclusions about the extent to which each of these factors actually influences ED use (for a discussion on the potential multifactoriality of linguistic phenomena, see e.g., Gries, 2018; Paquot and Plonsky, 2017). This may also in part explain the somewhat contradictory results on the influence of L2 proficiency on the use of EDs, as it could for example be that learners respond to different tasks in different ways depending on their level of proficiency. In the present study, all three variables (i.e., level of proficiency, task type, and individual author) are used in the analysis to examine the extent to which each of these contributes to the frequency of EDs (i.e., hedges and boosters) both independently and in interaction with the other variables.

3. Material and methods

The data used in the present study are drawn from the English subsection of the *Finnish National Certificates Learner Corpus* (YKI), which is one of the largest modern learner corpora in Finland and contains responses from the National Certificate of Language Proficiency examination,[1] a foreign language proficiency test officially recognized by the National Agency for Education of Finland. The results of the examination can be used for a variety of education- and employment-related purposes (e.g., job or university applications). The examination itself consists of three subtests (writing, reading, and speaking), each assessed on a six-point scale corresponding with the CEFR levels A1–C2.

The data used in the present study contain altogether 1,773 texts rated at levels B1 to C2 in the writing subtest (a total of 296,587 tokens), written by 591 writers who reported Finnish as their first language (see Table 1 for an overview of the data). The age of the respondents varies from 15 to 59, the mean age being 32.2 years. The minimum length of essays included in the data was set to 100 tokens, and the mean length of the texts is 167 tokens.

The corpus contains three responses from each writer, which makes it possible to examine the influence of both task type and authorial style across the different tasks. As shown in Table 1, the three tasks are labeled 'letter', 'complaint', and 'opinion'. Of these, 'letter' refers to a short message sent to a friend, a colleague, or an instructor with the purpose of asking a question, giving directions or setting up a meeting, 'complaint' to a message sent to a company with the purpose of complaining about a faulty product or bad customer service (e.g., a broken microwave oven or a delayed flight), and 'opinion' to an argumentative essay written on topics ranging from public transport to internet safety and politics, typically fashioned as a letter to the editor. Thus, within each task type, there are multiple subtopics that share a similar communicative purpose (i.e., to inquire, to complain, or to express one's opinion). More specifically, the letter tasks entail multiple kinds of requests, the complaint tasks cover multiple kinds of dissatisfactory consumer experiences, and the opinion tasks consist of a wide array of different kinds of argumentative topics. Each task type, with its multiple subtopics, is nevertheless distinct from the others in its intended communicative goal, meaning there is no overlap between the three task types in this sense.

To locate as many epistemic expressions as possible in the data, a preliminary list of EDs was compiled based on previous literature (Biber, 2006; Gray and Biber, 2014; Hyland, 2005; Hyland and Milton, 1997). In addition to this, 10 randomly sampled texts from each proficiency level (i.e., altogether 40 texts) were manually analyzed to identify any additional EDs. After this, an additional sample of 40 texts was manually analyzed to evaluate the accuracy of the list-based query. At this point, the list-based query located 502 out of 533 EDs (94%) in the texts; all the additional EDs found at this stage (e.g., *partially, nearly*) were also included as search terms in the finalized list used in the automatic extraction (for the complete, lemmatized list, see Appendix A), which was done using AntConc 3.5.9 (Anthony, 2020).

As the use of EDs is often contextually determined (e.g., Kärkkäinen, 2003, p. 21), all of the entries returned by the automatic retrieval were manually examined to determine whether their use was epistemic, as in (3), or not, as in (4), and all non-epistemic uses of the target forms were removed from the data.

(3) I find this rather unacceptable. (66784-2-C2)
(4) Breakfast was filling but I couldn't find any scrambled eggs. (57348-2-B2)

All remaining EDs were also analyzed as hedges or boosters according to Hyland (2005). The identification and classification of EDs was done by one annotator; to ensure the quality of the annotation, a subsample of the data was coded on two separate occasions to identify any possible inconsistencies. After the manual analysis, altogether 6,344 EDs (3,694 hedges and 2,340 boosters) were found in the data.

The present study applies a text-based approach (see e.g., Ivaska, 2022) in which the unit of observation is one text. Thus, to examine the use of EDs in individual texts, the frequencies of both hedges and boosters were computed for each text file separately and normalized to occurrences per 100 words (phw). These frequencies were then used as response (or dependent) variables in two separate mixed-effect models, where task type, level of proficiency and the author ID are used as predictor (or independent) variables. More specifically, both task type and level of proficiency are included in the model as fixed effects, and the interaction between the two is also examined. The CEFR levels were converted into numeric values (i.e., B1 = 1, B2 = 2, C1 = 3, C2 = 4; see e.g., Paquot, 2018, p. 36). In addition to the fixed effects, random intercepts are computed for each individual writer ID, meaning the model accounts for writer-specific variability across the different tasks and the fact that each writer contributes multiple data points in the data (see e.g., Cvrček et al., 2020, p. 475; Gries, 2021, p. 771; Winter and Grice, 2021).

All variables used in the analyses, along with their roles and possible values are listed in Table 2. To examine the statistical significance of the fixed effects, likelihood ratio tests were performed by comparing the full models to partial models in which one of the predictors (or the interaction term for two predictors) was removed from the model (Winter, 2020, pp. 262–263); this was done with respect to all fixed effects separately. All statistical analyses were carried out in *R* (R Core Team, 2021); the associated data and R code are also available online.[2]

[1] https://www.oph.fi/en/national-certificates-language-proficiency-yki

[2] https://osf.io/p4kf6/

Table 2
Variables used in the analyses (1,773 observations).

Variable	Role	Type	Values
HEDGE_FREQ: Frequency of hedges (phw) in one text file	response	numeric	mean (sd): 1.2 (1.1) min ≤ median ≤ max: 0 ≤ 0.9 ≤ 7.3
BOOSTER_FREQ: Frequency of boosters (phw) in one text file	response	numeric	mean (sd): 0.7 (0.8) min ≤ median ≤ max: 0 ≤ 0.6 ≤ 5.6
LEVEL: CEFR level	predictor	numeric, fixed	B1 (291, 16.4%) B2 (534, 30.1%) C1 (651, 36.7%) C2 (297, 16.8%)
TASK: Task type	predictor	categorical, fixed	complaint (591, 33.3%) letter (591, 33.3%) opinion (591, 33.3%)
ID: Writer ID	predictor	categorical, random	591 individual IDs (e.g., 12,345-B1)

Table 3
Frequency of hedges across task types and levels of proficiency (total number, mean and standard deviation of hedges phw).

	Complaint			Letter			Opinion		
	Total	**mean (phw)**	**sd**	**Total**	**mean (phw)**	**sd**	**Total**	**mean (phw)**	**sd**
B1	95	0.71	0.87	103	0.81	0.83	265	1.74	1.06
B2	187	0.72	0.82	211	0.81	0.83	596	1.90	1.22
C1	354	0.94	0.86	289	0.79	0.73	823	1.91	1.27
C2	196	1.07	0.79	137	0.77	0.80	438	1.88	1.29

Table 4
Coefficients for the fixed effects.

	Estimate	Std.error
(Intercept)	0.51	0.11
LEVEL	0.14	0.04
TASKletter	0.32	0.16
TASKopinion	1.26	0.16
LEVEL:TASKletter	−0.15	0.06
LEVEL:TASKopinion	−0.09	0.06

Table 5
Likelihood tests for each (fixed) variable and interaction.

Variable	Degrees of freedom	χ^2	P-value
LEVEL	1	5.15	0.023*
TASK	2	65.28	<0.001***
LEVEL:TASK	2	6.79	0.033*

4. Results

This section presents the results of the present study. The section is divided into two subsections, of which the first examines the frequency of hedges as a function of the predictor variables outlined in the previous section, and the second does the same with respect to the frequency of boosters.

4.1. The frequency of hedges

The total number and mean frequency of hedges phw across the tasks and levels of proficiency is shown in Table 3, which shows that the standard deviation related to the use of hedges is very high with respect to all task types and levels of proficiency, indicating that there seems to be quite a bit of variability in the frequency of hedges across the board.

To tease apart the estimated influence of each individual variable (or their interaction), a mixed-effects model was used to model the frequency of hedges as a function of LEVEL, TASK, and their interaction. To account for writer-specific variability, a random intercept for each individual writer was also included in the model. In other words, the following model was fitted on the data:

HEDGE_FREQ ~ LEVEL * TASK + (1 | ID).

The coefficients of the model are shown in Table 4. The intercept represents the estimated frequency of hedges when the CEFR level is B1 and TASK is complaint, while the other coefficients illustrate how changes in the fixed predictors influence the frequency of hedges in relation to the intercept. In other words, the estimate for level B1 when TASK is 'letter', for example, is 0.51 + 0.32 = 0.83 hedges phw, and for 'opinion', the same estimate is 0.51 + 1.26 = 1.77 hedges phw. Table 4 also shows that while the main effects of LEVEL and TASK are positive (i.e., changes in both task and level alone result in an increase in hedge frequency), the interaction between the two yields a negative coefficient. This, in turn, suggests that when changes in both LEVEL and TASK occur simultaneously, their individual (main) effects are slightly diminished.

The $R^2_{marginal}$ of the model is 0.211 and the $R^2_{conditional}$ 0.213, meaning that fixed effects alone can describe 21.1% of the observed variation in the frequency of hedges, while the inclusion of the random intercept for individual authors increases the explanatory power of the model only by little, to 21.3%. In other words, it seems that, in this model, the role of individual differences is not integral in explaining the frequency of hedges in the texts.

Table 5 shows the results of the likelihood ratio tests performed to evaluate the significance of the fixed predictors and indicates that both the main effects of TASK and LEVEL and the interaction between the two have a statistically significant effect on the frequency of hedges.

The combined effects of TASK, LEVEL, and their interaction (Table 4) are visualized in Fig. 1, which also shows that there are clear task-related differences in the frequency of hedges in the texts. The highest number of hedges are found in the opinion task, regardless of the level of proficiency. As the variable LEVEL is numeric, the coefficient given for its main effect in Table 4 can be interpreted as follows: for each one-unit increase of LEVEL (e.g., moving from B1 to B2), the frequency of hedges increases by 0.14 hedges phw. The interactions between LEVEL and TASK, on the other hand, are given a negative coefficient in Table 4,

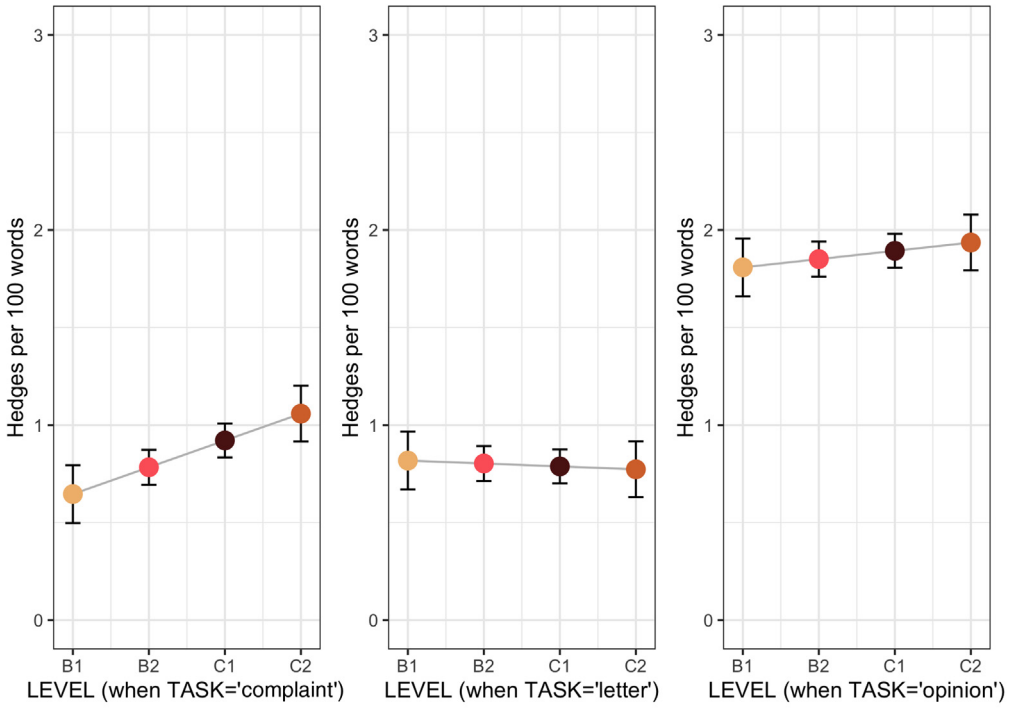

Fig. 1. The frequency of hedges as a function of TASK interacting with LEVEL.

Table 6
Most common hedges in the 'letter' task (raw frequency/frequency phw).

B1	B2	C1	C2
1. think (21, 0.16)	1. think (72, 0.12)	1. would (35, 0.10)	1. would (15, 0.08)
2. would (20, 0.15)	2. maybe (35, 0.10)	2. think (27, 0.07)	2. think (11, 0.06)
3. maybe (12, 0.09)	3. would (31, 0.09)	3. quite (23, 0.06)	3. quite (11, 0.06)
4. say (15, 0.06)	4. could (28, 0.07)	4. might (18, 0.05)	4. seem (8, 0.04)
5. probably (8, 0.04)	5. quite (25, 0.07)	5. seem (17, 0.05)	5. probably (7, 0.04)

indicating that the level-related increase observed for the intercept is slightly reduced when TASK is either 'letter' or 'opinion'.

This notion is also corroborated by Fig. 1, which visualizes the combined effect of the model coefficients in Table 4 and shows that the slope of the line drawn between the different levels of proficiency when TASK is either 'letter' or 'opinion' differs from 'complaint'. Especially 'letter' stands out in this regard, as here, there is actually a minute decline in the use of hedges with proficiency, as the estimated frequency of hedges in the letters is 0.83 when LEVEL is B1 and 0.80 when LEVEL is C2. Some small signs of decrease can also be seen when examining the observed frequencies of the most common hedges in the letters across the proficiency levels in Table 6, which also shows that the frequencies of the individual hedges seem to be fairly low across the board in the letter task.

Of the hedges listed in Table 6, *would* and *think* seem to be preferred by all groups. In this task, *think* is often used to frame statements in which the writer is speculating about their schedule, for example, as in (5), and *would*, on the other hand, is used alongside requests such as (6), where the writer is attempting to emphasize how much they would benefit from the assistance of the recipient. In addition to this, adverbial hedges such as *maybe* and *probably* are typically used in the letters to make it possible for the recipient to decline a request, as in (7), or to reassure them when giving directions, as in (8).

(5) I think that there will be few hours free time. (57409-1-B2)
(6) This would help me to catch up with the reading. (76193-1-C1)
(7) It's being held in Oxford, so maybe you could help me to find accommodation. (55831-1-B2)
(8) You'll probably figure what to do with the keys then. (106737-1-B2)

When TASK is 'opinion', however, each one-unit increase of LEVEL decreases the frequency of hedges by 0.09, meaning the interaction between LEVEL and TASK slightly decreases the positive main effects given for the two variables in Table 2. The combined effect of all the variables remains positive, however, as the estimated frequency of hedges in this context is 1.77 for level B1 and is 1.92 for level C2. The most common hedges in the opinion texts are shown in Table 7, which shows that *would* and *think* are the most frequently used hedges in this task as well, although the groups B1 and B2 show a much stronger tendency for the use of *think* than C1 and C2.

In this task, *think* is used in a very traditional manner; the speaker claims something about a state of affairs and then qualifies the claim so as to make room for disagreement, as in (9). Depersonalized expressions wherein the writer qualifies a claim proposed by a third party, as illustrated by the use of *say* in (10), also seem to be particularly frequent in the opinion texts when compared to the other tasks; this is most likely related to the communicative purpose of the texts, as the writer is asked to argue for or against a certain stance, which entails the underlying assumption that there are two possible points of view to consider.

(9) I think the right to have free medical care belongs to everybody. (84712-3-C1)
(10) People often say that they haven't had time to eat all day. (72708-3-B2)

Table 7
Most common hedges in the 'opinion' task (raw frequency, frequency phw).

B1	B2	C1	C2
1. think (92, 0.71)	1. think (128, 0.50)	1. think (112, 0.31)	1.would (43, 0.23)
2. would (25, 0.19)	2 would (74, 0.29)	2. would (98, 0.27)	2. think (40, 0.22)
3. maybe (24, 0.18)	3. opinion (47, 0.18)	3. opinion (66, 0.18)	3. seem (32, 0.17)
4. opinion (22, 0.17)	4. say (42, 0.16)	4. can (59, 0.16)	4. can (27, 0.15)
5. say (17, 0.13)	5. seem (28, 0.11)	5. seem (64, 0.12)	5. might (25, 0.14)

Table 8
Most common hedges in the 'complaint' task (raw frequency, frequency phw).

B1	B2	C1	C2
1. think (32, 0.25)	1. think (29, 0.11)	1. would (48, 0.13)	1. would (24, 0.13)
2. maybe (10, 0.08)	2. would (17, 0.07)	2. think (37, 0.10)	2. seem (16, 0.09)
3. quite (9, 0.07)	3. maybe (16, 0.06)	3. say (28, 0.08)	3. say (13, 0.07)
4. would (9, 0.07)	4. quite (15, 0.06)	4. seem (22, 0.06)	4. believe (10, 0.05)
5. might (6, 0.05)	5. find (13, 0.05)	5. believe (21, 0.06)	5. quite (9, 0.05)

Table 9
Frequency of boosters across task types and levels of proficiency (total number, mean and standard deviation of boosters phw).

	Complaint			Letter			Opinion		
	Total	mean (phw)	sd	Total	mean (phw)	sd	Total	mean (phw)	sd
B1	48	0.37	0.57	58	0.44	0.58	155	0.98	0.99
B2	101	0.38	0.61	204	0.77	0.94	308	0.96	0.69
C1	223	0.59	0.68	261	0.63	0.82	488	1.16	0.90
C2	101	0.52	0.52	98	0.49	0.71	295	1.31	1.07

Finally, Fig. 1 also shows a proficiency-related increase in the use of hedges in the complaint task. As 'complaint' (when LEVEL = B1) serves as the intercept, the estimated frequency of hedges on the different levels of proficiency is directly calculable from the main effect estimate given for LEVEL, meaning the difference between levels B1 and C2, for example, is 3 × 0.14 = 0.42 hedges phw. The most common hedges in the complaints are illustrated in Table 8, which shows that while groups B1, B2 and, to some extent, C1, seem to rely on *think* in the complaints as well, it is not included among the most common hedges on level C2 at all.

In this task, hedges are often used by the writers in when shifting blame, as in (11) and (12), thus leaving the matter of liability open for debate; it also seems that *think* and *seem* are used for similar purposes here although particularly the latter seems to be more characteristic of the higher proficiency groups. The modal verb *would* is also particularly frequent in the higher proficiency groups, and it is often used to accompany statements of consequence, as in (13).

(11) And the reason is your company, I think. (41714-2-B1)
(12) It seems that your operator has forgot to send it! (37886-2-C1)
(13) This would cause extra costs for both of us. (35990-2-C1)

Coupled with the observed increase of hedges with proficiency in the complaints, the uses of *think* and *seem* suggest that although the learners do, in general, seem to be sensitive for the need to qualify their claims in situations where the integrity or responsibility of the recipient is questioned, more proficient learners may do so more often and by using a different wording (i.e., *seem*). In addition to this, the increase of *would* suggests that more proficient learners are also more likely to refer to potential future events when writing complaints; in this sense, the use of *would* here is similar to the letter task (see example (6)), although the intended tone is often very different.

4.2. The frequency of boosters

Table 9 shows the observed number and mean frequency of boosters across the different tasks and levels of proficiency; in this case as well, the standard deviations are very high across the board.

Table 10
Coefficients for the fixed effects.

	Estimate	Std.error
(Intercept)	0.28	0.09
LEVEL	0.08	0.03
TASKletter	0.37	0.13
TASKopinion	0.49	0.13
LEVEL:TASKletter	−0.09	0.05
LEVEL:TASKopinion	0.05	0.05

The second mixed-effects model models booster frequency as a function of LEVEL, TASK, and their interaction, and computes a random intercept for each individual writer. The resulting model is thus the following:

BOOSTER_FREQ ~ LEVEL * TASK + (1 | ID).

The coefficients for the fixed effects in the model are shown in Table 10. The intercept estimate for booster frequency (i.e., when LEVEL = B1 and TASK = complaint) is 0.28 boosters phw, which is noticeably lower than the estimate given for hedges in the same context (i.e., the intercept estimate in Table 4). The same can be said of the other tasks as well, as the estimated frequency of boosters (when LEVEL = B1) is 0.28 + 0.37 = 0.65 boosters phw for 'letter', and 0.28 + 0.49 = 0.77 boosters phw for 'opinion'. It is also noteworthy that this time only one of the coefficients of the model is negative, indicating that only the interaction between LEVEL and the task 'letter' seems to reduce the frequency of boosters, while changes in the other predictors result in an increase instead.

The $R^2_{marginal}$ of the model is 0.111 and the $R^2_{conditional}$ 0.162; the random intercept for writers thus contributes slightly more to the explanatory power of the model than was the case for the model related to hedges. The results of the likelihood ratio test (Table 11) also show that all fixed predictors have a statistically significant effect on the frequency of boosters in the texts.

The combined effect of TASK, LEVEL, and their interaction (as shown in Table 10) is illustrated in Fig. 2, which shows that in this case as well, the opinion task seems to yield the highest frequencies of boosters

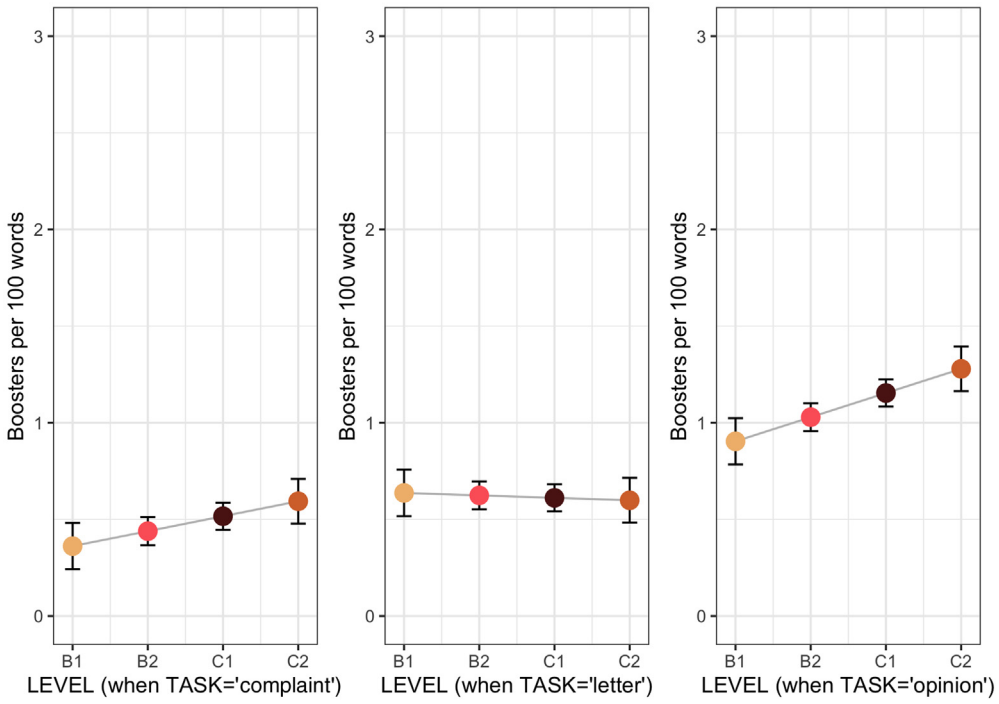

Fig. 2. The frequency of boosters as a function of TASK interacting with LEVEL.

Table 11
Likelihood tests for each (fixed) variable and interaction.

Variable	Degrees of freedom	χ^2	P-value
LEVEL	1	9.38	0.002**
TASK	2	16.56	<0.001***
LEVEL:TASK	2	9.09	0.011*

across all proficiency levels, although the differences between the different tasks are typically smaller than was the case with hedges. The coefficients in Table 4 also show that the main effect of LEVEL is positive, indicating an increase of boosters across the levels of proficiency; the interaction between LEVEL and TASK presents a slightly more complex picture, however. More specifically, the frequency of boosters seems to increase with proficiency when TASK is either 'complaint' or 'opinion'; the tendency is especially pronounced with respect to the latter. For 'letter', however, the line drawn between the different levels of proficiency in Fig. 2 is, again, a slightly declining one, indicating that in this particular task, the number of boosters shows signs of slight decrease across the proficiency levels.

As 'complaint' serves as the intercept in the model, the estimated frequency of boosters in this context can be calculated based on the main effect of LEVEL. Thus, the estimated increases of booster frequency for each one-unit increase of LEVEL is 0.08 boosters phw in this task. The most common boosters used in the complaints are listed in Table 13.

Based on Table 13 it seems that adverbial boosters (e.g., *really, always, of course, clearly*) seem to be particularly common in the complaints. Of these, for example *really* is used on all levels of proficiency to a varying degree, and is typically used to underline the hurt caused by the subject of the complaint, as illustrated by (14). *Clearly*, on the other hand, seems particularly characteristic of C1 and C2, and it is often used to emphasize how obvious a mistake made by the service provider is.

(14) This really was the worst holiday ever. (41607-2-B2)
(15) […] they are clearly translated by someone who does not know the language. (57451-2-C2)

In the letter task, the frequency of boosters shows a very slight decline across the different levels of proficiency, as the estimated frequency of boosters in the task is 0.65 on level B1 and 0.62 on level C2. The most common boosters in the letters listed in Table 12 also show that there are remarkable similarities across the groups in the kinds of boosters preferred, although it does also seem that while the groups B1 and B2 rely on the adverb *really*, C1 and C2 writers more often opt for the modal auxiliary *will*.

Of these, the modal auxiliary *will* is often used to reassure the recipient of something turning out for the best rather than to claim authority as illustrated by (16). On levels B1 and B2, the most common booster is *really*, which in turn is often used for emphasis, as in (17).

(16) It will surely be fun to work with you (58981-1-C2)
(17) I really fell in love with that city. (37573-1-B2).

In the opinion task, the estimated frequency of boosters increases with proficiency – in fact, as both the main effect of TASK and its interaction with LEVEL are positive, there is a noticeable increase of booster frequency across the levels of proficiency: when TASK = opinion and LEVEL = B1, the estimated frequency of boosters is 0.77, while the same estimate for level C2 is 1.16. Interestingly, the most common boosters used in the task are, nevertheless, remarkably similar across the levels of proficiency (Table 14), as *will, really, always* and *of course* are among the most common boosters in all groups.

In the opinion tasks, boosters are mainly used to frame assertions in a very prototypical manner, which echoes the observations made with respect to the use of hedges in the same task (see Section 4.1); when using boosters, the purpose of the writer is mainly to underline the writer's confidence in the claim presented, as illustrated in (18), (19) and (20).

(18) It is known that smoking kills. (37577-3-B2)
(19) […] and they will definitely change the world-for the better. (37718-3-C1)
(20) It is, of course, clear that a healthier life-style, will lead to less disease. (85630-3-C2)

The opinions are also characterized by the use of multiple boosters within a single sentence, as shown in (21) and (22); this tendency seems to be particularly prominent on the advanced levels of proficiency, which is likely, in turn, to be related to the increased booster frequency on these levels as well.

Table 12
Most common boosters in the 'letter' task (raw frequency, frequency phw).

B1	B2	C1	C2
1. really (21, 0.16)	1. really (72, 0.28)	1. will (92, 0.25)	1. will (32, 0.17)
2. know (14, 0.11)	2. know (33, 0.13)	2. know (36, 0.10)	2. really (20, 0.11)
3. will (8, 0.06)	3. will (28, 0.11)	3 really (33, 0.09)	3. sure (10, 0.05)
4. of course (5, 0.04)	4. sure (20, 0.08)	4. always (32, 0.09)	4. know (10, 0.05)
5. sure (4, 0.03)	5. always (11, 0.04)	5. sure (26, 0.07)	5. actually (8, 0.04)

Table 13
Most common boosters in the 'complaint' task (raw frequency, frequency phw).

B1	B2	C1	C2
1. really (21, 0.16)	1. never (9, 0.04)	1. really (36, 0.10)	1. clearly (13, 0.07)
2. know (14, 0.11)	2. of course (8, 0.03)	2. will (19, 0.05)	2. never (9, 0.05)
3. will (8, 0.06)	3. really (7, 0.03)	3. clearly (19, 0.05)	3. will (9, 0.05)
4. of course (5, 0.04)	4. always (5, 0.02)	3. fact (16, 0.04)	4. really (7, 0.04)
5. sure (4, 0.03)	5. will (5, 0.02)	4. know (14, 0.04)	5. fact (7, 0.04)

Table 14
Most common boosters in the 'opinion' task (raw frequency, frequency phw).

B1	B2	C1	C2
1. will (55, 0.42)	1. will (106, 0.42)	1. will (125, 0.34)	1. will (74, 0.40)
2. really (23, 0.18)	2. really (44, 0.17)	2. always (64, 0.17)	2. really (34, 0.19)
3. always (17, 0.13)	3. always (31, 0.12)	3 really (46, 0.13)	3. always (29, 0.16)
4. of course (12, 0.09)	4. of course (22, 0.09)	4. of course (46, 0.13)	4. fact (23, 0.13)
5. true (11, 0.08)	5. know (17, 0.07)	5. fact (38, 0.10)	5. of course (13, 0.07)

5. Discussion

The present study set out to examine the frequency of both hedges and boosters in written L2 English in relation to task type, level of L2 proficiency and individual authorial style of the writer. In this context, it seems that the frequency of EDs is largely governed by task type, which has been found to be a defining factor in the use of EDs in previous studies as well (e.g., Gablasova et al., 2017; Gray and Biber, 2014; Pérez-Paredes and Díez-Bedmar, 2019; Yoon, 2020). This is true for both hedges and boosters; in both cases, the main effect of task type elicited the biggest differences in the models. It also seems that hedges are used more frequently than boosters across the board, which is also in line with previous research (e.g., Gablasova et al., 2017; Hyland and Milton, 1997; Kärkkäinen, 2003; McEnery and Kifle, 2002).

In addition to task type, the level of proficiency was found to be a significant predictor of the frequency of both hedges and boosters, both in its own right and in interaction with task type; proficiency-related differences typically seemed to be much smaller than those elicited by task type alone, however, at least with respect to the levels of proficiency examined in the present study (i.e., B1-C2). The present study thus lends support to previous studies reporting task-related differences in the use of EDs (e.g., Crosthwaite and Jiang, 2017; Gablasova et al., 2017; Hinkel, 2009; Pérez-Paredes and Díez-Bedmar, 2019), while also showing that task and proficiency level interact in a meaningful way. In this sense, the results are thus in partial contrast to for example Yoon (2020), who examined learner writing on CEFR levels A1-B1.2 and found that while topic was a statistically significant predictor for the frequency of both hedges and boosters, no such effect could be observed for the proficiency level of the writer. Nevertheless, as discussed above, it could be that perhaps the use of EDs (both hedges and boosters) becomes a more defining feature on the higher levels of proficiency (i.e., B1 onwards). This interpretation also gains support from the findings of for example Pérez-Paredes and Díez-Bedmar (2019), who reported an increase in the frequency of the certainty adverbs *really, actually* and *obviously* in spoken L2 English across the levels B1-C1.

In the present study, the highest frequencies of both hedges and boosters were found in the opinion task, which is not particularly surprising, as the task is designed to directly invite the writers to argue for or against a certain stance and is thus likely to elicit more stance expressions as well. The effect of the interaction between level and task was also most substantial in the opinion task, as the frequencies of both hedges and boosters showed the greatest proficiency-related increases in this task. The complaint and letter tasks, on the other hand, showed lower frequencies for both hedges and boosters, but the two tasks also differed in how learners on different levels of proficiency responded to them; while the frequencies of both hedges and boosters showed little change across the different levels of proficiency in the letter task, both ED types showed a steady increase with proficiency in the complaint task. A similar trajectory of proficiency-related increase was also found with respect to the opinion task, although in the opinion texts, the greater increases were observed with respect to boosters, while in the complaints, the increase was particularly prominent with respect to hedges.

The results of the present study also suggest that hedges and boosters are used for different purposes across the different tasks, which lends support to the notion that EDs are used to express a multitude of different meanings in addition to the general expression of (un)certainty (e.g., Hyland, 2005, p. 133). In the complaints, for example, hedges were often used to soften accusations, while in the letters, they were used to make it possible for the recipient to deny a request. Boosters, on the other hand, were typically used to for example underline the hurt caused in the complaints, while in the opinions, they were used to indicate certainty of a certain event taking place in the future, for example. Of the tasks examined here, the complaints can be seen as particularly tricky; on the one hand, the writer needs to maintain a positive relationship with the recipient to ensure cooperation, often resulting in the use of tentative

expressions, but on the other hand, they also need to seem authoritative enough to warrant the compensation they are looking to receive. The different uses of EDs thus underline the need to modulate one's expression to fit the communicative needs of each task, and the increases in observed with proficiency indicate that perhaps more proficient learners are more adept in adapting their expression in this sense.

The present study also analyzed the role of individual differences in the form of random effects in the mixed-effects models. Based on the results, the variation introduced by individual writers in the data is very small despite the high standard deviations observed in the use of both hedges and boosters. More specifically, the introduction of the random effect for each individual author increased the explanatory power of the model used to model the frequency of hedges by very little, meaning the variable accounted for a very small portion (i.e., 0.2 percentage points) of the variance observed. As for boosters, the increase in the explanatory power of the model following the inclusion of individual authors was more noticeable (i.e., 5.1 percentage points), which suggests that individual authorial style might be more influential in explaining the variation related to the use of boosters than that of hedges. The results obtained here thus echo the results of previous studies that have reported high intra-group variation in the use of EDs (e.g., Aijmer, 2011; Uccelli et al., 2013), while also showing that 1) individual differences do not seem to have an important role in explaining this variation and 2) the influence of individual differences may vary with respect to different kinds of expressions (i.e., hedges vs boosters). The results obtained here also question the role of individual authorial styles in the use of EDs as suggested by Gablasova et al., 631), who examined the phenomenon using correlational analyses to examine the use of individual EDs in spoken L2 English. It could, however, be that the influence of individual differences is stronger in spoken language and/or is tied to the use of individual expressions rather than the general frequency of EDs examined here (see also Pérez-Paredes and Díez-Bedmar, 2019).

On a more general level, the statistical models examined here were able to describe roughly 21% of the variation related to hedges and 16% of the variation related to boosters, meaning there is still quite a bit of unexplained variation with respect to both types of EDs. This can perhaps be partly attributed to the inherently "messy" nature of linguistic data, which often results in comparatively low explanatory values (Winter, 2020, p. 177), but it also indicates that task type, level of proficiency and individual authorial styles can only account for a part of the variation observed, meaning there is still a need for future studies examining the influence of additional variables. For example, the present study employs a cross-sectional (versus longitudinal) design, making it impossible to examine the author × proficiency interaction, which could also play an important role in how EDs are used across the levels of proficiency. As previous studies also indicate that there are differences in how different L1 groups use EDs (e.g., Hyland and Milton, 1997; McEnery and Kifle, 2002) and that topic and L1 also seem to interact in a meaningful way (e.g., Hinkel, 2009; Yoon, 2020), it would be interesting to more closely examine whether proficiency seems to play a part in this interaction in as well (i.e., L1 × proficiency), especially in a longitudinal setting (for a cross-sectional examination, see e.g., Yoon, 2020). In the present study, each writer also contributed only one text per each task type examined, making it impossible to examine how individual authorial style influences the use of EDs within multiple texts of the same task type, which might also be an interesting avenue for further studies.

The present study is subject to some limitations. Of these, the most prominent is the issue of the multiple topics within each task type; despite the shared communicative purpose within each task type (complaint, letter, opinion), there is some variation in the micro-level topics within each subcorpora (as discussed in Section 3). It is important to note, however, that there is no overlap between the different task types with respect to the communicative purpose despite the variation of topic within each task. It could, however, be that even fairly fine-grained topic differences introduce additional noise in the data (see e.g., Hinkel, 2009; Yoon, 2020), which may also, in turn, be related to the relatively low explanatory power of the statistical models examined here. The results presented here are also mostly limited to the overall frequency of hedges and boosters; an interesting point of further study would be one where the types and choice of EDs are examined from a more variationist point of view or with respect to the variety of EDs used in the individual essays.

6. Conclusion

The results of the present study have implications for (learner) corpus and research design. First, as the most notable differences observed here were all related to task type, the results of the present study support the notion that task type should be carefully controlled for when comparing different groups of writers. More specifically, some of the differences observed between different (L1 and/or L2) groups in previous studies could simply be due to differences in task type or the communicative purpose of the texts between the two subsets of data. Secondly, the results of the present study also suggest that the influence of authorial style matters a great deal less than for example the type of task at least from the point of view of ED frequency, and the present study thus lends support to the observation made by Cvrček et al. (2020, p. 484), who note that "in corpus design, representative corpora should be built with an emphasis on diversity of registers rather than diversity of authors, as the former introduces more variation than the latter".

The results of the present study also relate to the importance of register awareness in foreign language teaching and learning, which is also reflected in the CEFR Companion Volume with respect to sociolinguistic appropriateness, for example, where the key concepts include the appropriate use of politeness forms and language functions and "adopting an appropriate register (from B2)" (Council of Europe, 2020, p. 137). In this respect, the implications of the study are twofold: on the one hand, the increase in the frequency of EDs with proficiency suggests that, at least in the case of the complaint and opinion tasks, the learners seem to become more sensitive for the need to modulate their expression with respect to both hedges and boosters. On the other hand, it also seems that level-related differences in the preferred ED types (i.e., the most common hedges/boosters) are fairly subtle, which indicates that in this sense, the learners seem to be fairly equally equipped to deal with the different communicative needs related to different task types.

The results also suggest that EDs are used for various micro-purposes across the different tasks in addition to signaling a general sense of (un)certainty, which is directly related to the ability of conveying "finer shades of meaning" mentioned in the CEFR. This should also be accounted for in language teaching by drawing the students' attention to the different roles that EDs can play in different kinds of texts. Previous studies also show that there are differences between different L2 groups how EDs are used, which further underlines the importance of understanding and disentangling the role of proficiency from other possible variables related to the matter in this sense. This, in turn, would allow a more targeted approach as to how communicative competence is taught to different groups.

Declaration of Competing Interest

None.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Appendix A. A lemmatized list of the epistemic devices examined

Type	Expression
Modal auxiliary (12)	*can, could, may, might, must, shall, should, will, would, have to, ought to, be going to**
Lexical verb (32)	*appear, argue, assume, believe, claim, consider, contend, doubt, estimate, expect, feel, find, guess, imagine, indicate, know, predict, presume, propose, reckon, say, see, seem, sound, speculate, suggest, suppose, suspect, tend, think, understand, wonder*
Adverbial (57)	*absolutely, according to, actually, almost, all the time, always, apparently, approximately, arguably, as far as*, assumedly, basically, certainly, clearly, commonly, correctly, definitely, doubtless, essentially, evidently, frequently, generally, hardly, hypothetically, indeed, in general, in theory, largely, likely, maybe, necessarily, never, normally, mainly, mostly, nearly*, obviously, of course, often, partially*, perhaps, possibly, practically, presumably, probably, quite, rarely, really, relatively, roughly, some, sometimes, surely, totally*, truly, undoubtedly, usually*
Adjective (35)	*accepted, actual, apparent, arguable, aware, bound, certain, clear, confident, convinced, correct, doubtful, evident, false, guaranteed, hypothetical, inevitable, liable, (un)likely, obvious, positive, (im)possible, (im)probable, plain, prone, proved, reliable, seeming, sure, true, well-known*
Noun (44)	*argument, assessment, assumption, belief, certainty, chance, claim, conclusion, conviction, discovery, doubt, estimate, estimation, evidence, explanation, fact, feeling, idea, inclination, knowledge, likelihood, observation, opinion, possibility, potential, prediction, presumption, principle, probability, proposal, realization, rumor, sense, sign, statement, suggestion, suspicion, tendency, theory, thesis, threat, understanding, view, viewpoint**

*added to the list after the manual analysis.

References

Aijmer, K., 2011. Well I'm not sure I think… The use of well by non-native speakers. Int. J. Cor. Linguist. 16 (2), 231–254. doi:10.1075/ijcl.16.2.04aij.

Anthony, L., 2020. *Antconc* (3.5.9) [Macintosh OS X]. Waseda University.

Bartley, L., Hidalgo-Tenorio, E., 2016. Well, i think that my argument is…," or modality in a learner corpus of English. Rev. Espan. Linguist. Apli. 29 (1), 1–29. doi:10.1075/resla.29.1.01bar.

Bax, S., Nakatsuhara, F., Waller, D., 2019. Researching L2 writers' use of metadiscourse markers at intermediate and advanced levels. System 83, 79–95. doi:10.1016/j.system.2019.02.010.

Biber, D., 2006. Stance in spoken and written university registers. J. Eng. Acad. Purp. 5 (2), 97–116. doi:10.1016/j.jeap.2006.05.001.

Biber, D., Conrad, S., 2009. Register, Genre, and Style. Cambridge University Press.

Biber, D., Johansson, S., Leech, G.N., Conrad, S., Finegan, E., 2021. Grammar of Spoken and Written English. John Benjamins Publishing Company doi:10.1075/z.232.

Chen, Y.H., Baker, P., 2016. Investigating criterial discourse features across second language development: lexical bundles in rated learner essays, CEFR B1, B2 and C1. Appl. Linguist. 37 (6), 849–880. doi:10.1093/applin/amu065.

Council of Europe, 2001. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge University Press.

Council of Europe, 2020. Common European Framework Of Reference For Languages: Learning, Teaching, Assessment – Companion Volume. Council of Europe Publishing.

Crosthwaite, P., Jiang, K., 2017. Does EAP affect written L2 academic stance? A longitudinal learner corpus study. System 69, 92–107. doi:10.1016/j.system.2017.06.010.

Cvrček, V., Laubeová, Z., Lukeš, D., Poukarová, P., Řehořková, A., Zasina, A.J., 2020. Author and register as sources of variation: a corpus-based study using elicited texts. Int. J. Cor. Linguist. 25 (4), 461–488. doi:10.1075/ijcl.19020.cvr, John Benjamins.

Gablasova, D., Brezina, V., McEnery, T., Boyd, E., 2017. Epistemic stance in spoken L2 english: the effect of task and speaker style. App. Linguist. 38 (5), 613–637. doi:10.1093/applin/amv055.

Gray, B., Biber, D, 2014. Stance markers. In: Corpus Pragmatics: A Handbook. Cambridge University Press, pp. 219–248. doi:10.1017/CBO9781139057493.012.

Gries, S.T., 2021. Generalized Linear) mixed-effects modeling: a learner corpus example. Lang. Learn. doi:10.1111/lang.12448.

Gries, S.T., 2018. On over- and underuse in learner corpus research and multifactoriality in corpus linguistics more generally. J. Second Lang. Stud. 1 (2), 276–308. doi:10.1075/jsls.00005.gri.

Hinkel, E., 2009. The effects of essay topics on modal verb uses in L1 and L2 academic writing. J. Pragmat. 41 (4), 667–683. doi:10.1016/j.pragma.2008.09.029.

Hyland, K. (2005). *Metadiscourse*. Continuum.

Hyland, K., Milton, J., 1997. Qualification and Certainty in L1 and L2 Students' Writing. J. Second Lang. Writ. 6 (2), 183–205. doi:10.1016/S1060-3743(97)90033-3.

Ivaska, I., 2022. Regression Analysis. Handbook of Pragmatics Online In J.-O. Östman & J. Verschueren (Eds.). John Benjamins Publishing Company doi:10.1075/hop.24.reg2.

Kärkkäinen, E., 2003. Epistemic Stance in English Conversation: A Description of Its Interactional functions, with a Focus on I think. John Benjamins Publishing Company.

Kirkham, S., 2011. Personal style and epistemic stance in classroom discussion. Lang. Literature 20 (3), 201–217. doi:10.1177/0963947011413505.

Larsson, T., 2019. Grammatical stance marking across registers. Regist. Stud. 1 (2), 243–268. doi:10.1075/rs.18009.lar.

McEnery, T., Kifle, N.A., 2002. Epistemic modality in argumentative essays of second-language writers. In: Academic Discourse. Pearson Education, pp. 182–215.

Murakami, A., 2016. Modeling systematicity and individuality in nonlinear second language development: the case of english grammatical morphemes: modeling individual nonlinear development. Lang. Learn. 66 (4), 834–871.

Paquot, M., 2018. Phraseological competence: a missing component in university entrance language tests? Insights From a Study of EFL Learners' Use of Statistical Collocations. Lang. Assess. Q. 15 (1), 29–43. doi:10.1080/15434303.2017.1405421.

Paquot, M., Plonsky, L., 2017. Quantitative research methods and study quality in learner corpus research. Int. J. Lear. Corpus Res. 3 (1), 61–94. doi:10.1075/ijlcr.3.1.03paq.

Pérez-Paredes, P., Bueno-Alastuey, M.C., 2019. A corpus-driven analysis of certainty stance adverbs: obviously, really and actually in spoken native and learner English. J. Pragmat. 140, 22–32. doi:10.1016/j.pragma.2018.11.016.

Pérez-Paredes, P., Díez-Bedmar, M.B., 2019. Certainty adverbs in spoken learner language: the role of tasks and proficiency. Int. J. Learn. Corpus Res. 5 (2), 257–279.

R. Core Team, 2021. R: a Language And Environment For Statistical Computing. R Foundation for Statistical Computing.

Staples, S., Egbert, J., Biber, D., McClair, A., 2013. Formulaic sequences and EAP writing development: lexical bundles in the TOEFL iBT writing section. J. Engl. Acad. Purp. 12 (3), 214–225. doi:10.1016/j.jeap.2013.05.002.

Uccelli, P., Dobbs, C.L., Scott, J., 2013. Mastering academic language: organization and stance in the persuasive writing of high school students. Writt. Commun. 30 (1), 36–62. doi:10.1177/0741088312469013.

Winter, B., 2020. Statistics For Linguists: An Introduction Using R, 1st ed. Routledge doi:10.4324/9781315165547.

Winter, B., Grice, M., 2021. Independence and generalizability in linguistics. Linguistics 59 (5), 1251–1277. doi:10.1515/ling-2019-0049.

Wulff, S., Gries, S.Th, 2021. Exploring individual variation in learner corpus research: methodological suggestions. In: Learner Corpus Research Meets Second Language Acquisition. Cambridge University Press; Cambridge Core, pp. 191–213. doi:10.1017/9781108674577.010.

Yoon, H.-J., 2020. Interactions in EFL argumentative writing: effects of topic, L1 background, and L2 proficiency on interactional metadiscourse. Read. Writ. doi:10.1007/s11145-020-10085-7, *0123456789*.