

Developing a coding scheme for annotating opinion statements in L2 interactive spoken English with application for language teaching and assessment

Yejin Jung – Dana Gablasova – Vaclav Brezina – Hanna Schmück
Lancaster University / United Kingdom

Abstract – Evaluative meanings are known to be difficult to identify and quantify in corpus data (Hunston 2004). The research in this area has largely drawn on the annotating schemes offered by the frameworks of Appraisal (Read and Carroll 2012; Fuoli 2018) or stance (Simaki *et al.* 2019). However, these annotation schemes have been applied predominantly to written production and to first language use. This study, therefore, proposes an annotation scheme for identifying and classifying linguistic expressions of opinion with particular application for second language (L2) language teaching and language assessment contexts. In addition, the coding scheme also specifically deals with spoken interactive communication, with particular attention paid to aspects such as the co-construction of opinion statements (Hovarth and Eggins 1995). The paper outlines the components of the coding scheme along with their theoretical underpinning, addresses some of the challenges in applying the codes and annotating real-life data, and discusses future possibilities and considerations related to the application of the coding scheme.

Keywords – evaluative language; linguistic expression of opinion; coding scheme; L2 pragmatic ability; spoken corpora; learner corpora

1. INTRODUCTION¹

Expressing evaluative meanings, defined as language which serves to express “a speaker’s attitude, stance, viewpoint, or feelings on entities or propositions” (Hunston and Thompson 2000: 5), is an integral part of human communication. Indeed, as Thompson and Alba-Juez (2014: 5) point out, “finding a text or even a sentence without any trace of evaluation is a very challenging, if not impossible task.” Evaluative language has been analysed in different genres, such as legal communication (Goźdz-Roszkowski 2018), academic writing (Jiang and Hyland 2015), and media discourse (Bednarek 2006),

¹ We would like to thank the ESRC Centre for Corpus Approaches to Social Science (CASS) at Lancaster University and the Trinity College London, for permitting data access for this study. We are also very grateful to the anonymous reviewers and the editors, Robbie Love and Carlos Prado-Alonso, for their valuable comments.



and across different modes of communication (written, spoken, and online), see Greenberg (2000) or Mullan (2010). Due to its central role in communication, the ability to express views also features prominently in the domains of language teaching and assessment, with a number of courses, textbooks and exams highlighting this language function. For example, the ability to state and support an opinion is included as an indicator of L2 communicative competence in widely used language proficiency frameworks such as the *Common European Framework for Reference* (Council of Europe 2020) and the *American Council on the Teaching of Foreign Languages* oral proficiency guidelines (ACTFL 2024). In addition to this, many standardised language proficiency tests —such as the *International English Language Testing System* (IELTS 2019), the *Graded Examinations in Spoken English* (GESE; Trinity College London 2024), the *Test of English as a Foreign Language* (TOEFL; Educational Testing Service 2018)— ask test takers to express opinions in order to evaluate their linguistic ability.

Linguistic expression of opinion represents a complex language function, which draws on interaction of social, cognitive and linguistic resources, making it challenging to identify and classify different types of opinion statements. Several coding frameworks have been developed (e.g., Martin and White 2005; Wiebe *et al.* 2005; Gray and Biber 2012) to operationalise different types of evaluative language, including expressions of opinion. Building on this research, the current study proposes an annotation scheme for the identification and classification of linguistic expressions of opinion with particular application in second language (L2) teaching and assessment contexts. Such research is crucial for setting curricular goals in the teaching of communicative skills as well as when assessing different stages of communicative and interactional competence in language tests (Roever 2011; Galaczi 2014). In addition to contributing to a better understanding of L2 pragmatic ability, the coding scheme also addresses interactive spoken communication, a genre characterised by frequent exchange of views between interlocutors (Biber *et al.* 2021). This study, therefore, seeks to develop a coding scheme that is applicable in contexts characterised by a high degree of turn-taking and co-construction of discourse between two or more interlocutors. To reflect these aims, the data used in the study were selected from the *Trinity Lancaster Corpus of spoken interactive L2 English* (TLC; Gablasova *et al.* 2019). The study will first present the broader framework and rationale for the coding scheme and then introduce the specific components of the scheme along with examples of the coding. It will next focus on an

empirical evaluation of the coding scheme and discuss the challenges of applying it to L2 data and interactive spoken production.

2. THEORETICAL FRAMEWORK FOR THE ANNOTATION SCHEME

2.1. *Principles of defining linguistic expressions of opinion*

As a first step in the coding scheme development, it is necessary to identify the major principles for defining opinion statements, which will then serve as a general framework for the annotation scheme. The construct of evaluative language has been investigated and/or operationalised in a number of widely-used frameworks, representing different theoretical and methodological approaches. For example, the Appraisal framework (Martin and White 2005), based on Systematic Functional Linguistics, has been influential in identifying and categorising different types of evaluative (emotive and attitudinal) statements, which include evaluation of people's characteristics and the qualities of objects/entities. In addition to Appraisal, Wiebe *et al.* (2005) created a coding scheme for application in an NLP context which focuses on automatic identification of internal mental and emotional states such as opinions, beliefs, thoughts, emotions, sentiments, and speculations, with particular attention paid to the classification of intensity and polarity of these attitudes. Using corpus linguistic methodology, research on stance such as Gray and Biber (2012) categorised different types of speaker position adopted towards a statement or an entity, with a particular focus on the lexical and grammatical resources used for indexing perspective. Beyond this, Hyland (1998, 2005) developed a categorisation for different expressions indicating different types and degrees of speaker/writer stance and engagement (e.g., hedges and boosters).

Drawing on this body of research from different fields, several principles for defining opinion and distinguishing it from other types of evaluative language can be formulated. First, when investigating language related to expressing opinions, it is important to distinguish between two phenomena (Bednarek 2009): 1) an expression of opinion that refers to the psychological reality of forming and expressing an evaluative judgment, and 2) the linguistic expression of opinion by speakers/writers. While acknowledging the complexity of cognitive, psychological, social, and linguistic processes involved in the formation and expression of opinions, this study focuses only on the second category.

Second, in order to identify an instance of evaluative language, previous frameworks relied on a combination of two key approaches: 1) the presence of explicit linguistic expressions that mark values, subjectivity, and stance, and 2) contextual clues. For example, Wiebe *et al.* (2005) detect private states using three types of linguistic markers: explicit mentions of private states, speech events expressing private states, and expressive subjective elements. The markers that have been closely related to evaluative language include value-laden words (e.g., *great*, *horrible*) and stance markers indicating (un)certainity towards a proposition (e.g., *maybe*, *really*), see Biber *et al.* (1999). While the presence of such words may provide explicit clues to the occurrence of evaluative statements, evaluative judgments can be also implied via other words and contexts (see Section 3.2.1 for examples).

Finally, it is necessary to distinguish expressions of opinion from other related types of evaluative language, especially that of linguistic expression of affect, since the two constructs have been interlinked, to some degree, in several coding frameworks (e.g., Martin and White 2005; Wiebe *et al.* 2005). These approaches reflect the complex relationship between emotions and opinions, and the fact that many evaluative statements (e.g., *Violeta is a fine person* or *I like Violeta*) can be linked to a speaker's positive/negative emotions toward an entity/proposition. However, a distinctive feature of an opinion statement is that it is not a purely emotional reaction to an entity/proposition, but it also involves a cognitive process (Bednarek 2006). It has been argued that an opinion involves a value judgment, which entails an (implicit or explicit) comparison between the object of evaluation and a norm (Labov 1972; Martin and White 2005). For example, the evaluative statement *Violeta is a fine person* involves the process of comparing Violeta against the normative principles of 'being fine'. On the contrary, an emotional state or process does not necessarily involve a comparison against (implied or perceived) norms. As a result, statements which include an explicit reference to what is traditionally considered an emotion (e.g., *I love it*), see Bednarek (2008) or Mackenzie and Alba-Juez (2019), can be distinguished from the expression of opinion, although it is acknowledged that a linguistic expression of opinion can be based on both a cognitive and an emotive response to the entity/proposition that is being evaluated. Examples (1)–(3) illustrate the type of evaluative statements that include linguistic cues (underlined) indexing emotion, which were thus excluded from the coding of opinion statements. All examples in this paper are taken from the *Trinity Lancaster Corpus* (TLC), described in

Section 4.1. The individual ID of each text is given in the square brackets after the extract. In the transcripts, Speaker 1 (S1) always denotes an L1 speaker, while Speaker 2 (S2) refers to an L2 speaker.

- (1) S2: Venice is my favourite city [6_SP_1]
- (2) S2: and and I I love the culture of New York because there's a lot of s=
every of people [2_7_SP_47]
- (3) S2: I would like to design modern architecture because I like how the
architects er er play with er geometric shapes and forms [2_7_SP_8]

2.2. *Defining linguistic expression of opinion in language teaching and assessment contexts*

The ability to express opinions is a key part of L2 users' communicative and interactional competence, reflecting the stage of their linguistic/pragmatic development (Galaczi 2014). This ability, therefore, represents a central concern in many language teaching and testing contexts with a focus on L2 communicative strategies. For example, the task specification for the spoken component in the IELTS exam states that “the ability to communicate opinions and information on everyday topics and common experiences and situations” is one of the main skills assessed in the task (IELTS 2019: 5). In the spoken part of the Aptis General exam, “the candidate gives opinions and provides reasons and explanations” (O’Sullivan and Dunlea 2015: 22). In the GESE test, the test takers are expected to “communicate facts, ideas, opinions and attitudes about a chosen topic” (Trinity College London 2024: 38). In these speaking tests, L2 users are typically asked to state and support their opinions as well as engage with the opinions expressed by other interlocutors.

Despite the role played by this particular language function (i.e., the expression of opinion), there seems to be only a limited body of research on how to reliably identify and evaluate opinion statements in these contexts. Previous studies addressed different aspects of evaluative language use by L2 speakers such as stance-taking and expressions of (dis)agreement (Iwasaki 2009; Fordyce 2014; Galaczi 2014; Bardovi-Harlig *et al.* 2015; Gablasova *et al.* 2017; Fogal 2019; Pérez-Paredes and Bueno-Alastuey 2019). However, this research mostly focused on forms related to linguistic evaluation and provided only a limited insight into the nature of communicative strategies related to opinion stating by L2 speakers at different proficiency levels. The annotating scheme

presented in this study has therefore been developed to enable researchers and practitioners to capture and assess different aspects of communicative strategies employed by L2 speakers in interactive communication. In particular, it allows researchers to 1) measure the frequency of opinion statements expressed by the speakers, 2) measure the complexity of their expressions of opinion, and 3) record whether L2 speakers are able to state their opinions independently in the course of interaction or whether additional support from the teacher/examiner may be required to elicit the views. The components in this scheme allow researchers to distinguish, for example, L2 speakers who may have the linguistic resources to formulate a simple opinion statement, but who may not have appropriate pragmatic knowledge to manage intersubjective relations (e.g., employ politeness strategies to mitigate the impact of their expressed views) or who may struggle to express their opinions in an ongoing, fast-paced conversation. Such findings may support previous research outcomes that showed that highly proficient L2 speakers engage in opinion exchange in more collaborative and reciprocal manners than intermediate-level speakers (Galaczi 2014).

3. CODING SCHEME

3.1. General approach to the annotation methodology

The main aim of this paper is to introduce and evaluate a scheme that pays special attention to L2 production in an interactive context, with a direct application in language teaching and assessment. The current scheme builds on previous research on annotation of evaluative language and linguistic expression of opinion, while taking into consideration the specific pedagogical and assessment concerns discussed above. To reflect this purpose, the scheme addresses three dimensions: 1) stating of an opinion, 2) providing a support for the opinion statement, and 3) the interactional pattern in which the opinion statement/support was produced.

3.2. Annotation categories

3.2.1. Opinion statements

A linguistic expression of speaker opinion, or an opinion statement, is defined in this study as a speech act/language function that informs the listener about a speaker's opinion

expecting no specific response or action (as opposed to a question or a directive), see Biber *et al.* (2002). To code an utterance as an opinion statement, the following three conditions must be satisfied:

1. An opinion statement involves a marker or markers of social values, subjectivity, or comparison (Hunston and Thompson 2000).
2. An opinion statement is attributed to the speaker, not someone/something else (Sinclair 1986; Hunston 2000).
3. An opinion statement is syntactically declarative (Biber *et al.* 2002).

First, an opinion statement should include a value judgement on the object or situation being evaluated; that is, it includes an evaluation in terms of good/bad, positive/negative, important/unimportant, reliable/unreliable, certain/uncertain (Hunston and Thompson 2000; Bednarek 2006). An opinion statement thus must contain at least one grammatical or lexical item that indexes values (e.g., *great*), subjectivity (e.g., *my*) or comparison (e.g., *just*, *never*), see Hunston and Thompson (2000). There is no exhaustive list of such items and thus the interpretation may vary depending on the context. While some lexical and grammatical items may be more explicitly evaluative in meaning (e.g., *best*, *beautiful*), others may be more context-dependent such as modal verbs (e.g., *may* and *could*). (Bednarek 2006; Thompson and Alba-Juez 2014). Example (4) demonstrates the use of the word *special*, which in some contexts can serve an evaluative purpose (in the sense of ‘extraordinary’ or ‘remarkable’), but in the context of this example fulfils a descriptive function without implying a judgement (i.e., it refers to something with a specialised purpose).

- (4) S2: in my free time I work in a shop of jumping clay [...] okay jumping clay is a special clay which you can mould and create a lot of things like objects for the kitchens er would you like to see something made with jumping clay? [2_7_SP_4]

Example (4) illustrates the fact that the presence of a lexical item by itself may not be sufficient to identify an opinion statement and that the context of the utterance needs to be analysed for further evidence of evaluation, such as markers of comparison to an implied norm or a reference point.

Second, an opinion is inherently personal (Myers 2004) and, therefore, to distinguish an expression of opinion from other types of (evaluative) statements, only the evaluative statements whose communicative context allows them to be attributed to the

speaker as the source of the judgement will be considered an opinion statement. There are various ways of (not) indicating that a value judgment is from the speakers themselves, described using the constructs of ‘evidentiality’ (Chafe and Nichols 1986), ‘subjectivity’ (Traugott 2010), or ‘epistemic stance’ (Biber *et al.* 1999). For example, speakers may mark the source of information (e.g., *according to x*) when making an evaluative comment to indicate how certain the speaker is about the truth or validity of the comment (Biber *et al.* 1999). Speakers may also choose to adjust the extent to which they appear responsible for their value judgment by attributing it to themselves (e.g., *from my perspective*) or to other sources (e.g., *the government said*), see Hunston (2000) or Sinclair (1986). Thus, if a speaker makes an evaluative statement but attributes the value judgement or perspective to a different entity, the comment may be classified as an expression of value, but not as a linguistic expression of the speaker’s opinion (e.g., *Spanish people think that the new law has a positive influence*). A range of linguistic markers can be used to determine whether a statement satisfies the second criterion and whether the speakers are expressing their own opinion rather than reporting views attributable to another source. For example, these are linguistic items such as first-person pronoun (*I, my*) and/or epistemic stance markers (e.g., *I think, I believe, from my perspective* and *in my opinion*). When the statement explicitly identifies a source of view other than the speaker, it is not coded as an opinion statement. Examples (5)–(6) demonstrate such use, with the source of the view underlined.

(5) S2: they say is really difficult to pass [2_7_SP_8]

(6) S2: oh er in Galicia it’s er it’s said that er recycling’s not er an important question [2_7_SP_49]

However, it should be also noted that it is relatively common for speakers not to explicitly indicate the source of information. Without any explicit marking, the opinion statement is therefore attributed to the speaker. This applies also in the cases when the opinion statement occurs as a direct response to the other interlocutor’s question/request for an opinion as in such situations, the speakers are considered to express their own opinions.

Finally, the last condition requires the opinion statement to be realised in the form of a declarative sentence (Biber *et al.* 2002). As a result, the expression of opinion potentially implied in different types of questions will not be coded as opinion statements.

Examples (7)–(9) illustrate such uses in different types of yes-no questions (7), wh-questions (8), and tag questions (9).

(7) S2: Do you think advertising is necessary? [SP_112]

(8) S2: What do you think of advertising? [SP_108]

(9) S2: Lovely isn't it? [2_6_SP_66]

In addition, an evaluative comment realised in a directive/imperative manner (e.g., *be careful when you buy clothes*) is not considered as an expression of opinion.

3.2.2. Opinion supporting statements

A supporting statement is defined as a statement that provides a supportive or background information for an opinion statement. A speaker may make a supportive move in consideration or anticipation of a listener's response (Edmondson 1981). The discourse function of these supporting statements is two-fold. First, they provide additional information about the nature of the opinion statement and, second, they play a role in the intersubjective and interactional dimension of the communication. For example, they can serve as face-saving devices used to mitigate the social impact of an opinion statement (e.g., to weaken or strengthen it), see Blum-Kulka *et al.* (1989).

In the current study, two conditions are required for a statement to qualify as a supporting statement. First, it has to satisfy the condition of providing additional information about the opinion statement that preceded it. Second, a functional link between the two statements—the opinion statement and the supporting statement—has to be identifiable (e.g., from the presence of an explicit marker or contextual clues). It is also possible for a supporting statement to contain an opinion statement. However, if the two conditions listed above are fulfilled (i.e., the second opinion statement provides additional information about the preceding statement and there is evidence of a functional link between the statements), the second statement will be coded as a supporting statement.

Based on the previous literature (Carlson and Marcu 2001; Galaczi 2014) and on a small-scale grounded analysis of a sample of the data, five main types of supporting statements have been included in the coding scheme. These are supporting statements

expressing 1) reason, 2) elaboration, 3) contrast, and 4) evidence. The fifth category, ‘other’, is used where none of the previous types of supporting statement can be applied.

The category of ‘reason’ involves supporting statements in which the speaker provides a cause, motivation, or background for the opinion previously stated. Many supporting statements in this category are explicitly marked by the use of *because*, as illustrated in example (10).

- (10) S1: yeah okay so it was list A designer goods okay so what do you think about erm designers who spend a lot of energy and time into designing beautiful expensive clothes and then having them copied and sold in the streets?

S2: like fakes

S1: fake yeah mm

S2: I think that it's not fair because someone is spending a lot of time doing an exclusive product for people who can afford it and then if er if other persons make fake of tho-those product they increase at no er the value of those first er products will become lower [2_SP_1]

The second major category, ‘elaboration’, involves the speaker providing additional information, or specific details about the opinion statement. In this category, the links between the information included in the opinion statement and the supporting statement can reflect different relationships (e.g., general-specific, whole-part, or object-attribute). This type of a supporting statement usually provides an example of what is stated in the opinion statement, as shown in example (11), or paraphrases it with the aid of different lexical items, as in example (12).

- (11) S1: ah yeah yeah and what about the alphabet?

S2: er that's very difficult

S1: yeah

S2: they have for example four As [6_SP_31]

- (12) S1: you have to trust the site yeah

S2: yes it's er er secure like buying on Zara or shops like that you know it's safe to do it but not er in a strange er shop [2_SP_5]

If the supporting statement involves ‘contrast’, the third major category, it includes a statement that contradicts the opinion statement or offers a different perspective on the evaluative judgement from the opinion statement, as demonstrated in example (13).

- (13) S1: she had a little bit of fat on her knee so she had an operation on her knees
- S2: I think that’s silly but I think <unclear text="this"/> kind of people have a lot of pressure and because they are all the time er in the media and everyone is looking at them I think they want to feel that I’m perfect
[2_SP_9]

The next category involves providing ‘evidence’; in these cases, the supporting statement indicates the source of information that the evaluation in the opinion statement is based on. Evidence differs from ‘reason’ in that it explicitly states external evidence such as data, sources, or numbers that demonstrate the validity of the statement. This use is shown in example (14).

- (14) S2: because alcohol really damage your health it’s prove by medicine
[SP_107]

Finally, the supporting statements that do not fall into these four major categories are coded as ‘other’. These include statements with some evidence that the speaker is attempting to provide a support for an opinion statement, but it is impossible to determine the nature of the support (e.g., as the statement remained incomplete). Reasons for incomplete supporting statements include interruption from the other interlocutor who then shifts the interaction to a different topic, as illustrated in example (15).

- (15) S1: It’s very difficult to explain this question because er it's e-er-ex=
- S2: because
- S1: <voc desc="laugh"/> <unclear text="bus driver"/>
- S2: to to to for me the this situation is out of control the military are going to come in [2_SP_32]

Following the definition of an opinion statement (see Section 3.2.1) and a supporting statement (Section 3.2.2), this study distinguishes two main types of opinion statements: 1) an opinion statement without a supporting statement, referred to as ‘simple opinion

statement’, and 2) a ‘complex opinion statement’, which consists of an opinion statement followed by a supporting statement.

3.2.3. Interactional context of the opinion/supporting statements

The third annotation category relates to the interactional context of the opinion or supporting statements. The main communicative function recognised in this category is whether an opinion statement/supporting statement was expressed spontaneously by the speaker (referred to as ‘unprompted’) or whether it followed directly from the other interlocutor’s question or request (referred to as ‘prompted’), see Degoumois *et al.* (2017). This dimension recognises that, while the frequency and type of opinion/supporting statements provide insights into L2 proficiency, the interactional context in which these statements occurred can provide additional understanding when interpreting the nature of these expressions of opinion. Previous studies have suggested that overall speaking proficiency might have an effect on L2 speaker’s ability to respond to interlocutors in ongoing interactions, with advanced speakers demonstrating a greater tendency to build on the utterances of previous speakers appropriately (e.g., Watanabe 2017; Abe and Roever 2019). This dimension is therefore crucial 1) to further our understanding of the nature of L2 speakers’ proficiency and ability to express opinions, and 2) to describe the nature of an opinion expression in an interactive communication better. In addition to providing insights into L2 communicative abilities, this dimension is important in order to understand the co-constructed nature of opinion expression in language assessment contexts. For example, it can contribute to a better understanding of Oral Proficiency Interviews as a genre of language assessment discourse and its communicative features related to the test takers at different proficiency levels.

The prompted opinion/supporting statement is identified if the other interlocutor directly asks the speaker to express their views (e.g., *what do you think about this?*) or to express a support for an opinion statement (*why do you think so?*) and the speaker provides such statement as a response, as shown in example (16). When no such preceding prompt is identified in the conversation, the opinion/supporting statement is coded as unprompted.

(16) S2: very difficult

S1: yeah so why why is it difficult? how is it difficult?

S2: because it's completely different to Spanish [2_6_SP_31]

It should be noted that in some contexts it may be difficult to determine whether an opinion/supporting statement is prompted or not. First, a request for an opinion can be stated in syntactically varied ways (Green 2014). For example, while it most commonly takes the form of an interrogative (e.g., *do you think it's an important skill?*), it can also occur as an imperative (e.g., *tell me about Euro*), a declarative (e.g., *we'll talk about tourism and particularly the negative side of tourism okay some people say that tourism can do more harm than good*), or a declarative with a tag question (e.g., *very hard, isn't it?*). Thus, for example, in the case of the declarative sentence above, it remains uncertain whether the speaker is asking for an opinion or just stating a proposition. In the example including an imperative, it is unclear whether the speaker is asking for an opinion about Euro or for something else (e.g., facts).

The coding scheme had to consider also the instances where one speaker asks for an opinion related to a particular topic, but the other speaker offers an opinion about a different topic, as shown in example (17):

(17) S1: how im= how important is it for you to earn a good salary?

S2: It's a good question [2_6_SP_67]

In example (17), S1 is asking for an opinion about a specific proposition (i.e., earning a good salary) but S2 responds with an opinion statement about the question itself. The coding schemes regards such responses as an unprompted opinion statement. An opinion/supporting statement is considered prompted if the object of evaluation (either an entity or a proposition) is introduced by the first speaker. On the other hand, the statement is considered unprompted if a new object of evaluation is introduced by the second speaker in their response.

As stated above, one of the aims of developing this coding scheme was to make it applicable in spoken interactive communication. Considering that it can be quite challenging to segment streams of L2 spoken data due to repetitions, ellipses, or false starts (Foster *et al.* 2000), the current coding scheme clarifies how to deal with a number

of challenges related to syntactic or interactional realisations of opinion such as elliptical responses and coordinated phrases.

First, a common pattern that occurs specifically in interactive communication is that of elliptical responses (Chia and Kaschak 2023). In the coding scheme, affirmative expressions such as *yes*, *no*, *of course*, and other similar variants are considered (and coded as) instances of opinion statements if they occur as a direct response to an interlocutor's prompt, as illustrated in S2's response in example (18).

- (18) S1: but do you do you think that it's a good idea to have pocket money maybe whe= maybe in the future?
- S2: yes because for example if you go with your friends or with any other people to anywhere and you have to for example buy something or take a taxi to ... [2_6_SP_31]

Second, in terms of a range of syntactical types of opinion/supporting statements, these can be realised by noun phrases (example 19) and relative clauses (example 20) if it is clear from the context what entity or proposition is being evaluated.

- (19) S2: and okay problems with the prices problems with the mortgages [SP_35]
- (20) S2: you have to be able to talk to the audience, which is very difficult [2_6_SP_5]

Next, a common issue in interactive communication is presented by coordinated clauses or words. In utterances where evaluative meanings are coordinated (usually by a connector *and*), an opinion statement should evaluate a single entity or proposition. When two entities or propositions are evaluated in a coordinated clause, they are considered as two instances of opinion statements, as shown in examples (21) and (22):

- (21) S2: I I think that is a job very interesting (okay) (opinion statement 1) and I think that I a= I am a a person qualificate for this job (opinion statement 2) [6_SP_29]
- (22) S2: so that it's difficult to ask a question to the teacher (opinion statement 1) and and speak to the teacher alone (opinion statement 2) [2_6_SP_62]

Finally, in interactive communication, it is important for the scheme to systematically address co-constructed meaning. This is a situation where two speakers collaboratively

develop an evaluative statement, which thus cannot be attributed to only one of the speakers. Example (23) demonstrates such pattern, which is not treated as an opinion statement in the proposed scheme.

- (23) S1: oh yes
 S2: that
 S1: it's safer
 S2: is safer
 S1: isn't it??
 S2: than than other ways to
 S1: okay okay
 S2: to pay and with children er I think that's people have to to be
 S1: be a bit more careful
 S2: to be care
 S1: don't they?
 S2: yes [6_SP_60]

4. EVALUATION OF THE CODING SCHEME: AN EMPIRICAL STUDY

4.1. Data

The study used a subset of the *Trinity Lancaster Corpus* (Gablasova *et al.* 2019), reflecting the aim to make the scheme applicable to teaching/testing contexts. The corpus consists of 4.1 million words from the transcriptions of over 2,000 dyadic interactions recorded as part of the GESE, an international exam of spoken English, developed and administered by Trinity College London (Trinity College London 2024). The corpus contains data from L2 speakers from different L1 backgrounds and three main proficiency levels of the Common European Framework of Reference: B1 (pre-intermediate), B2 (intermediate), and C (advanced, comprising C1 and C2 levels).

For the evaluation of the coding scheme 29 transcripts from the TLC were selected. Each speaking exam in the GESE involves one L2 speaker (the test taker) and one L1

speaker (the examiner). Data from two speaking tasks were used in the study: 1) the conversation task, in which the interlocutors exchange and discuss their views on general topics selected by the examiner, and 2) the discussion task, in which the topic is selected and introduced by the test taker. Both tasks are highly interactive and thus offer a communicative environment in which opinions on a number of topics are stated and discussed by the two speakers. Only the data from the L2 speakers were used in this study, even though the contributions from the L1 speakers were taken into consideration in the interpretation and coding of the production of L2 speakers. The 29 transcripts represent data from three proficiency levels: ten transcripts at the B1, ten at the B2, and nine at the C level of proficiency. The transcripts were also selected to represent L2 English speakers from two L1 backgrounds —Chinese (14) and Spanish (15)— in order to include L2 production from typologically different L1s and cultural backgrounds, which could affect communicative preferences and strategies of L2 speakers.

4.2. Procedure

4.2.1. Coding procedure

Initially, two coders independently coded the sample to identify opinion statements. Between the two coders, 320 opinion statements were identified in the utterances produced by L2 speakers in 29 transcripts. The first coder was an expert coder involved in the development of the scheme; the second coder was then trained to apply the scheme. The training involved multiple rounds of practice coding sessions, in which the second coder applied the scheme and compared the results to the first coders' results, and differences between the codes were discussed. During these sessions, both typical (model) examples of each component of the scheme were used as well as more borderline cases. During each stage, the second coder compared their annotations to the first coder. The differences were discussed with reference to the coding scheme and the criteria were further clarified. After this process, the second coder proceeded to code the 29 transcripts independently. Inter-rater agreement statistics (agr, AC1, and Cohen's kappa) were calculated to evaluate the consistency of the application of the coding scheme.

The coding procedure first involved each coder identifying an instance of an opinion statement and, second, deciding whether it is a case of a simple or a complex opinion statement. The simple opinion statement consists only of the main opinion

statement (example 24), while a complex opinion statement consists both of the opinion statement and a supporting statement (example 25).

(24) S2: that was horrible [2_SP_1]

(25) S2: it was very tiring we had to go like seven hours a week [2_6_SP_5]

The instances of complex opinion were then coded further for the type of supporting statement following the categories presented in Section 3.2.2. As a final step, the context in which both simple and complex opinions were produced, was further explored and both types of opinion expressions were coded as prompted or unprompted. Figure 1 provides an overview of the main categories of the scheme; an overview of all types of opinion and supporting statements involved in the scheme is available in the Appendix.

| | | |
|-----------------------------|--------------------------------------|------------|
| Expression of opinion | Simple opinion statement | Prompted |
| | | |
| | Complex opinion statement | Unprompted |
| | | |
| | | Prompted |
| | | |
| | | Unprompted |
| | | |

Figure 1: Coding scheme for expression of opinion: Main categories

4.2.2. Inter-rater agreement: general considerations

Identifying and classifying expressions of opinion is a challenging task due to the complexity and, in many cases, also the subjectivity of the decision-making involved. Identifying opinion statements is considerably different from the coding schemes where the choices are limited (e.g., when the annotators select one option from a closed set) or when the coding is applied to entities (units) that have already been identified (e.g., the annotators are asked to select a category to which a certain word belongs). When identifying opinion statements in a continuous interactive discourse, annotators have to make decisions based on a potentially unlimited list of expressions that can act as markers of evaluative judgements in given contexts. Further, there are no pre-set boundaries or a finite number of entities to code (e.g., an utterance could contain several opinion statements or none). Such coding, therefore, involves high-inference decisions and has been observed to result in a greater difference in agreement than the situations when

annotators deal with well-defined (or pre-defined units) and closed sets of attributes to apply in coding (Allwood *et al.* 2007; Read and Carroll 2012).

Statistically, the agreement measure *agr* seems to be the most appropriate to quantify inter-rater agreement in these situations. *Agr* is a directional measure of agreement, using one rater as a baseline for the other rater's performance (Wiebe *et al.* 2005: 196). *Agr* is particularly suitable in cases such as pragmatic annotation of corpora, where raters are tasked with the identification of linguistic features in the flow of discourse rather than applying a coding scheme to a given set of cases (e.g., Wilson *et al.* 2006). The measure allows evaluating consistency of application of a particular coding scheme in the following way from the perspective of different raters:

Rater A taken as a baseline:

$$agr_1 = \frac{A \text{ matching } B}{A}$$

Rater B taken as a baseline:

$$agr_2 = \frac{B \text{ matching } A}{B}$$

Two values of *agr* are therefore typically reported and sometimes interpreted as precision and recall. They can be combined in a single F1 measure (Van Rijsbergen 1979; Fuoli and Hommerberg 2015) as follows:

$$F1 = \frac{2 \times agr_1 \times agr_2}{agr_1 + agr_2}$$

In addition, inter-rater agreement about further classification of opinion statements, once identified, can be measured using standard statistics such as Cohen's kappa (κ) and AC1 (Brezina 2018: 90–91). Kappa and AC1 are appropriate in this case, as once the raters identified an opinion statement, they are then selecting from a closed set of two options: presence or absence of a supporting statement/prompt. While AC1 represents a recent and sophisticated measure of inter-rater agreement, estimating agreement by chance more precisely in extreme cases (Brezina 2018), Cohen's kappa has been used in this study as well in order to allow for comparison with other earlier studies that opted for this measure. In Cohen's kappa, values close to 0 indicate that the agreement is most likely due to chance, while the values close to 1 indicate a very strong to full agreement between the

raters. In order to interpret the level of agreement, we follow Rietveld and van Hout (1993), who interpret the values in the following way: 0–0.20 indicate a small degree of agreement, 0.21–0.40 means fair agreement, 0.41–0.60 are considered to represent a moderate agreement, while 0.61–0.80 suggest a strong agreement between the raters. AC1 operates on the same scale and has therefore a similar interpretation.

4.3. Empirical results

4.3.1. Agreement on the identification and type of opinion statements

Table 1 provides an overview of the results, showing the per cent agreement, agr, Cohen's kappa, and AC1 based on the rating of 320 cases identified by two coders.

| Inter-rater agreement | | | | | |
|---|-------------|-------------|------|---------------|----------------|
| | % agreement | Agr | F1 | K | AC1 |
| Presence of opinion statement | N/A | 0.52/0.64* | 0.57 | N/A | N/A |
| Presence of opinion statement (intersecting cases counted as agreement) | N/A | 0.59/ 0.74* | 0.66 | N/A | N/A |
| Presence of a supporting statement | 79.1 | N/A | | 0.5 (p<.001) | 0.64 (p<.001) |
| Presence of a prompt | 83.58 | N/A | | 0.66 (p<.001) | 0.068 (p<.001) |

Table 1: Inter-rater agreement on three dimensions of the coding scheme for expression of opinion.

(*The first value refers to the agr metric results based on the first coder taken as the baseline and the second value uses the second coder as the baseline)

We will first examine the level of agreement regarding whether a stretch of text was considered to qualify as an opinion statement. Overall, in the strictest interpretation, the agr metric shows that, taking the first rater as the baseline, over 50 per cent agreement between the raters (0.52) was reached; taking the second rater as the baseline, even a stronger agreement (over 0.64) was reached between the raters. The combined F1 score for this situation is 0.57. While there was a relatively good agreement between the raters, in addition to the opinion statements identified by both raters, each rater also identified additional statements that went unnoticed by the other rater. A closer inspection of the results revealed that there were two systematic issues that affected inter-rater agreement. The first involved distinguishing the nature of the evaluative statement (i.e., an opinion statement vs. a supporting statement) when dealing with interconnected expressions of opinion. The second issue involved dealing with boundaries of opinion statements, that

is, a situation where both raters identified the presence of evaluative language but differed in how they determined the boundaries of an opinion statement. Several coding schemes that dealt with classifying utterances or expressions in larger stretches of naturally-produced discourse reported similar issues that affected the rate of agreement regarding the identification and counting of the target units (e.g., Wiebe *et al.* 2005; Allwood *et al.* 2007; Read and Carroll 2012). These studies approached the issue at the level of coding and considered “intersecting text as agreement” (Read and Carroll 2012: 433; see also Wiebe *et al.* 2005). When intersecting cases were considered as agreement (following Wiebe *et al.* 2005 and Read and Carroll 2012), the agr metric reported that, taking the first rater as the standard, a fairly strong agreement between the raters (nearly 0.6) was reached; taking the second rater as the standard, a substantial agreement (over 0.7) was reached between the raters. The combined F1 score was 0.66 in this case. The issues of establishing the boundaries and distinguishing between interconnected opinion statements are further discussed in Section 4.3.2 below.

Regarding the presence of supporting statements, a relatively high per cent agreement (nearly 80 per cent) was achieved between the raters, with kappa and AC1 indicating a fairly strong agreement. When inspected more closely, some of the disagreement between the raters was related to the issue of interconnected opinion statements described above. In these cases, both raters recognised a stretch of discourse as being related to an opinion statement; however, while one of the raters coded the two utterances as separate opinion statements, the other rater coded the following utterance as a supporting statement.

Finally, focusing on the presence of a prompt for an expression of opinion, the results showed a high per cent agreement (above 80 per cent), while kappa and AC1 also indicated a substantial agreement between the raters. Further analysis revealed that the disagreement between the raters was mostly due to the distance of a prompt from the expression of opinion statement, with prompts separated by several turns from the expression of opinion being most often coded differently by the coders.

4.3.2. Discussion of results and implications

Overall, the study identified the strengths as well as challenges to be further addressed when applying the coding scheme for identifying expressions of opinion for pedagogical and language testing purposes.

First, it demonstrated the challenges of identifying instances of opinion statements in continuous, interactive discourse, characterised by chaining of utterances and dynamic topic development. In addition to dealing with the typical features of spoken production such as unfinished or rephrased utterances, a further difficulty was presented by some features of learner language (e.g., utterances with semantic, lexical, or grammatical errors). These in some cases created additional challenges for understanding the meaning of an utterance, as demonstrated in example (26) taken from an L2 speaker in the data.

(26) well that I I that I I'm <pause> well I am a very late very late person

In some cases, therefore, these features affected the ability of the raters to understand and interpret the meaning of the utterances and to apply the coding scheme with confidence. It is possible that these challenges —present in spoken interactive communication— led each rater to identify a set of additional cases in line with their specific interpretation of the flow of the discourse. Many of these were, upon a review by the other rater, considered to be acceptable instances of opinion statement. This pattern suggests that it would be beneficial for several coders to code each transcript highlighting ‘candidate’ opinion statements according to the scheme criteria. As a second stage, either an expert coder could review these candidate statements and make a final decision to accept/reject, or the two (or more) raters could further review each other’s candidate statements and, if needed, discuss the rationale behind their decision. This approach would likely result in a higher recall rate of opinion statements marked by explicit as well as more subtle clues.

Another systematic issue that affected the inter-rater agreement on the identification of the opinion and supporting statements was related to the ability to determine the boundaries and relationships between individual opinion statements. Given the nature of the discourse in this study (i.e., interactive communication), it was relatively common for a speaker to express an opinion and follow it by a supporting statement, which would then lead to another opinion statement, often on a related topic. This pattern appeared especially typical of advanced L2 speakers in the data, whose turns tended to be longer and more complex, making it more challenging to distinguish between individual

opinion/supporting statements. This issue is illustrated in example (27), which is characterised by the proximity of several opinion statements as part of connected speech.

- (27) S1: okay thank you very much now let's talk about erm money =
 S2: money er a money in my life it is important but not much but is important because I I have to buy everything with money it's necessary
 S1: okay do you have a job? [6_SP_1]

In this example, both adjacent statements satisfy the criteria for identifying an opinion statement, as discussed in Section 3.2.1: 1) *a money in my life it is important but not much but is important* and 2) *it's necessary*. The first of these statements is also followed by a supporting statement of reason (*because I I have to buy everything with money*). While the annotators agreed on the first statement and the supporting statement, disagreement tended to appear when dealing with statements such as *it's necessary*. A possible difference was related to the question of whether this statement represents a continuation of the preceding utterances (i.e., of the supporting statement, expressing justification for the opinion statement) or a new opinion statement, closely related to the previous topic. On the one hand, it could be considered as a new opinion statement, as it satisfies the relevant criteria; on the other hand, it could be viewed as part of the supporting statement in that it further elaborates on the supporting statement and provides additional justification.

Classifying such adjacent evaluative statements is challenging as they are often related to the same conversational topics, as shown in example (27) above. When two propositions evaluate the same entity, they often contain pronouns that refer back to the entity mentioned previously (e.g., *it*) or semantically related lexis (e.g., *important* and *necessary*). As a result, such stretches of thematically connected utterances make it difficult for annotators to determine at what point one instance of an opinion statement ends and another one starts. Example (28) further demonstrates the issues involved in distinguishing between different evaluative comments about the same entity.

- (28) really because they have an singing teacher okay er is called [name] that she has a very beautiful voice she's really a she's a really good person and she helps us to sing really really well [2_6_SP_71]

In this example, the speaker is offering three propositions that qualify as an opinion statement according to the coding scheme: 1) *she has a very beautiful voice*, 2) *she is a really good person*, and 3) *she helps us to sing really really well*. These statements express value judgements about a teacher which are coordinated (e.g., they use the connector *and*). What is notable is that each statement evaluates the entity according to different attributes, for example, the voice quality, personality, and profession. This type of discourse raises the question of whether—if different statements evaluate an entity from three different perspectives—they should be viewed as three different instances of opinion. In its current form, the scheme does not explicitly instruct the coders on how to deal with such inter-connected statements. However, as the study demonstrated, such guidance is necessary in order to reliably address such utterances that are relatively common in interactive speech, where the meaning of utterances develops dynamically.

In general, the results suggested that the scheme can be applied with sufficient degree of reliability to the target communicative settings (i.e., L2 interactive production). However, following the findings of this study, it is recommended that the areas that represent systematic difficulties are further addressed in the coding scheme and in the guidance for the raters. In particular, the guidance for the raters should include several examples of evaluative statements that demonstrate stretches of discourse with several inter-connected opinion/supporting statements as well as examples with multiple opinion statements related to the same entity. Such guidance will likely result in a more reliable application of the coding scheme and greater agreement between the raters. However, it should be acknowledged that the high degree of subjectivity involved in investigating evaluative language will naturally result in a certain degree of difference between individual raters.

While the study brought encouraging results in terms of investigating expression of opinion in L2 production, there are also some limitations involved in the development and application of the scheme. First, the coding was restricted to the expression of opinion produced by L2 speakers (test takers), excluding the language produced by the L1 speakers (examiners) in the study, although their production was taken into consideration when classifying the L2 utterances, reflecting the co-constructed nature of the discourse. Second, the coding system was only evaluated in relation to the dyadic production involving two speakers taking turns, in a semi-formal environment. Its usability in different environments (e.g., with multiple L2 speakers, in informal conversations) should

be further investigated. Finally, the coding was based mainly on the linguistic information available in the transcribed speech, without access to paralinguistic (e.g., pauses, laughs) or non-linguistic features (e.g., gestures indications of speaker turn). Such information would be valuable in interpreting the nature of the utterances. It should be noted, however, that working with transcribed speech may reflect the resources available in the testing and teaching context.

5. CONCLUSION

This study described and evaluated a coding scheme for the expression of opinion, which was specifically designed for the identification of this pragmatic feature in spoken interactive discourse, although its use can be extended to other settings including written communication. The scheme thus complements and further broadens the resources available for investigating evaluative language in different contexts and for different purposes. A particular contribution of the scheme lies in its application to L2 production and to pedagogical/assessment settings, where expressing views plays a major role. The annotation scheme proposed in the study can lead to a better understanding of L2 pragmatic knowledge by providing a tool for systematically recording different types of opinion expressions and relating them to L2 proficiency (e.g., Jung 2024). The findings based on the scheme can be then applied in the development of teaching materials and for evaluating the validity of (speaking) tasks that focus on eliciting and assessing expressions of opinion.

REFERENCES

- Abe, Makoto and Carsten Roever. 2019. Interactional competence in L2 text-chat interactions: First-idea proffering in task openings. *Journal of Pragmatics* 144: 1–14.
- ACTFL. 2024. *ACTFL Proficiency Guidelines 2024*. American Council on the Teaching of Foreign Languages. https://www.actfl.org/uploads/files/general/Resources-Publications/ACTFL_Proficiency_Guidelines_2024.pdf
- Allwood, Jens, Loredana Cerrato, Kristiina Jokinen, Costanza Navarretta and Patrizia Paggio. 2007. The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation* 41: 273–287.
- Bardovi-Harlig, Kathleen, Sabrina Mossman and Heidi E. Vellenga. 2015. The effect of instruction on pragmatic routines in academic discussion. *Language Teaching Research* 19/3: 324–350.

- Bednarek, Monika A. 2006. Evaluation and cognition: Inscribing, evoking and provoking opinion. In Hanna Pishwa ed. *Language and Memory: Aspects of Knowledge Representation*. Berlin: Mouton De Gruyter, 187–221.
- Bednarek, Monika A. ed. 2008. *Emotion Talk Across Corpora*. London: Palgrave Macmillan.
- Bednarek, Monika A. 2009. Language patterns and attitude. *Functions of Language* 16/2: 165–192.
- Biber, Douglas, Susan Conrad and Geoffrey Leech. 2002. *Longman Student Grammar of Spoken and Written English*. London: Longman.
- Biber, Douglas, Jesse Egbert, Daniel Keller and Stacey Wizner. 2021. Towards a taxonomy of conversational discourse types: An empirical corpus-based analysis. *Journal of Pragmatics* 171: 20–35.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Blum-Kulka, Shoshana, Juliane House and Gabriele Kasper eds. 1989. *Cross-cultural Pragmatics: Request and Apologies*. New York: Ablex.
- Brezina, Vaclav. 2018. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press.
- Carlson, Lynn and Daniel Marcu. 2001. *Discourse Tagging Reference Manual*. https://web.archive.org/web/20170808131213id_/https://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf
- Chafe, Wallace and Johanna Nichols eds. 1986. *Evidentiality: The Linguistic Coding of Epistemology*. New York: Ablex.
- Chia, Katherine and Michael P. Kaschak. 2023. Elliptical responses to direct and indirect requests for information. *Language and Speech* 67/1: 228–254.
- Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Companion Volume*. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>
- Degoumois, Virginie, Cécile Petitjean and Simona Pekarek Doehler. 2017. Expressing personal opinions in classroom interactions: The role of humor and displays of uncertainty. In Simona Pekarek Doehler, Adrian Bangerter, Geneviève de Weck, Laurent Fillietaz, Esther González-Martínez and Cécile Petitjean eds. *Interactional Competences in Institutional Settings: From School to the Workplace*. Berlin: Springer, 29–57.
- Edmondson, Willis. 1981. *Spoken Discourse: A Model for Analysis*. London: Longman.
- Educational Testing Service. 2018. *The Official Guide to the TOEFL Test*. New York: McGraw-Hill Education.
- Fogal, Gary G. 2019. Tracking microgenetic changes in authorial voice development from a complexity theory perspective. *Applied Linguistics* 40/3: 432–455.
- Fordyce, Kenneth. 2014. The differential effects of explicit and implicit instruction on EFL learners use of epistemic stance. *Applied Linguistics* 35/1: 6–28.
- Foster, Pauline, Alan Tonkyn and Gillian Wigglesworth. 2000. Measuring spoken language: A unit for all reasons. *Applied Linguistics* 21/3: 354–375.
- Fuoli, Matteo. 2018. A stepwise method for annotating appraisal. *Functions of Language* 25/2: 229–258.
- Fuoli, Matteo and Charlotte Hommerberg. 2015. Optimising transparency, reliability and replicability: Annotation principles and inter-coder agreement in the quantification of evaluative expressions. *Corpora* 10/3: 315–349.
- Gablasova, Dana, Vaclav Brezina and Tony McEnery. 2019. The Trinity Lancaster Corpus: Development, description, and application. *International Journal of*

- Learner Corpus Research* 5/2: 126–158.
- Gablasova, Dana, Vaclav Brezina, Tony McEnery and Elaine Boyd. 2017. Epistemic stance in spoken L2 English: The effect of task and speaker style. *Applied Linguistics* 38/5: 613–637.
- Galaczi, Evelina D. 2014. Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics* 35/5: 553–574.
- Goźdź-Roszkowski, Stanislaw. 2018. Values and valuations in judicial discourse. A corpus- assisted study of (dis)respect in US Supreme Court decisions on same-sex marriage. *Studies in Logic, Grammar and Rhetoric* 53/1: 61–79.
- Gray, Bethany and Douglas Biber. 2012. Current conceptions of stance. In Ken Hyland and Carmen Sancho Guinda eds. *Stance and Voice in Written Academic Genres*. London: Palgrave Macmillan, 15–48.
- Green, Anthony. 2014. *Exploring Language Assessment and Testing: Language in Action*. London: Routledge.
- Greenberg, Joshua. 2000. Opinion discourse and Canadian newspapers: The case of the Chinese “Boat People.” *Canadian Journal of Communication* 25: 517–537.
- Horvarth, Barbara and Suzanne Eggins. 1995. Opinion texts in conversation. In Peter H. Fries and Michael Gregory eds. *Discourse in Society: Systemic Functional Perspectives*. New York: Ablex, 29–45.
- Hunston, Susan. 2000. Evaluation and the planes of discourse: Status and value in persuasive texts. In Susan Hunston and Geoff Thompson eds. *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press, 176–207.
- Hunston, Susan. 2004. Counting the uncountable: Problems of identifying evaluation in a text and in a corpus. In Alan Partington, John Morley and Louann Haarman eds. *Corpora and Discourse*. Berlin: Peter Lang, 157–188.
- Hunston, Susan and Geoff Thompson eds. 2000. *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press.
- Hyland, Ken. 1998. Boosting, hedging and the negotiation of academic knowledge. *Text* 18/3: 349–382.
- Hyland, Ken. 2005. Stance and engagement: A model of interaction in academic discourse. *Discourse Studies* 7/2: 173–192.
- IELTS. 2019. *Information for Candidates Introducing IELTS to Test-takers*. Cambridge: University of Cambridge Local Examinations Syndicate.
- Iwasaki, Noriko. 2009. Stating and supporting opinions in an interview: L1 and L2 Japanese speakers. *Foreign Language Annals* 42/3: 541–556.
- Jiang, Feng and Ken Hyland. 2015. ‘The fact that’: Stance nouns in disciplinary writing. *Discourse Studies* 17/5: 529–550.
- Jung, Yejin. 2024. *Examining L2 Speakers’ Expression of Opinion in the Trinity Lancaster Corpus*. Lancaster: Lancaster University Dissertation.
- Labov, William. 1972. *Language in the Inner City: Studies in the Black English Vernacular*. Pennsylvania: University of Pennsylvania Press.
- Mackenzie, Lachlan and Laura Alba-Juez eds. 2019. *Emotion in Discourse*. Amsterdam: John Benjamins.
- Martin, James and Peter White eds. 2005. *The Language of Evaluation: Appraisal in English*. London: Palgrave Macmillan.
- Mullan, Kerry. 2010. *Expressing Opinions in French and Australian English Discourse*. John Amsterdam: John Benjamins.
- Myers, Greg. 2004. *Matters of Opinion*. Cambridge: Cambridge University Press.

- O'Sullivan, Barry and Jamie Dunlea. 2015. *Technical Report: Aptis General Technical Manual*. Version 1.0. English Language Assessment Research Group: British Council.
- Pérez-Paredes, Pascual and M. Camino Bueno-Alastuey. 2019. A corpus-driven analysis of certainty stance adverbs: Obviously, really and actually in spoken native and learner English. *Journal of Pragmatics* 140: 22–32.
- Read, Jonathon and John Carroll. 2012. Annotating expressions of Appraisal in English. *Language Resources and Evaluation* 46/3: 421–447.
- Rietveld, Toni and Roeland van Hout. 1993. *Statistical Techniques for the Study of Language and Language Behaviour*. Berlin: Mouton De Gruyter.
- Roever, Carsten. 2011. Testing of second language pragmatics: Past and future. *Language Testing* 28/4: 463–481.
- Simaki, Vasiliki, Carita Paradis and Andreas Kerren. 2019. A two-step procedure to identify lexical elements of stance constructions in discourse from political blogs. *Corpora* 14/3: 379–405.
- Sinclair, John M. 1986. Fictional worlds. In Malcolm Coulthard ed. *Talking about Text: Studies Presented to David Brazil on His Retirement*. University of Birmingham: English Language Research, 43–60.
- Thompson, Geoff and Laura Alba-Juez eds. 2014. *Evaluation in Context*. Amsterdam: John Benjamins.
- Traugott, Elizabeth C. 2010. (Inter)subjectivity and (inter)subjectification: A reassessment. In Hubert Cuyckens, Kristin Davidse and Lieven Vandelanotte eds. *Subjectification, Intersubjectification and Grammaticalization*. Berlin: Mouton De Gruyter, 29–71.
- Trinity College London. 2024. *Examination Information: Graded Examinations in Spoken English (GESE)*. <https://www.trinitycollege.com/resource/?id=5755>
- Van Rijsbergen, Cornelis. 1979. *Information Retrieval*. London: Butterworth and Co.
- Watanabe, Aya. 2017. Developing L2 interactional competence: Increasing participation through self-selection in post-expansion sequences. *Classroom Discourse* 8/3: 271–293.
- Wiebe, Janyce, Teresa Wilson and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation* 39: 165–210.
- Wilson, Theresa, Janyce Wiebe and Rebecca Hwa. 2006. Recognizing strong and weak opinion clauses. *Computational Intelligence* 22/2: 73–99.

Corresponding author

Yejin Jung

Lancaster University

Department of Linguistics and English Language

County South

LA1 4YL

Lancaster

United Kingdom

Email: y.jung@lancaster.ac.uk

received: August 2023

accepted: April 2024

**APPENDIX: OVERVIEW OF THE DIFFERENT TYPES OF OPINION STATEMENTS (OS) AND
SUPPORTING STATEMENTS IN THE CODING SCHEME**

| Opinion Statements Categories | | |
|---|---------------------------|---|
| | Presence of prompt | Supporting statement Type of Opinion Statements |
| 1 | Not present | Not provided Unprompted Simple Opinion Statement |
| 2 | Not present | Provided Unprompted Complex Opinion Statement |
| 3 | Present | Not provided Prompted Simple Opinion Statement |
| 4 | Present | Provided Prompted Complex Opinion Statement |
| Supporting Statements Categories | | |
| 1 | Giving a reason | |
| 2 | Giving elaboration | |
| 3 | Giving a contrasting idea | |
| 4 | Giving evidence | |
| 5 | Others | |