CHAPTER 2

Testing Short-Text Multi-Dimensional Analysis
Approaching Learner Corpus Data at the Micro-Structural Level

2.1 Introduction

This chapter presents the first of our six chapters adopting a short-text MDA-driven approach to studying discourse in learner language. The primary aim of this chapter is to interrogate the efficacy of the short-text approach to MDA, discussed in Chapter 1, through the use of a severe test (McEnery and Brezina, 2022: 95–96). To do so, we undertake a micro-structural analysis of discourse at the turn level, which, in our view, constitutes a severe test as turns provide less data than would be available if we took a whole text, task or discourse unit view of the data. Moreover, if we carry out an analysis of discourse at the level of the largest micro-structure – the turn – we can gain a view of discourse at that level which we can then draw upon when undertaking a macro-structural analysis at the level of the discourse unit. When exploring the data to discern discourse functions at the turn level, we will bring a wide range of metadata into play to see to what extent the short-text MDA may give us insights into the impact of present variables on learner performance.

A secondary aim of this chapter is to provide a detailed outline of the short-text MDA procedure. For this first analysis, we provide a lot of detail regarding the dimensions themselves because, as well as exploring the dimensions observed, we also want to be clear about our method and provide to readers the data that will allow them, to a reasonable extent, to critically evaluate our findings. Subsequent short-text MDAs in this book are focused more on 'use' than 'proof', as they are the result of a similar procedure. As such, in the interests of space, we will not produce a step-by-step discussion of the steps undertaken in those later analyses, and we will trim back some of the reporting of features of the dimensions, relying on this first analysis as a template and justification for the analyses to come.

While the primary focus of this chapter is on learner discourse, it must be noted that examiners play a key role in the examination. Not only do

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

they contribute to discourse in the corpus, they often shape it. For example, we might expect them to produce explanatory sequences when they are introducing tasks, when they tell the student that the task is changing and when they introduce a new task. The social context warrants the examiner to perform those functions, not the student. Hence, we might expect, on these grounds alone, that the discourse functions of the examiner may differ from those of the examinee at the micro-structural level. This is not the case in macro-structural analyses – there the relative contribution of the examiner and examinee may vary, but the discourse units are always co-constructed. They are not monologic, the macro-structures are a blend of the examiner and examinee discourse functions at the micro-structural level. To permit a clearer view of this, while the main focus of this chapter will be on learner language, at the end of the chapter we will briefly introduce the result of the short-text MDA of examiner turns. To contextualise the analysis to come, we begin with a discussion of existing MDA-based research of discourse in learner language.

2.2 MDA of Learner Language

For reasons outlined in Chapter 1, MDA-based studies of learner language are the closest, methodologically, to what we may achieve with short-text MDA. There are, broadly, two approaches to such studies. The first approach is a full MDA. To recap, this is where dimensions of linguistic variation are computed by subjecting the relative frequencies of numerous lexico-grammatical features across the texts of a corpus to factor analysis. This analysis uncovers a series of dimensions comprising the major patterns of linguistic co-occurrence across that corpus. These dimensions are then interpreted as continuums of functional variation based on the notion that frequent patterns of co-occurring linguistic features tend to be motivated by at least one underlying communicative function (Biber, 1988). As discussed in Chapter 1, Biber (1988) used this approach on a corpus of spoken and written English and discovered six major dimensions of linguistic variation.

The second approach to MDA is not aimed at computing new dimensions of linguistic variation. Rather, it is concerned with comparing and projecting new texts or registers onto existing dimensions of linguistic variation by measuring the relative frequencies of the linguistic features used in the original analysis against those in new texts or registers. Then, based on the mean and standard deviation scores of the original analysis, factor scores of the new texts and registers are calculated (Berber-Sardinha, 2014),

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

enabling multi-dimensional functional descriptions of the ways in which the new texts vary with respect to the registers and texts used in the original analysis.

Both approaches to MDA have been applied to corpora of written and spoken learner language. For the most part, previous research has tended to favour projecting corpora of learner language onto existing dimensions of linguistic variation, especially dimensions of spoken and written English, as a way of investigating various research questions, such as (i) how the language produced in learner interviews compares to other spoken and written English registers (e.g. Aguado-Jiménez, Pérez-Paredes and Sánchez, 2012); (ii) how spoken learner language varies according to different elicitation tasks and in different test conditions (e.g. Connor-Linton and Shohamy, 2001); (iii) how native and non-native speakers vary in their use of particular communicative functions in essay writing (e.g. Van Rooy and Terblanche, 2006) and during elicitation tasks in interviews (e.g. Pérez-Paredes and Sánchez-Tornel, 2015); (iv) how writing varies in different kinds of learners of English (e.g. Pakistani, English as a Foreign Language (EFL), English as a Second Language (ESL), English as a Native Language (ENL)) with different cultural backgrounds (e.g. Abdulaziz, Mahmood and Azher, 2016); and (v) how the language produced by learners varies over time after English for Academic Purposes (EAP) instruction (e.g. Crosthwaite, 2016).

These studies demonstrate that language use varies according to a variety of factors, including task (written or spoken). For example, Connor-Linton and Shohamy (2001) projected the oral production interview of L1 Hebrew EFL students onto Biber's (1988) dimensions of linguistic variation of spoken and written English. They examined whether the communicative functions employed by the learners varied according to different elicitation tasks: talking about oneself and two role-playing tasks (complaining about noise and requesting an extension to a deadline). They found that complaints were less narrative and much more associated with situation-dependent reference than the other tasks, as complaints typically involve talking in the present tense about things in the immediate context. Requesting or talking about the self often concerned events in the past or future. Complaints and requests were more persuasive than telling about the self. While many of the linguistic co-occurrence patterns were used by the learners for the same communicative function as the written and spoken English registers in Biber's (1998) original analysis, Connor-Linton and Shohamy (2001) suggest that the features associated with Dimension 5 (abstract v. non-abstract) were potentially being used differently by the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

learners in comparison to native speaker registers. This is in part a motivation for the study in this book, as the finding indicates the need to investigate exactly how learners use language from a bottom-up perspective by conducting a full MDA as opposed to projecting texts onto existing dimensions based on native speaker data.

However, our study is not the first to take this bottom-up approach. Although less common, full MDA has been applied to explore corpora of written (e.g. Ascención-Delaney and Collentine, 2011; Friginal and Weigle, 2014) and spoken (e.g. Friginal and Polat, 2015; Friginal et al., 2017; Staples et al., 2017) learner language. For example, Ascención-Delaney and Collentine (2011) investigated the major communicative functions employed in the various writing tasks of second- and third-year university-level learners of Spanish. They found three major dimensions, including narrative vs. expository, descriptive expository prose and expository prose with a stance. They then explored the overall association of the different written text types to these dimensions. For example, they found that argumentative essays were more strongly associated with 'expository prose with a stance', whilst narrative texts were the text type least associated with this dimension.

In another study of learner writing, Friginal and Weigle (2014) conducted a longitudinal MDA of writing from L2 students in three time periods during a semester (early in the semester, midway through and late in the semester). They identified four dimensions and then assessed the relationship between both the average assessment scores and the date when the essay was written with the average dimension score of each essay. They found patterns between the average dimension scores of the learners' writing varied over time. For example, they found that the average score of the learners' first piece of writing was more involved. However, over time their writing got less involved and more informational. They found that this also correlated with the students' assessment scores. Essays from students with lower assessment scores were more involved, whilst essays that had higher assessment scores were informational.

A full MDA of spoken learner language was conducted by Friginal and Polat (2015). They used the Louvain International Database of Spoken English Interlanguage, which allowed them to use data from speakers with eleven different L1 backgrounds. Each speaker in the corpus performs three tasks – a set topic discussion, a free discussion and a picture description task. The set task is drawn by the L2 speaker from three options. Friginal and Polat aimed to investigate language variation in relation to differences in both the speakers' L1 background and the interview

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

tasks. They identified four dimensions of linguistic variation and then compared speakers of different L1 backgrounds by averaging the factor scores of the different groups. One finding was that, overall, Swedish learners of English were considerably more involved and conversational, whereas Japanese learners were more informational. Additionally, they identified that Dutch, Swedish, Chinese and Spanish speakers were the groups most strongly associated with the communicative function 'complex statement of opinion'. Italian was the least strongly associated with this function. Despite operationalising how speakers of different L1 backgrounds compare by averaging the factor scores of the different groups, Friginal and Polat's comparison of the different interview tasks in relation to the dimensions of linguistic variation is far less systematic, tending to be much more subjective. This is because they have factor scores for each interview as opposed to factor scores for the interviews separated by task. Accordingly, they were not able to compute average factor scores for the individual tasks, so any association of task to function is masked. They were only able to claim that particular linguistic co-occurrence patterns were more associated with the interview tasks through close reading of a sample of interview texts. So, while this study sheds some light on the relationship between L1 background and the use of particular linguistic repertoires, it also revealed a strong need to examine the major communicative functions at a more specific level than at the whole interaction level.

Friginal and Polat's study provides an important motivation for our use of short-text MDA to explore the Trinity Lancaster Corpus (TLC). As noted, their study suffers from a degree of data aggregation that masks an important dimension along which functions in their data may vary – task. In the TLC, tasks are clearly demarcated, allowing us to make a more systematic observation of how function varies according to task as well as other variables. By looking at this corpus of conversations between L1 examiners and L2 examinees, we can thus achieve the primary aim of this chapter by investigating the major communicative functions of learner language at a low level of aggregation – the turn level – in the corpus.

2.3 Corpus Analysis

We began our analysis by carrying out the short-text MDA procedure outlined in Chapter 1. That procedure relies, of course, on reasonably accurate part-of-speech tagging, so we undertook a short test of tagging

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

accuracy on the data. The precision of the tagger on the learner's turns is 0.97, which was calculated from an investigation of 100 random instances of each linguistic feature, whilst the recall rate is 0.91, which was calculated by assessing the number of false positives (incorrect tags) and false negatives (missed tags) in 100 random transcript lines. The overall F-score (a score which combines precision and recall) of the tagger on learner's turns is 0.94.[1] We deemed these results sufficiently accurate to proceed with our analysis, though in interpreting our results we were, of course, aware that errors in tagging could, in principle, impact our results. In particular, throughout this book we were mindful of this when undertaking the close reading of examples to interpret the dimensions that we explore.

After tagging, each turn in the training and test datasets was automatically analysed for the presence or absence of the 130 linguistic features encoded for our analysis.[2] These results were recorded in a categorical data matrix, where each row represented a turn in a transcript and each column was a linguistic feature. Le Roux and Rouanet (2010) advise that very infrequent features (e.g. those that occur in <5 per cent of the data) either need to be pooled with other related features or they might need to be discarded because infrequent features can overly influence the results of the Multiple Correspondence Analysis, as they contribute more to the overall variance. Thus, in line with Le Roux and Rouanet (2010), features that occurred in fewer than 5 per cent of the turns were either pooled with a more general grammatical category, if appropriate, or removed from the data matrix altogether, leaving forty-three linguistic features (see Appendix A for a description of the feature set and pooling decisions). Time adverbs, for example, did not occur in more than 5 per cent of the turns. As a result, these were pooled with other adverbs into a 'General Adverbs' category. In circumstances where a more general category did not exist, such as with reflexive pronouns, these features were discarded from the feature set. While we will not return to a discussion of this procedure, note that it is used in each of the short-text MDAs presented in this book.

We identified a number of features that may be productive in distinguishing the language use of the learners, including the speaker's L1, country of origin, grade/proficiency level and overall mark, as well as

[1] The F-score is used here in combination with the separate precision and recall scores. For a critical assessment of the use of the F-score, especially in a context where the other scores are not considered separately, see Powers (2011).

[2] The full tagset is given in Appendix B.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

various combinations of this information. Each turn from a learner was then analysed for the presence or absence of these features, as well as the interview task the turn came from (Greeting, Listen, Conversation, Discussion, Interactive Task, Presentation) and the length of the turn (in word tokens). This metadata was added to the data matrix. The final data matrix for the whole TLC training corpus is made up of individual turns, each assessed for length (in word tokens) and the presence or absence of 191 variables (43 linguistic features and 148 pieces of metadata). This data matrix was then subjected to Multiple Correspondence Analysis in R using the 'FactoMineR' package (Husson et al., 2020).

For our analysis we also needed to specify any supplementary variables; these can be qualitative and quantitative. Turn length was specified as a supplementary quantitative variable, whilst the other 147 metadata variables were qualitative. The reason for specifying turn length as a supplementary variable is owing to the possibility of text length confounding the analysis because it has not been controlled for in the analysis of the presence/absence of linguistic features. In particular, the short-text version of MDA used here does not analyse the relative frequencies of features. The relative frequencies of features are measured in standard MDA as a way to control for texts of different lengths. Measuring the relative frequencies of features, as opposed to their absolute frequency, means that texts of different lengths can be compared reliably as the frequencies of features are relative to the length of the text. Thus, by only measuring the presence or absence of features in this version of MDA, text length is not controlled for and could confound the analysis, as the more words a turn has the more likely it is to contain a variety of different linguistic features. Defining turn length as a supplementary variable enables the assessment of the degree to which turn length is correlated to the results of the analysis. Supplementary qualitative variables and turns were assigned coordinates revealing their association to the dimension patterns, enabling the assessment of the degree to which the task type and the speaker's L1, country of origin, grade, proficiency and overall mark are associated with the communicative functions of the turns.

In the corpus, the short-text MDA revealed forty-three dimensions of linguistic variation in descending order of importance. Each category of a linguistic feature (e.g. *presence* of Nouns and *absence* of Nouns) was assigned a positive or negative coordinate on each dimension. Additionally, each category of a linguistic feature was also assigned a contribution value for each dimension. Contributions show which categories

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

of features are the most important contributors to the dimensions. In this way, they are similar to factor loadings in factor analysis. Le Roux and Rouanet (2010) suggest that the categories of variables with a higher than average contribution should be interpreted, as these represent the patterns of variation with the most discriminatory power. All the contributions assigned to the categories of linguistic features for each dimension equal 100 and there are 2 categories for each of the 43 active variables (2 × 43 = 86). Therefore, the average contribution of a category is 1.16 (100/86). Categories of features with contributions above 1.16 were interpreted for their underlying function based on the notion of linguistic co-occurrence (Biber, 1988). Unlike factor loadings, the contributions do not have polarity. Thus, to interpret the dimensions, we used the coordinates of the categories of features in combination with their contributions, which reveal the distribution of the features across the turns. In this way, features that are distributed in similar ways have coordinates close to each other, and features that are not distributed in similar ways have coordinates that are far apart from each other, that is, in opposite quadrants (Le Roux and Rouanet, 1984, 2010). Specifically, we consider the linguistic features with above-average contributions, but we interpret those features with positive coordinates in opposition to those with negative coordinates as a continuum of variation.

To assist interpretation, we examined the turns strongly associated with the dimensions to view the co-occurring features in context. The short-text MDA also assigned contributions and coordinates to each turn on each dimension. Similar to the categories of linguistic features, the turns with high positive and negative coordinates that most strongly contribute to the dimension were then interpreted along with the features associated with the corresponding side of the dimension for their underlying communicative function. Throughout the book, we illustrate the functions we discuss by drawing on such examples, which we will call prototypical examples. To interpret our dimensions, starting with Dimension 1, each dimension was interpreted until the dimensions were no longer readily interpretable, as is common practice in MDA. In total, for the learner turns, we interpreted six dimensions. The sixth dimension was uninterpretable, indicating the end of our analysis. The first dimension primarily represented text length and so it was excluded, leaving four meaningful dimensions that characterised our data. These dimensions account for 93 per cent of the variance in our dataset, calculated using the modified rates for short-text MDA (Benzecri, 1992: 412).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

2.4 Results

Each interpreted dimension is presented in the following sub-sections. Each dimension includes a table consisting of the features most strongly contributing to it and a table showing the turns most strongly associated with the linguistic co-occurrence patterns. Each example given is followed by the name of the corpus file it is taken from. Following the description of the communicative function linked to the dimension's linguistic co-occurrence patterns, the association of the supplementary variables to the communicative function is also presented. Whilst the TLC consists of interviews from learners from over thirty-seven linguistic and twenty-six cultural backgrounds, the corpus does not always contain enough learners in a particular group to assess whether the patterns identified are predominantly a result of the learners' linguistic/cultural backgrounds, meaning that the patterns cannot be extended to other similar learners of this linguistic/cultural background (a known issue with learner corpora, as discussed by McEnery et al., 2019). For example, our corpus only includes one learner whose first language is Czech. The language produced by this learner will not only be influenced by their first language but also their grade and proficiency, among many other factors. As a result of so few learners being in this group, it would be impossible to assess if the overall association of this learner's turns were only a result of their first language. Therefore, our assessment of variation according to the different groups of learners will be limited to those groups that have enough learners and enough turns produced by those learners to sustain our analysis. Groups were included if their turns comprised more than 5 per cent of the corpus. So, for instance, whilst there is an adequate number of learners from grade 7 that scored an overall pass, those learners do not produce enough turns in the interactive task to be assessed and compared to other groups of learners in this task, whereas they produce enough turns in the discussion task to be included and compared. As will become apparent in the next section, at the micro-structural level this still leaves us with a wide degree of variation to consider.

Although this approach factors in the absences of features, which are arguably just as important as what is present, it is hard to interpret the function of the absence of features in the context of turns as absent features are unobservable. Moreover, features which are absent and are strongly contributing to one side of the dimension tend to also be present and strongly contributing to the other side of the dimension. Thus, to avoid repetition (i.e. discussing the potential function of the absence of features

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

on one side of the dimension and then discussing the function of their presence on the other), the absences of features are presented in this chapter, but ignored for the purpose of discussion and only the observable presences of features are interpreted.

In what follows, for each dimension explored we consider that the variation is caused by a series of variables, including task. The task variable is particularly important to consider as, should we see variation of discourse functions by task, it shows that there is an interaction between the functional analysis of utterances and the task-related functional purposes of the texts in the corpus. Given that previous studies have not been able to study this level in a bottom-up fashion well, if we are able to use the short-text MDA approach to analyse the impact of task, that would cast fresh light on a relatively neglected feature influencing learner performance.

A final note before we begin the discussion of our dimensions. Throughout the book, while a dimension is introduced with reference to its number, when discussed in the text the labels for either side of the dimension are used, rather than calling them 'positive Dimension 1' or 'negative Dimension 2' or some such. This is because we want to shift the discussion to the functional level and away from a simple discussion of rather abstract polarities and dimensions as quickly as possible. This will be done with every short-text MDA in this book. Later, when we are discussing and comparing short-text MDAs from different analyses, we also clearly note from which short-text MDA the function hails, for example, we may preface a function name with 'learner turns' or some other descriptive label. We do this not only because it will help readers refer back, where they wish, to the description of the function in question, but also, and importantly, because we will find that, as the book progresses, the same, or similar, functions arise in different analyses, and the short descriptive premodification of the function label thus becomes vital in distinguishing the same function relating to different analyses.

2.4.1 Dimension 1: Long Turns versus Short Turns

Positive Dimension 1 (Long Turns) is characterised by the *presence* of thirty-five linguistic features, whereas negative Dimension 1 (Short Turns) is characterised by the *absence* of five features. This suggests that Dimension 1 overall is reflecting variation in turn length, as typically the more words a turn has the more likely it is to have the presence of numerous linguistic features, as opposed to shorter turns, which will more likely be linked

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.1 *Results of the Pearson correlation between turn length (in word tokens) and turn coordinate for each dimension.*

	Dim. 1	Dim. 2	Dim. 3	Dim. 4	Dim. 5
Correlation between dimension coordinates and turn length	0.67	0.03	0.05	0.13	0.09

to the absence of features.[3] This interpretation is supported by the turns most strongly associated with Dimension 1. Those turns associated with the Long Turns are considerably longer than those classed as Short Turns. For instance, the turn most strongly associated with Long Turns is 304 words long, whilst the turn most strongly associated with Short Turns is 3 words long.

As mentioned, given that we did not measure the relative frequencies of features, we included turn length as a supplementary quantitative variable as a way of assessing whether turn length had confounded the analysis. Table 2.1 presents the results of the Pearson's correlation between turn length and the turn coordinates for each dimension. This table shows that Dimension 1 is most strongly positively correlated with turn length, thereby supporting the interpretation that Dimension 1 is largely reflecting turn length.

The strong correlation of turn length to Dimension 1 is because, as noted, the length of the turn is the strongest likely influence on the presence or absence of features. In other words, as turns get longer, they provide more opportunity for linguistic features to occur. The longest turn in the data (304 words) provides much more opportunity for a linguistic feature to occur than the shortest turn (3 words), in which the opportunity for any linguistic feature to occur is, necessarily, highly constrained. Yet the strong correlation with turn length is linked to turn length not being controlled for in the short-text MDA, which measures the presence or absence of linguistic features, as opposed to their frequency relative to the length of the text, as in standard MDA. Relative frequencies of features are analysed in standard MDA to compare texts of different lengths reliably. Given that considering the presence or absence of features alone does not control for text length, it is not surprising that turn length influences the results. However, apart from a slight positive association to

[3] This is a point we return to and explore a little more when considering Dimension 1 of the discourse unit analysis in the next chapter.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.2 *The linguistic features most strongly associated with Dimension 2.*

Dim. 2	Features (coordinates, contributions)
+	Non-Initial_Filler_A (0.178,1.262), Preposition_A (0.219,1.726), Initial_Filler_A (0.222,1.785), Subject Pronoun_P (0.285,1.522), General_Noun_A (0.369,3.626), Contrastive Conjunctions_P (0.563,1.207), BE as main verb_P (0.627,4.386), Private Verb_P (0.669,3.097), Analytic Negation_P (0.828,3.594), Auxiliary DO_P (0.845,2.424), Amplifiers_P (0.87,3.687), Complementation_P (adjective/verb + that complements + adj + to complement clauses + WH clauses) (0.929,2.54), Pronoun *it*_P (1.041,6.993), Contractions_P (pronoun, WH-word) (1.326,8.854), Predicative Adjective_P (1.415,10.319)
–	General_Determiner_P (–0.623,2.359), Non-Initial_Filler_P (–0.587,4.165), Proper Noun_P (–0.536,1.777), Definite Article_P (–0.517,2.661), Initial_Filler_P (–0.506,4.067), Preposition_P (–0.485,3.813), General_Noun_P (–0.386,3.79), Coordinating Conjunction_P (–0.37,1.523), General_Verb_P (–0.282,1.391), BE as main verb_A (–0.171,1.194)

Dimension 4 (discussed later), turn length only notably correlates with Dimension 1, suggesting that turn length has been largely controlled for in the first dimension. Because of this correlation and because the length of turn is not a functional interpretation, Dimension 1 is excluded from further linguistic interpretation here, though features of this discussion of it will have echoes in other discussions of similar dimensions in this book.

2.4.2 *Dimension 2: Involved versus Informational*

Dimension 2 is interpreted as opposing turns that have an Involved communicative function on the positive side with turns that are more Informational on the negative side. Turns that are Involved tend to encode one's personal stance and exhibit a more conversational style, whereas turns that are Informational are more informationally dense and tend to concern things other than personal feelings.

Table 2.2 shows the features associated with Dimension 2. In all such tables, P denotes a feature that is present and A one that is absent. Many of the features with positive coordinates in Table 2.2 are used in the turns to encode personal stance, such as BE as a main verb, private verbs, predicative adjectives and complementation. For instance, BE as a main verb (*be* and *that's*), predicative adjective (*passionate*) and private verb (*know*) with complement clause ('if that's good') are used in Example 1 in Table 2.3 to encode the learner's view about whether the people from their country are

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.3 *The turns most strongly associated with positive and negative Dimension 2.*

Dim. 2		Coordinate
1	I say that we could be more passionate but I don't know it's it's if that's good (file 2_AR_3)	1.19
2	I th= I don't know but I think it's not im= very important (file 2_6_IT_106)	1.11
3	and erm in May nineteen ninety four for the first time in South Africa's history all the races voted in democratic election (file 2_6_IT_100)	−0.79
4	with er a lot of big and small countries and the nationa= nations and nationalities er only one example in the republic of Dagestan live more than thirty nationalities with their own language and let alone national local products (file 2_7_RU_2)	−0.78

passionate and also to express uncertainty about whether being passionate is good.

Other features associated with the Involved function include subject pronouns, pronoun *it* and contracted forms, which are used in the turns in order to refer to a previously mentioned person or entity. Pronouns tend to be associated with a more interactive function as the context is shared and so the names of the people or entities do not need to be repeated. For instance, without access to the context, it is impossible to deduce what the pronoun *it* is referring to in Example 2 in Table 2.3. Given the nature of the interview, all of the turns are inherently interactive; however, the use of pronouns with the verb contracted in the turns strongly associated with the Involved function marks an involved and informal communicative style (Biber, 1988).

Finally, the Involved function is characterised by contrastive conjunctions, which are used in the turns in order to draw a contrasting opinion. For example, the contrastive conjunction *but* is used in Example 2 in Table 2.3 to encode the author's stance that her lack of knowledge is justified because that particular thing is not important. Overall, all the linguistic features contribute to an Involved communicative function whereby the learners encode their personal stance and attitude, often about a previously mentioned event or entity.

By contrast, the features most strongly associated with the Informational function (see Table 2.2) are associated with the careful integration of information. There are various nouns and nominal modifiers, such as proper nouns, general nouns, general determiners, definite articles and prepositions. These features are used to integrate a high degree of information

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

and are associated with the formation of complex noun phrases. Example 3 in Table 2.3 includes these features to integrate the month (*in May)*, year *(nineteen ninety four)*, the historical significance (*for the first time in South Africa's history*) and who (*all the races*) voted in the democratic election. Such densely packed and integrated complex phrases are associated with language that is produced when there is time to carefully plan and edit, such as in academic writing (Biber, 1988). Whilst the language produced is, understandably, not necessarily planned and edited in the same way as academic writing, the informational function is characterised by initial and non-initial fillers – features typically used to hold the floor whilst the speaker formulates their responses. Thus, it can be argued that the turns associated with the informational function display a more careful style. Finally, coordinating conjunctions are also strongly associated with the informational function. Coordinating conjunctions are often used to combine two sentences together in order to incorporate additional information or they are used to list two related referents, such as in Example 4 in Table 2.3 (*big and small countries*). Overall, the features strongly associated with the Informational function co-occur in the turns in order to efficiently integrate information.

Dimension 2 is therefore interpreted as opposing turns that have an Involved function with turns that are Informational. This dimension has been observed in nearly all studies employing MDA, even in those investigating only spoken or only written discourse (Biber, 2014). The present analysis is only investigating spoken language and yet it is possible to observe turns which not only exhibit a conversational style, but also appear to be produced much more carefully, like planned written language, such as those associated with the negative side of Dimension 2.

How do these functions interact with some of the variables present in the TLC? Tables 2.4–2.7 present the associations of the different groups of turns to Dimension 2. Tables 2.4 and 2.5 present the Dimension 2 associations of the turns distinguished by the learners' overall mark and proficiency in the Conversation and Discussion tasks, respectively. Both tables show that the turns from learners who received a higher overall mark are more closely associated with an Involved function. For example, the turns in the Conversation task from learners in the proficiency level B2 grade 7 that received an overall mark of Distinction are associated with an Involved function, whilst the turns from learners who received Merit and Pass in proficiency level B2 grade 7 are less Involved and more Informational. A similar pattern can be observed in Table 2.4, with the turns from learners from proficiency B2 grade 8 learners who received a higher overall mark more

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.4 *The Dimension 2 association of the turns in the Conversation task from groups of learners defined by proficiency, grade and overall mark.*

Conversation_Distinction_proficiency.B1_grade.6	0.068
Conversation_Distinction_proficiency.B2_grade.7	0.092
Conversation_Distinction_proficiency.B2_grade.8	0.172
Conversation_Merit_proficiency.B1_grade.6	0.001
Conversation_Merit_proficiency.B2_grade.7	−0.065
Conversation_Merit_proficiency.B2_grade.8	0.018
Conversation_Pass_proficiency.B1_grade.6	−0.033
Conversation_Pass_proficiency.B2_grade.7	−0.095
Conversation_Pass_proficiency.B2_grade.8	0.096

Table 2.5 *The Dimension 2 association of the turns in the Discussion task from groups of learners defined by proficiency, grade and overall mark.*

Discussion__Distinction_proficiency.B1_grade.6	−0.105
Discussion__Distinction_proficiency.B2_grade.7	0.062
Discussion__Distinction_proficiency.B2_grade.8	0.015
Discussion__Merit_proficiency.B1_grade.6	−0.186
Discussion__Merit_proficiency.B2_grade.7	−0.11
Discussion__Merit_proficiency.B2_grade.8	−0.064
Discussion__Pass_proficiency.B1_grade.6	−0.185
Discussion__Pass_proficiency.B2_grade.7	−0.168
Discussion__Pass_proficiency.B2_grade.8	−0.134

associated with the Involved communicative function. In general, Table 2.4 also shows that the turns in the Conversation task from the most proficient learners (either in terms of grade of exam or score awarded) are more Involved, whilst the least proficient learners' turns are more Informational. The main exceptions in both tables are learners in grade 6 who received an overall mark of distinction. These learners produce more Involved turns than some learners in higher grade exams with lower overall marks.

Overall, this suggests that the more involved turns the learner produces in the Conversation and Discussion tasks, the more likely they will have received a higher overall mark than other learners in the same proficiency and grade level. Additionally, in general the more proficient a learner is, the more likely they will be to produce more involved turns in these tasks.

Table 2.6 presents the Dimension 2 associations of the turns distinguished by the learners' overall mark and proficiency in the Interactive

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.6 *The Dimension 2 association of the turns in the Interactive task from groups of learners defined by proficiency, grade and overall mark.*

Interactive_Distinction_proficiency.B2_grade.7	0.163
Interactive_Distinction_proficiency.B2_grade.8	0.237
Interactive_Merit_proficiency.B2_grade.7	0.027
Interactive_Merit_proficiency.B2_grade.8	0.127
Interactive_Pass_proficiency.B2_grade.7	−0.042
Interactive_Pass_proficiency.B2_grade.8	0.117

Note: Grade 6 students do not take this task.

task. It shows that, overall, the turns produced by each group of learners in the Interactive task are more Involved rather than Informational. This suggests that this task may encourage a more involved and interactive style. Similar to the previous tables, Table 2.6 shows that the turns from learners receiving the highest overall mark are more Involved than those from learners in the same proficiency level who receive lower overall marks.

Table 2.7 shows the overall association of the turns from the learners, distinguished by their L1 and cultural background, with Dimension 2. Table 2.7 shows that Portuguese learners produce turns that are, overall, most strongly associated with an Involved communicative function, whereas Italian learners produce turns that are, overall, most strongly associated with an Informational communicative function. Table 2.7 shows that the turns from Spanish learners are, overall, Involved. However, Spanish L1 speakers from Spain produce English L2 turns that are more Informational, whereas those from Mexico and Argentina (all of whom have a variety of Spanish as their L1) produce English L2 turns that are more associated with an Involved communicative function. Additionally, turns produced by learners from Hong Kong are more Informational, whereas turns by learners from the rest of China are slightly more Involved. It is tempting, when presented with such data, to conclude that we are seeing a potential cultural, or L1 transference, issue at work. While we may hypothesise either or both, we do not have the corpus resources to explore this – ideally we would need conversational corpora, matched for task, of L1 conversations in these languages to begin to explore these issues. We do not have those corpora, hence any L1 interference noted in this book will be mentioned briefly and should be viewed as a hypothesis formation designed to encourage future research and corpus building. As part of that work, however, researchers should be aware that other systemic issues may be at play which may explain these results – differences

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.7 *Cultural and linguistic background associations with Dimension 2.*

Italy	−0.131
Italian	−0.13
Hong_Kong	−0.037
Sri_Lanka	−0.036
Spain	−0.024
China	0.012
Chinese	0.013
Argentina	0.02
Spanish	0.021
Russia	0.022
Russian	0.022
Hindi	0.023
Sinhala	0.025
India	0.04
Gujarati	0.045
Marathi	0.06
Tamil	0.083
Mexico	0.089
Brazil	0.194
Portuguese	0.211

in teaching practice in different countries could, quite easily, produce these results, for example. That would certainly be a hypothesis to explore if a future study showed no clear L1 transference effect that could explain these results.

2.4.3 *Dimension 3: Irrealis (Unknown) versus Realis (Known)*

Dimension 3 is interpreted as opposing turns on the positive side which mark a situation or event not known to have happened (Irrealis) with turns on the negative side that state something that the author considers to be factual or a known state of affairs (Realis). Specifically, Irrealis turns often involve the speakers encoding their opinions and making suggestions, such as articulating what they might do or say if they were in a particular problematic situation. These turns also comprise descriptions of unknown events through personal desires, probabilities, presumptions and hypothetical situations and conditions. By contrast, Realis turns tend to characterise and describe known, or directly observable, attributes of a particular subject. These interpretations are supported by the linguistic features most strongly contributing to Dimension 3.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.8 *The linguistic features most strongly associated with Dimension 3.*

Dim. 3	Features (coordinates, contributions)
+	General_Verb_P (0.264,1.279), BE as main verb_A (0.264,2.986), Private Verb_P (0.603,2.631), Complementation_P (adjective/verb + that complements + adj + to complement clauses + WH clauses) (0.645,1.279), Object_Pronoun_P (0.688,1.327), Modal of Possibility_P (0.749,1.966), Second-Person Pronouns_P (0.761,4.334), Infinitive_P (0.777,3.094), WH_word_P (0.806,1.982), Public Verb_P (0.881,2.223), Analytic Negation_P (1.113,6.803), Stance_Verb_P (1.153,4.71), Auxiliary DO_P (1.826,11.825)
–	BE as main verb_P (–0.97,10.969), Predicative Adjective_P (–0.967,5.044), Contractions_P (pronoun, WH-word) (–0.962,4.881), Third Person Singular Verb_P (–0.791,5.331), Pronoun *it*_P (–0.701,3.317), Proper Noun_P (–0.531,1.824), Indefinite Article_P (–0.41,1.257), Attributive Adjective_P (–0.334,1.294)

In particular, Irrealis, as presented in Table 2.8, is characterised by a variety of verb forms, such as general verbs, infinitives, modals of possibility, private verbs, public verbs, stance verbs and auxiliary DO. These features are often used to 'think aloud' about unknown events and/or to propose a potential course of action. For example, modals of possibility are often used in the turns to mark ability and possible action in order to make recommendations, give advice or suggest/propose a course of action, and public verbs are often used in the turns to introduce speech. These features co-occur in Example 5 in Table 2.9 in order to introduce speech that the hearer could say as a possible course of action to solve a problem (e.g. *you can tell him…*).

Many of the other verbs strongly associated with Irrealis are used to encode personal stance, thoughts and feelings, such as private verbs, stance verbs and also complementation. These features are often used to encode personal opinions and desires. For instance, private verb and complementation in the form of a WH-clause occur in Example 6 to encode that the speaker does not know a particular word (*I don't know how you s= you call that*). Additionally, in Example 5 the stance verb *want* is used with the infinitive *to let,* which completes the meaning of the stance verb in order to express a personal desire. Auxiliary DO and analytic negation are also strongly associated with positive Dimension 3 and are used in the turns often in order to negate an action or event to express the unknown, such as the phrases *I don't want* or *I don't know* in Examples 5 and 6 respectively. In addition to verbs, positive Dimension 3 is characterised by object

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.9 *Turns most strongly associated with positive and negative Dimension 3.*

Dim. 3		Coordinate
5	so you can tell him that y= I don't want to let you stay here you have to move away or (file 2_7_CH_22)	1.34
6	well er I would give them <pause length='short'/> er I don't know something to play some er <pause length='short'/> I don't know how you s= you call that but some games that keep your mind working (file 2_7_ME_5)	1.15
7	er by definition animal hunting involves hunting or trapping any animals or poaching it for the sake of <unclear/> it has existed for a long time since the rise of Homo sapiens and it's a very very important characteristic of the hunter gatherers <unclear/> only a few contemporary societ-societies are <unclear/> erm hunter gatherers and remains of it are still present in North Africa <pause length=\\'short\\'/> (file 2_RUM_1)	−0.91
8	only on the Mexico City it's legal under some conditions like when it's under three months and when it's because of erm <unclear/> or or the foetus has a genetic problem or things like that yeah (file 2_ME_8)	−0.89

pronouns and second-person pronouns. Object pronouns *him* and *them* are used in Examples 5 and 6 respectively to mark the semantic patient that is possibly going to be acted upon. Second-person pronouns mark an addressee, often the examiner, such as Examples 5 and 6.

Overall, the features strongly associated with Irrealis are connected by an underlying function of expressing the unknown (i.e. situations or events that are not known to have happened). This is often realised in turns that are describing possible future events or action, such as by learners talking about what they would do in a situation or giving advice about what their addressee can do in the future to solve their problem.

By contrast, the features most associated with Realis turns, presented in Table 2.8, are associated with describing and characterising a known subject, entity or state of affairs. The verb most strongly associated with Realis is BE as a main verb, which is used in the turns to identify or describe an attribute of the subject matter-of-factly. In addition, Realis is characterised by both adjectival forms: attributive and predicative. These are used in the turns to provide concrete detail on the subject. For example, the attributive adjective *contemporary* and the predicative adjective *present* in Example 7 in Table 2.9 add detail about the subject's nature and existence. Additionally, the predicative adjective *legal* and attributive adjective *genetic* in Example 8 are used to add detail on the subject. Realis is also

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.10 *The Dimension 3 association of the turns in the Conversation task from groups of learners defined by proficiency, grade and overall mark.*

Conversation_Distinction_proficiency.B1_grade.6	0.098
Conversation_Distinction_proficiency.B2_grade.7	−0.006
Conversation_Distinction_proficiency.B2_grade.8	−0.032
Conversation_Merit_proficiency.B1_grade.6	0.1
Conversation_Merit_proficiency.B2_grade.7	−0.037
Conversation_Merit_proficiency.B2_grade.8	0.047
Conversation_Pass_proficiency.B1_grade.6	0.093
Conversation_Pass_proficiency.B2_grade.7	−0.084
Conversation_Pass_proficiency.B2_grade.8	0.097

characterised by third-person singular verb forms like *has* in Example 8 and *involves* in Example 7, which mark that the subject is external to the people in the conversation. Negative Dimension 3 is also characterised by proper nouns and indefinite articles, which mark the introduction of new subjects. For example, the proper noun *Mexico City* is introduced in Example 8 as a place where abortion is legal in particular circumstances. These circumstances are then introduced through the indefinite article (*or the foetus has a genetic problem*). In addition to the introduction of new subjects, Realis is characterised by the pronoun *it*, which is often used in the turns to refer back to something previously mentioned. Overall, these features are used in the turns to describe and characterise new and known subjects.

How do Irrealis and Realis respond to the variables in our analysis? Tables 2.10–2.12 present the associations of the different groups of turns to Dimension 3. Tables 2.10 and 2.11 present the Dimension 3 associations of the turns distinguished by the learners' overall mark and proficiency in the Conversation and Discussion tasks, respectively. These tables show that the production of Irrealis and Realis turns do not systematically interact with overall mark or learner proficiency. For example, Table 2.10 shows that learners in proficiency level B2 grade 7 who received the higher mark Distinction produce fewer Realis turns than learners who received lower overall marks of Merit and Pass. However, the opposite pattern is found for learners in proficiency level B2 grade 8, whereby learners who received the higher mark produced fewer turns associated with Irrealis turns than learners in the same proficiency grade who received lower marks.

Additionally, Table 2.11 shows that at grade 7 there is a clear inverse relationship between the mark awarded and use of Realis – the use of

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.11 *The Dimension 3 association of the turns in the Discussion task from groups of learners defined by proficiency, grade and overall mark.*

Discussion__Distinction_proficiency.B1_grade.6	−0.148
Discussion__Distinction_proficiency.B2_grade.7	−0.048
Discussion__Distinction_proficiency.B2_grade.8	−0.108
Discussion__Merit_proficiency.B1_grade.6	−0.147
Discussion__Merit_proficiency.B2_grade.7	−0.138
Discussion__Merit_proficiency.B2_grade.8	−0.084
Discussion__Pass_proficiency.B1_grade.6	−0.134
Discussion__Pass_proficiency.B2_grade.7	−0.169
Discussion__Pass_proficiency.B2_grade.8	−0.152

Table 2.12 *The Dimension 2 association of the turns in the Discussion task from groups of learners defined by proficiency, grade and overall mark.*

Interactive_Distinction_proficiency.B2_grade.7	0.278
Interactive_Distinction_proficiency.B2_grade.8	0.294
Interactive_Merit_proficiency.B2_grade.7	0.265
Interactive_Merit_proficiency.B2_grade.8	0.299
Interactive_Pass_proficiency.B2_grade.7	0.264
Interactive_Pass_proficiency.B2_grade.8	0.315

Realis declines with proficiency at this grade. This is not true at the other grades in the table, where the picture is more mixed.

Table 2.12 presents the Dimension 3 associations of the turns in the Interactive task, distinguished by the learners' overall mark and proficiency. Table 2.12 shows that the production of turns associated with Irrealis and Realis in the Interactive task does not systematically interact with overall mark or learner proficiency. Rather, Table 2.12 shows that, overall, the turns produced by each group of learners in the Interactive task are more associated with expressions of the unknown. The task itself is undoubtedly the driver of this as it is oriented towards the Irrealis – the examiner describes a situation, which does not involve the learner, and the learner must ask questions to find out more information as they typically move towards making some observations about the situation. In short, the student discusses a situation that they have no knowledge of and projects forwards to produce contingent solutions. Thus, by nature, the task encourages talk characterised by Irrealis.

Table 2.13 shows the overall Dimension 3 association of the turns from the learners distinguished by their L1 and cultural background. Table 2.13

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.13 *The Dimension 3 association of the turns from groups of learners defined by their linguistic and cultural backgrounds.*

Sinhala	−0.141
Sri_Lanka	−0.137
Italian	−0.06
Italy	−0.06
Russia	−0.031
Russian	−0.031
Hindi	0
Marathi	0.004
Spain	0.005
India	0.006
Gujarati	0.01
Mexico	0.012
Spanish	0.015
Brazil	0.026
Hong_Kong	0.032
China	0.035
Portuguese	0.035
Tamil	0.036
Chinese	0.039
Argentina	0.046

indicates that the turns from learners whose L1 is Chinese, Portuguese, Spanish, Gujarati or Marathi are more associated with Irrealis, whereas turns from learners whose L1 is Sinhala, Italian or Russian are more associated with Realis.

2.4.4 *Dimension 4: Infer versus Reveal*

Dimension 4 is interpreted as opposing turns that Infer with turns that Reveal. In particular, the turns which Infer tend to incorporate reasoned discussions of entities or people other than the self. They often discuss advantages, disadvantages, and particular scenarios, situations and possibilities. By contrast, the turns which Reveal often disclose the personal details, actions and desires of the speaker. These interpretations are supported by the linguistic features most strongly contributing to Dimension 4.

The features most strongly associated with Infer, presented in Table 2.14, are used to refer to a subject that is not the self, such as second-person pronoun, demonstrative pronoun, nominalisation and third-person singular verb. These features often occur in order to introduce and discuss

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.14 *The linguistic features most strongly associated with Dimension 4.*

Dim. 4	Features (coordinates, contributions)
+	Third Person Singular Verb_P (0.351,1.347), First-Person Pronoun_A (0.358,6.225), Subject Pronoun_A (0.377,6.357), General_Interjection_P (0.464,2.934), Auxiliary DO_P (0.557,1.408), Complementation_P (adjective/verb + that complements + adj + to complement clauses + WH clauses) (0.571,1.283), Nominalisation_P (0.727,1.998), Demonstrative_Pronoun_P (0.755,2.285), Modal of Possibility_P (0.988,4.377), WH_word_P (1.25,6.093), Second-Person Pronouns_P (1.33,16.96)
–	First-Person Pronoun_P (–0.825,14.353), Subject Pronoun_P (–0.674,11.359), Object_Pronoun_P (–0.638,1.464), Modal of Prediction_P (–0.613,1.264), Stance_Verb_P (–0.577,1.511), Second-Person Pronouns_A (–0.212,2.703)

entities, people or situations and make an inference about them. For example, second-person pronouns are often used in the turns most strongly associated with Infer to refer to the addressee or the universal or generalised *you* in order to discuss particular situations or scenarios and make an inference (e.g. *it doesn't matter if you're a celebrity or how much money you have…*). Nominalisations are used to introduce a subject or event, such as *advertisement* in Example 10 in Table 2.15. Demonstrative pronouns and interjections are often used to refer to something in the immediate context. For example, *that* in Example 9 is used to refer back to something that the author has previously said in order to make an overall evaluation and inference of that topic. The interjection *oh* in Example 9 is used to mark a sudden thought or reaction to an answer to a previous question. Third-person singular verbs are used to deal with topics of immediate relevance (Biber, 1988). For instance, *thinks* in Example 10 is used to make an inference about what the speaker has just heard from another speaker.

Other features strongly associated with Infer are used to discuss possibilities. For example, the modal of possibility *can* is used to discuss the possibilities and abilities of entities, such as Example 9, where *can* marks the potential of saving lives by donating organs. Complementation is used to elaborate. For example, the complement clause in Example 10 is used by the speaker to elaborate on what they have inferred about what the speaker thinks. Other features discuss particular scenarios, such as WH-word. For example, *when* is used to talk about a scenario in Example 10, and *what* in Example 9 is used to introduce a scenario that matters to the people who decide who gets an organ from a donor. Finally, auxiliary DO is often used to discuss an action or event. It often co-occurs in the turns with analytic negation (e.g. *don't*,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.15 *Turns most strongly associated with positive and negative Dimension 4.*

Dim. 4		Coordinate
9	er here are some pros oh and for statistics what matters on the transplant list it doesn't matter if you're a celebrity or how much money you have what matters is the severity of your illness <unclear text=\\'depends tha=\\'/> waiting on your blood type so for pros a single donate a single donor can save up to eight different can save up to eight lives <pause length='short'/> <unclear/> okay er <laugh/> <pause length='short'/> a single donor can save up to eight lives and can donate over twenty-five different organs so that's pretty incredible (file 2_ME_23)	1.12
10	er the speaker thinks that er when that when you spend money on on advertisement the other effect is that you don't spend on your on being wealthy and having a good life (file 2_SP_20)	1.08
11	yeah now I'm going to Salvation's Army to give them the word of god (file 2_ME_11)	−0.68
12	I used to study bones and I really like them and I would like to (file 2_7_SP_23)	−0.68

doesn't) to infer an event (or the lack of an event) given a scenario. Overall, these features co-occur in the turns in order to introduce and discuss a topic, person or scenario in order to infer something about that entity.

By contrast, the features most strongly associated with Reveal, presented in Table 2.14, are used to encode and reveal personal stance, events and plans. For example, subject pronouns, especially first-person pronouns, are used to involve the self and mark the self as the agent in order to reveal something about the self. Stance verbs, such as *like* in Example 12 in Table 2.15, are used to reveal personal stance and judgements. Modals of prediction like *would* in Example 12 or BE + *going to* in Example 11 are used to reveal personal desires and personal plans. Finally, object pronouns are also strongly associated with Reveal and these are used to mark the object or patient acted upon (Quirk et al., 1985), suggesting that the revelation often concerns the speaker doing something to something or someone else that has been previously mentioned. These features co-occur often in the turns in order to reveal something.

Tables 2.16–2.19 present the associations of the different groups of turns to Dimension 4.

Table 2.16 presents the Dimension 4 associations of the turns in the Conversation task distinguished by the learners' overall mark and proficiency. The plot shows that generally the turns from learners who received

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.16 *The Dimension 4 association of the turns in the Conversation task from groups of learners defined by proficiency, grade and overall mark.*

Conversation_Distinction_proficiency.B1_grade.6	−0.102
Conversation_Distinction_proficiency.B2_grade.7	−0.007
Conversation_Distinction_proficiency.B2_grade.8	−0.097
Conversation_Merit_proficiency.B1_grade.6	−0.089
Conversation_Merit_proficiency.B2_grade.7	−0.011
Conversation_Merit_proficiency.B2_grade.8	−0.059
Conversation_Pass_proficiency.B1_grade.6	−0.065
Conversation_Pass_proficiency.B2_grade.7	0.001
Conversation_Pass_proficiency.B2_grade.8	−0.039

the higher overall mark in their proficiency level and grade group were less associated with Infer in the conversation task than the turns from the learners who received the lower overall mark, except for those who received Merit in proficiency level B2 grade 7 (although the difference is marginal).

Table 2.17 presents the Dimension 4 associations of the turns in the Discussion task distinguished by the learners' overall mark and proficiency. It shows that the turns from the higher grade and scoring exams shift progressively towards Infer. Although Table 2.17 shows that the turns from learners who received the highest mark of Distinction are generally more associated with Infer and less associated with Reveal than the learners in the same exam grade who received the lowest mark of Pass, learners who received Merit in the same exam grade are not between the two, as would be expected if, overall, mark was systematically associated with Dimension 4.

Table 2.18 presents the Dimension 4 associations of the turns in the Interactive task distinguished by the learners' overall mark and exam grade. Table 2.18 shows that, overall, the turns produced by each group of learners in the Interactive task are more associated with Infer as opposed to Reveal. This indicates that this task in the L2 interview encourages more reasoned discussions of particular entities, as opposed to personal revelations. Table 2.18 shows an association between the turns of the learners in proficiency level C1 (grade 10 exam) and the overall mark, whereby the turns from learners who received higher overall marks were less associated with Infer. However, the strength of this pattern does not extend to the turns of learners from proficiency level B2 (grade 8 exam), as learners who received Merit produced turns that were less associated with Infer than those learners who received Distinction.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.17 *The Dimension 4 association of the turns in the Discussion task from groups of learners defined by proficiency, grade and overall mark.*

Discussion__Distinction_proficiency.B1_grade.6	−0.092
Discussion__Distinction_proficiency.B2_grade.7	−0.118
Discussion__Distinction_proficiency.B2_grade.8	0.03
Discussion__Merit_proficiency.B1_grade.6	−0.108
Discussion__Merit_proficiency.B2_grade.7	−0.063
Discussion__Merit_proficiency.B2_grade.8	−0.115
Discussion__Pass_proficiency.B1_grade.6	−0.087
Discussion__Pass_proficiency.B2_grade.7	−0.141
Discussion__Pass_proficiency.B2_grade.8	−0.05

Table 2.18 *The Dimension 4 association of the turns in the Interactive task from groups of learners defined by proficiency, grade and overall mark.*

Interactive_Distinction_proficiency.B2_grade.7	0.327
Interactive_Distinction_proficiency.B2_grade.8	0.333
Interactive_Merit_proficiency.B2_grade.7	0.319
Interactive_Merit_proficiency.B2_grade.8	0.332
Interactive_Pass_proficiency.B2_grade.7	0.376
Interactive_Pass_proficiency.B2_grade.8	0.342

Table 2.19 shows the overall Dimension 4 association of the turns from the learners distinguished by their L1 and cultural background. Table 2.19 indicates that the turns from learners whose L1 is Portuguese, Chinese, Gujarati, Tamil and Sinhala are more associated with Reveal, whereas turns from learners whose L1 is Russian, Marathi, Italian and Spanish are more associated with Infer.

2.4.5 *Dimension 5: Narrative versus Non-Narrative*

Dimension 5 is interpreted as opposing turns that have a Narrative function with turns that are Non-Narrative. This finding is also the first strong indicator that we need to account for narrative in the study of learner language. As discussed at the start of the book, narrative will come into focus later in the book. For now, we will proceed with a discussion of narrative as it is revealed through the short-text MDA of the learner turns.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.19 *The Dimension 4 association of the turns from groups of learners defined by their linguistic and cultural backgrounds.*

Brazil	−0.091
Portuguese	−0.084
China	−0.067
Chinese	−0.064
Sri_Lanka	−0.063
Gujarati	−0.056
Tamil	−0.044
Sinhala	−0.039
Hong_Kong	−0.031
Spain	−0.011
India	−0.009
Hindi	−0.003
Argentina	0
Spanish	0.008
Mexico	0.039
Italy	0.043
Italian	0.044
Marathi	0.077
Russia	0.119
Russian	0.119

The Narrative turns tend to narrate and give accounts of events about particular people and entities. By contrast, the Non-Narrative turns tend to express facts about subjects and encode personal thoughts and intellectual states. These interpretations are supported by the linguistic features most strongly contributing to Dimension 5.

The Narrative function is characterised by numerous pronouns, including object, third-person, second-person and possessive pronouns (see Table 2.20). These pronouns are often used in the turns to introduce and refer to particular people or entities in order to give an account of their actions or of events happening to them. For example, third-person pronouns (*he*, *them*) and third-person possessive determiners ('their parents') are used in Example 13 in Table 2.21 in order to refer to a boy and his parents. Second-person pronouns are often used in the turns as second-person narration in order to involve the listener(s) in the story or provide instructions (e.g. *you can cho-choose to support the boy*).

In addition to pronominal forms, Narrative is characterised by numerous verb forms, including past tense verbs, public verbs, general verbs, modals of prediction and infinitives. These are used to narrate and elaborate past,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.20 *The linguistic features most strongly associated with Dimension 5.*

Dim. 5	Features (coordinates, contributions)
+	Non-Initial_Filler_A (0.175,1.726), General_Verb_P (0.221,1.205), Initial_ Filler_A (0.233,2.775), Past_Tense_P (0.427,1.477), General_Subordinator_P (0.455,1.775), Second-Person Pronouns_P (0.469,2.226), Possession_P (determiner, noun, pronoun, proper noun) (0.49,1.984), Infinitive_P (0.718,3.563), WH_word_P (0.744,2.277), Third Person Pronoun_P (0.759,5.441), Modal of Prediction_P (0.869,2.684), Public Verb_P (1.487,8.553), Object_Pronoun_P (1.555,9.16)
–	Auxiliary DO_P (–1.525,11.142), Analytic Negation_P (–1.122,9.335), HAVE as main verb_P (–0.745,1.978), Contrastive Conjunctions_P (–0.582,1.819), Non-Initial_Filler_P (–0.577,5.698), Private Verb_P (–0.567,3.146), Initial_Filler_P (–0.531,6.322)

Table 2.21 *Turns most strongly associated with positive and negative Dimension 5.*

Dim. 5		Coordinate
13	with their parents like you can choice to be with them to agree with them or you can cho-choose to support the boy and maybe he could get the thing that he wants that is going to <unclear/> music department (file 2_7_AR_29)	1.23
14	so today I'd like to talk to you about how the language we speak affects the way that we see the world around us I've divided my talk into three <unclear/> sections I'll begin by elaborating <unclear/> language includes this thought we'll then progress to the cultural impacts on language <unclear/> contributes to the language that they speak and we conclude by looking at the impact of language on relationships and expression or in other words how language affects the individuals the route <unclear/> to society to the individual (file 2_IN_10)	1.15
15	I think this year I don't know because er I'm I haven't book er yet (file 2_6_IT_23)	–0.96
16	er but I don't know because I don't have any brother in primary (file 2_7_AR_3)	–0.9

present and future events. For example, the modal of prediction *would* and infinitive form of the public verb (*talk*) is used in Example 14 to introduce and narrate the content of the learner's upcoming talk (e.g. *I'd like to talk*…).

Other features strongly associated with Narrative turns include general subordinators and WH-words, which are often used to expand or

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

elaborate on an event. For example, subordinators can elaborate on when or where an event happened or should or could take place (Quirk et al., 1985). Additionally, WH-words can expand an idea unit (Chafe, 1982, 1985), such as with *which* in Example 14 (e.g. *I'll begin by elaborating ways which language includes this thought…*). Overall, these features co-occur in the turns most strongly associated with positive Dimension 5 in order to give an account of events and actions.

By contrast, Non-Narrative turns are characterised by features associated with expressions of intellectual states and ideas, such as private verbs (e.g. *know*) and contrastive conjunctions (e.g. *but*), which are used to introduce a different or contrasting idea. Additionally, there are features associated with expressions of possession, such as HAVE as a main verb (e.g. *I don't have any brother in primary*). Non-Narrative is also characterised by analytic negation and auxiliary DO, which often co-occur in order to negate an event or intellectual state. For example, auxiliary DO and analytic negation co-occur with the private verb *know* in Examples 15 and 16 in Table 2.9 in order to express uncertainty. Finally, initial and non-initial fillers are also strongly associated with Non-Narrative. Fillers are generally used in the turns in order to hold the floor whilst the speaker thinks of an appropriate response. Overall, these features are connected by an underlying Non-Narrative function as they co-occur in the turns most strongly associated with Non-Narrative in order to express ideas, thoughts and facts about particular subjects.

Tables 2.22–2.25 present the associations of the different groups of turns to Dimension 5. Table 2.22 presents the Dimension 5 associations of the turns in the Conversation task distinguished by the learners' overall mark and proficiency. Table 2.22 shows that the turns of the learners who received the higher marks among the other learners in the same proficiency and grade level are less associated with Non-Narrative in the conversation task.

Table 2.23 presents the Dimension 5 associations of the turns in the Discussion task distinguished by the learners' overall mark and proficiency. Table 2.23 shows that the turns from learners receiving the higher marks are generally more associated with a Narrative function in the Discussion task than learners receiving the lowest mark of Pass, suggesting that a Narrative style may be rewarded with a higher mark in this context.

Table 2.24 presents the Dimension 5 associations of the turns in the Interactive task distinguished by the learners' overall mark and proficiency. Table 2.24 shows that, at any given exam grade, the turns from learners

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.22 *The Dimension 5 association of the turns in the Conversation task from groups of learners defined by proficiency, grade and overall mark.*

Conversation_Distinction_proficiency.B1_grade.6	−0.035
Conversation_Distinction_proficiency.B2_grade.7	0.031
Conversation_Distinction_proficiency.B2_grade.8	−0.014
Conversation_Merit_proficiency.B1_grade.6	−0.157
Conversation_Merit_proficiency.B2_grade.7	−0.09
Conversation_Merit_proficiency.B2_grade.8	−0.03
Conversation_Pass_proficiency.B1_grade.6	−0.192
Conversation_Pass_proficiency.B2_grade.7	−0.203
Conversation_Pass_proficiency.B2_grade.8	−0.063

Table 2.23 *The Dimension 5 association of the turns in the Discussion task from groups of learners defined by proficiency, grade and overall mark.*

Discussion__Distinction_proficiency.B1_grade.6	0.048
Discussion__Distinction_proficiency.B2_grade.7	0.139
Discussion__Distinction_proficiency.B2_grade.8	0.045
Discussion__Merit_proficiency.B1_grade.6	−0.062
Discussion__Merit_proficiency.B2_grade.7	0.055
Discussion__Merit_proficiency.B2_grade.8	0.058
Discussion__Pass_proficiency.B1_grade.6	−0.099
Discussion__Pass_proficiency.B2_grade.7	−0.018
Discussion__Pass_proficiency.B2_grade.8	−0.024

Table 2.24 *The Dimension 5 association of the turns in the Interactive task from groups of learners defined by proficiency, grade and overall mark.*

Interactive_Distinction_proficiency.B2_grade.7	0.352
Interactive_Distinction_proficiency.B2_grade.8	0.336
Interactive_Merit_proficiency.B2_grade.7	0.262
Interactive_Merit_proficiency.B2_grade.8	0.297
Interactive_Pass_proficiency.B2_grade.7	0.198
Interactive_Pass_proficiency.B2_grade.8	0.219

in the interactive task who received an overall higher mark are more associated with Narrative than learners who received a lower overall mark in the same proficiency and grade category. The table also suggests that more proficient learners produce fewer turns associated with Narrative in the interactive task than less proficient learners.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.25 *The Dimension 5 association of the turns from groups of learners defined by their linguistic and cultural backgrounds.*

Italy	−0.224
Italian	−0.223
Spain	−0.067
Russia	−0.066
Russian	−0.066
Spanish	−0.012
China	−0.011
Chinese	−0.005
Hong_Kong	0.008
Mexico	0.023
Portuguese	0.062
Brazil	0.065
Argentina	0.068
Sri_Lanka	0.13
Gujarati	0.213
Sinhala	0.227
Hindi	0.232
India	0.24
Tamil	0.279
Marathi	0.29

Table 2.25 shows the overall association of the turns from the learners distinguished by their L1 and cultural background with Dimension 5. Table 2.25 shows that Italian, Russian, Spanish and Chinese learners produce turns that are more associated with a Non-Narrative function, whereas Marathi, Tamil, Hindi, Sinhala, Gujarati and Portuguese learners produce turns that are more associated with a Narrative function. Spanish-speaking learners from Spain produce more turns associated with a Non-Narrative style than Spanish-speaking learners from Mexico and Argentina.

2.5 Examiner Turns

Having outlined a turn-based functional characterisation of the learner turns in the corpus, we now move on to consider examiner speech. Of course, the focus of this book is learner language, so our discussion of examiner speech will be correspondingly brief. It cannot, however, be omitted. This is because, as noted, the data in all of the corpora used in this book are generally dialogic. The functions outlined so far are not being

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

produced by the L1 speakers at random or exclusively as a monologue. The functions are being selected with communication with an interlocutor in mind. We may assume that the choice of the functions in any turn in the L1 is produced with a variable range of intents – a response to the examiner's speech, an intention to elicit a response from the examiner or as a move in macro-structure, such as a narrative, for example. In such a context it would be folly to try to understand learner speech without also understanding the speech of their interlocutor. This will become clearer in the next chapter. For now, we will present a summary study of examiner turns, based on a short-text MDA of them, to give a sense of how the speech of the examiners and the examinees, when viewed this way, is similar to, or differs from, each other. The analysis is summarised in Table 2.26. However, in discussing this table we will foreshadow some of the findings of the next chapter by focusing on a subset of the functions present in the speech of the examiners, which will later be shown to have a particular role to play in influencing the short-text MDA of the TLC discourse units. In the table, functions which are also present in the turn level short-text MDA of the L2 speakers are underlined.

One finding is very clear from this table. The functions of the examiner speech and those of the learner speech are quite distinct at this microstructural level. With the exception of Narrative, there are no shared functions between the two at the level of the turn. As hypothesised at the start of this chapter, the social roles of the interlocutors in the data lead to them producing distinct repertoires of discourse functions. Those functions may interact, but they are not congruent – in the context of an exam in which they have quite distinct roles, this functional differentiation between the two speakers is as understandable as it is striking. Yet to assume that the speakers were simply distinct and to leave our analysis there would be to misconstrue the nature of discourse and to mischaracterise the interdependencies between examiner and examinee functions which, while they may appear distinct, could be mutually dependent. To begin our transition to a consideration of interaction, let us examine two dimensions, 2 and 7, from the examiner short-text MDA, which will be of importance in the next chapter. Dimension 2 splits between the descriptive and the information seeking. The following examiner turn, from the discussion task of a grade 7 exam taken by an Indian student (file 2_7_IN_11), shows an example of a turn which is associated with a Descriptive function:

(17) E: mm it sounds good erm yeah I've never really watched anything similar to that I've haven't watched a lot of cartoons but some things yeah I like to see yeah

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

Table 2.26 *Examiner discourse functions.*

Dimension	Label	Features present
Dim. 1 +ve	Long turns	Prediction modal, Infinitive, Private verb, Nominalisation, Quantifier, Auxiliary DO, Predicative adjective, Subordinator, Definite article, General determiner, Past tense, Analytic negation, Indefinite article, Contracted forms, Pronoun *it*, Third-person singular verb, Coordinating Conjunction, General adverb, Preposition, Stative forms, First-person pronoun, WH-word, Question mark, Subject pronoun, General verb, Second-person pronoun, Attributive adjective, General noun.
Dim. 1 –ve	Short turns	Only absent features (Subject pronoun, General noun, General verb)
Dim. 2 +ve	Descriptive	Predicative adjective, Pronoun *it*, Contracted forms, Analytic negation, Stative forms, Demonstratives, Amplifier, Third-person singular verb forms
Dim. 2 –ve	Information seeking	WH-word, Auxiliary DO, Question marks, Second-person pronoun
Dim. 3 +ve	Guide future action	Nominalisation, Modal of prediction, Public verb, Quantifier, First-person pronoun, Infinitive, General adverb, Definite article, Preposition
Dim. 3 –ve	Discovering stance	WH-word, Auxiliary DO, Third-person singular verb, Question, Predicative adjective, Past tense verb, Pronoun *it*, Stative form
Dim. 4 +ve	Stating stance	Analytic negation, Auxiliary DO, Negative interjection, Private verb, Third-person pronoun, Infinitive, Subject pronoun, General verb
Dim. 4 –ve	Discussing the here and now	Third-person singular verb, Demonstrative, General determiner, Proper noun, Stative, Definite article, Preposition, Attributive adjective, General noun
Dim. 5 +ve	Interjection (positive)	Positive interjection
Dim. 5 –ve	Interjection (other)	Negative interjection, General interjection
Dim. 6 +ve	Past orientation	Negative interjection, Indefinite article, Third-person singular verb, Analytic negation, Third-person pronoun, Past tense verb, Quantifier, Definite article, Attributive adjective, Amplifier, Preposition, General noun
Dim. 6 –ve	Future orientation	Modal of prediction, Contracted form, Infinitive, Predicative adjective, Demonstrative, General determiner, Second-person pronoun, Subordinator, Positive interjection
Dim. 7 +ve	Narrative	Third-person pronoun, Negative interjection, Past tense verb, Public verb, Third-person singular verb, General determiner, Subordinator
Dim. 7 –ve	Stance seeking	Amplifier, Attributive adjective, Quantifier, Auxiliary DO, Private verb, Indefinite article, Second-person pronoun

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

The features which combine to perform the Descriptive function are apparent in this example, for instance, stative forms (*sounds*), contracted forms (*I've*, *haven't*), demonstratives (*that*) and the pronoun *it*. The function is used by the examiner to simply outline a factual statement, in this case regarding the examiner's viewing habits. By contrast, we find examiner turns with a function of Information Seeking on the negative side of Dimension 2. The following is one such turn from the conversation task of an Argentinian student taking a grade 6 exam (file 2_6_AR_27):

(18) E: okay and what did you have to pack <pause length='short'/> before you went? ha= you had a lot of preparation

The question, seeking a statement of information from the student, orients the turn towards the Information-Seeking function, with features which combine to perform that function present including a WH-word (*what*), Auxiliary DO, a question mark and a second-person pronoun (*you*). It also tries to guide the selection of the function for the next turn by the student – given an Information-Seeking turn from the examiner, it is easy to find examples of the student being guided towards a Reveal function, even if they fail initially to produce such a turn. Consider the following exchange from a Mexican student taking a grade 6 exam (file 2_6_ME_108):

(19) E: oh alright okay and erm so er I'm can't read that one what does that say?
S: balls
E: balls so what's what's important about the balls?
S: well I think that is very important if I if the ball have a good contents and

In this sequence the examiner uses an Information-Seeking function, eliciting a bare Informational response from the student. The examiner then prompts the student with another Information-Seeking response and this time, instead of using the Informational discourse function from their repertoire of turn-level discourse functions, the student produces a Reveal function instead. In the transcript the examiner then allows the student to continue with the Reveal function (the following two examiner turns are phatic). So we may, at this point, presume that examiner turns and functions, while distinct, may influence and shape a student turn, and vice versa.

What of Dimension 7? This is composed of two functions, Narrative on the positive side of the dimension and Stance Seeking on the negative side of the dimension. The Narrative function relates to turns that do not necessarily constitute a full narrative in themselves but which, in context,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

contribute elements to a narrative. Consider the following examiner turn, from the interactive task of a grade 7 Argentinian student:

(20) E: he said that er I'm not his father and that he can do what he wants to do <pause/>

Again, we see ample evidence of the grammatical features which bundle to produce this function. For example, a subordinator (*that*), third-person singular verb (*wants*) and a past tense verb (*said*). There is clearly a past orientation here, and some suggestion of a flow of events through time – this is clearly a response to a statement of some sort. When we look at the broader context, this is one of a series of such turns which together constitute a narrative. This is the discourse unit in which the turn occurs (from file 2_7_AR_16):

(21) S: er <pause/> do you like er <pause/> how dress how they dress?
E: no I hate it
S: why?
E: he just looks very er scruffy it means er he doesn't look neat he doesn't look tidy he wears old jeans and old T-shirts
S: you can <pause/> erm <pause/> tell <pause/> tell about how you feel
E: mm
S: about er his dress
E: I did <pause/>
S: and what di= what did <pause/> what did er <pause/> she say?
E: he said that er I'm not his father and that he can do what he wants to do <pause/>

This makes the example slightly clearer in terms of function. The student is eliciting, across a number of turns, a narrative in response to a prompt for the task from the examiner, in this case 'my nephew used to dress very well but now he's totally changed his appearance I'm not sure what to think about it'. In terms of narrative, as will be explored later in the book (see Chapter 8), this is a complicating action—the genesis of the narrative. In the discourse unit just given, we have proceeded from that complicating action to the first discussion of the issue with the nephew. Following this discourse unit, the discussion proceeds through a series of interactions and issues until we arrive at a resolution, which is a, perhaps impractical, suggestion by the student that the examiner should oblige the nephew to dress better. Throughout, turns abound, which are key elements of narrative. We will not explore further here whether the Narrative function exists beyond the turn level – we have argued for one such example here, but this is certainly suggestive of the possibility that, when we analyse

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

discourse units in Chapters 3–7, Narrative will extend beyond the micro- to the macro-structural level. What is important for now, however, is to note that a Narrative function exists at the turn level in the examiner turns.

2.6 Conclusion

This initial exploration of the data using short-text MDA at the micro-structural level (the turn) has established that the technique can be used to group turns such that, when we investigate turns, we are able to discern discourse functions that are associated with the grouped data. This indicates a link between form (the basis on which the grouping occurs) and function (the discourse functions we have assigned to the learner and examiner speech). Reflecting on our findings, we might hypothesise that the likelihood of congruence between the turn-level analysis in this chapter and the discourse unit analysis to be presented in the next chapter is low – the discourse units contain both learner and examiner speech and are co-constructed in a dialogic exchange. We saw that the set of functions used by the examiners and examinees were largely different. This, in turn, allowed us to reflect on how the role of the interlocutor, as warranted by the social context in which the interaction occurs, can influence the discourse functions they employ. Likewise, we showed, for the examinees, that task was a major context within which variation occurs. While not a principal focus of this book, we also saw how cultural background seemed to be linked to variation.

It might, at this point, be possible to argue that we can gain an adequate view of the functions of discourse by taking this micro-structural approach. However, we wish to explore, in the next two chapters, how, and whether, taking a macro-structural approach can reveal that, sitting above the micro-structural level, separate macro-structure discourse functions may be discerned. If there is a high degree of congruence between the micro- and the macro-structural analyses, then we may doubt that the discourse unit level is adding much to our understanding. If the functions at the turn level are the same as those at the discourse unit level, then while we might be able to comment on how the micro-structure level meshes with the macro-structural level to achieve this – for example, which repertoire of discourse functions at the turn level coalesce to form a specific function at the macro-level – our view of the functions of discourse would not change. We would have observed those functions at the micro-level. On the other hand, our second goal will be achieved if we find there is no, or imperfect, congruence between the functions revealed in our micro- and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002

macro-structural analyses, a point indicated by the differences identified in examiner discourse.

Such a view would help not merely to show how the levels coalesce, it would also show how higher, macro-level functions are projected from the functions at the micro-level, potentially revealing that the two may be relatively distinct functionally, but they are closely linked structurally. If we do discern such functions, and can argue that they are independent wholly or by degree of the functions at the micro-level, we may also see whether those discourse functions in turn are sensitive to the role of interlocutor, proficiency and task. Exploring these ideas will be the focus of the following two chapters.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:16:54, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.002