CHAPTER 3

Investigating Discourse Units
Approaching Learner Corpus Data at the Macro-Structural Level

3.1 Introduction

The previous chapter demonstrated that short-text MDA, when applied to the turn as the highest level of discourse micro-structure, produced valuable insights into the discourse of learner language. The analysis revealed that learner discourse functions varied by task and varied, at least from a distant reading perspective, by proficiency. Moreover, through concordance analysis, the findings in Chapter 2 demonstrated that there is an interplay between the discourse functions used by learners and examiners. Through this analysis, a repertoire of discourse functions for learners emerged at the level of discourse micro-structure and these functions were shown to be almost entirely distinct from those of the examiners, at least at this level.

Developing a better understanding of these emergent discourse functions is central to the focus of this chapter, which primarily investigates the learner–examiner relationship at a macro-level. To do so, this chapter is guided by one key research question: Does the discourse unit level of analysis produce a distinct set of functions from the turn-level analysis and if not, what functions does it produce? Given that discourse units (macro-structures) are composed of multiple turns (micro-structures) and given that discourse functions in learner and examiner turns vary, a reasonable hypothesis would be that discourse units composed of turns (with distinct discourse functions) would result in another distinct set of macro-discoursal discourse functions. If we discover that the discourse functions at the macro-structural levels are the same as the turn level, then it would suggest, at the very least, that near identical processes were at work in each case or, as is perhaps more likely, that the smaller unit, the turn, is actually controlling, and hence diagnostic of, the larger unit—the discourse unit. It is the aim of this chapter to identify how the evidence supports or refutes this hypothesis.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

In looking at discourse units, however, we need to highlight once again a simple, yet important, distinction between the analysis at this level and the analysis at the turn level. The turn-level analysis was focused at the level of the individual speaker, whereas the discourse units studied in this chapter are co-constructed. So, while at the turn level we are looking at, and accounting for, choices made by an individual who is selecting certain types of turn to perform specific functions, sometimes, but not always, in response to a turn from an interlocutor, at the discourse unit level, those choices always interact with one another as the speakers select and co-construct a discourse unit that performs a function in which they are both involved. It is easy to conceive of how this choice of function may be subjected to many pressures – it may be determined by task, it may be the choice principally of one speaker or it may be explicitly jointly negotiated, for example. So, in what follows, as well as considering the role that the discourse units have to play in the interaction between student and examiner, we will also be mindful of these pressures, taking different perspectives on the association of discourse units and functions to illustrate this, where it happens. We will also, where relevant, note how the roles of the examiner and student may vary with regard to the selection and production of a discourse unit.

In pursuing the goal of this chapter then, the primary analysis presented pertains to a short-text MDA of the discourse units in our TLC data, firstly generating and interpreting dimensions of functional variation before turning to the question of how, if at all, the resulting dimensions at the discourse unit level vary from those we have seen already in Chapter 2 at the turn level. To support this investigation, this chapter draws only on the first three dimensions, while the remaining two dimensions are unpacked in Chapter 4. Based upon the analysis for the first three dimensions, we look at variation and task afresh and, in doing so, present examples which help to illuminate the interaction between the micro- and the macro-structural levels. As such, this chapter offers a more in-depth reflection on the nature of variation in learner discourse by task and proficiency level while also further unpacking the relationship between the learners' turns and those of the examiners.

Our focus from this point onwards will be on the tasks in the examinations. This is for three reasons. Firstly, task was shown in Chapter 2 to be an important dimension of variation in our data. This makes a lot of sense, in terms of a form to function analysis. Secondly, the general preoccupations of learner corpus research, which often focuses on L1 background and in particular on errors, has led to less focus on tasks. Yet, at the level of discourse, we

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

think that task is the dominant variable – it is what determines whether the student is thought to have communicated successfully and appropriately or not. It will be the main focus of our analyses in this and the following four chapters where we will look at how task and proficiency interact at the level of macro-structure. We will return to the observation of the relationship of errors at the lexical level to discourse later in the book, but for now we will note that our overall position is that, at the macro-structural level, the focus on errors, especially grammatical errors, does not seem to be as fruitful at the micro-structural level. Finally, we set aside L1 background from the point of view of necessity. In aggregating our data into discourse units, we are greatly reducing the scale of observations we can make in the data. For example, while we might be able to rely on 1,737,822 observations if we focused on lexis (i.e. the size of the corpus used in Chapter 2, in words) in our annotated TLC data, when we shift to the discourse level we find that we are dealing with a much smaller set of observations – there are only 21,148 discourse units in the section of the TLC marked with this annotation. These then have to be viewed from a perspective of task, which spreads the set of possible observations more thinly. Within the tasks we are also interested in proficiency, reducing the possibility of making observations further as the data distributes for each task across the marking categories and the exam grades. The analyses in this chapter and the next show that we are able to make observations combining mark and task for exams at grades 6–8 in the corpus. Hence this is where we focus our discussion and claims, as we did in Chapter 2. To further subdivide by L1 background would scatter the data too much. However, in Chapter 9 we will return to consider L1 background in a subset of discourse units in a qualitative study.

To begin our investigation, we processed the TLC data using the same short-text MDA as before but focused upon the discourse unit level introduced in Chapter 1, not turns. The result of the analysis was five dimensions, the first three of which are outlined in the subsections here.

3.2 Dimension 1: Long versus Short Discourse Units

This dimension, as with the study in the previous chapter, distinguishes short and long units of analysis. Given that this is the second time that this has been observed, it is worth pausing to consider why this dimension is present again. *Prima facie*, we may assume that it has nothing to do with the proficiency of the speakers in the corpus, other than reflecting that sometimes such speakers produce units of analysis that are longer or shorter than average. However, it is worth understanding why the technique is so sensitive

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

to this that it routinely assigns the first dimension, the one where evidence is clearest and strongest, to this distinction. A clue lies in another feature of this dimension which is not clear to the reader. While the dimensions introduced in the previous chapter, and other dimensions in this chapter, are combinations of present and absent features, the first dimension contrasts short units of analysis with longer units of analysis exclusively on the basis of absences on the short unit side and presences on the longer unit side.

How does this represent a clue? It does so because the short-text MDA technique is most useful in sparse data environments – indeed, that is why we are using it rather than full MDA in this book. The shorter the unit of analysis, the more it is characterised by absences, as there is simply not enough of an opportunity for features to appear. The longer the unit of analysis, the more, relatively speaking, the analysis is dominated by presences, as the features observed have more chance to appear. In essence, the greatest influence on the presence of features is the length of the text. Yet absence may be overwhelming. Consider the absences in the following turn, which occurs 6,392 times in the TLC corpus, by way of illustration:

(22) s: yes

Of course, the interplay of frequency and opportunity has other predictable effects. For example, the frequency of a word will also impact on its propensity to appear given a limited number of observations. So, those words which have a propensity to appear frequently – many function words or fillers, for example – are ones which we should expect to see more often in short text sequences than long ones. For the analyses carried out here this observation has some relevance. Consider pronouns – a limited, relatively closed class of words which not only have a high propensity to occur in English but which are even more likely to occur in speech. We should expect to see them more than, for example, an ordinal noun, which, while also from a closed class of items, has neither a high propensity to occur nor is particularly indicative of speech. To illustrate this argument, the words *I* and *you* are the first and second most frequent words in the TLC corpus used in this chapter with frequencies of 62,975 and 58,912, respectively. By contrast, *first*, *second* and *third* occur in that data 1,952, 320 and 120 times, respectively. So, both the category *pronoun* and the individual members of the category examined have a higher propensity to occur in the data than the category *ordinal noun* and the words from that category. This is to be expected as the personal pronouns examined 'are by far most common in conversation' (Biber et al., 2021: 332) and *I* and *you* show the same propensity to be the most frequent personal pronouns in this corpus as they do in the findings reported by Biber et al. (2021: 333).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

Likewise, the lower frequency of the ordinals is to be expected as they are 'least frequent in conversation' (Biber et al., 2021: 280). So, like the other analyses we have presented, the first dimension is related to the length of the textual unit analysed. In particular, it opposes all the presences of features on one side of the dimension with all the absences of features on the other side of the dimension. This leads to an opposition between long and short textual units, because as the objects under study get longer in terms of word length, they are more likely to have the presence of particular grammatical features. Conversely, shorter units are more likely to exhibit the absence of features.

Note that this complicates our view of Dimension 1. We passed over it in the previous analysis, but now we can see that, to some extent, Dimension 1 is, broadly speaking, showing us features of the interaction which are markedly frequent in speech versus those which are not. The dimension, through presence, is showing that important factors relating to mode of production and function – conversation – emerge from the data. Hence the dimension might be best characterised as data in which there is enough evidence that frequent features of spoken language become visible (on the –ve) versus data in which the frequent features of spoken interaction are not frequent enough to be seen (on the +ve). As may be predicted by that characterisation, the –ve end of the dimension is characterised exclusively by presences – and these are commonly of the type of features discussed so far, for example object pronouns (occurring with a coordinate of –0.424 on this dimension) which have a propensity to occur frequently enough in conversation to emerge from our data (there are fourteen such features overall). Likewise, the preponderance of absences relative to presences may also be understood as a predictable quality of the +ve end of the dimension (this is characterised by thirty-seven absences). The same argument holds for Dimension 1, as explored in Chapter 2. So while long and short sequences are one way of looking at this dimension, it is important to remember that the mode of the interaction in particular also has a powerful effect on the dimension. We should also accept the possibility that, if we were to break the corpus into tasks and analyse those, when we compare them, we may also find that function would cause similar features to occur in Dimension 1.

For the purpose of the analysis presented here we should note that Dimension 1 may hold more interest than is at first apparent. These insights have consequences for our consideration of the data produced in the exam – the results from Dimension 1, for both the turn and discourse unit analysis, indicate strongly that the speech-like nature of the data is strong. In a spoken language examination designed to elicit spontaneous

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

conversations, this is an in part a predictable, but also reassuring, finding. While the students may have prepared for the examination, there is no indication here that what we are seeing in the corpus is not speech-like, whether we view it at the turn or discourse unit level. So whatever preparation the students may have undertaken, this does not seem to result, for example, in the regurgitation of pre-prepared and memorised written material. The results from this dimension thus give a provisional attestation of the conversational nature of the data. We will consider further the insights into conversational interaction that Dimension 1 may indicate in Chapter 7.

3.3 Dimension 2: Descriptive and Affective versus Informative and Instructive

Dimension 2 shows how, on the one hand, description and the encoding of emotions on the positive side of the dimension are set against discourse units principally concerned with information and instructions on the negative side of the dimension (Table 3.1).

To begin with the Descriptive and Affective function, within such discourse units we see features such as copular verbs, predicative adjectives and comparatives used to encode description, leading to turns such as 'it's more beautiful'.[1] Passive constructions and cause subordinators are also associated with discourse units with a Descriptive and Affective function, with these being used to describe and explain a particular phenomenon as evidenced in turns within discourse units such as:

(23) S: and it is exported to the a lot of countries
E: mm
S: er because it's very famous in Spain[2]

Features associated with the Informative and Instructive function have a future orientation, for example prediction modals, which are used to talk about things that will happen in the future. Likewise, suasive verbs are used, and these imply intentions to bring about some change in the future. Consider the example that follows, which is the most strongly associated discourse unit to negative Dimension 2 (contribution: 0.066, coordinate: −0.657), taken from the exam of a grade 7 Indian student (file 2_7_IN_21):

[1] TLC file 2_8_AR_15.
[2] TLC file 2_7_SP_43.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

Table 3.1 *The linguistic features strongly associated with Dimension 2.*

Dim. 2	Features (coordinates, contributions)
+	Time Adverb_A (0.15, 0.707), Nominalisation_A (0.163, 0.744), Object Pronoun_A (0.17, 0.79), Stance Verb_A (0.183, 0.93), Demonstrative Pronoun_P (0.186, 0.696), WH-Word_A (0.206, 0.828), Contrastive Conjunction_P (0.225, 0.927), Prediction Modal_A (0.242, 1.68), Third-Person Singular verb_P (0.245, 1.781), Phrasal Verb_A (0.262, 2.055), Pronoun *it*_P (0.262, 1.727), Cause Subordinator_P (0.27, 1.094), Infinitive_A (0.328, 2.247), Passive_P (0.386, 0.789), Predicative Adjective_P (0.394, 3.68), Question Mark_A (0.494, 3.054), Comparative_P (0.508, 1.404), Existential there_P (0.52, 1.807), General Verb_A (0.537, 1.268), Copular Verb_P (0.558, 4.964), Second-Person Pronoun_A (0.8, 4.642), First-Person Pronoun_A (0.823, 3.095), Subject Pronoun_A (0.857, 0.723)
–	BE as main verb_A (0.845, 3.915), Suasive Verb_P (0.836, 2.815), Phrasal Verb_P (0.568, 4.457), Third-Person Singular verb_A (0.515, 3.743), Predicative Adjective_A (0.465, 4.344), Prediction Modal_P (0.465, 3.231), Time Adverb_P (0.38, 1.79), Pronoun *it*_A (0.354, 2.333), Copular Verb_A (0.321, 2.853), Stance Verb_P (0.317, 1.607), Infinitive_P (0.3, 2.059), Nominalisation_P (0.288, 1.313), Object Pronoun_P (0.286, 1.334), Indefinite Pronoun_P (0.261, 0.947), Public Verb_P (0.253, 1.042), Amplifier_A (0.222, 0.838), Question Mark_P (0.198, 1.222), Second-Person Pronoun_P (0.159, 0.923)

(24) E: move on now to the er interactive task
S: mm
E: and as you know in this task erm I will tell you something and then you have to ask me questions to find out more information and make comments you need to keep the conversation going and after four minutes I'll end the conversation
S: okay
E: are you ready?
S: yeah

This discourse unit is notable because the function here is clearly selected by the examiner and the student has an entirely passive role in this interaction. If we look at the top 100 discourse units most associated with negative Dimension 2, we see the same – a function dominated by the examiner.[3]

[3] Note that throughout the book, where we have analysed 100 discourse units, our procedure has been as follows. We began by analysing the fifty most prototypical discourse units. That analysis was then replicated on the next fifty most prototypical units. Our aim was both to repeat and attempt to falsify the initial analysis. In all cases reported in this book, the replication was successful and the attempt to falsify failed.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

This happens because this function is important in discourse management. Such discourse units are primarily instructive. They are often used to bring about change within the interaction, such as moving onto the next task in the examination. Such discourse units are linked to the presence of stance verbs and infinitives. Within the corpus, they have an important part to play when the examiner guides the candidate and advises them that 'you need to keep the conversation going'.[4] Infinitives, such as *keep* in this case, are an important feature that pushes discourse units containing exam management interactions onto the positive side of Dimension 2. An example of such a discourse unit, which shows the broader sequence within which the transition to a new task is set, follows. Note that infinitives (*find out*, *keep*) and stance verbs (*need*) are present:

(25) E: well now we'll move on to the next part
S: okay
E: alright er for this part I'll tell you something
S: okay yes
E: then you have to ask me questions to find out
S: yes
E: more information and
S: okay
E: make comments you need to keep the conversation going after four minutes I'll end the conversation
S: okay
E: are you ready?
S: yes
E: okay
S: of course

The examiner is providing instructions relevant to the exam situation and the interaction is metadiscoursal. However, not all of the discourse units with an Informative and Instructive function relate so clearly to discourse management – the example that follows is from an Italian student speaking towards the end of the Conversation task in a grade 6 exam.[5] While the discourse unit begins as discourse management, it switches to an attempt at gathering information. The Informative and Instructive function becomes pronounced because of the examiner using information seeking as a means of scaffolding the interaction with the student:

[4] As a result, this phrase occurs 392 times in the discourse unit annotated TLC corpus, always being spoken by the examiner.
[5] TLC file 2_6_IT_60.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

(26) E: let's move on then let me ask you about er rules and regulations okay
S: erm
E: so let me ask you about er what do you have to do if you want to get a driving licence in Italy? what do you have to do?
S: er you'll you s=
E: yeah how do I get a driving licence in Italy? what do I do?
S: erm
E: okay
S: about er abou= er about

The identification of the dimensions, and an understanding of their relationship both to turns and to discourse units, is the beginnings of an analysis. But to better understand how the functions revealed by the dimensions, as they relate to the discourse units, are subject to different pressures, requires further exploration. We saw in the previous chapter that clear relationships exist between the dimensions at the turn level and variables present in the corpus, which would suggest that the turn-level dimensions themselves represent important functions which help us to understand the students represented in the corpus and their attainment. To see if the same holds true for the co-constructed discourse units, for each dimension explored in this chapter we will carry out a similar study, exploring the relationship of the discourse units to attainment and task. If the functions vary by task, this is an important signal that certain functions are important for a task, while others may be a clear signal that a task is associated with low performance. This, if true, in turn would suggest that we should see some variation by dimension with reference to attainment. In the analyses that follow, we will consider, for the main tasks, whether the dimension seems to vary according to attainment – the study will be repeated for each Dimension analysed. Table 3.2 shows how attainment in the Conversation, Discussion and Interactive tasks varies, from the perspective of the discourse unit, on Dimension 2.

What we see on Dimension 2 are participants who receive an A in the Conversation task using the Descriptive and Affective function more. As discourse units which place themselves on the Informative and Instructive (negative) part of the dimension increase, so the attainment of the student declines. A good reason for this has been suggested already – the Informative and Instructive function is associated with the provision of direction to the students. The more that this is needed, it seems, the poorer the performance of the student. A good example of this comes from file 2_6_CH_22. At the start of the Conversation task, across two discourse units an examiner tries to initiate a conversation on a set topic – the extract

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

Table 3.2 *The association of the discourse units in the Conversation, Discussion, and Interactive tasks with Dimension 2 according to the mark learners received.*

Conversation_A	0.036
Conversation_B	−0.017
Conversation_C	−0.051
Conversation_D	−0.248
Discussion_A	0.162
Discussion_B	0.129
Discussion_C	0.102
Discussion_D	−0.023
Interactive_A	−0.181
Interactive_B	−0.236
Interactive_C	−0.221
Interactive_D	−0.792

that follows presents the examiner turns across two discourse units, showing how the instruction must be restated and reformulated. The problem continues beyond this extract, but the effect is obvious – the examiner is initiating discourse units associated with the Informative and Instructive function in an attempt to encourage the student to engage with the task.

(27) E: now we're going to talk about something different
S: okay
E: let's talk for a moment about travel
S: yeah
E: what's your opinion of travel?
S: I have been to Shanghai
E: But what's your opinion of travel
S: er I don't know this do= this question but I know
E: what do you
S: just
E: think about travel
S: travel is you can play in a place and you
E: but what do you think about travel itself? what's your thoughts about the travel? we're talking about travel
S: oh yeah
E: not Shanghai but travel

Following on from the previous example, it is perhaps understandable how, in the Discussion task, candidates who receive higher marks (A–C) are able to engage more with discourse units which are placed on the Descriptive and Affective part of this dimension (see Table 3.2); this is a function they draw upon as a resource to participate in and sustain the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

discussion. This is in large part determined by the nature of the prompts from the examiner – the discussion itself is kept going by requests for information, and often the response to those requests includes descriptive and affective components. Consider the following discourse unit from file 2_6_AR_15:

(28) E: so what are we going to talk about?
S: well I'm going to talk about New York because
E: okay
S: I think it's an amazing place I loved it I went there two times and I loved it so much
E: what what were you doing in New York?
S: well erm we went well there was a like <unclear text='place'> you can go shopping and my father and I love shopping buy clothes or electronic devices
E: mm
S: and so we went there and we go to the Fifth Street and Broadway
E: okay
S: and we bought lots of things
E: alright and erm you said that you never stay at the same hotel
S: no we never stayed at the same hotel because we like to like go to different places and know how they are because they're all different there lots of hotels there because <unclear text='there are lots of tou='>
E: lots of
S: hotels there
E: hotels yeah
S: because there are lots of tourists

This discourse unit is a perfect example of a Descriptive and Affective discourse unit. The examiner's prompts instruct the examinee to produce replies in response to a topic introduced by the student. The student situates the response in the past by describing features of visits to New York that they have had. In response to the prompts from the examiner, the student produces many replies which include affective language such as *amazing*, *loved*, *love* and *like*.[6] By contrast, the students awarded a D are associated with discourse units which incline to the Informative and Instructive side of the dimension (see Table 3.2). The reason is much the same as with the Conversation task – minimal responses from the learners,

[6] In total, the discussion runs across eight discourse units. The top-down coding of the discourse units here is revealing – the affective nature of the language throughout is clear in five of the eight discourse units, which are coded to include feelings. Six of the eight discourse units are also either coded as relating to the past or as being primarily descriptive, both driven by the student describing aspects of their travels in the past in response to examiner prompts.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

associated with a lack of comprehension, force a series of scaffolding turns from the examiner which draw on the features defining the Informative and Instructive function, as can be seen in this discourse unit from the beginning of the Discussion task in file 2_6_IN_30:

(29) E: so we're gonna begin with your topic and what are you going to talk to me about?
S: er ma= I'm going to talk about my er football
E: oh right you like football not cricket
S: no
E: mm <pause/> and what's your favourite club?
S: my favourite club is erm <pause length='short'/> Barclays Premier League
E: yes but that's not a club that's a league what's your favourite club?
S: club er <pause length='short'/> FC Barcelona
E: okay so not the Barclays Premier League that's
S: yes
E: something different okay why do you like Barcelona so much?
S: pardon

Table 3.2 also shows that the link of the negative side of the dimension to low attainment continues in the Interactive task. This task more naturally draws upon the Informative and Instructive function as the Interactive task requires learners to provide advice to the examiner regarding what to do in a specific situation. However, the location of discourse units on this side of the Dimension is much more marked for the poorest performing students. When we look at the discourse units in the Interactive task, we can see that the poorest performing students evoke many questions from the examiner, who tries to repair the interaction. This leads to discourse units, such as the following from file 2_7_CH_18, which is associated with the Informative and Instructive function:

(30) E: okay <pause/> <anon/> my friend is staying with me at the moment <pause/> we've had quite a few problems I'm not sure what to do about it
S: erm <pause/> er w= er would er would you ask the other friends er er what would you stay with he?
E: I'm sorry can you repeat that?
S: erm er I I I change another questions
E: alright

Of course, the analysis so far has simply looked at grade and task type. If it is the case that student proficiency is a key driver of this dimension, then we should expect the level of the exam to also be an important variable: if

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

Table 3.3 *The Dimension 2 association of the discourse units in the Conversation task from groups of learners defined by proficiency, grade and overall mark.*

Conversation_A_B1_grade6	−0.121
Conversation_A_B2_grade7	0.231
Conversation_A_B2_grade8	0.199
Conversation_B_B1_grade6	−0.242
Conversation_B_B2_grade7	0.218
Conversation_B_B2_grade8	0.12
Conversation_C_B1_grade6	−0.296
Conversation_C_B2_grade7	0.272
Conversation_C_B2_grade8	0.07
Conversation_D_B1_grade6	−0.361
Conversation_D_B2_grade7	0.027
Conversation_D_B2_grade8	0.048

students are entered for exams that are appropriate to their level of proficiency, then we might expect, with increasing proficiency across the grades, that the degree of scaffolding from the examiner may reduce. In other words, the likelihood of a speaker at grade 6 requiring scaffolding in a given task should be higher than speakers at grade 7 in the same task and, in turn, we should see less of this still in that task at grade 8. However, the possibility exists that scaffolding persists, but that what is scaffolded changes. Table 3.3 explores the hypothesis that scaffolding declines with proficiency. What is remarkable about this table is that, for the Conversation task, discourse units with an Informative and Instructive function are only dominant at the lowest grade examined, grade 6. At that grade, the placement on grade 6 declines from most (grade D) to least (grade A) quite smoothly. Beyond grade 6, the impact of this need for scaffolding seems to be less uniform, leading to all subsequent combinations of grade, task and attainment placing themselves on the Descriptive and Affective side of the dimension to different degrees. While less marked, the same pattern generally holds for the Discussion task, with the exception of students awarded an A at grade 6, as shown in Table 3.4. From this we can certainly support the view that grade 6 students need more scaffolding than higher grade students, regardless of whether we consider the Conversation or Discussion tasks.

A separate discussion is needed for the Interactive task, as that is part of the examination for grade 7 onwards only. The task naturally draws on the Informative and Instructive function because of its nature. However, if scaffolding is needed, we should presumably see a greater weight of discourse units being drawn from this function than the needs of the task itself

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

Table 3.4 *The Dimension 2 association of the discourse units in the Discussion task from groups of learners defined by proficiency, grade and overall mark.*

Discussion_A_B1_grade6	0.077
Discussion_A_B2_grade7	0.226
Discussion_A_B2_grade8	0.366
Discussion_B_B1_grade6	−0.023
Discussion_B_B2_grade7	0.321
Discussion_B_B2_grade8	0.147
Discussion_C_B1_grade6	−0.018
Discussion_C_B2_grade7	0.223
Discussion_C_B2_grade8	0.19
Discussion_D_B1_grade6	−0.154
Discussion_D_B2_grade7	0.092
Discussion_D_B2_grade8	0.207

require. If our hypothesis about this tendency to draw on the Informative and Instructive function lessens as proficiency develops, we should see, as students progress from grades 7 to 8, the weakest students at both levels being associated with more Informative and Instructive discourse units, but the level 8 students at any level being associated with fewer such discourse units than the level 7 students. Table 3.5 shows a slightly different picture. Both of these statements hold true for students awarded a D or C at levels 7 and 8. However, students awarded a B or A break this trend. At grade 7, students awarded an A or B seem to associate equally with the Informative and Instructive function, though both are associated less noticeably than students awarded a C or D at either grade. At grade 8, however, the association increases, rather than decreases, for both, and the association of the Informative and Instructive function with students awarded a B at grade 8 is so marked that it exceeds that with students awarded a C grade at either level 7 or 8.

The answer for the persistence of a pull from the Instructive and Informative function beyond grade 6 is, in some ways, obvious. While the scaffolding that is very evident at grade 6 may have faded, the need for the examiner to give instructions to the examinee has not. In fact, of the ten discourse units most associated with the Informative and Instructive function mentioned earlier, nine are from grade 7 and 8 exams and relate solely to the provision of instructions to the student regarding a change of task, for example. Tellingly, the tenth discourse unit, and eleventh, in this sequence comes from grade 6 students and in both cases the instruction is given with scaffolding – the students cannot understand the instructions

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

Table 3.5 *The Dimension 2 association of the discourse units in the Interactive task from groups of learners defined by proficiency, grade and overall mark.*

Interactive_A_B2_grade7	−0.164
Interactive_A_B2_grade8	−0.208
Interactive_B_B2_grade7	−0.161
Interactive_B_B2_grade8	−0.396
Interactive_C_B2_grade7	−0.227
Interactive_C_B2_grade8	−0.211
Interactive_D_B2_grade7	−0.844
Interactive_D_B2_grade8	−0.697

and they have to be reformulated, as in the following example of an Italian student taking a grade 6 exam (file 2_6_IT_60):[7]

(31) E: let's move on then let me ask you about er rules and regulations okay
S: erm
E: so let me ask you about er what do you have to do if you want to get a driving licence in Italy? what do you have to do?
S: er you'll <unclear/> you s=
E: yeah how do I get a driving licence in Italy? what do I do?
S: erm <pause length='short'/>
E: okay
S: about er abou= er about

Here we have the function used to provide direction about the examination, but the function is also used to scaffold – the student does not understand the prompt. They have difficulty in communicating – with false starts and signs of uncertainty and hesitation throughout the discourse unit. They do not signal understanding. The prompting from the examiner, accordingly, continues in the following discourse unit. This is quite unlike the other discourse units drawn from the higher grade exams where instruction is given, but no scaffolding is needed. If we expand our examination of the discourse units most associated with the Informative and Instructive function, and shift the micro-structural focus to the lexical level, the domination of that function by the examiner becomes overwhelmingly clear. Looking at the top 100 discourse units which distribute that way (the discourse units which are prototypical of this function), we

[7] Note this discourse unit appears earlier as Example 26. It is repeated here to make a different point, with the change of numbering permitting the possibility to separate out the different points being made when referring to the example.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

find 6,165 words spoken. Of these words, only 642 (10.41 per cent versus a proportion of 57 per cent in the whole corpus) are spoken by the student. These are produced typically as one-word utterances and are overwhelmingly either phatic or signal agreement. The four most frequent words spoken by students, *okay* (150), *yes* (118), *mm* (59) and *yeah* (45) account for 372 of the words spoken (57.94 per cent) by students in these units. By contrast, the examiners speak 5,523 of the words in these discourse units, meaning that 89.59 per cent (versus a proportion of 43 per cent in the whole corpus) of the words in these discourse units are spoken by the examiner. When we look at the examiner's turns, however, we see immediately that they are heavily formulaic – there are fixed expressions aimed at managing the transition to a new task (*are you ready*, eighty-nine examples), preparing the student for what is to come (*after four minutes I'll end the conversation*, sixty-nine examples) and explaining to them their role in the task (*you need to keep the conversation going*, ninety examples). The input from the examiner is generally formulaic, the response from the student is minimal and passive, and these examples are almost all from the first discourse unit of a task. The exceptions in that group of discourse units are produced within a task, not at the start. In such cases the switch to the Informative and Instructive function at that point is a sign of a repair because that behaviour, dominated by the examiner, usually introduces a task; it does not generally form part of the task itself. Where it does, it is because, in essence, the examiner is having to intervene to ensure the smooth running of the task in the face of the inability of the student to perform the task.

By contrast, if we look at the 100 prototypical discourse units on the Descriptive and Affective side of the dimension, we see a function dominated by the student in which formulaic speech has little role to play. The discourse units contain 6,180 words, 1,844 spoken by the examiner (29.84 per cent versus a proportion of 43 per cent in the whole corpus) and 4,336 spoken by the student (70.16 per cent versus a proportion of 57 per cent in the whole corpus). While dominated by the student, the dominance here is less marked than the difference seen on the opposite side of the dimension. If we set a threshold of a minimum frequency of four for an n-gram of three or more to be considered frequent, this dimension barely registers any formulaic speech – no n-grams in the examiner speech meet this threshold, which is far from exacting, while few meet the threshold in the student speech. None of these are greater than three words in length, and the most frequent occurs seventeen times (*a lot of*). The n-grams themselves relate to the realisation of the discourse unit function,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

mainly through being expressions of quantity or degree such as *a lot of* and *it was very*.

Of course, we must bear in mind, when considering our groups of 100 discourse units to characterise a function, that we are looking at what we might consider to be prototypical examples. These are the discourse units most strongly associated with the function in question. As we move away from the prototypes we may see, as the fuzziness of prototypicality ensues, the features which may be seen clearly in the prototype worn away as different blends of discourse functions come into play and the salience of the dimension we are interested in diminishes. Nonetheless, considering the prototypical cases is of importance as it allows us to consider a characterisation of the discourse unit function that points, as far as is possible, to the functional nature of the discourse unit and the conditions of its creation in as clear a way as possible. We must not lose sight of the fact, however, that away from the prototypical cases any given observation we make of the prototype may, or may not, be in play for any given example we explore.

Returning to the finding, the Informative and Instructive function, in contrast to that on the positive side of the dimension, prototypically shows a discourse unit in which, in essence, an examiner is using such a preponderance of Information-Seeking turns (as identified in the short-text MDA of examiner turns in Chapter 2) that the micro-level does, in fact, converge functionally with the macro-level. The macro-structure here has a function built from a series of micro-structures which have largely the same function. So in one dimension, we can see that a micro-structure may be nearly coterminous, functionally, with the macro-structure within which it sits but also, on other occasions, the macro-structure is built from a range of micro-structural functions. In part this is determined by dominance – in co-constructed turns, macro-structures are built from inputs from interlocutors (the student and examiner) who have distinct repertoires of discourse functions licensed by the tasks they are engaged in. The result is almost bound to be different as a result – it is a blend of different sets of functions. However, where one speaker dominates and the function they are producing essentially requires a sequence of mono-functional utterances, as we see with the Information-Seeking function with the examiner when it occurs at the start of a task, the conditions for a high degree of congruence between the micro- and macro-structures of discourse are met.

The analysis of the 100 discourse units most associated with each side of the dimension yielded, in the context of prototypical examples, a helpful, principally qualitative, insight into the relationship between the micro- and macro-structures at work, as well as the dynamics of the interaction

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

between the two interlocutors. Accordingly, we will look at these 100 prototypical discourse units in our exploration of each dimension in this chapter and that following.

3.4 Dimension 3: Unknown (Irrealis) versus Known (Realis)

Dimension 3 balances talking about the unknown on the positive side (Irrealis) against talking about the known (Realis) on the negative side (Table 3.6).

In the Irrealis function numerous features are used to describe action that is not known to have happened. For example, necessity modals can be used to indicate what should be done as opposed to what has been done. Conditional subordinators are used to introduce a possible situation. Likewise, stance verbs can be used to indicate desired action that has not happened, such as *I would like to do some exercise*. Overall, these features co-occur in discourse units which discuss events that are not known to have happened and are thus in the irrealis mood. The following discourse unit shows an Irrealis function discourse unit (file 2_7_SP_38):

Table 3.6 *The linguistic features strongly associated with Dimension 3.*

Dim. 3	Features (coordinates, contributions)
+	Cause Subordinator_P (0.206, 0.706), Numeral Noun_A (0.221, 1.93), Third-Person Pronoun_P (0.23, 1.131), Private Verb_P (0.234, 1.58), Auxiliary DO_P (0.268, 1.846), Subordination_A (0.275, 1.642), Stance Verb_P (0.288, 1.474), Negative Interjection_P (0.288, 0.79), Contrastive Conjunction_P (0.305, 1.89), Possession_A (0.308, 2), Question Mark_A (0.318, 1.397), Analytic Negation_P (0.339, 2.848), Third-Person Singular verb_A (0.412, 2.65), Proper Noun_A (0.415, 4.069), That + verb complement_P (0.431, 1.24), Conditional Subordinator_P (0.474, 2.685), Contracted Forms_A (0.49, 3.383), Attributive Adjective_A (0.504, 1.795), Necessity Modal_P (0.51, 1.407), BE as main verb_A (0.857, 4.464)
–	Numeral Noun_P (–0.998, 8.731), Adjective + to complement clause_P (–0.593, 1.732), Preposition_A (–0.486, 1.169), Numeral Determiner_P (–0.462, 2.298), Progressive Aspect_P (–0.448, 2.148), Proper Noun_P (–0.396, 3.887), Analytic Negation_A (–0.357, 3), Private Verb_A (–0.344, 2.324), Auxiliary DO_A (–0.301, 2.072), Time Adverb_P (–0.261, 0.935), Possession_P (–0.237, 1.539), Subordination_P (–0.224, 1.341), Contrastive Conjunction_A (–0.22, 1.361), Phrasal Verb_P (–0.215, 0.703), Contracted Forms_P (–0.201, 1.383), Third-Person Singular verb_P (–0.196, 1.261), Third-Person Pronoun_A (–0.183, 0.902), Stance Verb_A (–0.167, 0.853), Conditional Subordinator_A (–0.155, 0.877)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

(32) E: mm mm I mean I would I would like to do some type of exercise with
S: yeah
E: with him
S: yeah
E: but I don't know what exercise we could do

In this sequence the examiner talks about some poorly specified activity in the future.

By contrast, the features on the negative side combine to create a Realis function – they are associated with describing known events and entities, realised through numerous noun types and noun modifiers such as numeral nouns, numeral determiners and proper nouns. Possession is also associated with the Realis function and this is used to describe what is known to belong to someone. For example, in the opening discourse unit of the Greeting task of file 2_6_ME_96, which has Dimension 3 coordinates of –0.579, the student shows their identification card to the examiner and says 'my identification' and then states 'My name it is.' After that the examiner notes 'your … identification … is here'.[8] All of these are linked to the Realis function through possessive determiners. Overall, the features which constitute the negative side of Dimension 3 co-occur in discourse units where the speakers are engaged in talking about things they know to have happened/exist and or are happening. The discourse units most associated with this side are often greeting exchanges where the speakers exchange names, ID information and general background information, as in the following discourse unit (file 2_7_IN_23):

(33) E: hello come in <pause/>
S: good morning sir
E: good morning come and sit down <pause/> so how are you today? okay
S: yeah I'm really good thank you
E: good so can I check your I-ID please?
S: <unclear text='yeah'/>
E: okay thank you very much
S: <unclear text='okay'/>
E: so your name is it's a very long name
S: <laugh/> <anon/>
E: <anon/> oh
S: <laugh/>
E: okay and my name is <anon/>
S: mm
E: and you're doing grade seven

[8] The ellipses are used in place of backchannels vocalised by the student at these points.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

S: yeah
E: yeah
S: yeah

Here is a discourse unit in which the discussion is focused on requests for information in the here and now – information about the candidate and their feelings. Possessive determiners, once again, anchor the discourse unit to the Realis function.

If we look at the 100 prototypical discourse units, on either side of the dimension, with coordinates furthest towards either pole, we see a clear confirmation of the description given so far. On the positive side (Irrealis), we find a function dominated by the examinee. Overall, the discourse units contain 5,997 words: 4,224 (70.44 per cent) spoken by the students, 1,773 (29.56 per cent) by the examiners. Both do produce some formulaic speech, but, using our threshold of three-plus word sequences produced at least four times, we find few, though those that we do find clearly relate to the Irrealis function, such as *I think that* (five examples) for the examinee and *I don't know* (four examples) for the examiner. So the n-grams show that both examiner and examinee are using the Irrealis function.

On the negative side, we see a function dominated by the examiner as the discourse units which are most markedly Realis are those, as the example given suggests, which are at the start of the examination where the examiner is checking a series of facts about the student. For this function, our 100 discourse units are composed of 6,363 words: 1,647 (25.88 per cent) from students, 4,716 from examiners (74.12 per cent). However, formulaic speech is common in the speech of both student and examiner, all of which relates to information presentation, aligning well with the Realis function. For the examiner, formulaic sequences relate to politeness (*how are you*, seventy examples, *nice to meet you*, sixty examples, *thank you very much*, forty-one examples), information seeking (*how are you today*, fifty-four examples, *what's your name*, fifty-eight examples), statements of fact (*you're doing grade …*, forty examples) and directions (*come and sit down*, twenty-nine examples, *please sit down*, seven examples). The student also produces such sequences, with a politeness formula (*nice to meet you*, twenty-two examples) and statements of fact (*I'm fine thank you*, five examples, *my name is*, fifty-four examples) giving rise to frequent n-grams. It is interesting to note that the n-grams alone strongly indicate that, with reference to the micro-structural analysis in the previous chapter, there is not as strong a link between micro-structure discourse functions and the macro-structural functions for the Realis function. While there are n-grams which clearly align with the Realis function, many of the n-grams

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

Table 3.7 *The Dimension 3 association of the discourse units in the Conversation, Discussion and Interactive tasks from groups of learners defined by overall mark.*

Conversation_A	0.149
Conversation_B	0.217
Conversation_C	0.217
Conversation_D	0.259
Discussion_A	−0.117
Discussion_B	−0.08
Discussion_C	−0.091
Discussion_D	−0.234
Interactive_A	0.414
Interactive_B	0.463
Interactive_C	0.452
Interactive_D	0.442

observed are not clearly linked to Realis, indicating a more complex mapping of micro- to macro-structure here than the one we observed in the Descriptive and Affective function.

Returning to consider our macro-structural functions, to what extent, if any, is the ability to use the two functions associated with the tasks in the corpus? Let us begin by considering the Conversation task.

Table 3.7 presents the association of the discourse units in the tasks according to the mark that was received. It shows that the Conversation task is linked, for all four grades, with the Irrealis function. However, there is a clear difference across the grades, with the coordinates for students receiving the lowest score being more strongly associated with the Irrealis function than those with better grades. The highest grade is linked to discourse units which are least associated with the Irrealis function.

By contrast, the Discussion task is linked to the Realis function. As with the Conversation task, however, the lowest grade exams have coordinates which are more strongly associated with the Realis function (see Table 3.7). Hence the same question applies to both the Discussion and Conversation task findings – what links lower grades to the pronounced reliance on these functions?

In the case of the Realis function, there are thirty-eight students in the corpus who were awarded a D grade for their Discussion task. All of the students draw heavily on the Realis function – in the Discussion task they generate 313 discourse units collectively. Only 119 (38.02 per cent) of these discourse units have coordinates which associate them with the Irrealis function. By contrast, students with marks A–C produce 8,140 discourse

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

units as part of the Discussion task, with 3,807 (46.77 per cent) being associated with the Irrealis function. There is a marked preference for Realis in the weaker students, a finding supported if we consider the Log Ratio score for the difference between the two, which is 4.07 when we compare the ratio of Irrealis in the students scoring A–C to those using Irrealis at grade D, showing a marked effect.

What linguistic processes are driving this difference in the discourse? The drivers are distributed across the range of features that form the dimension. One example is possessive determiners and other pronouns, both parts of the Realis function. For example, for some students these are linked to a limited repertoire of reference and cohesion in their speech. For instance, one Indian student (in file 2_6_IN_39), scoring D for the Discussion, has a chosen topic of 'My favourite book'. The discussion stretches over 5 discourse units amounting to 460 words. All of the discourse units are Realis. An important organising feature of the Realis function, possessive determiners, is strongly present in this student's speech, with twenty possessive determiners used in the task. Eight are used by the examiner (*your* six times, *my* once and *his* once). Twelve are produced by the student (*my* ten times, *your* twice). The student repeats the phrase *my favourite book* four times. The student fails to exhibit variety in the formation of cohesive chains of reference to the book, relying instead on reiteration—a form of lexical cohesion rooted in repetition rather than, for example, using near synonyms, the most direct form of cohesion (Halliday and Hassan, 2013: 278). They refer to the book by using possessive determiners in the phrase *my favourite book*, by using the pronoun *it* or by simple repetition of the word *book*, as the following excerpt shows (from 2_6_IN):

(34) E: good aren't they? good movies excellent okay and would you like to give this book to someone?
S: no I can't give as it is my favourite book and er and <pause/> I can give my f= my father as he only gave me the book
E: okay alright and if you got a new book what will you do with the new book?
S: er I will keep it aside as I will read the other new book erm and I will keep it put in a locker as it is my favourite book

The sequence shows a repeated use of the phrase *as it is my favourite book* by the candidate in each of their utterances. The second turn by the student also contains an error where the student fails to use a suitable referring expression to sustain a chain of reference *I can give my f= my father as he only gave me the book* – presumably a reference to the book should have appeared prior to *my father*. Reference to the book is not re-lexicalised,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

for example to *novel* or *story*. It is also notable that, in this discussion, the actual content of the book is never introduced. The student returns to a discussion of *my favourite book* or asks the examiner about their favourite book. By contrast, a Mexican student, in file 2_6_ME_13, discusses their favourite book in a discussion awarded an A. The first discourse unit following the introduction of the task, in which the book is named as *A Thousand of Splendid Suns*, is clearly different from the previous example from the Indian learner:

(35) E: I see so erm <pause/> <clears/> er <pause/> er how would you say this book is organised?
S: in four parts
E: okay <pause/>
S: and do you <unclear/>
E: yes yeah ple-ple-please
S: I'll tell you the argument it's about a girl that is very poor and she's the illegitim= child of a very rich man
E: uhu
S: so <pause/> she with erm because of life circumstances
E: mm
S: she has to erm get married with a man you know I it's a book <unclear/> if the erm events are in Afghanistan you know
E: mm
S: with Islam and all that thing
E: <clears/>
S: she has to go and get married with a man that is like forty years old
E: uhu
S: than her
E: yeah
S: and erm in the first days he's very nice and all that but then she get pregnant and but she abort
E: uhu
S: naturally she's like she can't have childs
E: ah

The student does use *book* as a lexical cohesive device through repetition without varying reference to it through near synonyms or synonyms. However, as the student refers directly to the book less often, this may not be a great issue, in part because they focus on the content of the book rather than simply referring frequently to the book. For example, in the sequence *it's a book <unclear/> if the erm events are in Afghanistan* – the student refers to the contents of the book and mentions the book by implication, rather than by repeating to achieve a reiteration effect. The student also demonstrates less reliance on a limited repertoire of possessive

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

determiners across seven discourse units, five of which are Realis and two of which are Irrealis. Only nine possessive determiners are produced across 643 words, 6 by the examiner (*your* 5 times and *her*), 3 by the student (*my*, *her*, *you*). So the student actually uses a wider range of possessive determiners than the Indian student who only produced two types of possessive determiners, even though they produce many more possessive determiners in an interaction which overall was much shorter. The Realis discourse units—the first, second, fourth and fifth in the task—are focused either on the introduction of the task (discourse unit one) or the novel itself (second, fourth and fifth). The third discourse unit is Irrealis as it deals with a hypothetical – how the examiner would feel if they read the book. The final discourse unit is Irrealis and deals with the examiner having forgotten what the student's favourite book is about, with their uncertainty pushing the discourse unit into the Irrealis function. So, this example shows how one student from India, over-relying on a limited repertoire of possessive determiners and other pronouns, while engaging in a shallow discussion of a book which largely revolves around the assertion that this book is their favourite, draws heavily on the Realis function in a way which is indicative of poor performance. This contrasts with the Mexican student who does not draw frequently on possessive determiners, yet demonstrates a broader repertoire of them, and discusses their book in a way that focuses on what the book is about, using the Realis function, while also dealing with hypotheticals with the Irrealis function. This is the type of difference that is captured when we look at Table 3.7.

What of the Irrealis function in Conversation? Exploring this casts more light on the findings from the Discussion. The key to the use of either Realis or Irrealis is function – where the task calls for a specific function, that is the function that should be used. This means that, for some students scoring an A in the Conversation task, they do well because the choice of Conversation task requires Realis and they draw on the Realis function, lessening the presence of Irrealis accordingly. Where there is a good match between the needs of the conversation and Realis, then a high ratio of Realis features can align with a high score.

A good example of this comes from a Chinese student, awarded an A for the Conversation task, in file 2_6_17_CH. The choice of topic in this conversation is the student's experiences of travelling around the UK. The conversation is rooted in lived experience. In a conversation of twelve discourse units, eight of them are Realis and four Irrealis. The Realis discourse units relate, for example, to explaining that they have visited both Stonehenge and a Harry Potter experience. Yet Irrealis is called on in the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

ninth discourse unit when the student takes the initiative and asks if the examiner will talk about their travel experiences. This unit may have taken a Realis turn if the examiner had done so, but instead the examiner talks about the possibility that they may soon run out of time. So some questions in the conversation task may develop an orientation to Irrealis depending on the reaction of the examiner, or the examinee, to how the conversation develops. But note also that initiating an interaction requiring a specific function does not always lead to it being produced. For students this may be a question of proficiency but, as we can see in this example, the power imbalance between the student and examiner can lead to a context where the student initiates a function, but the examiner declines the offer to pursue that function, showing the control that they are licensed to exert over the direction of the conversation.

A further example of how functions respond to the direction of the interaction, this time by the examiner, comes from another Chinese student awarded an A for the Conversation task. In file 2_6_CH_35 they are asked about money in a conversation over four discourse units. Three of these are Irrealis, one Realis. The first discourse unit, which is Realis, focuses on what money the student has and how they earned it. The second discourse unit marks a shift to Irrealis, triggered by the examiner, who asks what the student intends to do with the money. The student then responds to this and the rest of the conversation continues in the Irrealis function. So the key to the performance of the student in both cases was not an avoidance of the use of Realis, it is in using Realis when it is appropriate and Irrealis when that is appropriate in response to the prompt from the examiner. Likewise in the Discussion task, the over-reliance on Realis by the students awarded D came about from a failure to produce a sustained engagement with the Irrealis function when required (as was seen with the Indian learner in Example 34). An example of that for the Conversation task comes from file 2_6_SP_6. A Spanish student, awarded a D for the Conversation task, begins with a focus on fashion. This task spans nine discourse units, all of which are Realis. The student's score is clearly linked to their inability or unwillingness to engage with Irrealis where appropriate, as the following discourse unit shows:

(36) E: yes okay alright then so i-i-if you go to a wedding what what kind of clothes do you wear?
S: if you go to the big s= er st= er shopping centre centre
E: yeah
S: Marineda City
E: yeah

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

S: are a lot of shops
E: <pause/> <unclear text='just'/> to to buy elegant clothes
S: yes

The opening utterance from the examiner clearly begs a response engaging with the Irrealis function. What emerges, however, moves towards the Realis via a *non-sequitur* – that if you go to Marineda City shopping centre there are a lot of shops. The examiner then has to effect a repair to make this statement relevant to the question asked. So a drift from what is required is evident both in terms of the choice of function by the student and because of the repair from the examiner. This continues into the next task when the examiner again tries to get the student to shift to Irrealis by asking whether the student thinks that clothes reflect personality. The student responds with a further apparent *non-sequitur* 'jeans and T-shirt' and the examiner has to effect a further repair, asking them if they think that wearing jeans and a T-shirt reflects personality. So, in this case we recognise the overall pattern that sees this D grade student shift in the Conversation task into Realis – a failure to produce the Irrealis when appropriate. The consequence of that is the Gricean maxim of relation is breached – the student is not producing relevant responses, they break the maxim that contributions 'be relevant' (Grice, 1975: 46). This is different from Example 34 – there the student tried to pull the function back to Realis by saying that the new book *is my favourite book*, moving from the hypothetical to the real. This student fails at the pragmatic level as well as the functional level. The most likely explanation for this is aptitude – students with a higher grade have some mastery of the features which help them to deliver either function. Some students receiving a lower score are able to produce the Realis, but may on occasion either fail to understand that they should switch to Irrealis or are incapable of doing so. However, note that other tactics are available to students with proficiency issues that at least mean that, pragmatically, the exchange may be acceptable, as shown in Example 34.

However, for students graded D the likelihood of them being incapable of Irrealis is difficult to support, precisely because in the Conversation task it is the same group of students who were more likely to produce discourse units with an Irrealis rather than Realis function. An alternative hypothesis rests on comprehension rather than production – the students are not taking the cues to produce speech relying on the function that best fits the context, or abiding by conversational maxims, because they are experiencing a failure of comprehension. A possible way of exploring the nature of the issue of proficiency here is to turn to the Interactive task, which

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

Table 3.8 *The Dimension 3 association of the discourse units in the Conversation task according to learners' proficiency, grade and overall mark.*

Conversation_A_B1_grade6	0.135
Conversation_A_B2_grade7	0.232
Conversation_A_B2_grade8	0.037
Conversation_B_B1_grade6	0.209
Conversation_B_B2_grade7	0.252
Conversation_B_B2_grade8	0.17
Conversation_C_B1_grade6	0.187
Conversation_C_B2_grade7	0.241
Conversation_C_B2_grade8	0.26
Conversation_D_B1_grade6	0.242
Conversation_D_B2_grade7	0.314
Conversation_D_B2_grade8	0.294

does not include the lowest grade of students. With the Conversation and Discussion tasks we saw a clear difference between the students marked D and the rest. Does this persist for the interactive task? As can be seen in Table 3.7, the answer to this question is 'no'. While it is possible that the nature of the task may occasion a different behaviour, we do at least see a very different pattern of distribution of the students in this table. Like the Conversation task, the Interactive task draws mainly on the Irrealis function for all students, but unlike that task, for the Interactive task the students graded D are not the obvious outliers relative to A.

To try to get a clearer view of whether the grade of the exam, the grade awarded and the task interact to produce results which are different from the overall view provided, Tables 3.7–3.10 show how the relationship of the Realis and Irrealis functions vary when those features are combined.

Again when we look at how the task, mark and exam level interact with Dimension 3, we find that, regardless of grade of exam or mark awarded, discourse units in the Conversation task are more associated with Irrealis. With reference to the students awarded a D, however, we do not see any sign that, with increasing proficiency, their selection of the Irrealis function in this task declines. By contrast, the students awarded an A do show a decline at grade 8 in relation to their association with the Irrealis function in this task. A similar pattern is visible for the D students in the Discussion task (see Table 3.9) – as they progress through the exam grades, their use of Realis is more pronounced, and this contrasts with students awarded an A who are markedly less reliant on Realis for this task beyond grade 6.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

Table 3.9 *The Dimension 3 association of the discourse units in the Discussion task from groups of learners defined by proficiency, grade and overall mark.*

Discussion_A_B1_grade6	−0.192
Discussion_A_B2_grade7	0
Discussion_A_B2_grade8	−0.041
Discussion_B_B1_grade6	−0.159
Discussion_B_B2_grade7	0.054
Discussion_B_B2_grade8	−0.182
Discussion_C_B1_grade6	−0.15
Discussion_C_B2_grade7	−0.018
Discussion_C_B2_grade8	−0.073
Discussion_D_B1_grade6	−0.15
Discussion_D_B2_grade7	−0.277
Discussion_D_B2_grade8	−0.404

The hypotheses put forward already for this seem to apply when we look at higher level data. If we look at students in grades 6–8 awarded a D in the Conversation task, we get an insight into the increased level of Irrealis in this grade – we see the student avoiding Irrealis and the examiner prompting them to produce Irrealis utterances, as we saw in Example 34. This scaffolding of the conversation is more persistent at grade 8, as can be seen in the following discourse unit, from the Conversation task of an Italian student awarded a grade D, in corpus file 2_6_IT_62:

(37) E: oh fantastic okay good luck erm do you have pocket money from your parents?
S: er yes sometimes
E: yeah and do you have to erm work for pocket money?
S: er so=
E: do you help around the house?
S: ah
E: or do you need to do
S: er
E: some jobs

While the overall discourse unit is Irrealis, it is the examiner who is driving that, not the student. The student produces short Realis responses, which fail to elaborate on the examiner's questions. It is the scaffolding from the examiner that seems key here.

The key which allows us to understand the sustained relationship between the Conversation task and the Irrealis function for grade D students lies at the micro-level. If we return to the micro-structural level

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

and consider the examiner language micro-structural functions from Chapter 2, we see that the Information-Seeking function is important. This is closely related to the Irrealis function as it is composed of questions, typically about the future or imaginary scenarios. When we look at the relationship of this micro-structural dimension to learner performance, we find that learners in the Conversation task received a lower overall mark if the examiners produced more turns associated with Information-Seeking as opposed to Descriptive turns. When the examiner produced fewer Information-Seeking turns and more turns associated with the Descriptive communicative function, the learners received a higher overall mark across all grades and marks. This is a perfect example of how the micro-structural and macro-structural approach we have taken provide insight when applied together.

The same is true, though less markedly, when we consider the Discussion task. When we consider task, grade of exam and mark together, discourse units produced in interactions where students were scored a D at grade 6 are only weakly linked to the Realis function, relative to the other grades. In part, this is attributable to the scaffolding behaviour of the examiner. When we consider examiner language alone in this task, we find that the only clear pattern that emerges from the short-text MDA of micro-structural functions in examiner speech is that, at grade 6, students scoring D elicit a markedly higher proportion of Information-Seeking turns from the examiner. As those turns contribute to the Irrealis side of the dimension, a point of gravity is established for the discourse units of the D grade students that pulls the coordinates of their discourse towards the Irrealis. However, unlike the Discussion task, there is clear evidence that, as proficiency develops, the need for scaffolding through questioning declines, as the association between Information-Seeking turns from the examiner and the student's grade applies to no other combination of grade of exam and mark for this task other than grade 6.

In the Interactive task, grade 7 learners across all marks communicate in a way more markedly linked to the Irrealis function than learners receiving the same mark in grade 8 exams (see Table 3.10). The result for students scoring a D at grade 8 look anomalous – and when we explore the corpus we discover that, when the variables are combined in this way, there is relatively little data for students with this score for the interactive task at grade 8 – only three such students exist in the corpus. However, when we explore the language of those students, features of poor function selection that we have seen before are evident. For example, in file 2_8_IN_16, an Indian student, awarded a D for the Interactive task, is asked by the examiner to

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

Table 3.10 *The Dimension 3 association of the discourse units in the Interactive task from groups of learners defined by proficiency, grade and overall mark.*

Interactive_A_B2_grade7	0.425
Interactive_A_B2_grade8	0.386
Interactive_B_B2_grade7	0.487
Interactive_B_B2_grade8	0.413
Interactive_C_B2_grade7	0.479
Interactive_C_B2_grade8	0.4
Interactive_D_B2_grade7	0.748
Interactive_D_B2_grade8	−0.115

'ask me questions to find out more information and make comments you need to keep the conversation going'. The examiner then reveals that a seventeen-year-old nephew of theirs is about to get married. The student then proceeds largely to comment on the situation, as in the following exchange, which is a discourse unit occurring after it has been revealed that the girl the nephew wants to marry is sixteen:

(38) S: sixteen
E: yeah yeah
S: oh
E: mm
S: that's a big problem <laugh/>
E: well yes I think so it is legal it's okay you can marry at sixteen <pause/> okay alright well thank you very much for your advice on that

This is an almost exclusively Realis discourse unit. The candidate does not use the Irrealis function to seek information, only asking two questions in one turn in the first discourse unit of a four-discourse unit interaction. Instead, they simply state what should be done and it is the examiner who effectively does the information seeking on behalf of the candidate, for example by giving the age of the girl involved. So, while it is likely that for grade 8 students awarded a D grade we have too little data to make a firm statement about trends, we can at least see the discourse dynamics explored so far playing out faithfully in the small volume of data that we have.

3.5 Conclusion

In outlining this chapter, we raised a question that guided our investigation of whether the functional analysis of the micro-structural level of the previous chapter would simply be reflected at the macro-structural level in the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

discourse unit analysis presented in this chapter. Through the analysis of discourse unit functions in the second and third dimensions, we have gained some provisional answers to that question. The answer being that, for the most part, discourse functions at the discourse unit level do not map directly onto those at the turn level, unless we find a macro-structure function which is both dominated by one of the speakers and which is mono-functional in nature at the micro-structural level, as was the case with the Informative and Instructive discourse unit function. Generally, however, the function at the macro-level is distinct from those at the micro-level. So, while we might argue for the Informative and Instructive function that similar functions are performed by the examiner at the micro-structural level to constitute the macro-structural level, this is not the case for the other functions explored. Hence the micro-structural level does not necessarily select the macro-structural level – while Information-Seeking turns from the examiner may be assembled to perform the Informative and Instructive function, those micro-structures may also appear in discourse units with other functions. So it is the higher level goal – the formation of a discourse unit with a specific function – that governs the selection of the micro-structural functions. As we saw, the setting of that higher level goal may be contested. That contestation is subject to the asymmetries of power in the exam setting. Further, we also saw that proficiency intervenes, manifesting at the pragmatic level on occasion, to either assist or frustrate the fulfilment of the higher level.

Importantly, the selection of function at the macro-level is strongly tied to task and to proficiency. Different levels of proficiency may make the examiner, in particular, select a function that is required to sustain an interaction, for example. The tasks also were shown to have preferences for different functions and the co-constructed nature of the functions also became apparent. With the exception of the Informative and Instructive function, we found that while the learners typically dominated and led the construction of the discourse, the examiners were not passive interlocutors – they shared some of the features of the function they were participating in (evidenced, for example, by the n-grams used by the examiners in the Irrealis function) and would select functions, on occasion, as a way of effecting repairs. Moreover, the grade for which the assessment was designed appears to exhibit variation, where tasks designed to meet the same assessment criteria across different grades seem to elicit discourse functions to different degrees.

Importantly, we saw also that short-text MDA, which worked well at the micro-structural level in Chapter 2, seems to function just as well at the macro-structural level.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003

In the next chapter we start by looking at a feature we have already mentioned in this book – narrative. By now, we know that narrative plays an important role at the micro-structural level. The next dimension will show that it is also present at the macro-level. This therefore raises the question once again of the relationship, or lack thereof, between the role of narrative at micro- and macro-levels.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 15:43:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009208932.003