

Michael McCarthy, England/University of Nottingham &
Anne O’Keeffe Ireland/Mary Immaculate College, Limerick

Competencies Explored and Exposed: Grammar, Lexis, Communication and the Notions of Levels

This chapter will explore the notion of language competencies. It will look briefly at the widely used and accepted Common European Framework of Reference (CEFR) and its different competence descriptors and can-do statements. We will explore how competencies can best be measured and will give details of the English Profile programme. To this end, we will use real learner language to provide an empirical basis for the description of competencies, which is core to the English Profile programme. The chapter will be based on examples from the Cambridge Learner Corpus, a collection of over 30 million words (about 95,000 scripts) of student writing from the Cambridge exams. In this corpus, errors have been marked. The corpus contains the work of 125,000 students, with 130 different first languages, from 190 different countries. In addition, we will draw on spoken data, including oral exam recordings and non-exam spoken material in an attempt to explore competence in different contexts of use. In particular, the chapter will examine linguistic and interactive competencies in spoken and written language and will discuss some of the implications for language pedagogy.

1. Introduction

The Common European Framework (CEFR) (Council of Europe 2001) has, in a relatively short period of time, achieved currency in a wide range of countries within and without Europe. Its system of levels, ranging from the basic A1 to its highest current level, C2, is widely and commonly used by teachers, syllabus designers, examination boards and publishers of teaching materials alike. It has become a convenient shared language among these several groups of professionally involved bodies and individuals for the assignment of proficiency levels to learners. The CEFR presents, for each level (A, B, C) and sub-level (A1, A2, B1, B2, C1, C2) sets of statements designed to capture what a learner at that level *can do* with the language repertoire at their disposal. An example would be the statement that a typical B2 learner can “interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party” (Council of Europe, 2001: 24), while the C2 learner should be able to speak “so smoothly that the interlocutor is hardly aware of it” (2001: 28).

The labels and the common understanding of the can-do statements on the part of those who use them have evolved based on the pooled expert judgement of decades of

professionals and, as such, cannot be said to lack a solid foundation or to be the whim of some obscure official agency. However, in the time span since the inception and spread of the CEFR, substantial corpora of native- and non-native user data for languages such as English, along with learner corpora, have become more readily available and accessible. With the benefit of corpus data, it becomes possible to add empirical observation to the basis upon which learners are assigned to the different levels. In this chapter, we explore some of the ways in which corpus data can be interrogated to gain a better understanding of the emergence and consolidation of the different levels of competence the CEFR attempts to describe and the typical competencies that learners display at each level. This will naturally have an effect on materials design and the targets teachers set in their classes.

We focus here mainly on grammatical competence, knowledge of collocation and speaking skills. These three areas are chosen because, in the first place, grammar is always likely to be a principal preoccupation in the language classroom and materials. Secondly, the ability to use appropriate collocations is a competence that becomes more and more important as learners pass the basic vocabulary threshold and need to consolidate their knowledge with more 'depth' (e.g. how words are used) rather than 'breadth' (e.g. the number of words one knows). Finally, less work has been done on learner spoken corpora up to now because of a lack of good data, but this is being remedied and a number of projects are now underway under the umbrella of the English Profile project to gather learner data in different contexts. However, we acknowledge that empirical investigation of a wide range of language knowledge, skills and abilities in other contexts (e.g. pronunciation and intonation, vocabulary size, communicative-functional skills, writing skills and so on) will be necessary before a complete picture of the typical learner at any given CEFR level can be achieved.

The present authors are both involved in the English Profile (EP) programme (see its website at www.englishprofile.org). EP is a large-scale, collaborative programme of research which brings together the work of scholars and practitioners in a range of countries and institutions, all of whom share the goal of achieving a more faithful and detailed description of the typical competencies of English language learners at the different CEFR levels. The programme is committed to the use of empirical data, and a huge data-collection effort involving the gathering of learner speaking and writing in a variety of contexts is underway in a number of countries at the time of writing. English Profile, therefore, sets out to *describe* what learners can do, rather than *prescribe* what they must or must not do. Researchers working in different branches of English language teaching and applied linguistics will all work towards the common goal of providing a more objective description of learner English at the A, B and C levels of the CEFR. Armed with such a description, it is hoped that teaching materials and the tasks learners are asked to undertake, syllabuses, examinations and teacher education will acquire a sharper focus and a more unified set of goals that any institution or individual can aim for.

2. How Can We Measure Competencies Using Learner Data?

Since 1993, Cambridge University Press, in collaboration with Cambridge ESOL, has been building the Cambridge Learner Corpus (CLC). This is a database of over 35 million words from Cambridge exam scripts, from over 135,000 scripts, written by learners from 190 different countries and spread across 130 different first languages. These scripts have been keyed into a computer and over 21 million words of data have, at the time of writing, been error-coded. The process of error coding involves the systematic marking of all errors in the students' scripts by teams of raters. The following convention is used to label errors within the corpus: <#CODE>wrong word|corrected word</#CODE>, as in Extract 1, from a B2-level exam script written by a learner whose first language was German:

Extract 1: Tense and collocation problems: a German-speaking learner

[codes illustrated here: TV = incorrect verb tense; RV replace verb]

I know it <#TV>costs|will cost</#TV> a lot of money to <#RV>do|make</#RV> these changes, but students could help to organise something.

The corpus software allows filtering of the corpus data according to a number of different categories, for example, exam taken, CEFR level, first language, year of examination, type of error tag, year, country, whether the student passed or failed their exam and so on. In Extract 1, from a corpus search for the verb *make*, we see an inappropriate verb-noun collocation (*do changes* instead of the more typical *make changes*) and an error of verb tense (the student wrote *costs* instead of *will cost*). In the screen shot below (Figure 1), also based on a search of *make*, we see examples of the reverse, i.e. the use of *make* used instead of *do*, which have been tagged with the code RV (replace verb). For any line selected, details about the learner who made that error appear in the grey bar at the top of the screen. In this case, the selected line represents a Certificate of Proficiency in English (CPE) student from Mexico. The exam is at the CEFR C2 level; in this example it was taken in the year 2000, Spanish was the student's first language and the overall result was *fail*.

Glancing down the screenshot in Figure 1, it is apparent that *make* and *do* present problems of collocation with the nouns that follow them. Students have created a range of inappropriate collocations, including *make a demonstration*, *make ... sport*, *make ... experiments*, *make ... activities*, all of which should have been constructed with the verb *do*, not *make*.

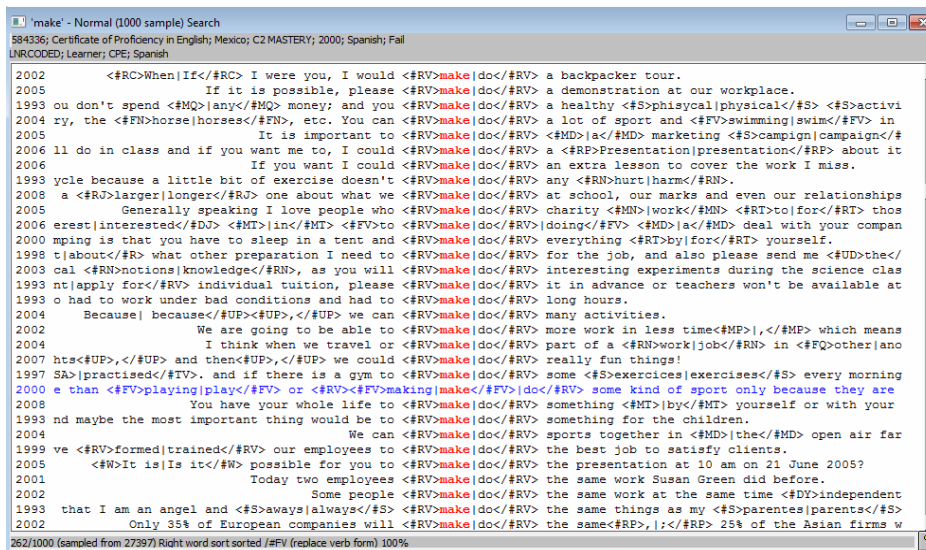


Figure 1: Screen shot from CLC search of verb *make*, showing *make/do* errors

3. The Value of Learner Error Data in Relation to EP

What value, then, does such learner error information have, and in particular, how can it be used to further the goals of the EP programme and to improve language pedagogy? For one thing, initial perusal of the sample in Figure 1 reveals that many of the learner errors with *make/do* occur in the context of sports and leisure activities. Few would question that this information could be very useful for teachers, learners, assessors, syllabus designers and materials writers. However, when we conduct more refined analyses and filter the data according to CEFR level, we can trace the *make-do* error in terms of the level at which it emerges, the level at which it peaks in its frequency of occurrence and the level at which it stabilises (e.g. the CEFR level at which most learners seem to be getting it right most of the time). Figure 2 illustrates that learners at A2 level struggle most with *make* versus *do*, whereas by B1 level, and above, the error rate is very low. From this we can say that learners at A1 and A2 attempt, often unsuccessfully, to control the *make/do* distinction, while learners from B1 level onwards typically do not confuse *make* and *do*¹. The consolidation phase, where one might posit that the greatest learning takes place, is when the learner is crossing the A-B threshold.

¹ The searches reported in figures 2-5 were carried out in 2010 by the present author Anne O'Keeffe and Geraldine Mark as pilot studies for the EP programme.

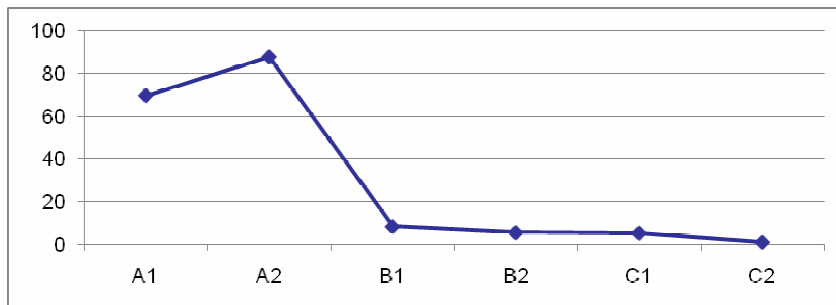


Figure 2: Errors with *make* and *do* across CEFR levels

By way of another example, typical of written examination contexts, if we search for the pattern *looking forward to* in the learner data, we find that most of the errors relate to the preposition. The most common error appears to be the use of *looking forward + for* rather than *looking forward + to*. For example, in a B1 level business English examination, a German learner wrote at the end of a business letter:

Extract 2: Preposition problem: *look forward*

*I am <#FV>looklooking</#FV> forward <#RT>for</#RT> your
<#RN>reply</#RN>*

We see other errors here too (the verb-ending on *look* and the spelling of *reply*) and we also learn from a closer look at the data that this student failed the examination. When we track the emergence, peak and decline of this particular error (Figure 3, below), we find that it emerges between A2 and B1, peaks at B2 and declines sharply between B2 and C1. This is a typical pattern: at first, the distribution of occurrences of any given item tends to be sparse (the item is unknown or very new to students) and so the error-rate appears to be low. Then, typically the item begins to be used widely, experimented with and, as a result, is often used wrongly or inappropriately, so the error-rate rises. There then typically follows a period of consolidation (perhaps through the combined benefits of error-correction and increased exposure to and practice with the item), where the item is increasingly used correctly and appropriately, till it stabilises and can be said to be a marker of competence at that CEFR level.



Figure 3: Trajectory of error pattern *looking forward for* across CEFR levels

However, such a conclusion, in the case of *looking forward to*, has to be tempered by the fact that B1 and B2 examinations may include more letter-writing tasks (which tend to generate the *looking forward to* pattern in valedictory sections) than C-level exams. Consideration of the prototypical level at which an item or pattern emerges and becomes consolidated, therefore, must always include scrutiny of the use of the item in a number of differing contexts (e.g. exams, free essay writing, classroom interaction, etc.) before we can confidently assert that it is absent, present, wrongly used or correctly used in the speaking or writing of the great majority of learners at any given level. The more varied the available learner data, the more reliable the statements can become. This means that teachers should always try to get as wide a range of speaking and writing tasks from their students as possible; otherwise, our assessment of our students may be distorted.

We can represent the use of an item and the rise and decline of errors in a slightly different way. Figure 4 shows the use of the affirmative future perfect structure *will have + -ed* in all its occurrences in the learner examination data, both correct and incorrect. As our intuition might predict, the future perfect structure is not used by A level learners, shows an increasing use among B level learners and is used most by C level learners. The drop in use from C1 to C2 may be attributable to task-related factors; on this evidence alone, then, the picture is far from clear.

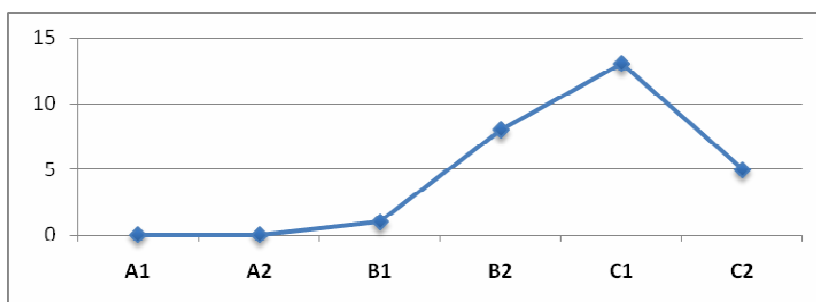


Figure 4: Frequencies of use of *will have + -ed* across CEFR levels

If we then consider where errors with *will have + -ed* happen most, we get a clearer internal picture. Figure 5 shows that 50% of B2 attempts at use of the future perfect are incorrect, even though use has risen rapidly from B1, so one can hardly state that mastery of the structure is a characteristic of B2 learners. By C1, however, the error rate has declined to a negligible level, and continues to be very low for C2. We can therefore say that learners, on the whole, typically know or ‘can do’ future perfect affirmative forms by C1 level. This sort of evidence can provide useful guidelines to teachers as to approximately what they can expect their learners to know at any given level and, conversely, whether we are teaching target items that are too high or too low for our particular classes.

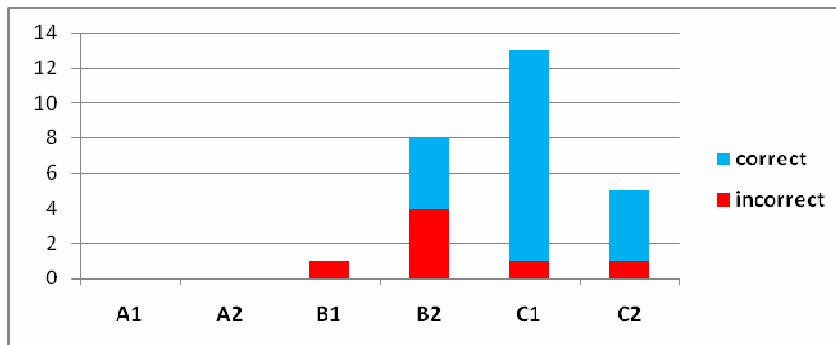


Figure 5: Error and correct use rates for *will have +-ed* across CEFR levels

Therefore, by looking at learner data in this empirical way, we can potentially profile what it is a typical A1, A2, B1, B2, C1 or C2 learner ‘can do’ in terms of grammatical competence. This type of corpus-based profile is underway as part of the EP programme. As we mentioned earlier, the statements that EP aims to produce and disseminate to the language teaching community will be rendered more reliable by the inclusion of a wide range of learner data, both spoken and written, collected both within and without institutional settings and from a wide range of ages and language backgrounds. Many issues still remain unresolved at the time of writing. Not least of the questions to be grappled with is whether the statements (or *Reference Level Descriptions* as they are officially termed) should be seen as lock-steps of achievement on a scale or whether something of a more developmental nature can be built into the system, that is to say, not only where a student is on the CEFR but how far he/she has ‘travelled’ to get there. This latter notion might include consideration of first-language differences and ‘distance’ from the target language, in this case English. For example, German-speaking learners may be expected to have fewer problems with German and English shared grammar features than speakers of Asian languages, whose grammars may be fundamentally different from Western European languages. Other factors include learning-cultures and curricular features and constraints (e.g. countries where speaking skills may be accorded a higher premium than writing skills, or vice-versa). The notion of the typical learner at any given level is certainly a complex one but the contribution of empirical data will without doubt be crucial to its proper exegesis.

4. The Learner Lexicon: the Case of Collocation

Native-speaker corpus data show that collocation, the way words co-occur in statistically significant ways, is a fundamental feature of language use and is central to the creation of meaning (Sinclair 1966, 1991, 2004). It is one of the competences we possess as language users. In the case of English, native- and expert users know that *fair* is more likely than *light* to collocate with *hair*, that *make* collocates with *effort* while *do* collocates with *duty*, and so on. One question of relevance to the EP programme, therefore, is whether learners develop collocational competence as they traverse the

CEFR levels and what difficulties they may encounter on the way. At a more general level, examination of learner data alongside the scrutiny of syllabuses and published course materials may reveal whether collocation is indeed a taught element, whether its teaching is effective and/or whether collocation is simply picked up or absorbed by some process of osmosis as the learner is exposed to more and more examples of the lexicon in use. It is important for teachers to be aware of these issues, especially when learners have crossed the threshold of the first 2,000 words and are struggling to put those words together into texts and conversations that sound reasonably natural.

In this section of the chapter we focus on actual examples of learners’ use of collocations rather than statistical trajectories across CEFR levels in order to illustrate the importance of looking at particular learner contexts of use and comparing those contexts with native-speaker contexts. This is not done to see whether learners are able to ape native speakers or whether we should brand learners who do not use the same collocations as native speakers as “failed monolinguals” (Cook 1998). Rather it is to see what kinds of collocations appear in the contexts in which learners use them (e.g. classrooms, examinations) and whether we can better construct those contexts to reflect more real-world contexts. Here we focus on the common set of everyday verbs often referred to as delexical verbs (Sinclair/Renouf 1988: 153). These include high-frequency verbs such as *get*, *go*, *do*, *have*, *take*, *make* which are amenable to varied interpretations depending on their immediate complements (e.g. *get a newspaper* is a different kind of *get* from *get the door*, *get angry* or *get home*). Our particular focus is spoken language and our learner data are drawn from Cambridge oral examinations and a corpus of classroom interactions collected at Shannon College of Hotel Management, National University of Ireland, the CLAS (Cambridge, Limerick and Shannon) corpus.

One initial comparison that may be made is the overall distribution of a set of delexical verbs in different kinds of data. It may be posited that the learner data we have at our disposal is more akin to spoken academic data than everyday conversation. A comparison is shown in Figure 6, based on a spoken academic sub-corpus of the Cambridge International Corpus (CIC) and a sub-corpus of everyday conversations from the CIC.

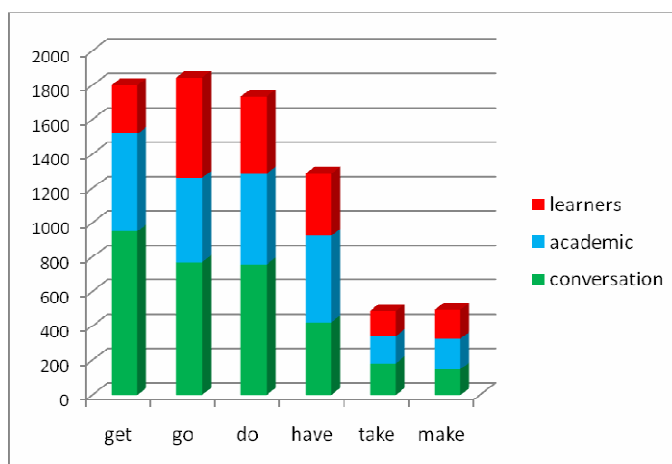


Figure 6: Distribution of delexical verbs across three spoken datasets

All other things being equal, we might hope to see equally-sized divisions for academic and learner speech in the column for each verb. In general, we do see a fairly balanced picture for *go*, *do*, *take* and *make* as between spoken academic contexts and the learner contexts. *Have* is somewhat imbalanced, while *get* shows the biggest discrepancy, with learners seemingly using it much less frequently than its distribution in the other two data types. Possible reasons for this may be (a) an underrating of the importance of *get* in classroom teaching and materials, (b) genuine difficulties with the meaning and use of *get* and its collocates, resulting in avoidance on the part of learners, (c) a prejudice against *get* so that it is seen as informal and inappropriate in the contexts in which learners typically operate (classrooms, exams, etc. – though the spoken academic data would suggest its use in institutional contexts is anything but rare), (d) insufficient learner data from a broad range of contexts, or (e) other reasons as yet unknown or unknowable.

The actual collocates of *get* in the data illuminate further some of the issues accruing to this area of competence. The most common collocates in the three datasets are shown in Table 1 (these include all the inflected forms of *get* but exclude the use of *get* as an auxiliary verb in the *get*-passive and *get*-causative structures). ‘[...]’ indicates intervening words such as determiners, adjectives before nouns, etc.

learners	conversation	academic
<i>get married</i>	<i>get [...] stuff</i>	<i>get rid of</i>
<i>get [...] job</i>	<i>get rid of</i>	<i>get [...] idea</i>
<i>get [...] chance</i>	<i>get home</i>	<i>get ready</i>
<i>get [...] information</i>	<i>get married</i>	<i>get [...] sense</i>
<i>get [...] money</i>	<i>get ready</i>	<i>get [...] feel</i>
<i>get [...] profit</i>	<i>get [...] hold</i>	<i>get [...] problem(s)</i>

Table 1: Most common collocates of *get* in three datasets

There is overlap here: *get married* appears in both the learner and the conversational data (where, in the case of oral exams and everyday interaction, speakers discourse on personal and social matters). *Get rid of* and *get ready*, prominent in the conversational and academic data, are absent from the learner data. The learner data is mostly dominated by nouns (*job*, *chance*, *information*, *money*, *profit* – reflecting discussion topics in oral examinations and business-oriented discussions in the CLAS corpus). Where nouns occur in the non-learner data we have somewhat more idiomatic collocations such as *get an idea of ...*, *get a real sense of ...*, *get hold of ...*, *get a feel for ...*, along with the ubiquitous, informal *get ... stuff* in conversation. It is at the level of idiomaticity that the native-user data differs from the learner data, perhaps through lack of attention in pedagogy to the more idiomatic chunks which the delexical verbs enter into or because chunking and idiomaticity may emerge only at higher levels of the CEFR and are thus not widely represented in a broad sample of user data such as the present one. Another possibility is avoidance of idiomatic chunks as associated with less formal, non-institutional contexts; once again though, the spoken academic data would tend to suggest that idiomatic chunks are by no means absent from the institutional context. What is notable in the learner data is that the *get*-collocations occur overwhelmingly at higher CEFR levels, with no examples showing at CEFR A level. However, we return again to our theme that learner data must be sought in a wide range

of contexts before we can confidently assert that learners do or do not display collocational competence at any given level of achievement.

5. Speaking Skills: the Case of Fluency

In section 1, above, we mentioned that the CEFR can-do descriptors included statements about speaking skills and, in particular the notions of fluency and smoothness, with the C2 learner being capable of speaking “so smoothly that the interlocutor is hardly aware of it.” (Council of Europe 2001: 28). In what ways, then, could the notions of “fluency” and speaking “smoothly” be given empirical substance and be amenable to observation in real data, whether native- or expert user data or learner data? Much of the past history of the study of fluency has centred on preoccupation with such features as speed of delivery, automaticity (the ability to retrieve linguistic units without thinking or hesitation) and placement of pauses (see McCarthy 2010 for a survey). This is very much a monologic view of fluency which takes little or no account of the role and influence of the interlocutor(s) in multi-party talk. Meanwhile, the CEFR statements rightly stress *interaction* and the presence of an interlocutor (see the quotes in section 1, above). In dialogue or multi-party talk, automaticity and the ability to construct one’s turn smoothly and efficiently may be seen as critical at speaker turn-boundaries, to as great or to a greater extent than anywhere else in the speaker’s turn. In other words, an *interactive* smoothness and fluency (or *confluence*) is what speakers and listeners try to achieve, and they do this jointly and cooperatively. The turn-boundary, where one speaker closes and another opens, therefore, suggests itself as an important locus for the understanding of how conversational flow is achieved.

Previous research has given us a good idea of what happens at turn-boundaries, both in terms of opening one’s turn and closing it. The apparently seamless way in which turn-taking unfolds in conversation, with pauses of only micro-seconds and few overlaps has been the subject of much study over many years (from the seminal paper of Sacks *et al.* 1974; to more recent examples such as Stivers *et al.* 2009). Normal conversational turn-taking displays a degree of smoothness and automaticity that ‘irons out the seams’ between speakers’ turns. But perhaps the most notable feature of conversational turns is the consistency in the manner in which they open, regularly and characteristically utilising one of a number of high-frequency free-standing items which provide links with what the previous speaker has just said (Tao 2003). In the conversational corpus, for example, the most common turn-openers (apart from laughter and other vocalisations) are shown in Table 2.

In addition to the highly interactive *I* and *you*, we find items which respond to and link with the previous speaker’s turn. These very small words create smooth transitions between one turn and another and, it is argued, contribute to the confluence of the talk – in the sense that it is the conversation which is fluent, not just the individual speaker.

1.	yeah
2.	mm
3.	oh
4.	and
5.	I
6.	no
7.	well
8.	yes
9.	but
10.	you

Table 2: Turn-openers in informal conversation

One thing that we can do, therefore, in the EP research context, is to track the emergence of turn-construction skills in much the same way that we have demonstrated for grammatical constructions in section 3, above. The body of EP spoken data is still in its early stages compared with the wealth of written data furnished by the Cambridge examinations over many years, but, already, with the data currently available, it is possible to see the growth of confidence in turn-construction at higher CEFR levels in oral examination data, with not only longer turns but the increasing use of interactive turn-opening items, especially in those task-contexts where some sort of genuine interaction is fostered. Extract 3 shows two B2-level exam candidates are doing a paired task, discussing with each other the merits and demerits of various places as destinations for English-language learning students to visit and spend time at (here an office versus a supermarket versus a factory). Highlighted in bold are turn-initial interactive items which, we would argue, contribute to the satisfactory sense of flow which the conversation achieves, despite apparent hesitancy in other parts of the speakers' turns.

Extract 3: Places to use one's English

[Square brackets vertically aligned indicate overlaps; ? = indistinct audio]

<Candidate2> If I go to a job in an office I must speak in English - that is the best way to improve my language.

<Candidate1> **Okay yeah I can see your point yeah I take your point** but you know it's I don't know [I can.

<Candidate2> **So** [in a way for me it's a supermarket because I usually when I go to when I go shopping for example in a supermarket some word I don't know it really well you know some English word I don't know.

<Candidate1> **Yeah.**

<Candidate2> **So** I have [to to look up in the dictionary in the to find out [to what it mean.

<Candidate1> [It's a **yeah**

[**Yeah okay.**

<Candidate2> **Exactly. So** [?.

- <Candidate1> [And what do you what do you think about factory?
 <Candidate2> **Fine.** Another kind **it’s another kind of job.**
 <Candidate1> **It’s another kind of job.**

Both candidates use with aplomb common linking words such as *and*, *yeah*, *okay*, *so*, to open their turns. Candidate 2 additionally uses *exactly* and *fine* as interactive response tokens which indicate a level of acknowledgement and engagement beyond that provided by simple *yeses* (McCarthy 2003; O’Keeffe *et al.* 2007: 140-158). Finally, the repetition of *it’s another kind of job* cements the interaction; Candidate 1 shows that proper attention is being paid to what Candidate 2 has just said. Once again, having data of this kind to observe and ratify the use of particular features on the part of learners requires the right task conditions which generate natural and normal features of interactive talk. How we construct and execute speaking tasks in class is crucial to what we can expect to get out in terms of apprehending our students’ competences. Apart from the obvious benefits of a robust underpinning of the CEFR levels in terms of the can-do descriptors, the EP research will, it is hoped, make a substantial contribution to the understanding of the relationship between tasks and their outcomes in linguistic terms and will lead to better task design, especially in the classroom and assessment contexts. The triangle of effects between tasks, the conditions under which they are performed and the scoring system which accompanies them in examination settings will provide inadequate evidence if the scoring system used measures the wrong things. In the case of fluency, for example, if the scoring system only considers monologic performance features, the assessment may overlook important features of interactivity (Barry O’Sullivan; personal communication).

6. Conclusion

The examples in sections 2-5, above, show how it is possible to use corpus evidence to track the emergence of, and attest the use of, particular features of the linguistic repertoire among learners at different levels of proficiency. As we have acknowledged, data from as wide a range of contexts of performance as possible are needed to get the optimum reliable picture of the learner’s level. Such data will include classroom performances, examinations and, ideally, non-institutional settings. Competence is not a monolithic affair, and we need to know not just how much grammar or vocabulary our learners can use, but how they are progressing towards greater competence in their use of collocation, their ability to create flow in dialogue, their reading, listening and writing competences, as well as their competence in using the language appropriately in social and cross-cultural contexts. With the powerful empirical evidence that corpora can provide, it will become possible to elaborate more reliably the substance of the CEFR can-do statements in a way that does not negate or necessarily undermine the pooled expert judgement of professionals over many decades who have developed an unerring ability to rate their learners’ proficiency levels. Rather it enables us to put flesh on the bones of a powerful system of description that gives learners and their educators practical, concrete goals to aspire to. The reference levels descriptions that will be the outcome of EP will not be a set of prescriptions of what to teach, but a set of

descriptions of what learners typically can do at any given level in the conditions under which we have been fortunate to be able to observe them.

Bibliography

- Cook, Guy (1998): The Uses of Reality: A Reply to Ronald Carter. In: *ELT Journal* 52(1), 57-63.
- Council of Europe (2001): *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Council for Cultural Cooperation: Committee for Education, Modern Languages Division. Cambridge: CUP.
- McCarthy, Michael J. (2003): Talking Back: 'Small' Interactional Response Tokens in Everyday Conversation. In: *Research on Language in Social Interaction* 36(1), 33-63.
- McCarthy, Michael J. (2010): Spoken Fluency Revisited. In: *English Profile Journal* 1 [<http://journals.cambridge.org/action/displayJournal?jid=EPJ> (February 2011)].
- O'Keeffe, Anne/McCarthy, Michael J./Carter, Ronald A. (2007): *From Corpus to Classroom*. Cambridge: CUP.
- Sacks, Harvey/Schegloff, Emmanuel/Jefferson, Gail (1974): A Simplest Systematics for the Organisation of Turn-Taking for Conversation. In: *Language* 50(4), 696-735.
- Sinclair, John (1966): Beginning the Study of Lexis. In: Bazell, Charles/Catford, John/Halliday, Michael/Robins, Robert (eds.): *In Memory of J. R. Firth*. London: Longman, 410-430.
- Sinclair, John (1991): *Corpus, Concordance and Collocation*. Oxford: OUP.
- Sinclair, John (2004): *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Sinclair, John/Renouf, Antoinette (1988): A Lexical Syllabus for Language Learning. In: Carter, Ronald/McCarthy, Michael J. (eds.): *Vocabulary and Language Teaching*. London: Longman, 140-158.
- Stivers, Tanya/Enfield, Nick/Brown, Penelope/Englert, Christina/Hayashi, Makoto/Heinemann, Trine/Hoymann, Gertie/Rossano, Federico/de Ruyter, Jan/Yoon, Keung-Eun/Levinson, Stephen (2009): Universals and Cultural Variation in Turn-Taking in Conversation. In: *PNAS (Proceedings of the National Academy of Sciences)* 106(26), 10587-10592.
- Tao, Hongyin (2003): Turn Initiators in Spoken English: A Corpus-Based Approach to Interaction and Grammar. In: Leistyna, Pepi/Meyer, Charles (eds.): *Corpus Analysis: Language Structure and Language Use*. Amsterdam: Rodopi, 187-207.

Internet

www.englishprofile.org (July 2011)