

1

What is English Profile?

Julia Harrison

Cambridge University Press

English Profile is a collaborative, multi-disciplinary programme of research, consultation and publication. It is designed to explain in detail what the Common European Framework of Reference for Languages (CEFR; Council of Europe 2001a) means for English language learning, teaching and assessment. The CEFR is neutral with respect to language, and describes learner proficiency in terms of what learners can do with the language and what functions they can perform. The various research projects supporting the English Profile Programme (EPP) aim to help teachers and learners of English everywhere to understand how the levels of the CEFR relate specifically to English, by describing the typical language that learners of English know and use at each level to carry out these functions.

This book provides an introduction to English Profile, but should be of interest for those already familiar with the EPP, as well as those who are new to it. This chapter examines the relationship between English Profile and the CEFR, explains the key role of corpora in English Profile research, and indicates where the Programme's focus has been thus far as it works to support the Council of Europe in its aim to help develop language learners' sociolinguistic and intercultural competence.

Chapters 2 and 3 look at resources which English Profile has completed (although both resources could also be considered as 'works in progress' in that they will need to be continually updated). Chapter 2 introduces the English Vocabulary Profile (EVP), including the rationale behind it, how it was developed, trialled and validated, and what methodology was used in its compilation. Chapter 3 similarly introduces the English Grammar Profile (EGP) and, using examples taken from the resource, provides some insights into how learners develop their grammar skills as they progress through the levels.

The remaining chapters each touch on other aspects of this wide-ranging programme. Chapter 4 looks at how teachers and learners can approach the question of functional language in English using the CEFR, while Chapter 5 provides portraits of learner language at each of the six CEFR levels. Chapter 6 looks at the way English Profile is being used by Cambridge University Press and Cambridge English Language Assessment to improve their teaching and testing materials. Chapter 7 reports on four new research projects which are shedding light on specific areas of learner language development,

Cambridge University Press

978-1-107-49398-8 – English Profile in Practice

Edited by Julia Harrison and Fiona Barker Michael Milanovic Nick Saville

Excerpt

[More information](#)

English Profile in Practice

such as metaphor, grammatical morphemes (e.g. *-ed*, *-ing*, articles) and formulaic language, and finally Chapter 8 discusses how and why you should get involved in English Profile and help to ensure that our findings are as accurate as we can make them.

Why is English Profile important for teachers?

The CEFR was designed to be a tool which could be used to promote ‘mutual understanding and cooperation’ in language education (Council of Europe 2001a:2). It can, for example, be used to set ‘feasible and worthwhile’ learning objectives and ‘enhance the transparency of courses, syllabuses and qualifications’ (Council of Europe 2001a:1–6). Since its publication in 2001, the CEFR has become increasingly influential in the language teaching and assessment community, not just in Europe but around the world, as a way of benchmarking language proficiency. Every country in Europe is formally signed up to the CEFR, and there are educational organisations in almost every country outside Europe that use the CEFR as the reference point for describing levels of English. One source of its appeal is its international quality; it means that learners can say that they have achieved ‘B1 in English’ or ‘C1 in German’ and expect that employers, as well as teachers and other language professionals, will be able to have a fairly clear understanding of what that means.

However, the CEFR can be difficult to use, not least because (as noted above) it is neutral with respect to the language being learned. It divides its description of language ability into six main levels ranging from A1 (Basic User) to C2 (Proficient User), and each level is defined using a series of ‘Can Do’ statements. These statements describe what learners’ can do with the language at each level, in terms of their language skills, and the communicative functions they are able to achieve, like agreeing, suggesting, or giving directions. Here is an example taken from the CEFR showing what A level learners ‘Can Do’ in the communicative activity ‘transactions to obtain goods and services’ (Council of Europe 2001a:80):

Table 1.1 Sample A level learner scale

A2	<p>Can deal with common aspects of everyday living such as travel, lodgings, eating and shopping. Can get all the information needed from a tourist office, as long as it is of a straightforward, non-specialised nature. Can ask for and provide everyday goods and services. Can get simple information about travel, use public transport: buses, trains, and taxis, ask and give directions, and buy tickets. Can ask about things and make simple transactions in shops, post offices or banks. Can give and receive information about quantities, numbers, prices, etc. Can make simple purchases by stating what is wanted and asking the price. Can order a meal.</p>
A1	<p>Can ask people for things and give people things. Can handle numbers, quantities, cost and time.</p>

What is English Profile?

These ‘Can Do’ statements, because they apply to all languages, are necessarily generalised, and can be difficult to interpret for specific languages and contexts. They reflect the skills-based approach of the CEFR, talking about what learners can do in terms of skills like listening, speaking, etc., rather than describing the actual language itself. For example, at A2 learners can ‘understand short, simple texts on familiar matters of a concrete type which consist of high frequency everyday or job-related language’ whereas B1 learners can ‘read straightforward factual texts on subjects related to his/her field and interest with a satisfactory level of comprehension’ (Council of Europe 2001a:69). But what is a ‘simple text’ in English? And what is the difference between a ‘simple text’ and a ‘straightforward factual text’? What exactly is ‘high frequency everyday’ English language?

To ensure that the CEFR can be adapted to local contexts and purposes, the Council of Europe has encouraged the production of Reference Level Descriptions (RLDs) for national and regional languages, and the EPP is carrying out this work for the English language. Ratified by the Council of Europe, the EPP aims to work from the generalised statements provided by the CEFR across to the specific words, phrases, grammar structures, etc. of English; to identify the specific language features that are actually used by learners of English at each level of the framework. There are several other projects which are doing or have done the same work for other languages, including *Profil Deutsch* for German, and *Profilo della lingua italiana* for Italian (see www.coe.int for further information).

Description, not prescription

Following the tradition established by the authors of the CEFR, English Profile describes what learners can do at each level in terms of positive achievement rather than negative deficiencies. While these descriptions can be set as objectives for learning, or used to establish whether a learner has attained the level of proficiency in question, it is important to stress that English Profile is providing a description, rather than a prescription. Like the Council of Europe, English Profile is not attempting to provide English Language Teaching (ELT) professionals with a *definitive* set of language points that they *should* teach at each level. As the CEFR authors emphasise in their ‘notes for the user’: ‘We have NOT set out to tell practitioners what to do or how to do it . . . It is not the function of the Common European Framework to lay down the objectives that users should pursue or the methods they should employ’ (Council of Europe 2001a:xi).

The exact choice of language points that are appropriate for a particular course will depend on a range of factors, such as:

- the range of levels of learners on the programme
- the age and educational background of the learners

Cambridge University Press

978-1-107-49398-8 – English Profile in Practice

Edited by Julia Harrison and Fiona Barker Michael Milanovic Nick Saville

Excerpt

[More information](#)

English Profile in Practice

- their reasons for learning English
- their areas of interest
- their first language(s)
- their experience of learning English so far
- other sources of input and opportunities to practise English.

The ELT professional will need to weigh these factors when making decisions about what to include in a course. English Profile provides resources to help with those decisions, and examples of curricula that have been drawn up using English Profile resources.

Corpus research

Throughout this book we will be referring to the use of a *corpus*, or (plural) *corpora* in English Profile research. A corpus is simply a collection of texts (nowadays this mainly means electronic texts), which can be searched for specific items such as words or grammar features, using dedicated software. For English Profile, these texts consist of written and spoken English produced by English language learners. These texts are held in a database together with information about the first language, age, nationality, etc. of the individuals who produced them.

Using corpora as the basis for their research means that English Profile researchers are using language that has been produced by real learners of English, to help us reach our conclusions. Previous language profiles have been produced by language specialists largely using their insights as expert users and teachers of the language. In contrast, because it uses corpora, English Profile's methodology is empirical in that it provides concrete evidence of what language learners throughout the world actually know at each level of the CEFR. Also, the fact that the corpus data the EPP uses has been provided by learners of English all over the world means that English Profile findings are 'non-linguacentric', that is (a) the research is not solely concerned with English as it is spoken by native speakers in the UK, USA or elsewhere, and (b) there is no first language bias, as the corpus contains language from such a wide variety of learners. This means that English Profile's findings can be referenced with far greater certainty than anything preceding it.

The main corpora that English Profile researchers use are the Cambridge Learner Corpus (CLC), the Cambridge English Profile Corpus (CEPC) and the Cambridge English Corpus (CEC).

The Cambridge Learner Corpus

The Cambridge Learner Corpus (CLC) was developed as a collaborative project between Cambridge English Language Assessment and Cambridge

Cambridge University Press

978-1-107-49398-8 – English Profile in Practice

Edited by Julia Harrison and Fiona Barker Michael Milanovic Nick Saville

Excerpt

[More information](#)

What is English Profile?

University Press. The CLC contains over 50 million words of written exam scripts produced by English language learners, making it the largest learner corpus in the world. It has been collected from learners of English all over the world taking Cambridge English exams, such as *Cambridge English: Preliminary* (also known as *Preliminary English Test (PET)*) or *Cambridge English: Proficiency* (also known as *Certificate of Proficiency in English (CPE)*).

A large proportion of this corpus has been coded for learner error. This means that researchers using the CLC can see if learners are making errors with, for example, a particular word or grammar feature. They can see what types of errors learners are making, including information about whether learners with different first languages make different errors; and they can also, crucially, see when learners have mastered an aspect of language, and are producing it accurately and appropriately.

For English Profile researchers, the exceptional thing about CLC data is that it is very clear which level of the CEFR the student who produced it has reached. Each of the Cambridge English exams is carefully calibrated to reflect a CEFR level, and we have information about whether a candidate has reached, exceeded or failed to reach a level. This means researchers using the CLC can be certain about the kind of language learners are capable of producing in exams at each level.

The CLC is also extremely useful for grammar research. Two English Profile grammar projects (see Chapter 3) used corpus research techniques as their principal method of investigating how learners acquire grammar. Both were able to take full advantage of the CLC's extensive annotations; as well as error information, the CLC is annotated with two more types of information, which makes it particularly useful for this kind of research:

1. Part of speech information: each word in the corpus has been automatically coded to identify which part of speech it represents. This makes it possible to search, for example, for the word *bear* as a verb rather than a noun.
2. Grammatical relations (parsing): a system called the Robust Accurate Statistical Parser (RASP)¹ was used to add information about the grammatical relationships between words, e.g. subject-verb-object relationships.

This means that researchers can easily and quickly search the corpus for a particular grammar feature, for example all the past participles of verbs, or all examples of comparative or superlative adjectives. They can see how many times this feature is used at a particular level, and, because of the error codes,

¹ The Robust Accurate Statistical Parser, developed by Ted Briscoe of the Cambridge Computer Laboratory, and John Carroll of the University of Sussex. For more information see: www.informatics.susx.ac.uk/research/groups/nlp/rasp/

Cambridge University Press

978-1-107-49398-8 – English Profile in Practice

Edited by Julia Harrison and Fiona Barker Michael Milanovic Nick Saville

Excerpt

[More information](#)

English Profile in Practice

they can also see to what extent the feature is being used correctly by real learners in exam contexts.

The Cambridge English Profile Corpus

The Cambridge English Profile Corpus (CEPC) is being collected specifically for English Profile, and has been designed to suit English Profile research needs. Like the CLC, it is a corpus of learner English produced by students all over the world. Also like the CLC, it will be aligned to the CEFR levels, allowing the study of the acquisition of English, and the development of teaching and assessment material, across proficiency levels. However, whereas the CLC is composed entirely of exam scripts and their related questions, the CEPC covers a wider range of learner output, including essays, coursework, and spoken data, collected in real or virtual classrooms or completed as homework. It includes responses to tasks designed by English Profile researchers for specific research purposes, to elicit features of learner production which are rarely captured in corpora, and it also includes more ‘free’ writing, where learners have the chance to express themselves without the constraints of an essay or exam format. CEPC data can therefore reveal learners’ capabilities beyond the specifics of what they have been taught and tested on.

The CEPC covers all six levels of the CEFR, and attempts to maintain a balance across a number of variables, including:

- educational contexts (e.g. primary or secondary, monolingual or bilingual)
- linguistic functions (e.g. informative, suasive, attitudinal, socialising and structuring discourse)²
- first languages of learners
- ages of learners, and other demographic information
- spoken data is also balanced across types of interaction e.g. casual conversation, formal presentation, oral exam, classroom discourse, role play etc.

The CEPC is being built by Cambridge University Press and Cambridge English Language Assessment, in collaboration with a network of participating educational establishments across the world, including schools, universities, research centres, government bodies (such as ministries of education) and individual education professionals. Written data is being collected via the online English Profile data collection portal, and the EPP welcomes

² The functional categories can be found in Appendix B of Green (2012): *Language Functions Revisited*. They are taken from van Ek and Trim’s ‘Threshold’ series. This series, also known as the T-series (van Ek 1975, van Ek and Trim 1991a/1998a, 1991b/1998b, 2001, Trim 2009), was produced by the Council of Europe and pioneered the function-oriented approach which the CEFR itself adopted.

Cambridge University Press

978-1-107-49398-8 – English Profile in Practice

Edited by Julia Harrison and Fiona Barker Michael Milanovic Nick Saville

Excerpt

[More information](#)

What is English Profile?

contributions from anyone involved in English language learning and teaching. More details about this can be found on the English Profile website (www.englishprofile.org).

The Cambridge English Corpus

Researchers also have access to a large amount of expert and native speaker data as, alongside the CLC, the Cambridge English Corpus (CEC) contains more than a billion words of written and spoken English from books, newspapers, advertising, letters, emails, websites, recordings of conversations, lectures, meetings, TV, radio, and many other sources (see www.englishprofile.org/index.php/corpus). This means that researchers can make comparisons between the way learners of English are using the language, and the way it is used in the wider context.

English Profile research: Directions

Initially, the focus of English Profile research has been on vocabulary, grammar and language functions.

1. Vocabulary

The English Vocabulary Profile (EVP) (www.englishprofile.org/index.php/wordlists) is an online resource which is freely available. It is the product of the research into vocabulary learning across the CEFR, and is now complete for all six levels, A1–C2. It allows the user to find out which words and phrases – and individual meanings of each word – are typically mastered by learners at each CEFR level. This is a really valuable tool for helping to guide decision-making around what to teach students as they progress, and Cambridge University Press authors and editors make extensive use of this research in developing their course materials. This resource will be discussed in detail in the next chapter.

2. Grammar

A separate research team has developed a similar resource, the English Grammar Profile (EGP) which describes learners' gradual mastery of grammar across the six CEFR levels. The research team, sponsored by Cambridge University Press, has examined which grammar features are mastered at which level of the CEFR, including how and where particular features are being used.

A parallel project, which focused primarily on grammar but also looked at aspects of lexis acquisition, was carried out by researchers based at the

Cambridge University Press

978-1-107-49398-8 – English Profile in Practice

Edited by Julia Harrison and Fiona Barker Michael Milanovic Nick Saville

Excerpt

[More information](#)

English Profile in Practice

University of Cambridge. Their work introduced the notion of ‘criteriality’ and involved identifying the language features which are characteristic (or ‘criterial’) for each level of the CEFR. Their results have been published as volume 1 of the English Profile Studies series (Hawkins and Filipović 2012).

Each of these two research projects has informed the other, and both will be further discussed in Chapter 3.

3. Functions

As we have seen, the CEFR describes language ability in functional terms, as a series of statements which describe what learners can do at each stage of the Framework. English Profile has built on this, looking at practical ways in which the Can Do statements of the CEFR can be expanded and refined, by providing additional detail with reference to contexts of use which are particularly relevant to learners of English. Research has focused on the C levels, because these were the least well-defined in the CEFR, and were not covered by the T-series (see the footnote on p. 6).

The results of English Profile research in this area have already been published as volume 2 of the English Profile Studies series (Green 2012). Chapter 4 of this volume looks at how teachers and related professionals can approach the question of functional language in English.

The value of language learning

The Council of Europe has consistently prioritised language learning as a social and political objective. They argue that ‘a further intensification of language learning and teaching in member countries is necessary in the interests of mobility’ (Council of Europe 2001a:5). Learning languages, they maintain, not only develops an individual’s linguistic skills, but also their intercultural understanding, or ‘interculturality’. Speakers of more than one language, whatever their proficiency level, are seen as fostering in themselves an understanding that other cultures’ norms can differ from their own, and that these differences should be treated with tolerance and respect. Becoming ‘plurilingual’, the Council of Europe argues, ‘enable[s] the individual to develop an enriched, more complex personality and . . . greater openness to new cultural experiences’ (Council of Europe 2001a:43). What begins with grammar and vocabulary leads to sociolinguistic and intercultural competence, and English Profile aims to contribute to and facilitate this process.

2 The English Vocabulary Profile

Annette Capel

*Consultant, Cambridge University Press and
Cambridge English Language Assessment*

The English Vocabulary Profile (EVP) is an interactive online resource that describes the words and phrases typically known and used by learners of English around the world. It is similar to an online dictionary in that it has *headwords* and detailed *entries*, and it can be searched in a number of different ways. This chapter covers the rationale and history of the EVP, its trialling and validation, the compiling methods used to create it, and finally, some practical suggestions as to how it can be fully exploited by ELT professionals such as materials writers, test developers and classroom teachers.

Rationale and history of the project

The EVP research team was led by Annette Capel, who carried out the corpus analysis and scrutiny of other ELT source materials to assign the CEFR levels, and then compiled the British English version with lexicographers Elizabeth Walter and Kate Woodford. Carol-June Cassidy adapted and added to this text in preparing the American English version. The complete A–Z text was read in draft by a number of reviewers and subsequently trialled and validated (see next section). The computing expertise of Daniel Perrett and Dominic Glennon has delivered an online resource that is both powerful and user-friendly.

The core objective from the outset of the EVP research back in 2007 has been to establish by means of electronic corpora and other reliable sources what vocabulary is commonly known and used by learners around the world at the different CEFR levels (A1 to C2). In other words, rather than providing a syllabus of the lexis that learners *should* know, the EVP project has concentrated on verifying the English lexical items they *do* know. This ties in with the CEFR emphasis on Can Do statements and also reflects research being done by other English Profile partners and research teams (see Chapters 3 and 4 for some examples).

What is meant by ‘know’ in this context? In early discussion at the sample stage, before the main compiling began, the issue of a possible disparity in level between receptive and productive language was raised, both internally and by some reviewers. This is an area that has often been researched over

English Profile in Practice

the years (see Melka 1997 for a fuller discussion of this area) and yet arguably remains intangible and unproven. Much will always depend on learning styles and, indeed, classroom dynamics – the extent to which opportunities are given for productive use. In exam classes, balanced preparation of all four skills has to take place if candidates are to succeed, and so it is advisable to encourage learners to actively use the words and meanings they are exposed to in textbooks and classroom teaching. In general, modern communicative classrooms provide more consistent opportunities for active use of the new language than a generation ago, and, especially in relation to vocabulary, the prevailing advice seems to be ‘use it or lose it’. So perhaps the gap between receptive understanding and productive use is not as wide as some people have claimed. The evidence in the CLC suggests this might be the case.

In the EVP, CEFR levels have been assigned at *meaning* and *phrase* level, in order to give as full a picture as possible of how a learner progresses up the six levels. At the Basic User level (A1 and A2), learners generally acquire new concrete words, mainly topic nouns and ‘action’ verbs, together with grammatical words like conjunctions, determiners and prepositions, as well as functional phrases. So, at Basic User level A1, a learner would know nouns such as *bag, board, breakfast, brother*, verbs like *come, go, meet, see*, grammatical words like *and, but, every, the, to, with*, and phrases such as *See you soon* and *Well done!*

The CEFR refers to the B levels as the Independent User levels, and this indicates a broadening of knowledge together with growing confidence in use. Our research has shown that learners are constantly adding to their store of meanings and phrases for the polysemous words they already know. Words with multiple meanings are highly frequent in English and are therefore extremely valuable for learners, as they feature in different everyday contexts. For example, there are 67 matches for the verb *keep* in the EVP, ranging from A2 up to C2 and including extended meanings like food *keeping*, i.e. remaining fresh, phrases such as *keep in touch* and *keep your cool*, phrasal verbs like *keep away*, and idioms such as *keep your eyes open*.

Trialling and validation of the A1–B2 pilot version

Once the first four CEFR levels of the EVP were compiled in draft, the prototype resource went through a process of validation and user trialling. This aimed to test the usability of the online platform, to verify the decisions taken on CEFR levels and to assess the coverage, with a view to adding anything relevant at A1–B2 that had been inadvertently omitted. To this end, password access was provided to known user groups, notably Cambridge University Press authors, editors and lexicographers, and Cambridge English item writers and exam developers, who worked with the resource over a 12-month period and submitted detailed comments via the feedback button. These