

The English Profile Network

Report on EPN Methods and their Implications for Reference Level Descriptions

Reference level Descriptions

The European Union's 2020 strategy seeks to promote economic mobility and foster social inclusion and cohesion. Communication in foreign languages is recognised as a key competence for the knowledge economy and society and improving language learning is seen to be a key element in the attainment of the 2020 flagship initiatives. Because it provides a shared basis for the international recognition of language qualifications, the Common European Framework of Reference for Languages – CEFR (Council of Europe 2001) facilitates mobility in terms of both education and employment for citizens of the EU and supports the call for a common language between the world of education and the world of work. However, it is acknowledged that the CEFR, partly because of its pan-linguistic scope, is necessarily underspecified with respect to any individual language.

To address this, the Council of Europe has called for the development of Reference Level Descriptions (RLDs) detailing characteristic linguistic and cultural features of a particular language at each of the six levels of the CEFR. Each RLD 'implements solutions and makes choices that are adapted to the language concerned' (Council of Europe 2005: 4). This will allow an individual's linguistic proficiency to be more accurately pinpointed in terms of the CEFR, but in a way that takes account of local contexts.

The English Profile Network sets out to develop RLDs for the English language. In this aim it has many precedents. First, the *Threshold* series of specifications which grew initially from the *Permanent Education* agenda of the Council of Europe during the 1970s. The series now extends to four levels (*Breakthrough*, *Waystage*, *Threshold* and *Vantage*). All of these have been made available for download from the English Profile website (www.englishprofile.org). The *Threshold* series volumes broadly correspond to the levels of the CEFR – *Breakthrough*: A1, *Waystage*: A2, *Threshold*: B1 and *Vantage*: B2 (and above). Counterparts to *Threshold* were produced for over 30 languages. Second, since the publication of the CEFR in 2001, RLDs have been developed for several languages, including, among others, the *Čeština jako cizí jazyk - Úroveň A1–B2* (Charles University in Prague 2005) for Czech, *Niveau A1-B2 pour le français* (Beacco and Porquier, 2007) for French and *Profile Deutsch A1 – C2* (Glaboniat et al. 2006) for German. Additionally, other contemporary initiatives have been launched to elaborate the CEFR for English (the *British Council/EAQUALS Core Inventory for General English* being one example).

The Council of Europe (2005) suggests that RLDs should draw on the *Threshold* series and benefit from the experience represented by them, but that they should be explicitly grounded in the CEFR. To this end, the Council lays down the following minimal requirements:

- 1) they should be developed using the CEFR descriptors in particular (as well as those of the European language Portfolios models) or other descriptors and explain exactly how to pass from the descriptors to the inventories of forms;
- 2) a description of the approach(es) used to establish the inventories of forms (survey, for example);
- 3) indications as to whether the proposed forms should be known for reception only or also for production;
- 4) inventories of the linguistic realisations of general notions, acts of discourse and specific notions / lexical elements and morpho-syntactic elements considered characteristic of this level.

(Council of Europe 2005: 3)

The call for inventories of 'general notions, acts of discourse and specific notions / lexical elements and morpho-syntactic elements' (see point 4) is, rather unhelpfully, expressed in terms that are not used in the CEFR. However, it seems to correspond to i) grammatical categories (general notions), ii) the lexicon (specific notions) and iii) communicative language functions, activities and texts (acts of discourse). With this in mind, the EPN has worked towards the development of the three key strands of the English Profile:

the English Grammar Profile (EGP)
the English Vocabulary Profile (EVP) and
the English Functions Profile (EFP).

Additional EPN work has been done on phonology, which we feel should not be excluded from consideration, although it has proved more challenging to scale in relation to the CEFR.

Established methodologies for developing RLDs have drawn heavily on the precedent of the *Threshold* level. This approach relies on expert linguistic judgement, such evidence as is available from Applied Linguistics research and the experiences of language teachers to determine appropriate linguistic exponents to realise 'communicative language functions' and 'general and specific notions' at each CEFR level. This can be characterised as a 'top-down' methodology because it begins from the interpretation of the CEFR descriptors by selected experts and works from these generalised statements down towards the specifics of language use in context. The objective is to elaborate the CEFR descriptors in terms of linguistic exponents: words and phrases that can be used by learners to carry out the activities described.

The EPN, in contrast, set out to develop a methodology which would be not only 'top-down' (involving expert interpretation of descriptors), but also – and more innovatively – 'bottom-up'. The top-down expert user perspective comes from the English Profile partners and their collaborators around the world who are able, like the developers of other RLDs, to draw on extensive interdisciplinary expertise and experience in language education and linguistic theory. The bottom-up aspect refers to the direct empirical study of learner language.

Fundamentally, the approaches and methods adopted by the EPN are based on the *socio-cognitive* approach to language in use first developed by Weir (Weir 2005, Shaw and Weir 2007 Khalifa and Weir 2009a and Taylor 2011). This framework emphasises the role of the mental processes that

learners are able to draw on in using a language, but also the constraints imposed by the nature of the tasks that they carry out and the social contexts in which they need to act.

In order to investigate learner language from the bottom up, or empirically, at each of the CEFR levels, very large samples of writing and speech are needed. A long-term aim of English Profile is therefore to build extensive and representative language corpora, useful for both the investigation of English Profile hypotheses about language levels and for ideas on their learning/teaching and assessment.

The EPN partners have cooperated in developing a corpus of learner English from students across Europe and beyond. This has allowed us to develop a methodology for describing language use which is empirical, but not solely concerned with the language as it is spoken in one country or region (the UK) or for a limited range of purposes. It is argued that this methodology is readily transferable to other languages, and provides an exemplar of innovative good practice in generating RLDs regardless of language: one that addresses the additional demands imposed when attempting to describe a major world language, but that could equally be applied to minority languages.

In this corpus building endeavour, the EPN has been able to draw on existing resources and to use the partners' experience with corpora in developing new resources that better meet the needs of the programme.

The Cambridge Learner Corpus (CLC) was developed as a collaborative project between Cambridge University Press and Cambridge ESOL. It has grown over the course of the EPN. It now contains around 50 million words of English from learners of English all over the world, 25 million of which has been coded for learner error.

The CLC has been very important to the English Profile, but does have some serious limitations. It is restricted to responses to test prompts and this may restrict the insights it can provide. An instance of the potentially distorting effect of relying on samples from examinations was the impact that words used in test prompts had on word counts in the development of the EVP. The word *deforestation*, to take one example, emerged as a relatively high frequency item in the data at the C levels, but this was traced to its use on the question paper in a particular test task. A further shortcoming of the CLC is that it mainly consists of written language. There are well-documented differences between spoken and written language use and it could be seriously misleading to base RLDs on written usage alone.

The Cambridge English Profile Corpus (CEPC) complements the CLC, but has been compiled with the needs of the developing RLDs in mind. Reflecting the impact of social variables on language use acknowledged in our socio-cognitive approach, 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. The aim is eventually to provide 10 million words of data, covering both spoken (20%) and written (80%) language. Both General English (60%) and English for Specific Purposes (40%) contexts are included. The corpus covers CEFR levels A1-C2, and attempts to maintain a balance across a number of variables, including CEFR level, first language, and educational setting.

Researchers using the CEPC in place of, or as a complement to the CLC are therefore better able to track learners' acquisition of a language feature across CEF levels, revealing their capabilities beyond

the specifics of what they have been taught and tested for in any given year. Like the CLC, the CEPC is aligned to the CEF levels through application of the CEFR descriptors which are used to assign learners and their performances to a CEFR level.

A third source of empirical data, although less developed at this stage, is the nature of the input that learners are able to process receptively. The EVP has drawn on native speaker corpus data as one indicator of this, as well as English language teaching materials. Progress towards compiling a corpus of input explicitly targeted to learners at different levels of the CEFR has begun in relation to the EFP, although it is clearly less straightforward to determine how input texts (rather than learner output) should be classified in relation to the CEFR (and even more complex for recordings used in teaching and assessing listening skills). The EPN has proved very helpful in sourcing such material from a wide range of educational contexts.

In transferring the empirical methodology of the EPN to other language contexts, an important first step would be to involve experts in corpus linguistics as well as pedagogy and language education, to survey the corpus resources already available for the language in question and to design corpus building activities around the needs of learners. The EPN experience should help to guide the compilation of corpora as a basis for analysis of learner production at the different CEFR levels.

The following sections outline how the RPN has addressed each of the four minimum requirements for RLDs laid down by the Council of Europe.

Methodologies for requirement 1: a) passing from the CEFR descriptors to the inventories of forms

Other RLD projects, such as *Profile Deutsch*, have drawn directly from the CEFR descriptor bank to generate contextualised profiles of learner language use. This is effectively an enhanced and reordered version of the *Kontaktschwelle*, a version of the *Threshold* level developed for German. A tri-national team of experts from Germany, Austria and Switzerland, was commissioned to reorder the material for levels A1 to B2, deciding on which material was appropriate to each level. The team then went on to add a specification for Levels C1 and C2. This was more challenging because the C levels are not fully specified in the CEFR. An issue (one that the EPN has also had to face) was that learners reaching the C levels tend to be more diverse and more specialised in their learning objectives than those at the lower levels. It was considered inappropriate to specify the grammar and vocabulary required at C1 and C2. Instead, the team set out a wide range of 'Can-do' descriptors to cover the activities and competences presented in the CEFR. An example of this can be seen in the treatment of one CEFR descriptor such as 'Can write clear, well-structured expositions of complex subjects, underlining the relevant salient issues' (C1). This descriptor was transposed and exemplified by the expert team of developers for increasingly localised contexts for learning German (including, where appropriate differences in usage between the three countries):

Can present his/her own standpoint in a commentary, giving prominence to the main points and supporting his/her views with detailed arguments

e.g. can, as a computer expert, write a critical review of a new software package , giving examples to illustrate its disadvantages

(Glaboniat et al. 2006 trans. John Trim)

At the lower levels, as in the *Threshold* series, examples are given of how generalised tasks can be achieved using words, phrases and grammatical structures believed to be available to the learners of German at different levels of proficiency. In this process, the expert judges related the CEFR descriptor to familiar language learning purposes. They considered the functional requirements of a particular, situated use of language. They then used their expert knowledge to decide on the linguistic exponents that learners judged to be at a given CEFR level might be able to use to carry out the tasks and activities described. The publication of the resource as a searchable CD-ROM served to show how RLDs could be made more flexible and useful through this form of presentation.

Building on the experience of other RLD developers, the EPN has sought innovative ways to pass from the CEFR descriptors to inventories of linguistic exponents. First, in collecting samples of learner language, judgements about the learner's general level are made in relation to the CEFR descriptors: each text or recording entered into the corpus is assigned a CEFR level. This is achieved through the expert judgement of a trained examiner either by using the CEFR descriptions as a rating scale or on the basis of the Cambridge examinations, which have been related to the CEFR. In this way, every sample of learner language used in developing the inventories has been assigned a CEFR level. It is also tagged by source. In this way, it is possible to trace the impact of the demands of the different text types and genres required by the assigned tasks.

Achieving representativeness is essential in corpus building in general and to the EPN in particular. It is clearly important to reflect the unique position of English as a global language and the EPN brings together English language learning, teaching and assessment stakeholders from around the world, avoiding a purely Anglo-centric approach. An important aspect of EPN corpus building efforts has been to collect samples of English language use from learners from a very wide range of different first language groups. This allows for consideration of the impact of different first languages on these features (through 'transfer' effects, but also in terms of the impact of cultural difference).

Methodologies for requirement 1: b) integrating new descriptors

The principle of including a wide range of voices extends to the sampling of the top-down expertise element of the EPN. Where *Profile Deutsch* employed a small team of experts who drafted new descriptors where necessary, the English Profile partners felt that the diversity of English language education would be better served by a broader-based survey approach. The higher levels of learner English were selected as starting point in this process as the *Threshold* series does not distinguish between B2, C1 and C2 and there are relatively few descriptors located at the C1 and C2 levels of the CEFR. The C levels are therefore the least well specified and least informative for the EPN projects.

Educational materials directed at higher level learners were identified and 'Can Do' statements associated with these were collected and incorporated into a database. Tools such as Key Words in Context (KWIC) concordances were used to identify the range and scope of language learning goals for higher level learners of English. Basic parameters for the design of meaningful descriptors were set and a set of new descriptors for higher level learners was synthesised from over 2,000 assembled statements.

The proposed elements for Can Do statements included the following:

Activity: Can... The social act (function) or related sequence of acts (activity) that the learner might be expected to accomplish by means of the language

Theme/ Topic: Concerned with... The themes, topics and settings in relation to which the learner might be expected to perform. In the CEFR, applicable themes are grouped under the four domains: educational, public, professional and personal

Input text: Based on... The nature of the text that the learner might be required to process as a basis for his or her own contribution or to demonstrate his or her comprehension

Output text: Producing... The nature of the text that the learner might be expected to produce or participate in producing to demonstrate (a specified degree of) understanding or to accomplish a task

Qualities: How well? The qualities that the learner would be expected to demonstrate in carrying out language activities. For production, these qualities might be grouped under the CEFR headings of Linguistic, Pragmatic, Sociolinguistic and Strategic competences

Restrictions: Provided that... Physical or social conditions and constraints under which the learner would be expected to perform

The new descriptors, collected from a wide range of sources, were integrated with the CEFR scales, not on the basis of the expert judgement of a small group of experts, but through replication of the CEFR survey methodology. This involved a survey of over 700 English language educators from around the world. Participants in the survey judged the difficulty of the new descriptors for their students. CEFR descriptors were included on the survey as anchor items so that they could be used in estimating the relative difficulty of the new descriptors in CEFR terms. For a more detailed description of this aspect of the research programme, refer to Green (2012).

Requirement 1: Lessons for RLD development

The principle of broad sampling established for the EPN could usefully be transferred to other languages. The careful selection of learner language for corpus building naturally plays a crucial role in the quality of the results. Sourcing samples of learner language from a representative range of contexts requires the development of a network of data collection sites and the development of suitable means of elicitation. Those engaged in data collection for the EPN include public schools, universities, and private language schools, along with research centres, government bodies (such as ministries of education) and individual education professionals. Obtaining samples that represent a wide range of functional language use implies devising a wide variety of carefully designed elicitation tasks. The EPN employed the framework of functions and notions in the *Threshold* series to build a table of specifications that guided the development of purpose-built tasks capable of eliciting a broad range of text types from learners.

Contributing data has been made as straightforward as possible and written data is submitted via a purpose-built online portal. However, appropriate permissions must be obtained and, reflecting the socio-cognitive approach (Weir 2005), there must also be appropriate and sufficient data collected about entries to capture and trace salient facts about the background of the learners involved and

the conditions under which each sample was elicited. The CEPC captures a number of such variables which may be used in filtering the language samples for targeted research:

- educational contexts (e.g. primary or secondary, monolingual or bilingual)
- task type e.g. letter, email, report, essay (written data)
- type of interaction e.g. casual conversation, formal presentation, oral exam, classroom discourse, role play etc (spoken data)
- specific domains (e.g. medical English, business English)
- first language of learners
- age range of learners, and other demographic information
- country of data collection

The CLC comprises data from learners from over 130 L1s, thus permitting an extensive study of L1 transfer effects across all major language families. Cross-linguistic differences by CEFR level is one of the main premises under investigation in the EPN, reflecting the objective of examining the extent of the learner's source language involvement in determining their linguistic profile.

Methodologies for requirement 2: the approach(es) used to establish the inventories of forms

The EPN favours criteriality over comprehensiveness in developing inventories. Rather than attempting a fully comprehensive inventory of the grammar, lexicon and communicative functions for each level of the framework, the Profile has exploited the corpora to seek out *criterial features* (Hawkins and Filipovic 2012): characteristics of learner language that are particularly associated with performances located at one level of the CEFR rather than another.

The criterial feature concept starts from the assumption that there are certain linguistic properties that are characteristic and indicative of English language proficiency at each CEFR level. The awareness of levels demonstrated by language educators is exemplified by the levels of agreement found between raters of spoken and written performance on English language tests. This consensus must be based on an implicit awareness of the specific properties that distinguish between levels and allow the performance to be matched to descriptors. The challenge for the EPN is to recover what these properties – or criterial features – really are. Making this criteriality more explicit for grammar, lexis, phonology and, ultimately, communicative functions provides the necessary specificity.

In the EPN, criterial features are defined in terms of the linguistic properties that have either been correctly or incorrectly attained at a given level. The former are 'positive properties of the CEFR level', the latter 'negative properties of the CEFR level'. Positive properties are features that are accurately used by learners at a given level. Examples of negative properties include the errors found to be typical of learners at the level. Properties can also be conceived in terms of patterns of usage and frequency found to be characteristic of the different levels. Again, learners may tend to show both 'positive usage distributions for correct properties of English' and 'negative usage distributions' that do not match those of proficient speakers (such as an over-reliance on a limited range of logical connectives).

The definition for what is criterial for a particular level (B1, for example) is a cluster of criterial features, each of which distinguishes B1 from other levels. Obviously, the more unique a feature is

to B1 alone, the more useful it will be as a diagnostic for that level. But a feature need not be unique to B1 to be criterial. In fact, it is an inevitable consequence of certain types of criteriality that many features will not be unique to a level. If learners of English acquire a novel structure at B1, that structure will generally persist through B2 and the higher C levels and will be characteristic of B1–C2 inclusive. It will distinguish them from all lower levels. The EPN is collecting evidence of criterial features of these different types for each of the CEFR levels.

The search for criterial features is facilitated by the parsing and annotation for errors of the CLC and CEPC. Cambridge University Press has long experience of manually annotating the CLC. This work has been carried out by a small team of trained annotators who tag the errors according to agreed procedures and propose corrections (see Table 1). All annotations are reviewed and errors and inconsistencies of annotation are identified and addressed.

Table 1 Sample error codes in the Cambridge Learner Corpus (from Salamoura and Saville 2009)

Error Code	Explanation	Exemplification
RN	Replace noun	Have a good travel (journey)
RV	Replace verb	I existed last weekend in London (spent)
MD	Missing determiner	I spoke to President (the) I have car (a)
AGN	Noun agreement error	One of my friend (friends)
AGV	Verb agreement error	The three birds is singing (are)

An innovation that has dramatically improved the rate of annotation for the EPN has been the introduction of semi-automated parsing – using the Robust Accurate Statistical Parsing (RASP) system (Briscoe, Carroll and Watson 2006) – and error annotation tools. Manual parsing and error annotation of learner corpora are time-consuming and error-prone, whereas existing automatic techniques cannot reliably detect and correct all types of error and do not always correctly assign parts of speech. The EPN solution has been to combine the strengths of the two approaches. The automated system identifies and corrects those features that are most amenable to automatic detection, enabling the human annotators to work on features that may be incorrectly identified and so require their intervention. The development of the semi-automatic error annotation approach is described in Andersen (2010).

Two approaches to the issue of criteriality have been adopted for the EGP. One, grounded in theoretical linguistics and based on an understanding of psycholinguistic constraints on processing, is hypothesis driven and involves expert predictions about features of learner language that are likely to vary by level. The other is computational and data-driven and involves machine learning techniques. In this case, criterial features are identified through competition.

In the former approach, the researcher identifies a feature that seems likely to vary with level – the use of relative clauses, for example. The researcher then looks for patterns of (accurate and inaccurate) relative clause use in the corpus. If a certain kind of relative clause usage is found to differentiate in the way predicted between performances identified with different CEFR levels, the hypothesis is confirmed and the structure in question is declared to be criterial. For a more detailed development of this aspect of the research programme, refer to Hawkins and Filipovic (2012).

Examples of criterial grammatical features identified for the B1 level include the following:

Inflection for person or not: I walk – he walk vs. I can –he* cans

Inflection for tense or not: *I walk – I walked vs. I must*

Finite and non finite forms: I cycle, I have cycled, I can cycle; I expect to cycle, I expect to have arrived; I expect to *can/be able to cycle

Questions: can/will Kim read? what does Kim read? *what reads Kim?

Negations: Kim can/will/does not drive, *Kim drives not.

The extent of the representation of so many different source languages in the CLC has allowed for differentiation between learners from different L1 backgrounds. Features such as subject-verb agreement and the syntax of questions for main verbs with lexical content (go, arrive, walk, drive, cycle) and for modal and auxiliary verbs ((will, can, must, have, be, do) were identified as criterial for L1 Spanish and German learners of English only (Parodi 2008).

In the computational approach, samples are presented in turn to a classifier (a statistical model programmed by the researchers) as representative of, for example, a B2 or B1 level performance (the level being determined by human judges on the basis of the CEFR descriptors). Thousands of automatically measured textual features are used in ‘training’ the system to distinguish between B2 and B1 responses. As large numbers of responses are presented to the system, the classifier comes to identify features and combinations of features that allow it to categorise the responses in the most accurate and efficient manner. If, for example, the word ‘and’ occurs with greater relative frequency in B2 responses than in B1 responses, it might emerge as one predictive feature. Other features that failed to improve the predictive power of the classifier would be eliminated from the model. The features or constellations of features that emerge as the strongest predictors of level are identified as criterial.

The advantage of the hypothesis-driven approach is that the results are readily interpretable in relation to theories of language learning. However, it can also be challenging because many word forms can fulfil a number of different grammatical functions. Searching the corpus for instances of particular syntactic patterns is therefore time consuming, severely limiting the number of features that can be investigated within a given period. The computational approach is very much faster and more inclusive, but the outcomes are not generally as straightforward to interpret. In practice the two approaches complement each other as each contributes to insights that might not be engendered through use of the other.

Requirement 2: Lessons for RLD development

The key lesson from the EPN for those seeking to emulate our approach may not be so much the specifics of the methodologies adopted as

1. The priority given to criterial, differentiating features rather than comprehensiveness. The focus on criteriality makes it considerably more practical to define a level as well as making the resulting inventories more manageable for users.
2. Interdisciplinarity. The interaction between researchers from different disciplines has been very important to the success of the EPN both in carrying out shared research and in extending the scope of the findings. For example,

computational approaches have facilitated the coding and searching of the corpus; educational perspectives have helped to guide and supplement the empirical research.

Methodologies for requirement 3: reception and production

Although direct empirical evidence from corpora is central to the EPN approach, it is not the only source. In developing the profiles, account must also be taken of what will be pedagogically useful, efficient and informative for practical users. This is well-exemplified in the most fully developed of the profile projects: the English Vocabulary Profile (EVP). Capel (2010) describes how the development of the EVP has not simply been a matter of mining corpora for words that occur in learner production at each level of the CEFR. Rather it has integrated expertise and judgement with evidence from lexicography based on both receptive and productive language use and the EP corpus data.

Most of the words and phrases covered in the EVP were derived in the first instance from lexicographic research on the Cambridge English Corpus (CEC): a billion words of written and spoken English text taken from a very wide range of sources. The texts in the CEC represent the kinds of authentic English text that language learners might need to understand receptively.

Research into the CEC involved counting concordance lines for the most frequently occurring 6,000 words of English and identifying where these words were being used with different senses. According to the number of occurrences of a given sense, the researchers assigned it to one of three relative frequency levels: *Essential*, *Improver* or *Advanced*, with *Essential* representing the highest frequency of occurrence. These word frequency classifications were used in the *Cambridge Advanced Learner's Dictionary* (CALD). At the outset of the EVP project the dictionary entries for all word senses tagged *Essential*, *Improver* or *Advanced* were placed in a database. This was used as the starting point for the investigation of the CLC and CEPC: productive learner language. The findings were then checked against language teaching practice through inspection of course books and other learning materials.

The investigation of different word senses is significant. Word frequency lists (such as those based on the British National Corpus) tend simply to locate words at a specific frequency level. However, informed by the weight given to the social use of languages in the EPN's socio-cognitive approach, it was recognised that the EVP would have greater value if different senses of a word were taken into account.

By way of illustration, the word *light* functions as an adjective, a noun, and a verb in the CEC. Both the adjective and the noun appear from the CLC data to be widely known at the A levels and meanings such as *EQUIPMENT (turn on the light)*, *NOT HEAVY (a light bag)* and *MAKE BRIGHT (fireworks lit up the sky)* are listed at A2 level, reflecting the CLC data. However, an issue in using corpus data for the Profiles is that it can never be entirely clear whether the absence of a feature is criterial, or just a matter of sampling. In the EVP, the first listing for *light* comes at A1. This is for the adjective: *PALE*, with reference to colour, as in *light blue*. The decision to list this sense at A1 was not based on corpus data: examples of A1 learners using *light* in this sense were lacking. However, it was found that this meaning of *light* often appears in text books targeting the A1 level and it was therefore given its A1 listing on this basis. A verbal meaning is listed at B1: *START FLAMES*. For this meaning there was, again, straightforward evidence of learner use in the corpus. Through other

EPN-related research (Martinez 2011), it also became increasingly clear that the EVP would need to take account of multi-word word expressions – phrasal verbs, idioms as well as individual words. Phrasal uses of *light*, such as *come to light* or *shed light on*, only occur in any volume at the C levels in the CLC data.

Although the corpus data was generally illuminating, expert judgment was often required when different sources of evidence were in conflict. For example, although the *Improver* verb *eliminate* is fairly frequent in native speaker corpora, its use at B2 level in the learner corpora was found, on the basis of the data collected on learner background, to be largely limited to first language speakers of Latinate languages. From the expert judgement perspective it seemed more appropriate to the C levels in terms of its register and use and was assigned accordingly.

Similarly, analysis based on native speaker frequency does not always capture words that have a high frequency in the language classroom and are particularly useful to learners. To reflect the importance of classroom language, another top-down source of evidence was consulted. Wordlists from English language text books and other materials for learners – such as the Cambridge English Lexicon (Hindmarsh, 1980) – were used to support the inclusion of words or senses in the EVP. Some examples of words identified through these sources are: *album*, *download*, *guidebook*, *haircut*, *questionnaire*, *skateboard*, *trainer*. Capel (2010) notes that most of these additions are nouns and either represent lifestyle choices that are functionally important to learners – downloading music or skateboarding, for example – or are words that come directly from the teaching and learning experience: words such as *questionnaire*.

Green (2012) suggests another approach to exploring criteriality in texts targeted at learner reception. This combines a tradition of research in language testing that has sought to identify features that make texts more or less accessible to learners with the English Profile concept of criterial differences. The study seeks text characteristics that distinguish between texts targeted at learners of contiguous CEFR levels. This should help to identify the distance between receptive and productive knowledge of features. In practice, Capel (2010) suggests that for vocabulary the gap often believed to exist between productive and receptive knowledge is not as wide as is generally believed: at least for learners at the A and B levels and for the ‘common core’ vocabulary included at the C levels. On this basis the EVP makes no distinction between receptive and productive use, but presents a single CEFR level for each sense of the words or phrases listed.

Requirement 3: Lessons for RLD development

Evaluation of the draft form of the EVP aimed to test the usability of the online platform, to verify the decisions taken on CEFR levels and to assess the actual coverage, with a view to identifying any omissions. Comments from users were acted on and potential level discrepancies were further researched, with revisions sometimes being made as a result. An online questionnaire was also completed by users, which largely focused on the usability of the system. In this way, the experience and needs of end users could be used to inform further development.

It is very apparent from our experience of the EVP that any RLD resource has to be regularly monitored and updated. There are several reasons for this. One is shifts in language over time brought on by technological innovation and social change. In the EVP, words such as *download* that have emerged over recent years are not reflected in the older corpora used in compiling word

frequency lists. Another is changes that may occur in the teaching and learning of languages: it is quite possible that with shifting approaches to teaching and changes in the populations of people learning languages, the nature of learner proficiency will alter. Learner corpora must therefore be constantly updated and expanded to integrate new populations of learners. A third is the expanding body of evidence that informs the Profiles. As more data is collected, new approaches to analysis emerge and more feedback becomes available from users, the Profiles can be refined and expanded as a result.

Methodologies for requirement 4: inventories

The Council of Europe presents a diagram illustrating how RLDs can help to relate the CEFR to teaching and assessment programmes 'for a given language, at a given level, in a given situation, for (a) given user/ learner group(s)' Council of Europe 2005, p.5). The EVP is already proving to be a very useful tool for users, but to achieve this end the inventories – the EVP, EGP and EFP (and other Profiles that may follow) need to be effectively integrated to form one coherent, integrated resource, explicitly linked to the CEFR.

Fleming (2009) suggests that descriptors can only be 'interpreted and understood in relation to concrete examples of what they mean in practice' (p.11). It is through elaboration, sharing and negotiating in relation to examples of practice that agreement in judgement is arrived at. RLDs should therefore provide adequate examples and sufficient basis for negotiation and refinement.

The CEFR itself outlines just such a system through the concept of an 'information pyramid,' explained in the following terms:

The user is presented with an information pyramid and can get an overview, a clear perspective, by considering the top layer of the hierarchy (here the "global" scale). More detail can be presented - *ad infinitum* -by going down layers of the system, but at any one point, what is being looked at is confined to one or two screens - or pieces of paper. In this way complexity can be presented without blinding people with irrelevant detail, or simplifying to the point of banality. Detail is there - if it is required. (CEFR p.40).

As the other Profiles develop to complement the EVP, the English Profile will provide this kind of information pyramid for English. Each layer of the overall system will be integrated, to the extent possible, with the others to provide related information, giving substance to the RLDs. Meeting the needs and expectations of users will involve the provision of a number of informational layers made up of components such as the following:

Illustrative Can Do statements for specific purposes – assessor – task designer – user

Frames setting out how the elements of the CEFR model may interact in shaping the difficulty of defined language tasks

Grids of criterial features: the inventories of criterial features and indications of how these impact on level definitions

Glosses: definitions and elaborations of key words used in the reference level descriptions

Commentaries: discussions of how the components of the reference level descriptions are interpreted in relation to specified illustrative tasks

Sample **tasks**: A growing database of examples of (receptive and productive) tasks that learners at different levels might be expected to carry out, with commentary explaining how these relate to criterial features and, ultimately to the CEFR descriptors

Sample **performances**: examples of learner performance on the tasks in 6 – recordings or written scripts – illustrating the interpretation of the level descriptions

The relative difficulty of the multiple means of realising language activities is considered both from a social perspective and from a cognitive perspective. A given activity involves the activation of (cognitive) language processes, the ‘many-to-many’ (Hawkins and Filipovic 2012) relationship between form and function opens a wide variety of choices to the user. Processing occurs ‘in relation to [social] themes, in specific domains’ (CoE 2001, p.9) which introduce sets of constraints on these choices. Only by understanding the interaction between the cognitive and the social does it become possible to get an adequate fix (for objective-setting, teaching and testing purposes) on the relative demands of language learning tasks. This understanding requires concrete exemplification.

Exploring the operation of language processes in the CEFR, we are presented with lists of different skills that may be engaged in processing the language needed to carry out a task (Figure 1). Language processing models, like the one presented in Weir’s (2005) socio-cognitive approach, suggest that different tasks will call on the learner’s language skills to differing degrees and in different combinations. Communicative purpose, affected by contextual variables, impacts on the nature of the processing that occurs. Shaw and Weir (2007) and Khalifa and Weir (2009) identify the higher levels of the CEFR with an increasing role for the semantic and cognitive skills, while perception, recognition and linguistic skills have become automatised for learners at higher levels.

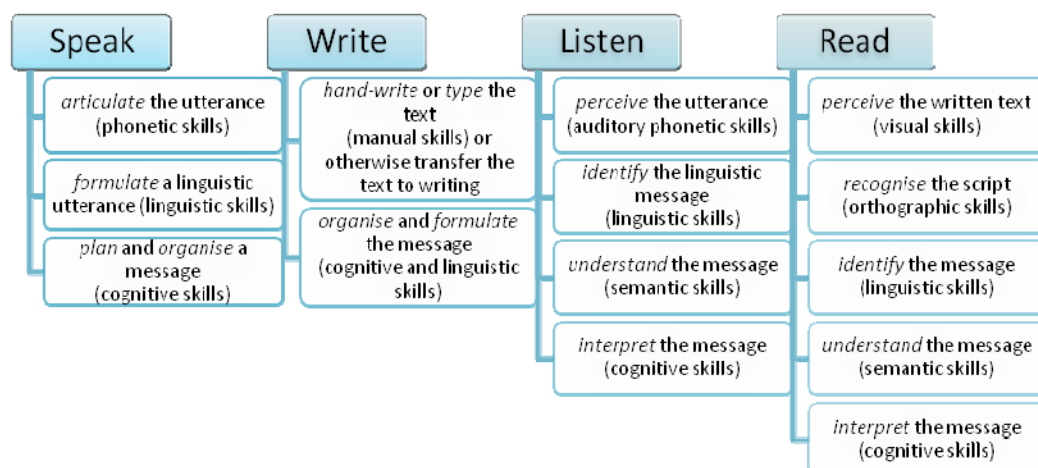


Figure 1 Language processes in the CEFR (CoE, 2001, pp.90-2)

In approaching a reading text, higher level learners are able to process the input more quickly (committing fewer cognitive resources to recognition and linguistic processing). In arriving with relative ease at an understanding of the message of the text, such learners have sufficient resources in reserve to be able to interpret and evaluate this message in relation to their prior knowledge or to the content of other texts. Learners with more limited competence may recognise the language and

even understand it linguistically, but this can require so much of their cognitive resource that without additional support they may struggle to form a coherent representation of the text as a whole.

All else being equal, receptive tasks that require retrieval of individual words (such as expeditious local reading or scanning tasks) are likely to prove easier than tasks requiring understanding of individual sentences (careful local reading), which in turn are likely to prove easier than tasks that require a detailed understanding of a text as a whole (careful global reading) or those that require integration of information from multiple texts (Khalifa and Weir, 2009). From this point of view, it becomes possible to identify the language processing demands made by different tasks and to estimate the potential impact of these demands on the learner.

Taking a social, or-contextual, category, it is equally possible to approach the description from the standpoint of topics and themes, navigating through the information pyramid from the more general view provided in the CEFR to access layers of increasingly finer detail. In the figures that follow, the boxes that shaped as arrows indicate that this is a strand that leads into the following sections; just as a hypertext link is followed in navigating from one screen to the next: in Figure 2 this is the personal domain.

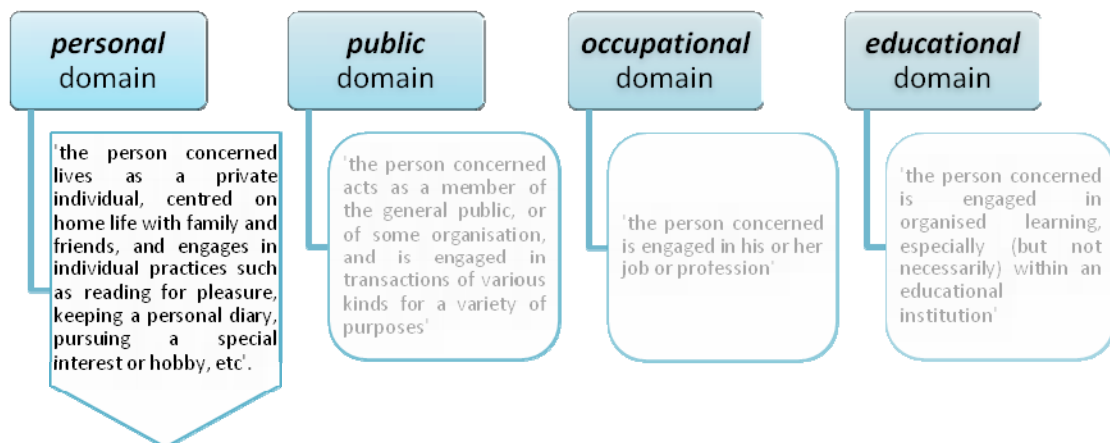


Figure 2 Domains in the CEFR (CoE, 2001, p.45)

In the CEFR, each domain is associated with certain themes or 'topics which are the subjects of discourse, conversation, reflection or composition' (CEFR p.51). For each domain, these themes are organised around seven categories: *locations, institutions, persons, objects, events, operations* and *texts*. The personal domain, following the strand we have taken as an example, has the following (Figure 3)

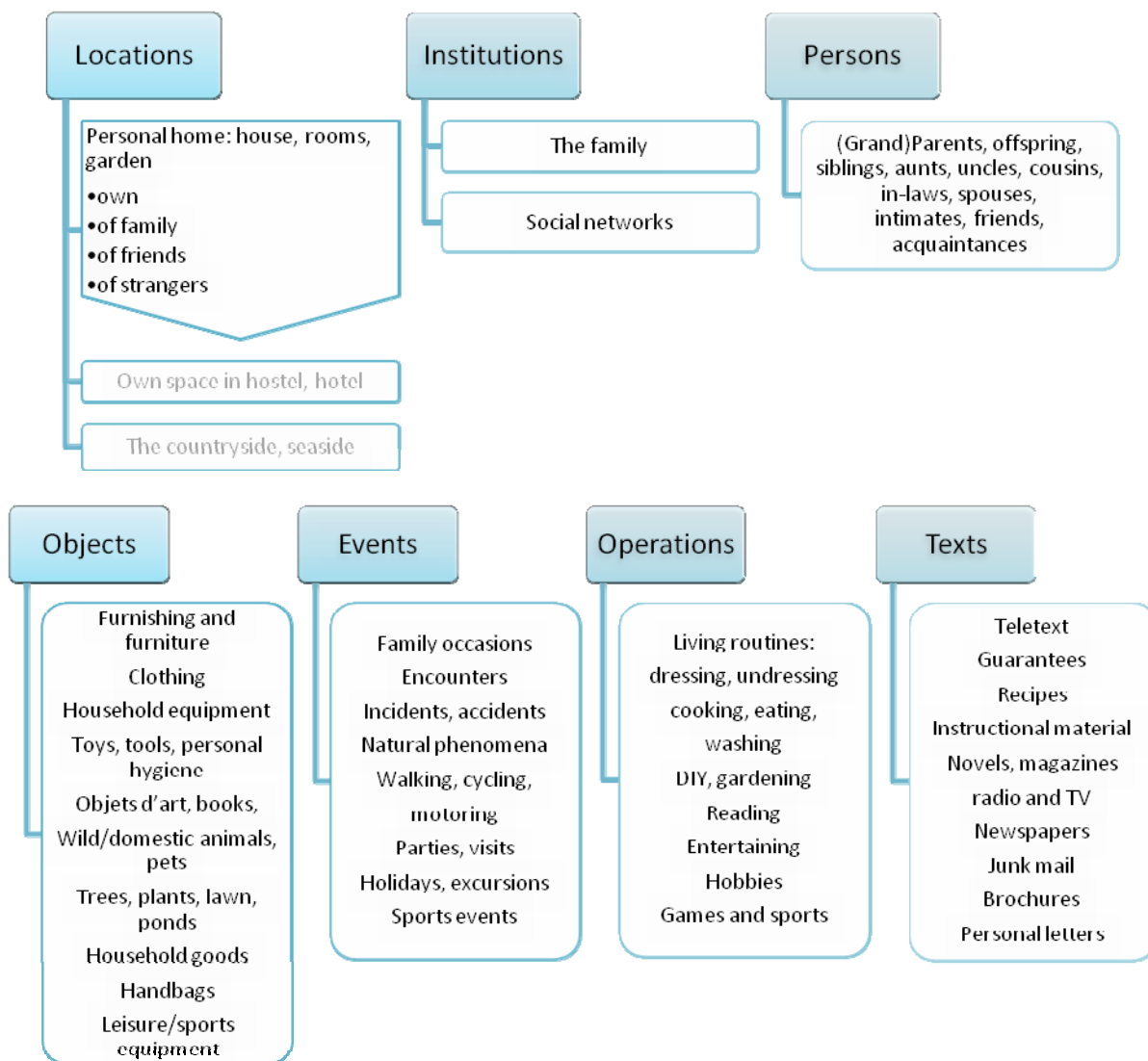


Figure 3 Locations, institutions and persons in the CEFR (CoE 2001, pp.48-9)

Moving beyond the CEFR, each theme might be pursued further to produce a set of topic-related 'specific notions.' Both the *Threshold* series vocabulary lists and the EVP (Capel, 2010) are organised thematically around semantic categories.

The EVP (Capel, 2010) provide a guide to common words and phrases that learners of English will need to know. Meanings of each word or phrase on the lists are assigned a level between A1 and C2 on the CEFR scale. Because of differences of classification, It is not possible to pursue the *house/home* theme from the CEFR directly into the EP wordlists, but the wordlist theme of 'buildings' is closely related and identifies the following words at each CEFR level (Figure 4).

A1	A2	B1	B2
<ul style="list-style-type: none"> •address •bath •bathroom •bed •bedroom •chair •dining room •door •floor •garden •key •kitchen... 	<ul style="list-style-type: none"> •apartment •armchair •barbecue •blanket •board •<u>bookcase</u> •bookshelf •building •carpet •ceiling •closet •curtain... 	<ul style="list-style-type: none"> •accommodation •air conditioning •architecture •balcony •basin •bin •block •candle •central heating •chest of drawers •cottage •cupboard... 	<ul style="list-style-type: none"> •barrier •basement •bench •brick •cabinet •cell •cellar •chimney •cloth (for cleaning) •column (building) •corridor •domestic...

Figure 4 EVP entries for 'Buildings' (www.englishprofile.org)

This brings us to the level of the individual word or phrase. Here is the entry for 'bookcase':

A2 **bookcase** /'bʊk.keɪs/

► **NOUN** [C] *a piece of furniture with shelves for putting books on*

Dictionary example:

We need a new bookcase.

🗣️ Learner example:

I left my passport at your house in the living-room on the bookcase, and my flight will be tomorrow evening, so I can't get my plan[e] if I don't have it.

Key English Test; A2; Arabic

Figure 5 EVP entries for 'Buildings' (www.englishprofile.org)

As illustrated by the word *bookcase* in Figure 4, the entry for each word in the list includes a definition, a dictionary example and, where available, an example taken from a learner's written work drawn from the relevant level of the CEFR.

Links can currently be made between the CEFR and the *Threshold* series specifications, although again there is scope for making these more direct and explicit in the English Profile. The theme of *personal house/ home* in the CEFR is reflected in the *house and home, environment* theme in Vantage (van Ek and Trim 2001, p.74). This theme introduces a number of related Can Do statements involving tasks based around the function of *describing* (Figure 6).

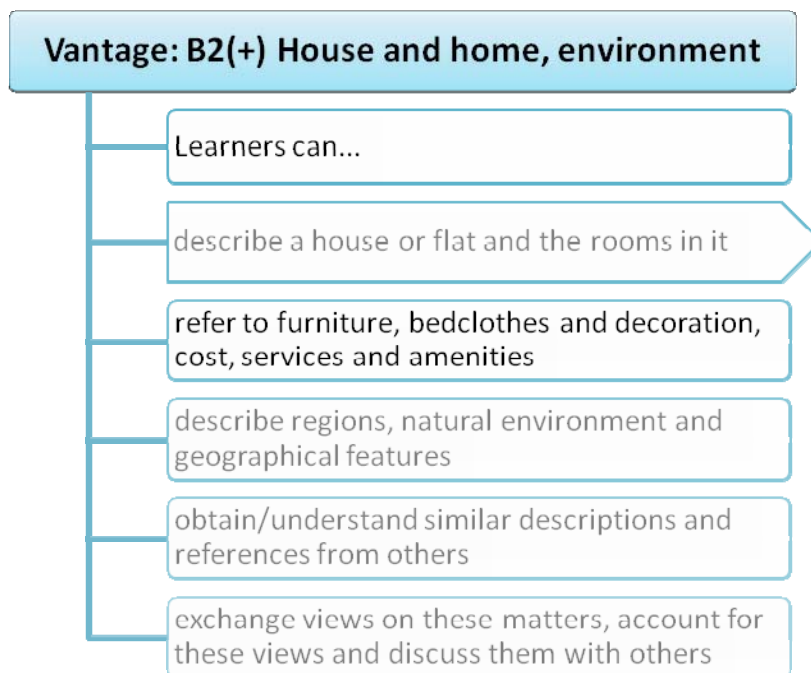


Figure 6 The theme of house, home and environment in *Vantage* (van Ek and Trim, 2000, p.

The grammatical resources that learners might be expected to draw on in realising these tasks can be traced in the listing of functions and linguistic exponents. *Describing* is linked in the *Threshold* series with *identifying* and *reporting* under the functional heading of *imparting and seeking information* (van Ek and Trim 2001, p.29). Examples of grammatical exponents, exemplifying the progression in these for the functions of *identifying (and specifying)* and *(stating and) reporting* are shown in Table 2 for the Breakthrough (A2) and Vantage (B2 and above) levels.

Table 2 identifying and reporting in *Breakthrough* (Trim 2009, p.20) and *Vantage* (van Ek and Trim, 2000, p.29)

Breakthrough	Vantage
1.1 identifying (with pointing gesture) (an object) this one, that one, these, those (a person) me, you, him, her, us, them (where pointing impossible) (a person) It + BE + me/ you/ him/ her/ us/ them. (a person or object) It + BE + NP (noun phrase). Examples This is the key	1.1 identifying and specifying (with indicating gesture, e.g. pointing, nodding) (an object) this (one)/ that (one)/ these/ those; (a person) me/ you/ him/ her/ us/ them <i>the</i> (adj) <i>one</i> + adjunct phrase/relative clause It's me you/ him/ her/ us/ them/ NP Pronoun/NP + BE +NP Examples the small one with the blue curtains

It is John's garden	Her office is at the end of the corridor on your left. This is the largest bedroom in the house.
1.2 reporting (describing and narrating) declarative sentences within the learner's grammatical and lexical competences (see 9.2.1) NB. This limitation applies wherever <i>declarative sentence</i> is specified.	1.2 stating and reporting (describing, narrating) (sequences of) declarative sentences NP+ say, think + complement clause NP+ ask/ wonder + indirect question there + be +NP + adjunct Examples The train has left. He says the shop is shut. He asked where they were going. There is a bank on the corner. There is a cow in our garden eating the plants.

The CLC informed work of Hawkins and Filipovic (2012) serves as a check on the progression in general notions and associated grammatical patterns mooted in the *Threshold* series. They note the importance of lexical triggers for grammatical patterns, but with this proviso generally find the *Threshold* series predictions are borne out by evidence from the learner data. This suggests that a flexible, searchable development of the *Threshold* specification might be a useful way to integrate the EVP, EGP and EFP at the B1 level. It is already possible to pursue domains, themes and topics in the ways outlined above. In future, a major contribution of the English Profile will be to bring together the available information in a more accessible and integrated form, to incorporate emerging findings and to provide a forum for debate on their implications.

To help build the consensus around the levels, it would no doubt be helpful for users to see additional material, such as more extended examples of learner production related to themes of this kind, and to share comments on how these relate to aspects of the level descriptions.

From the user perspective, it would be helpful to trace the language processes brought into play by a given task. The concept of illustrative performances can be exemplified through a task type that is well represented in the CLC: reviews of books and films. The following is an extract from a review that has been judged to be of B2 level.

Macbeth is the famous play by William Shakespeare and very exciting and dramatic. Someone like Macbeth who kills many people should be named bad but is this true?

Macbeth wants to become king and thinks he must kill the king to take his place. Furthermore his wife lady Macbeth is very strong and wants him to murder the king and is angry when she thinks he is so weak to do it. Macbeth is influenced by her almost and listens to her plan to kill the king. Later he murders the king and two guards but feels mad afterwards. ...

The EFP can be used to describe the rhetorical moves that the reviewer makes in constructing the review text:

Moral evaluation: 'should be named bad'

Raising doubt/ uncertainty: 'but is this true?'

The EVP can be used in classifying the use of the vocabulary employed in the review and the ways in which this contributes to its communicative function:

Someone like Macbeth who **kills** many people should be **named**, bad but is this **true**?

kill - verb - **A2**; **name** - noun - REPUTATION – **B2**; **name** – verb - SAY WHAT NAME IS – **B1**; **true** - adjective - NOT FALSE - **A2**; **bad** - adjective - UNPLEASANT - **A1**

The EGP can be employed to explore its grammatical features:

Someone like Macbeth **who kills many people** should be named bad but is this true?

Defining and non-defining relative clauses - **B2**; **Modal should** – **B1**; **Passives** – **B2**

Observing how these features interact in the text, can help users to build up a picture of what it is that informs the B2 evaluation and to make effective use of the profiles in their own practice.

The EPN provides an extensive and constantly growing library of such sample performances that can serve a range of purposes including familiarisation with the levels and the categories of description, communication between stakeholders working in different contexts and with different populations of learners as well as supporting ongoing research into learner language.

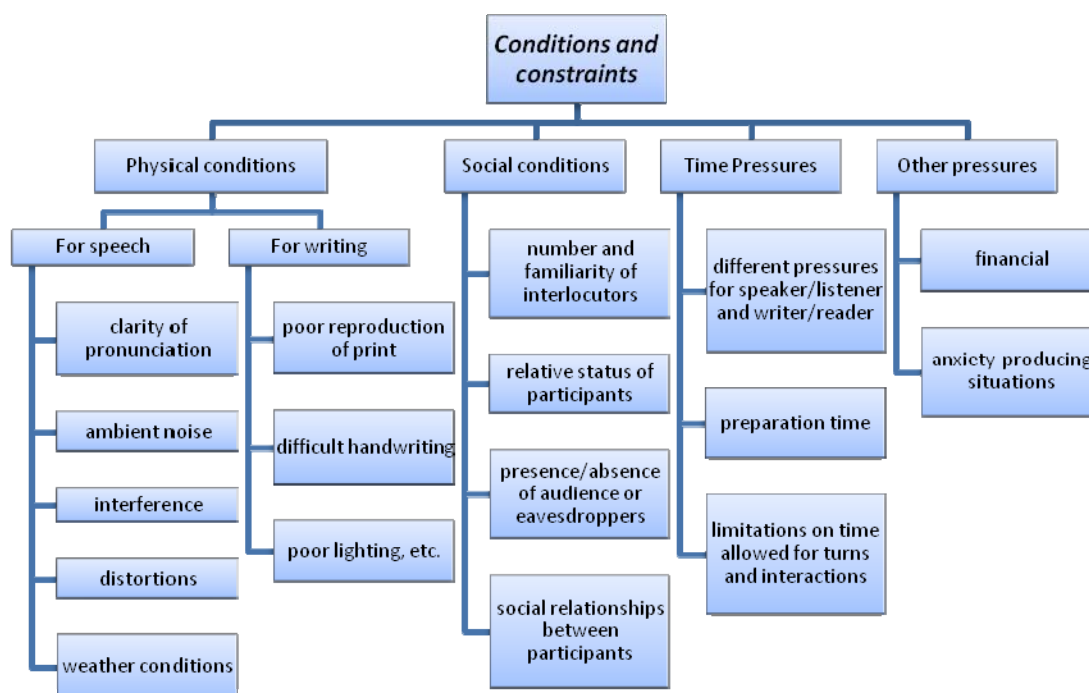


Figure 7 Conditions and constraints listed in the CEFR (CoE, 2001, pp.46-7)

Other elements in the CEFR model are not as well covered in existing materials as themes and topics and there is scope for more pioneering work. The ‘conditions and constraints’ (Figure 7) listed in the CEFR is one such area. These include *physical* and *social conditions* together with *time* and *other (financial and anxiety-producing) pressures*. While there is research available that addresses the impact of features such as preparation time or numbers of interlocutors on performance on certain kinds of task (typically test tasks in which these can be more easily controlled), this has served to

indicate just how complex such issues can be. More information on the impact of a range of contextual variables on performance is urgently needed and the English Profile should act as a spur to further research in this area.

Requirement 4: Lessons for RLD development

The value of RLDs depends in large measure on their clarity and usability for educators. As an outcome of the EPN, it is hoped that the CEFR itself can become a more useful tool for its intended purposes in relation to English. To achieve this, the interpretation of CEFR descriptors must be made more accessible to users and meet their everyday practical needs. This implies extensive field testing at every stage of the Profile tools and adaptation in the light of feedback.

The EVP has shown the importance of a dynamic approach to development, capable of adjustment in the light of new research evidence and in response to learner requirements. With this in mind, the choice of interactive, multi-media online presentation of outcomes is particularly suitable. Content can be constantly updated while access can be made flexible enough to meet the needs of different user groups. Future RLDs should consider the long term and present their outcomes in ways that are flexible enough to allow for ongoing changes to be made.

The EPN has proved invaluable as a vehicle for extending data collection, disseminating the Profiles and obtaining meaningful feedback from researchers and practitioners. Developers of RLDs for other languages would be well-advised to look for similar opportunities to build networks of collaborators who could contribute research expertise and practical cooperation.

Overall lessons and principles derived from the EPN experience

RLDs are an opportunity to bring research into language learning on the one hand and educational practice on the other into a productive relationship that can benefit both fields. Educational practice benefits the researcher because it can provide substantial quantities of data, located in relation to a coherent proficiency framework (the CEFR). Research benefits the practitioner because it helps to clarify objectives and provides evidence of how learners increase their functional abilities. RLDs based on empirical data have the potential to advance both our theoretical understanding of language acquisition and our ability to make use of this understanding in advancing learning.

Socio cognitive approach

The socio-cognitive approach, developed first by Weir (2005, Shaw and Weir 2007 Khalifa and Weir 2009 and Taylor 2011) and adopted by the EPN, substantially facilitates the development of a coherent research agenda for the project as a whole, and underpins the guidance it can offer for language teaching and assessment practice. This does not imply the use of any one specific research methodology, but if the research agenda is to serve the needs of teachers, it must take account of the social uses that language learners need in the real world as well as the psycholinguistic processes that they engage in. Linguistic forms are usually of central concern to researchers into second language acquisition, but function is central for teachers and learners. RLDs are directed at identifying the linguistic resources that learners may be able to bring to bear when they carry out tasks through language.

Empirical data

The collection of a learner corpus is clearly central to the development of any RLDs that follow the direction taken by EPN, but other forms of corpus building are also important. Representative corpora reflecting proficient language use are also vital as a point of reference and to provide indications of the language that learners are likely to encounter outside the classroom. Corpus resources are now available for many languages, but may not have all the features that the RLD developers might wish. The EPN has taken the pragmatic view that it is better to use what is available in the first instance, rather than delaying RLD development until a more tailored resource can be assembled.

Expert judgement

Although empirical data is important to the EPN, developing RLDs is not simply a research project. The products must communicate to a wide, non-specialist audience the best understanding we have managed to construct of the nature of the CEFR levels for English. This understanding is not straightforwardly provided by the data. Often the evidence that is available needs careful interpretation and elucidation before it can be useful for educators. In some cases, the evidence may appear contradictory. In others there is simply no evidence available. Over time, we may be able to fill the gaps, but in the interim, teachers and learners need the best guidance we can provide. Judgement and expertise is needed to build useful and comprehensive pedagogic tools from the partial and imperfect picture provided by our research.

Networking across disciplines

The EPN has involved an interdisciplinary team of researchers and a wide range of language teaching, publishing and testing organisations. The involvement of educators is important both to the collection of data, to ensuring the value of the Profiles that emerge and to disseminating outcomes. Building up a network of organisations that are willing to collect data, make experimental use of emerging tools and inform others of their value has been a major achievement. Feedback from users on their experiences with the emerging tools has contributed significantly to their quality.

Similarly, bringing together a network of researchers working in different fields has been fruitful. Opportunities for sharing research within the network, but across disciplinary boundaries has led researchers to adopt new perspectives on their work. Equally, researchers have benefitted from being able to interact directly with the educators who will make use of their findings. Setting up regular meetings and seminars where this sharing can occur has been important to the success of the EPN.

Technology and communication

The EPN has of course made extensive use of technology in its research methods: exploiting corpora, automated parsing, machine learning and statistical analysis techniques. Technology has also greatly facilitated data collection through the online portal and the ability to gather digital texts has reduced the burden of transcription. RLDs for other languages will, no doubt, benefit from similar tools and from further technological developments.

It is clear that technology can also offer promising opportunities for more useful and engaging presentation of RLDs. The EVP can be searched in a variety of ways, audio pronunciation guides are provided and there are options for filtering content. This is clearly a valuable step forward from the printed tables of the *Threshold* series and further possibilities for exploiting the EVP and linking it to

other Profiles are sketched out above. Developers of RLDs for other languages may consider adapting the interface used for the English Profile to their own contexts.

Continuity and innovation

A valuable legacy from the EPN to the development of all future RLDs would be to establish that RLDs are never complete, but can and should be open to constant refinement and revision. There is always more to be learned about the target language and how it is learned. More than that, the target language itself is constantly changing as are the people who seek to learn it. There is a constant need to keep pace with language change, technological change, theoretical developments and the shifting needs of users. Again, a flexible online approach to publication makes it easier to adapt quickly to a dynamic context.

References and additional sources on English Profile methods

Alexopoulou, T. (2008). Building new corpora for English Profile, *Research Notes*, 33: 15–19, Cambridge: Cambridge ESOL.

Briscoe, E., Carroll J. and Watson R. (2006). The second release of the RASP system. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*. Sydney, Australia.

Briscoe, E., J. Carroll and R. Watson (2006) *The Second Release of the RASP System*. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*, Sydney, Australia

Capel, A. (2009). A1–B2 vocabulary: Insights and issues arising from the English Profile Wordlists projects. Paper presented at the English Profile Seminar, Cambridge, 5–6 February 2009.

Capel, A. (2010). Insights and issues arising from the English Profile Wordlists project, *Research Notes*, 41: 2-7. Cambridge: Cambridge ESOL.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Council of Europe (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual*. Strasbourg: Language Policy Division.

Green, A. (2008). English profile: Functional progression in materials for ELT. *Research Notes*, 33: 19–25. Cambridge: Cambridge ESOL.

Green, A. (2012). *Language Functions Revisited: Theoretical and Empirical Bases For Language Construct Definition Across the Ability Range*. Cambridge: UCLES/Cambridge University Press.

Hawkins, J. A. and Buttery, P. (2009). Using learner language from corpora to profile levels of proficiency: Insights from the English Profile Programme. In Taylor, L. and Weir, C. J. (Eds). *Language Testing Matters: Investigating the Wider Social and Educational Impact of Assessment*, 158-175. Cambridge: Cambridge University Press.

Hawkins, J. A. and Buttery, P. (2010). Criterial features in learner corpora: Theory and illustrations, *English Profile Journal* 1 (1).

Hawkins, J. A. and Filipovic, L. (2012). *Criterial Features in L2 English*. Cambridge: UCLES/Cambridge University Press.

Hendriks, H. (2008). Presenting the English Profile Programme: In search of criterial features. *Research Notes*, 33: 7–10. Cambridge: Cambridge ESOL.

Khalifa, H. and Weir, C.J. (2009). *Examining Reading: Research and practice in assessing second language reading*. *Studies in Language Testing* 29, Cambridge: UCLES/Cambridge University Press.

Kurtes, S. and Saville, N. (2008). The English Profile Programme – An overview. *Research Notes*, 33: 2–4. Cambridge: Cambridge ESOL.

- McCarthy, M. (2010). Spoken fluency revisited, *English Profile Journal* 1 (1).
- Michael Fleming (2009) The use of descriptors in learning, teaching and assessment. DG IV / EDU / LANG (2009) 21. Strasbourg: Language Policy Division.
- North, B. (2000). *The Development of a Common Framework Scale of Language Proficiency*. New York: Peter Lang.
- North, B. (Ed.) (1992). *Transparency and Coherence in Language Learning in Europe: Objectives, Assessment and Certification*. Symposium held in Ruschlikon, Switzerland, 10–16 November 1991. Strasbourg: Council for Cultural Cooperation.
- O’Sullivan, B. and Weir, C. (2010). Test Development and Validation. In O’Sullivan, B. (Ed.) *Language Testing: Theory and Practice*. Oxford: Palgrave
- Parodi, T. (2008). L2 morpho-syntax and learner strategies. Paper presented at the Cambridge Institute for Language Research Seminar. Cambridge, 8 December 2008.
- Salamoura, A. (2008). Aligning English Profile research data to the CEFR. *Research Notes*, 33: 5–7. Cambridge: Cambridge ESOL.
- Salamoura, A. and Saville, N. (2009). Criterial features across the CEFR levels: Evidence from the English Profile Programme. *Research Notes*, 37: 34–40. Cambridge: Cambridge ESOL.
- Salamoura, A. and Saville, N. (2010). Exemplifying the CEFR: Criterial features of written learner English from the English Profile Programme. In Bartning, I., Maisa, M. and Vedder, I. (Eds). *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*, *Eurosla Monographs Series* (1), 101-132.
- Saville, N. and Hawkey, R. (2010). The English Profile Programme - the first three years, *English Profile Journal* 1 (1).
- Shaw, S. and Weir, C. J. (2007). *Examining Second Language Writing: Research and Practice*. In *Studies in Language Testing*. Volume 26. Cambridge: UCLES and Cambridge University Press.
- Taylor, L. (Ed.) (2011). *Examining Speaking: Research and practice in assessing second language speaking*. *Studies in Language Testing* 30, Cambridge: UCLES/Cambridge University Press.
- van Ek, J. and Trim, J. L. M. (1990a/1998a). *Threshold* 1990. Cambridge: Cambridge University Press.
- van Ek, J. and Trim, J. L. M. (1990b/1998b). *Waystage* 1990. Cambridge: Cambridge University Press.
- van Ek, J. and Trim, J. L. M. (2001). *Vantage*. Cambridge: Cambridge University Press.
- Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Oxford: Palgrave.