

Developmental stages in second-language acquisition and levels of second-language proficiency: Are there links between them?

Jan H. Hulstijn, J. Charles Alderson and Rob Schoonen
University of Amsterdam / Lancaster University / University of Amsterdam

The papers in this volume were written by European researchers loosely organized in the SLATE group: Second Language Acquisition and Testing in Europe (<http://www.slate.eu.org/>). The group was formed after two meetings, held in 2006, at the University of Amsterdam, organized by Jan Hulstijn, Charles Alderson and Rob Schoonen¹. This introductory chapter describes the rationale for the SLATE group's ambitions.

The first meeting, sponsored by the European Science Foundation (Ref: EW05-208-SCH) and held on 23-25 February 2006 at the University of Amsterdam, was called *Bridging the gap between research on second-language acquisition and research on language testing*. It brought together 19 researchers, based at eight universities in seven European countries, working in the fields of second language acquisition and language testing. The follow-up meeting, which was sponsored by the Netherlands Organisation for Scientific Research (NWO) (grant 254-70-100) took place 1-2 December in the same year and at the same venue. This meeting, attended by 20 researchers, affiliated to ten universities in eight European countries, was called *Stages of second-language acquisition and the Common European Framework of Reference for Languages*. The titles of these two SLATE meetings characterize the aim and the focus of the SLATE group. After a brief description of the Common European Framework (CEFR), this chapter shows how the SLATE group envisioned research, linking developmental stages in SLA with L2 proficiency levels as defined by the CEFR.

1. The Common European Framework (CEFR)

The Common European Framework of Reference for Languages (Council of Europe, 2001, in English and French; thereafter translations into 31 languages so far, including Arabic, Friulian and Japanese) contains proposals for formulating

¹ An earlier meeting had taken place in 2004, also in Amsterdam. Some of the people present at the two SLATE meetings in 2006 had also attended the 2004 meeting.

functional learning targets for language learning, teaching and assessment. Throughout Europe (and beyond), the CEFR has become a major point of reference for language in education, with both the ambition and the potential of bringing common standards and transparency across Europe. It has also become the point of reference for the formulation of objectives of foreign-language learning curricula and the certification of foreign-language proficiency skills of citizens continuing their educational or professional careers in other European countries.

The CEFR presents language and language learning within its social context, sees language users and learners as “social agents”, and in general advocates a notional/ functional, “sociolinguistic” approach to language use and therefore to (second/foreign) language development. The framework, distinguishing six “common reference levels” (Chapter 3), adopts a multifaceted approach to the concept of language proficiency. Acquisition of an L2 is said to be a matter of development along what is called a horizontal and a vertical dimension (Council of Europe, 2001, pp. 16–17).

The *horizontal dimension* specifies the “language activities” in which language users engage (pp. 44–57) in terms of (1) context of language use (e.g., in the personal and professional domain), (2) communication themes (e.g., travel and health), and (3) communicative tasks and purposes (e.g., making enquiries from employment agencies and reading safety instructions). Chapter 4 of the CEFR contains 40 scales specifying a large number of forms of oral and written language use (pp. 58–84), including several scales of strategic competence. The horizontal development also comprises dimensions of more general “communicative language competences” (p. 108), subdivided – as in the language proficiency model of Canale and Swain (1980) – in linguistic, sociolinguistic and pragmatic competences. Chapter 5 contains 13 scales for these competences (pp. 110–129).

The *vertical dimension*, outlined in Chapter 3 and applied to all the descriptor scales in Chapters 4 and 5, consists in “an ascending series of common reference levels for describing learner proficiency” (p. 16). The authors caution that “any attempt to establish ‘levels’ of proficiency is to some extent arbitrary, as it is in any area of knowledge or skill. However, for practical purposes it is useful to set up a scale of defined levels to segment the learning process for the purposes of curriculum design, qualifying examinations, etc.” (p. 17).

The development of communicative language competence can thus be seen both at the level of expanding one’s range of communicative activities and at the level of performing them in increasingly more complex and sophisticated ways.

It is important to emphasise that the 2001 version of the CEFR itself did not suddenly appear out of nothing. It is the natural result, an organic development, of the work in modern languages of the Council of Europe over three decades, from the initial development of notional-functional syllabuses (Wilkins, 1976.)

and *The Threshold Level* (van Ek, 1975), the latter being widely translated and revised for use in numerous other languages (*Niveau Seuil*, *Kontaktschwelle*, etc.), and then the later elaboration of further levels of the CEFR including a new version of *Threshold* (van Ek & Trim, 1998a), *Breakthrough* (Trim, 2001), *Waystage* (van Ek & Trim, 1998b), *Vantage* (van Ek & Trim, 2001) and the other levels published in the 1980s and 1990s. It can be argued that behind the CEFR of 2001 (and the two draft versions of 1996 and 1998) lies 30 years of experience in developing and implementing curricula, syllabuses and teaching materials at different “levels” of foreign language development. What is, however, in considerable doubt is whether the extensive use of such previous theoretical and practical perspectives was accompanied by empirical research into the constructs and claims of the curricula and their associated notions of level and development. What is perhaps most significant about the CEFR is that, unlike its predecessors, it was accompanied by numerous scales of foreign language development, which were the result of significant empirical research.

Although the bulk of the CEFR is the so-called Descriptive Scheme, which takes up the nine main chapters of the CEFR, most impact, based on the admission of the Council of Europe itself as a result of surveys it has recently conducted (Council of Europe, 2005), seems to be due to the various scales that were developed in parallel to the Descriptive Scheme. This is doubtless in part due to the fact that much of the Descriptive Scheme is couched in rather dense, academic and at times frustratingly opaque terms of little appeal to the likely average readership, whereas the scales are not only shorter and in more accessible language, they are also couched in ‘Can-do’ terms, since they indicate what learners at given “levels” are believed to be able to do. Indeed, the evidence is that the vast majority of those who claim familiarity with or knowledge of (not the same thing) the CEFR are referring to the scales, or at the very least, to the labels by which the six major levels are identified: A1, A2, B1, B2, C1, C2. In point of fact, as is clearly shown whenever CEFR familiarisation activities are conducted with language educationists (as recommended by the Manual for linking examinations to the CEFR, Council of Europe, 2009), detailed knowledge of even these six levels is frequently defective and uninformed. Given the complexity of the CEFR itself (not to mention the broad field of foreign language learning and use) it is not surprising that even those closely associated with the development of the CEFR and its implementation in various contexts do not claim complete knowledge or understanding of all that it contains.

A total of some 56 scales are contained within the CEFR publication, and these cover the macro skills (called “Communicative Activities” in the 1996/8 version of CEFR) of Reception (both written and spoken), Interaction, and Production (the macro skill/ activity of Mediation is not accompanied by any scales). In addition, scales of Strategies are subdivided into the same three macro

activities/skills, and scales of “Communicative Language Competence” are divided into Pragmatic (with separate scales of spoken fluency, flexibility, coherence and precision) and Linguistic (Range – both general and vocabulary range - and Control - grammatical accuracy, vocabulary control, phonological control and orthographic control).

Although the scales presented in the 2001 publication are in some sense a new compilation, it is important to be aware that their origin is very heterogeneous, and indeed Appendix B to the CEFR details this: a total of 30 sets of scales, from the USA, Canada, Australia, the UK, and Europe, dating from 1974 to 1993, including well-known scales like

- The Foreign Service Institute (FSI) and the Interagency Language Roundtable (ILR) Proficiency Ratings,
- The Australian Second Language Proficiency Ratings,
- The American Council on the Teaching of Foreign Languages (ACTFL) Proficiency Guidelines,
- The IELTS Band Descriptors for Speaking and Writing,
- The British Languages Lead Body National Language Standards
- The University of Cambridge/ Royal Society of Arts Certificates in Communicative Skills in English,

as well as lesser known scales like

- The Hebrew Oral Proficiency Rating Grid,
- Goteborg University Oral Assessment Criteria
- Fulcher’s Fluency Rating Scale.

Thus, in a very important sense, the CEFR scales themselves are *not* new: they are based upon decades of experience in building, using and, presumably, refining scales in the light of experience.

Secondly, it is also important to remember that the original scales were subjected to extensive analysis and deconstruction into pools of classified descriptors, and critical scrutiny, review, revision, adjustment and empirical testing before they resulted in their current form (see Appendix to 1996 CEFR draft 1996, p. 152ff, and CEFR 1998 draft p. 163ff for details). In practice, well over a thousand descriptors were taken from the source scales, filtered for overlap, clarity and focus, sorted by large groups of language teachers (basing themselves on their experience as both learners and teachers) into the categories the descriptors purported to describe, then critiqued for their clarity, accuracy and relevance and sorted into bands of proficiency. Many of these descriptors were applied to learners in the classes of participating teachers.

This exercise was repeated in a second round, this time with teachers of French and German as well as English, who inspected versions of the scales in

all three languages. The resulting draft scales and descriptors were then subject to Rasch analysis to determine their probabilistic rank order of “difficulty”. Those descriptors that ‘fit the model’ i.e. which could be scaled mathematically, were then calibrated, and those that could not be scaled were rejected (this was particularly the case with draft scales for socio-cultural competence).

The production of the scales was thus an extensive empirical exercise (resulting, in part, in a PhD thesis by North, 1996; see also North & Schneider, 1998). It is fair to say that the resultant scales are probably the best researched scales of foreign language in the world, although perhaps not (yet) the most widely used – that award probably goes to one of the source scales, the FSI, ILR, ACTFL family of scales.

It is important to separate consideration of the scales and the Descriptive Scheme of the CEFR from the use of either. It is clear that the CEFR and especially the scales have had and will continue to have enormous influence. Many ministries of education now require that their national exams be at one or more of the levels of the CEFR. Many institutions (including many universities and examination boards) claim that their exams are at particular levels of the CEFR. However, very few of these have produced any empirical evidence whatsoever that this is indeed the case. Although the Council of Europe has issued a manual for linking exams to the CEFR (Council of Europe, 2009; see also Figueras, North, Takala, van Avermaet, & Verhelst, 2005) and is currently encouraging the development of case studies in the use of the CEFR (Figueras & Noijons, 2009), there is no doubt that most claims of links lack appropriate empirical support. Similarly, many institutions, especially higher education colleges and universities, claim that their curricula are aligned to the CEFR and their graduates are at a given level of the CEFR. Again, there is virtually no empirical evidence that this is the case (see survey by Alderson, 2007a). Finally, some authors have pointed out that the CEFR levels are neither based on empirical evidence taken from L2-learner performance, nor on any theory in the fields of linguistics or verbal communication (Alderson, 2007b; Hulstijn, 2007), while some language testing experts have raised concerns with respect to the suitability of the CEFR for language testing purposes, arguing that the CEFR, in its current form, lacks specificity and consistency in its definitions and is insufficiently supported by empirical research to allow immediate implementation (Weir, 2005; Alderson et al., 2006).

Nevertheless, the scales have been used as the basis of developments in numerous contexts, even before they were published in their 2001 format. The DIALANG project (<http://www.dialang.org/>; Alderson, 2005; Alderson & Huhta, 2005) was one of the first language testing projects to use the CEFR scales as the basis of the specifications for their tests in 14 European languages, and the project translated these scales from English into the other 13 languages

through a careful quality control process, as well as making statistical comparisons during piloting to ensure the equivalence of the translations (Alderson, 2005; Kaftandjieva & Takala, 2002 report on these equivalences). In addition, since DIALANG contained tests of Grammar and Vocabulary, attempts were made to construct scales for those constructs for all 14 languages.

2. Calls for CEFR related acquisition research

Although there is still much work to be done to validate CEFR scales and their links to curricula and other statements regarding the development of communicative proficiency, we must equally recognise that especially SLA, but language assessment as well, have to date operated with notions of development and levels of development which have all too frequently been hopelessly imprecise. SLA, for example, has frequently simply taken groups of learners at supposedly different levels of ability, conducted cross-sectional research and claimed that the results show development. Yet the levels have been woefully undefined, often crudely labelled “intermediate” or “advanced”, or “first and second year university students” – which means little if anything in developmental terms – and which cannot therefore be interpreted in any meaningful way. It is our belief that, whatever its shortcomings, the CEFR has introduced a notion of levels of development that is far better – if only because it can be challenged – than the vague terms (not measures) used to date.

It should also be borne in mind that the CEFR is intended to be relevant to all languages, and is therefore not language specific, which means that most of the descriptor scales (those appearing in Chapter 4) concern the development of language-independent communicative competencies. Chapter 5 contains a few scales on the development of linguistic areas such as phonology, lexicon and grammar, but these are among the most problematic ones. The need for such scales to be language-independent, and thus be applicable to languages as different as Spanish, German and Finnish, makes them appear little more than a list of generic statements about growing accuracy and/or complexity in each linguistic domain.

A number of projects have begun to attempt to identify the features of particular languages at different CEFR levels - *Référentiel de Français Langue Étrangère*, *Profile Deutsch* (Glaboniat, Müller, Rusch, Schmitz, & Wertenschlag, 2005), *English Profile* (<http://www.englishprofile.org/>), but a number of theoretical and empirical questions remain concerning the componential structure of language proficiency at various CEFR levels. The CEFR authors acknowledge that language users may be placed at different levels of different scales, possessing “uneven profiles”: “Progress is not merely a question of moving up a vertical scale” (Council of Europe, 2001: 17). Furthermore, what the CEFR does not

indicate is whether learner performance at the six functional levels as defined in Chapter 4 actually matches the linguistic characteristics defined in Chapter 5, and, more specifically, *which linguistic features (for a given target language) are typical of each of the levels.*

According to the SLATE group, the question of which linguistic-communicative profiles could, and which profiles do, exist constitutes an empirical issue deserving to be investigated properly. Answers to these questions are essential if the CEFR is to be successfully implemented across Europe. Furthermore, as the CEFR is intended to be relevant to all and any language, it does not provide information for specific target languages, much less for specific first-second language combinations. In short, the CEFR is not yet capable of successful implementation. Additional research, linking functional and linguistic information, is urgently needed.

3. SLATE's overarching research question

At the first SLATE meeting in 2006, the following research question was seen as central to a collaborative enterprise: *Which linguistic features of learner performance (for a given target language) are typical at each of the six CEFR levels?* Initial investigation of this question can be conducted in at least two ways:

- Learner performance data that have been elicited and analysed in previous SLA research can be additionally rated with functional rating scales based on the scales presented in Chapter 4 of the CEFR (Council of Europe, 2001).
- Responses of learners who took language proficiency tests related to the CEFR (i.e., data already collected by language testers) can also be rated with functional CEFR-based scales and analysed linguistically.

However, for a more thorough investigation of this research question, new, international, cross-linguistic studies need to be conducted using a *common* research design, tasks, procedures, and analyses.

4. SLATE's more specific research questions and research goals

During the first SLATE meeting, the following points were mentioned as research questions and issues of *potential* relevance.

1. What are the linguistic profiles at every CEFR level for the two productive language skills (speaking and writing) and what are the linguistic features typical of the two receptive skills (listening and reading) at every CEFR level?

2. To what extent do common or different profile features exist across the seven target languages investigated by researchers in the SLATE group (Dutch, English, Finnish, French, German, Italian and Swedish)? Do the profiles differ along language-family lines - Finnish versus the two Romance languages represented in SLATE (French and Italian), and the three Germanic languages (Dutch, English and German)? To what extent do the profiles reflect learners' L1?
3. What are the limits of learners' performance of tasks at each of the CEFR levels? It is important not only to investigate what learners typically *do* at each of the CEFR levels (which requires elicitation of performance in rather "open" task formats), but also what learners *can* and *cannot do* (which requires the administration of tasks of a "closed" format, with and without time pressure, and the measurement of both accuracy and reaction times). In other words, in order to find features typical of a given CEFR level, we need to explore borderline features and to identify features shared by two or more levels and to construct appropriate scales.
4. Which linguistic features, emerging from our profiling research, can serve as successful tools in the diagnosis of learners' proficiency levels and of weaknesses that require additional attention and training? The investigation of this question is of considerable practical importance for all stakeholders (learners, teachers, curriculum and test designers and other language educationists). One desired outcome of investigation into this question would be the development of diagnostic tools as well as information which could be included in learners' language portfolios (Alderson, 2005).
5. Are there commonalities and differences between the linguistic profiles of foreign-language learners (learning the target language in the formal setting of a school curriculum or a language course) and those of second-language learners (learning the target language in the context of everyday use, with or without formal instruction)? This question is not only of theoretical importance but also of practical importance. For example, a feature such as (un)successful use of subject-verb agreement might have a predictive value different for foreign language learners from the value for second language learners.

At the second SLATE meeting (December 2006), additional targets were formulated. Some of these deserve to be reproduced here (quoted from the report of the second meeting):

- A framework for the linguistic analysis of learner performances at different CEFR levels.

- Suggestions for improvements of the CEFR and/or for extending it for use with young learners, L1 speakers, LSP (language for specific purposes) students, and the like.
- New knowledge that will help to design new teaching materials, curricula and diagnostic assessment instruments relating to the linguistic features that characterise different CEFR levels.

The papers in this volume, written by members of the SLATE group, resulted, either directly or indirectly, from the 2006 meetings in Amsterdam and later meetings in Aix-en Provence and Jyväskylä, demonstrating the fruitful nature of the issues raised and discussed.

References

- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Alderson, J. C. (2007a). The challenge of diagnostic testing: Do we know what we are measuring? In J. Fox et al. (Eds.), *Language Testing Reconsidered* (pp. 21–39). Ottawa: University of Ottawa Press.
- Alderson, J. C. (2007b). The CEFR and the need for more research. *The Modern Language Journal*, 91, 658–662.
- Alderson, J. C., & Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing*, 22(3), 301–320.
- Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2006). Analysing tests of reading and listening in relation to the Common European Framework of Reference: The experience of the Dutch CEFR Construct project. *Language Assessment Quarterly*, 3(1), 3–30.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1–47.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe. (2005). *Survey on the use of The Common European Framework of Reference for Languages (CEFR): Synthesis of results*. Strasbourg: Council of Europe.
- Council of Europe. (2009). *Relating language examinations to the Common European Framework of Languages: Learning, teaching, assessment. A Manual*. Strasbourg: Council of Europe.
- Figueras, N., & Noijons, J. (Eds.). (2009). *Linking to the CEFR levels: Research perspectives*. Arnhem: CITO/ EALTA
- Figueras, N., North, B., Takala, S., van Avermaet, P., & Verhelst, N. (2005). Relating examinations to the Common European Framework: a manual. *Language Testing*, 22(3), 261–279.

- Glaboniat, M., Müller, M., Rusch, P., Schmitz, H., & Wertenschlag, L. (2005). *Profil Deutsch*. Frankfurt: Langenscheidt.
- Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal*, 91, 663–667.
- Kaftandjiewa, F., & Takala, S. (2002). Council of Europe scales of language proficiency: a validation study. In J. C. Alderson (Ed.), *Common European Framework of Reference for Languages: Learning, teaching, assessment. Case studies* (pp. 106–129). Strasbourg: Council of Europe.
- North, B. (1996). *The development of a common framework scale of descriptors of language proficiency based on a theory of measurement* (Unpublished doctoral dissertation). London: Thames Valley University.
- North, B., & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing*, 15, 217–262.
- Trim, J. L. M. (2001). *Breakthrough*. Unpublished manuscript.
- Van Ek, J. A. (1975). *The Threshold Level*. Strasbourg: The Council of Europe.
- Van Ek, J. A., & Trim, J. L. M. (1998a). *Threshold 1990*. Cambridge: Cambridge University Press. (Original work published 1991).
- Van Ek, J. A., & Trim, J. L. M. (1998b). *Waystage 1990*. Cambridge: Cambridge University Press. (Original work published 1991).
- Van Ek, J. A. & Trim, J. L. M. (2001). *Vantage*. Cambridge: Cambridge University Press.
- Weir, C.J. (2005). Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing*, 22(3), 301–320.
- Wilkins, D.A. (1976). *Notional syllabuses*. Oxford: Oxford University Press.