



## ***POLICY FORUM***

### **The *Common European Framework of Reference for Languages (CEFR)* and the development of language policies: challenges and responsibilities**

*Organised by the Council of Europe Language Policy Division*

**Strasbourg, 6 – 8 February 2007**

*In honour of John L. M. Trim,  
Former Director of Council of Europe Modern Languages Projects*

### **The CEFR Common Reference Levels: validated reference points and local strategies**

**Brian North**

**6 February 2007**

This paper briefly addresses the following questions:

**What is the purpose of the CEFR?** The CEFR is a concertina-like reference tool that provides categories and levels that educational professionals can expand or contract, elaborate or summarise, according to the needs of their context. The aim is for users to adopt activities, competences and proficiency stepping-stones that are appropriate to their local context, yet can be related to the greater scheme of things and thus communicated more easily to colleagues and stakeholders.

**What are the Common Reference Levels and where do they come from?** The CEFR levels are based on the Council's work on objectives and are the result of process of deliberate convergence from the mid 1970s. The Swiss research project (1993-7) that produced the illustrative descriptors is briefly outlined and the salient characteristics of levels themselves explained.

**What claim to validity have the illustrative descriptors?** The reason why these particular descriptors have greater validity than those that preceded them is pointed out. The question whether a common framework can have validity both as a generic reference point and as a specific application in a local context will be addressed, together with the issue of use for the language of schooling.

**How can we ensure consistent interpretation of the levels?** Finally the draft Manual, intended to assist examination providers in relating their examinations to the CEFR, is briefly presented.

# The CEFR Common Reference Levels: validated reference points and local strategies

Brian NORTH  
Head of Academic Development, Eurocentres Foundation<sup>1</sup>  
Chair, EAQUALS<sup>2</sup>

## What is the Purpose of the CEFR?

The “*Common European Framework of Reference for languages: learning, teaching, assessment:*” CEFR (Council of Europe 2001) was developed between 1993 and 1996 by a Council of Europe international working party following the recommendation of an intergovernmental Symposium “*Transparency and Coherence in Language Learning in Europe*” held at Rüschtikon, near Zurich in November 1991. The CEFR was written with three main aims:

- To establish a metalanguage common across educational sectors, national and linguistic boundaries that could be used to talk about objectives and language levels. It was hoped that this would make it easier for practitioners to tell each other and their clientele what they wished to help learners to achieve and how they attempted to do so.
- To encourage practitioners in the language field to reflect on their current practice, particularly in relation to learners’ practical language learning needs, the setting of suitable objectives and the tracking of learner progress.
- To agree common reference points based on the work on objectives that had taken place in the Council of Europe’s Modern Languages projects since the 1970s.

In time, the existence of such a common reference framework would, it was hoped, help to relate courses and examinations to each other and thus achieve the “transparency and coherence” that had been the subject of the Rüschtikon Symposium.

The approach taken was to provide a conceptual framework made up of:

- A taxonomic *descriptive scheme*, summarised in Chapter 2 covering such issues as domains of language use, communicative language activities and strategies (in Chapter 4) plus the competences of the learner/user (in Chapter 5). The scheme can be seen as a further development of approach to objectives taken in Waystage, Threshold Level, Vantage Level and their equivalences for other languages.
- A set of *common reference levels*, defining proficiency in as many of these categories as possible at six levels (A1, A2, B1, B2, C1, C2) in empirically developed scales of illustrative descriptors (North 2000, North and Schneider 1998). This work made available a *descriptor bank*, validated and calibrated like an item bank, building on research with descriptor scales and item-banking in Eurocentres and elsewhere.

---

<sup>1</sup> The Eurocentres Foundation ([www.eurocentres.com](http://www.eurocentres.com)) has since 1968 had Consultative Status No 1 as an NGO to the Council of Europe. Eurocentres has 21 schools teaching English, French, Spanish, German, Italian and Japanese where those languages are spoken.

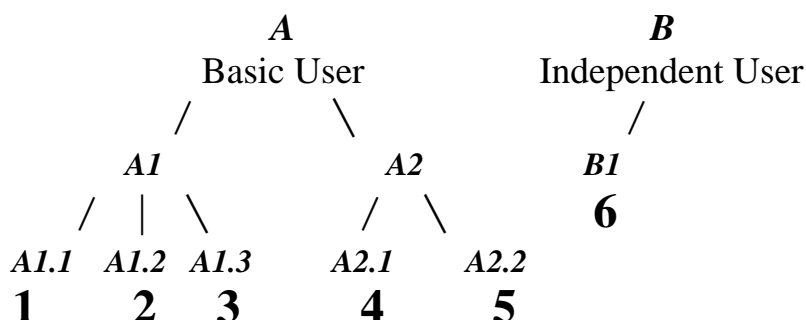
<sup>2</sup> EAQUALS: The European Association for Quality Language Services ([www.eaquals.org](http://www.eaquals.org)) has since 2005 had Participatory Status as an NGO to the Council of Europe. EAQUALS currently has 101 full member schools, plus 16 associate members.

The CEFR is not a harmonisation projects. The aim of the CEFR is to provide a mental framework that enables people to say where they were, not a specification telling them where they ought to be. Right at the very beginning of the CEFR, the authors emphasise:

“We have NOT set out to tell practitioners what to do or how to do it. We are raising questions not answering them. It is not the function of the CEF to lay down the objectives that users should pursue or the methods they should employ.” (Council of Europe 2001: xi Note to the User).

There is no need for there to be a conflict between on the one hand a common framework desirable to organise education and encourage productive networking, and on the other hand the local strategies and decisions necessary to facilitate successful learning in any given context. The aim of the CEFR is to facilitate reflection, communication and networking. The aim of any local strategy ought to be to meet needs in context. The key to linking the two into a coherent system is flexibility. The CEFR is a concertina-like reference tool that provides categories, levels and descriptors that educational professionals can merge or sub-divide, elaborate or summarise, adopt or adapt according to the needs of their context - whilst still relating to the common hierarchical structure.

The idea is for users to adopt activities, competences and proficiency stepping-stones that are appropriate to their local context, yet can be related to the greater scheme of things and thus communicated more easily to colleagues in other educational institutions and to other stakeholders like learners, parents and employers. For example in Finland, A1 is split into three levels to provide initial motivation. One might illustrate an approach like the Finnish one as a set of stepping stones towards independent language use, as shown in Figure 1.



Whilst the concept of “level” is useful for curriculum organisation and for giving a quick answer to the question “What level am I?” the rich descriptive framework provided by the CEFR illustrative descriptors is intended to facilitate profiling not levelling. In addition to the three summary scales introduced in Chapter 3, the CEFR offers a total of 54 sub-scales for different communicative language activities and strategies and for different aspects of communicative language competence. It is not the intention that anyone should use these 54 sub-scales operationally. Rather, their purpose is to help the user to consider the coverage of the learning programme or examination with which they are concerned: what are the priority areas and what level of proficiency is appropriate in each area? In setting such priorities, it is sensible to exploit the profiling possibilities by considering the setting of higher target levels for certain skills (e.g. receptive skills) – encouraging “partial competences.”

### What are the Common Reference Levels and where do they come from?

The CEFR levels (A1-C2) did not suddenly appear from nowhere. They have emerged in a gradual, collective process that started in 1913 with the Cambridge Proficiency exam (CPE) that defines a practical mastery of the language as a non-native speaker. This level has become C2. Just before the

last war, Cambridge introduced the First Certificate (FCE) – still widely seen as the first level of proficiency of interest for office work, now associated with B2. In the 1970s the Council of Europe defined a lower level called “The Threshold Level” (now B1), originally to specify what kind of language an immigrant or visitor needed to operate effectively in society. Threshold was quickly followed by “Waystage” (now A2), a staging point half way to Threshold.

The first time all these concepts were described as a possible set of “Council of Europe levels” was in a presentation by David Wilkins (author of “The Functional Approach”) at the 1977 Ludwighaven Symposium (Trim 1978). This symposium had represented the first - unsuccessful - attempt to move towards a common European framework in the form of a unit-credit scheme linked to common levels.

In 1992 the Council of Europe’s “Common Framework Working Party” adopted the following six “Common Reference Levels” for the future CEFR:

- Breakthrough (later A1),
- Waystage (later A2),
- Threshold (later B1),
- Vantage (later B2),
- “Effective Operational Proficiency” (later C1) exemplified by the new DALF (Diplôme Approfondi de Langue Française) from the CIEP,
- “Mastery” (later C2) exemplified by the Cambridge CPE (Certificate of Proficiency in English).

These six levels corresponded both to the seven levels that have been suggested by Wilkins in 1977 (minus the top level) and to the five levels adopted by ALTE: the Association of Language Testers in Europe (founded by Cambridge in 1991) – with the addition of A1.

In 1993-6, two Swiss members of the Working Party then used both qualitative and quantitative methodologies to develop illustrative descriptors for these six levels (North 2000, North and Schneider 1998, Schneider and North 2000). In a series of 32 workshops, teachers were given descriptors to sort into categories. This tested not only the clarity of the descriptors, but also the feasibility of the categories proposed for the sub-scales. The clearest, most relevant descriptors were then presented in descriptor-checklists that were used by some 300 teachers to assess learners in some 500 classes at the end of the school years 1994 and 1995. A statistical analysis of this data produced a scale on which each descriptor had a difficulty value (e.g. 176 or 053). The final step was to “cut” this continuous scale of descriptors to match the set of CEFR levels.

The Swiss research project actually identified 10 rather than 6 bands of proficiency. Between the criterion level for A2 and the criterion level for B1 there was found to be what came to be described as a “plus level.” The same was the case for between B1 and B2 (B1+) and between B2 and C1 (B2+). These “plus levels” were characterised by a stronger performance in relation to the same features found at the criterion level, plus hints of features that become salient at the next level. The “plus level” concept can be very useful in relation to school assessment because narrower levels allow learners to see more progress.

It is perhaps worth emphasising the salient features of the levels, as illustrated by the empirically calibrated descriptors:

**Level A1** is the point at which the learner can:

- *interact in a simple way, ask and answer simple questions about themselves, where they live, people they know, and things they have, initiate and respond to simple statements in areas of immediate need or on very familiar topics,*  
rather than relying purely on a rehearsed repertoire of (tourist) phrases.

**Level A2** reflects the Waystage specification with:

- the majority of descriptors stating social functions: *greet people, ask how they are and react to news; handle very short social exchanges; ask and answer questions about what they do at work and in free time; make and respond to invitations; discuss what to do, where to go and make arrangements to meet; make and accept offers,*
- plus descriptors on getting out and about: *make simple transactions in shops, post offices or banks; get simple information about travel; ask for and provide everyday goods and services.*

**Level B1** reflects The Threshold Level, with two particular features:

- maintaining interaction and getting across what you want to: *give or seek personal views and opinions in an informal discussion with friends; express the main point he/she wants to make comprehensibly; keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production,*
- plus coping flexibly with problems in everyday life: *deal with most situations likely to arise when making travel arrangements through an agent or when actually travelling; enter unprepared into conversations on familiar topics; make a complaint.*

**Level B2** reflects three new emphases:

- effective argument: *account for and sustain opinions in discussion by providing relevant explanations, arguments and comments; explain a viewpoint on a topical issue giving the advantages and disadvantages of various option,*
- holding your own in social discourse: *interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without imposing strain on either party; adjust to the changes of direction, style and emphasis normally found in conversation,*
- plus a new degree of language awareness: *correct mistakes if they have led to misunderstandings; make a note of "favourite mistakes" and consciously monitor speech for them.*

**Level C1** is characterised by access to a broad range of language that results in fluent, spontaneous communication:

- *express him/herself fluently and spontaneously, almost effortlessly; has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; there is little obvious searching for expressions or avoidance strategies - only a conceptually difficult subject can hinder a natural, smooth flow of language,*
- *produce clear, smoothly-flowing, well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices.*

**Level C2** represents the degree of precision and ease with the language of highly successful learners:

- *convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of modification devices,*
- *and a good command of idiomatic expressions and colloquialisms with awareness of connotative level of meaning.*

The illustrative descriptor scales in the CEFR make up a “descriptor bank.” In addition, the ELP Validation Committee has encouraged ELP developers to elaborate, extend and paraphrase the original illustrative descriptors. Experience suggests that this can be done in many ways, for example:

- editing the generic descriptor lightly to better suit the domain in question (academic study; professional areas, primary/secondary school classroom tasks);
- unzipping the tightly packed descriptors more typical of the CEFR into a more general header descriptor plus several sub-descriptors, which then sometimes take on a more specific orientation to the domain in question;

- adding new descriptors that mirror the concrete formulation style of the illustrative descriptors and exploit elements that appear in other descriptors at the level concerned.

A bank of CEFR-based descriptors for different age groups and educational sectors that were developed in these ways for different ELPs is provided on the [www.coe.int/portfolio](http://www.coe.int/portfolio) website.

### **What claim to validity have the illustrative descriptors?**

The illustrative descriptors address fundamental criticisms of previous descriptor scales in that they:

- cover a *communicative* descriptive scheme:
  - language activities and strategies (reception, interaction production)
  - language competences (linguistic, pragmatic, socio-linguistic)
- are each worded in positive terms, even for lower levels (Trim 1978)
- each present an independent, stand-alone criterion, defined independently from other descriptors (Skehan 1984: 217).

The descriptors are relatively unique in that they were not just written by an expert or committee, but were developed scientifically:

- from a comprehensive documentation of existing expertise in the area, with documented references to descriptors shared with pre-existing scales,
- in relation to both *what* learners do (communicative activities and strategies) and *how well* they do it (aspects of communicative competence),
- with a qualitative methodology that checked that teachers could relate to the categories used and that each descriptor employed was an unambiguous, stand-alone criterion,
- employing a sophisticated statistical methodology that enabled the descriptors to be scaled on the same mathematical scale as learners, (for some of whom video samples were also available),
- as a result of the use of the descriptors in a real, end-of-year assessment by practising teachers,
- in four educational sectors,
- in a multi-lingual environment,
- in relation to three foreign languages (English, French, German).

As a result, replication of the rank ordering of the descriptors has been high – with correlations between 0.92 and 0.97 to the original ordering:

- ALTE: 0.97 in conjunction with ALTE Can Do statements (Jones 2002:176),
- DIALANG: 0.96 in relation to listening and reading descriptors; 0.92 for writing (Kaftandjieva and Takala 2002: 114-121), This data is significant as it was for Swedish learners of Finnish, a non-Indo-European language.

Two fundamental questions remain, however, regarding limits to the validity of the descriptors:

- a) whether a framework can have validity both as a generic reference point and also as a specific application in a local context;
- b) whether a common framework developed for the teaching of foreign or second languages in mainstream education and the adult sector can be applied in primary school to the teaching of the language of schooling (the mother tongue of the indigenous population).

**Generic validity and specific application:** There are three sub-questions to this point:

- the practical application of theoretical concepts with generic validity;
- the common difficulty level of learning elements placed at different levels on a set of levels;
- the selection of a certain level to be a compulsory standard in a specific context.

**Practical application of theoretical concepts:** North (2000: 29) suggested that a framework scale ideally needs to be *context-free* in order to accommodate generalisable results from different specific contexts, yet at the same time the descriptors on the scale need to be *context-relevant*, relatable or translatable into each and every relevant context, and appropriate for the function they are used for in that context. In practical terms, this means that the descriptive scheme of the framework needs to:

- a) relate to the categorisation in theories of language competence, although the available theory and research is inadequate to provide a basis for it;
- b) be relevant to the contexts of the learning population concerned, though these cannot be predicted with any certainty;
- c) remain user-friendly and accessible to practitioners.

The CEFR illustrative descriptors attempt to do this; but it is a tall order. It would perhaps be fairer to say that the illustrative descriptors taken together with their further elaboration and amendment to specific contexts in the ELP descriptor bank discussed above meet this aim.

**Common difficulty level:** With regard to this issue, Spolsky considered:

“A functional set of goals exists in a social context.... Where this is consistent and common as in the Foreign Service, or in the Council of Europe notion of the Threshold Level for tourists and occasional visitors, it is not unreasonable to develop a scale that proceeds through the skills. ....If it cannot be based on a single social goal, a single set of guidelines, a single scale could only be justified if there were evidence of an empirically provable necessary learning order, and we have clearly had difficulty in showing this to be so even for structural items.”  
(Spolsky 1986: 154; 1989: 65)

A common, empirically proven *learning order* can probably never be demonstrated; second language acquisition research shows no signs of confirming such an order, despite what for a time appeared promising signs (e.g: Meisel et al 1981; Pienemann 1985). However, in the case of the bank of CEFR illustrative descriptors, an empirically proven interpretation of *difficulty* (i.e. the “Level” of a descriptor) can be claimed to be common across foreign languages, language regions, educational sectors (secondary, tertiary, and adult) and the learner/teacher divide. The “DIF” (differential item functioning) of the CEFR illustrative descriptors was extremely limited where it did occur it could easily be explained by not-necessarily-desirable curriculum practice (North 2000: 255-260).

**Selection of a certain level to be a compulsory local standard:** Educational standards must always take account of the needs and abilities of the learners in the context concerned. Unfortunately, educational standards often give the impression that they have been “plucked out of the air on the basis of intuition, which is frequently shown on closer examination to be wrongly conceived” (Clark 1987: 44). Norms of performance need to be definitions of performance that can realistically be expected, rather than relating standards to “some neat and tidy intuitive ideal” (Clark 1987: 46). This posits an empirical basis to the definition of standards.

The CEFR descriptors were in fact produced empirically in classroom contexts, relating learner achievement in the context (Switzerland) to the levels on the scale, as mentioned above. The European Language Portfolio (ELP) – with CEFR descriptors in adapted and elaborated forms appropriate to the educational context concerned as described above - provides an ideal exploratory tool to discover what level would be a realistic standard for any specific context. Yet unfortunately, in many contexts a



CEFR level (e.g. “B1”) continues to be “plucked out of the air” without an assessment of the realism of the objective or a consideration of the investment that would be necessary to achieve it.

That is not a limitation in the validity of the CEFR as such, but rather an unfortunately naïve use of it.

**Language of schooling:** As regards the second question, there is currently a lot of interest in the language of schooling. There are clearly areas of convergence between frameworks for the two purposes – for example the categories of the descriptive scheme itself would continue to be relevant. However, there will also be points where a framework to describe emerging competence in the language of schooling will put a different emphasis. For example, the CEFR illustrative descriptors define behavioural outcomes (results of learning), whereas a framework for the language of schooling may be more interested in describing emerging abilities and competences and in relating their development to the kind of educational scaffolding that encourages that development. Secondly, foreign language learners transfer to the foreign language competences they have already acquired in relation to the mother tongue and language of schooling. A framework for the latter would need to situate the development of language competences within the overall cognitive and social development of the children concerned. Many CEFR illustrative descriptors may thus prove unsuitable as they assume the cognitive and social competences of young adults at age 16 and above (e.g. the many descriptors at B2 emphasising the ability to develop and justify arguments).

These and other points of convergence and divergence are being considered by a new Working Party for a framework for the language of schooling that is expected to report in 2009.

### **How can we ensure consistent interpretation of the levels?**

North and Schneider (1998:243) emphasised that the production of a scale was only the first step in the implementation of a common framework, and that ensuring a common interpretation through standardised performance samples and monitoring data from tests was necessary.

The process of standardising the interpretation of the levels has been supported by the development of the draft manual “*Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)*” (Council of Europe 2003; Figueras et al 2005). The Manual proposes methodologies for relating the results from local assessments to the CEFR that do necessarily involve complex statistics or large-scale projects. Procedures are presented in three sets:

- **Specification** of the coverage of the examination in relation to the CEFR descriptor framework (categories described at different levels) in addition to a general description of the examination concerned.
- **Standardisation** of the interpretation of the levels through training with examples of performances calibrated with common criteria (CEFR Table 3) followed by benchmarking local performance samples to the framework levels. DVDs of performances illustrating the CEFR levels are available for the two official languages of the Council of Europe (English and French) and are almost completed for German and Italian.
- **Empirical Validation** of the results reported by tests through an independent corroboration of the linking claimed on the basis of specification and standardisation. To achieve this, one demonstrates a high correlation to an external criterion, which can be either a test already calibrated to the CEFR or judgements (by teachers or learners) with CEFR descriptors.

This scheme was adopted (a) because these categories are a good way of grouping linking methodologies found in the literature, (b) because they reflect the classic three stages of quality management (design, implementation, evaluation) and (c) because such broad concepts could thus be applied equally to formal, high-stakes assessment situations (examinations) and to lower-stakes school and teacher assessments.

The Manual has been piloted in some 20 case studies throughout Europe. Feedback considered at the start of the revision process (November 2006) indicates that the structure (specification, standardisation, validation) works well. The Manual appears to be a good way to critically review and evaluate the content and the statistical characteristics of an exam – and many stressed that the process is as important as the outcome. The Manual was felt to be a very effective way to mediate the CEFR, though several users reported on the difficulty of following the suggested procedures without a full set of calibrated performance samples and test items. In this respect the need for cross-language benchmarking during development was also stressed. Various suggestions have also been made to make the Manual more user-friendly and straightforward.

The definitive version of the Manual is expected to be available in January 2008. The documentation to the existing DVDs, plus writing performance samples are available on the website [www.coe.int/portfolio](http://www.coe.int/portfolio). Sample reading and listening items that are considered by the institutes they originate from to represent the CEFR levels have also been provided, and it is hoped that the actual case studies will produce a broader range of performance samples and test items that are calibrated to the levels.

The existence of sets of performance samples calibrated to the CEFR levels will also offer at least the possibility of further enriching the bank of illustrative descriptors. Techniques exist to identify salient features that appear to distinguish between levels; these techniques could be employed to double-check the content of the current descriptors and to elaborate areas that are currently under-defined (e.g. socio-linguistic competence; Level C2).

## **Conclusion**

Provided the descriptors in a common framework have a theoretically motivated development and provided that the framework presents a flexible structure, there is no need for a conflict between the desire to have a central framework to provide transparency and coherence and the need to have local strategies that provide learning goals specific to particular contexts. The main danger with regard to all common frameworks is a simplistic interpretation of them. The key to success is for users to appreciate that a common framework is a descriptive metasystem that is intended as a reference point, not as a tool to be implemented without further elaboration and adaptation to local circumstances.

## **References**

Alderson, J.C. (ed.) (2002): *Case studies in the use of the Common European Framework*. Strasbourg, Council of Europe, ISBN 92-871-4983-6.

Clark, John L. (1987): *Curriculum Renewal in School Foreign Language Learning*. Oxford: University Press.

Council of Europe (2001): *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge, Cambridge University Press.

Council of Europe (2003): “*Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)*” DGIV/EDU/LANG (2003) 5, Strasbourg, Council of Europe.

- Figueras, N., North, B., Takala, S., Verhelst, N. and Van Avermaet, P. (2005): Relating Examinations to the Common European Framework: a Manual. *Language Testing*, 22, 3, 1-19.
- Jones, N. (2002): Relating the ALTE Framework to the Common European Framework of Reference. In Alderson (ed.): 167-183.
- Kaftandjieva, F. and Takala, S. (2002): Council of Europe Scales of Language Proficiency: a validation study. In Alderson (ed.): 106-129.
- Meisel, J.H., Clahsen, H and Pienemann M (1981): On determining developmental stages in second language acquisition. *Studies in Second Language Acquisition* 3/2, 109–135.
- North, B. (2000): *The Development of a common framework scale of language proficiency*. New York, Peter Lang.
- North, B. and Schneider, G. (1998): Scaling Descriptors for Language Proficiency Scales. In: *Language Testing* 15/2: 217–262.
- Pienemann, Manfred (1985): Learnability and syllabus construction. In: Hyltenstam K. and Pienemann M. (ed.) (1985): *Modelling and assessing second language development*. Multilingual Matters, 23–75.
- Schneider, G and North, B. 2000: *Fremdsprachen können - was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit*. Chur/Zürich:
- Skehan, Peter (1984): Issues in the Testing of English for Specific Purposes. In: *Language Testing* 1/2, 202–220.
- Spolsky, B. (1986): A multiple choice for language testers. *Language Testing* 3, 147-158.
- Spolsky, B. (1989): *Conditions for second language learning. Introduction to a general theory*. Oxford, Oxford University Press.
- Stern, H.H. (1989): Seeing the wood AND the trees: Some thoughts on language teaching analysis. In: Johnson, R.K. (ed.): *The Second Language Curriculum*. Cambridge: University Press, 207–221.
- Trim, J. L.M. (1978): *Some Possible Lines of Development of an Overall Structure for a European Unit / Credit Scheme for Foreign Language Learning by Adults*. Strasbourg: Council of Europe, Appendix B.
- Wilkins, D. A. (1978): Proposal for Levels Definition. In: Trim 1978, 71–78, Appendix C.