Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

103

Linguistic Features at Different Levels of Language Proficiency: Some Facts, Assumptions and Suggestions

Ph.D., Formerly Research Professor
Sauli Takala
University of Jyväskylä, Center for Applied Language Studies
sjtakala@hotmail.com

ABSTRACT

The article deals with a topical and challenging theme in language education: what language properties are characteristic of different stages of language proficiency. The question of learning/teaching progression has always been a central concern in language education. It has acquired new interest with the increased use of the six levels of the Common European Framework of Reference for Languages (CEFR) for various purposes. The fifty-odd 6-point scales of various aspects of language use focus mainly on the uses of language and there are no language-specific descriptions as the CEFR is meant to apply to all languages. Only a few fairly general scales are available. The article addresses this issue of language-related properties of proficiency levels, presenting some existing work on such reference level specification, and presents the author´s own observations and assumptions about the grammatical and lexical resources of the proficiency continuum. A set of hypotheses are presented for testing.

Keywords: proficiency level, CEFR, SLA, grammar, vocabulary

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

105

INTRODUCTION

Language education has been a target of interest and a source of worry throughout the long history of formal education. Languages have tended to play an important role in education: the mother tongue in primary/elementary education and classical/modern languages (L2) in secondary education. It appears that there have been several pendulum swings in the approaches to L2 education and criticism of varying kind has been expressed at regular intervals (Kelly 1969).

Language education provision has always devoted considerable attention to the progression of the content.[1] In English, language proficiency has traditionally been categorized as elementary, intermediate and advanced (more recently also as basic, proficient and advanced). However, there has been no definite foundation for such a division. It has been based largely on tradition, experience and intuition. This is not to be interpreted as criticism or belittling as there is no reason to assume that these do not have some justification. In the absence of solid research foundation this was the best that could be done.

We entered a new stage when the European Common Frame of Reference for Languages (CEFR) appeared in 2001 after decades of development work under the aegis of the Council of Europe. The CEFR (2001) represents an action-oriented approach to language learning and use. It contains an extensive scheme of factors (sometimes referred to as the "horizontal dimension") which characterize language use for various purposes and various contexts, as shown by the construct definition (p. 9):

> Language use, embracing language learning, comprises the actions performed by persons who as individuals and as social agents develop a range of **competences**, both **general** and in particular **communicative language competences**. They draw on the competences at their disposal in various contexts under various **conditions** and under various **constraints** to engage in **language activities** involving **language processes** to produce and/or receive **texts** in relation to **themes** in specific **domains**, activating those **strategies** which seem most appropriate for carrying out

---

[1] One interesting source is "Selection and distribution of contents in language syllabuses" published by the Council of Europe in 1994.

the **tasks** to be accomplished. The monitoring of these actions by the participants leads to the reinforcement or modification of their competences.

The CEFR (2001) presents its specification of the various aspects of language use over a great number of pages in a linear form of presentation. In the interest of brevity and to help "seeing the forest for the trees" a pictorial overview of the category scheme is presented in Figure 1.[2] For all of the categories, except mediation, there are also a number of scales for a more detailed description of language use.
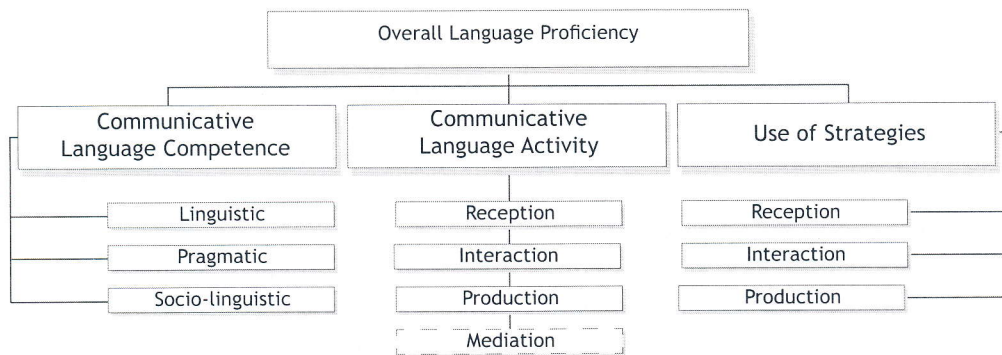

Figure 1. A pictorial representation of the CEFR construct.

However, the best known feature of the CEFR (2001) is the set of 53 6-point scales which illustrate the progression of language use from basic to proficient. In standard setting, this "vertical dimension" represents performance standards and the horizontal dimension the content standards (e.g. Cizek & Bunch 2007; Kaftandjieva 2004, 2005; Kane 2001; Zieky, Perie & Livingstone 2008).

The CEFR (2001) proficiency levels are labeled *basic user* (A1-A2; Fi: perustason kielenkäyttäjä, Swe: användare på nybörjarnivå, Ger: Elementare Sprachverwendung, Fr: utilisateur élémentaire), *independent user* (B1-B2; itsenäinen kielenkäyttäjä, självständig användare, Selbständige Sprachverwendung, utilisateur indépendent) and *proficient user* (C1-C2; taitava kielenkäyttäjä, avancerad användare, Kompetente Sprachverwendung, utilisateur expérimenté). North (2000) accounts for the sources and methodology used to produce the scales.

---

[2] *This figure has been constructed by Dr. Felianka Kaftandjieva (2005) and it is reproduced here with her kind permission.*

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

107

Most of the CEFR (2001) scales express pragmatic aspects of language use: what people can do with language. Section 5.2 in the CEFR (2001) presents six more directly language-oriented scales, one scale of socio-linguistic competence and six scales of pragmatic competences. These thirteen scales have been criticized to some extent for not providing a good enough basis for language-specific interpretations. In recent years, Reference Level Descriptions (RLD), with a more explicitly language-focused specification of the CEFR (2001) levels, have been developed for some languages, including e.g. French, German, Spanish, Italian and Estonian. It would seem, however, that they are not solidly based on empirical analyses of representative learner corpuses. However, it goes without saying that only truly representative corpora (several L2s, several contexts of use etc) can be used for robust generalizations.

In this article I will address some aspects related to the issue of the underspecification of the linguistic features of proficiency levels. I will present a selective review of research on this topic and some personal views, assumptions and recommendations based mainly on my extensive experience in rating learner texts. This means that the article represents a "mixed genre" and is mainly to be seen as programmatic.

2   SECOND LANGUAGE ACQUISITION:
FROM ELEMENTARY TO ADVANCED LEVELS

As pointed out in the above, the progression of language proficiency has been of central interest in the long course language education. Closer to our time the progression has been addressed by contrastive analysis, error analysis and performance analysis. The interest was to predict and prevent errors. When there was a shift in the paradigm towards mentalism and (creative) constructivism, and errors started to be seen as largely regular features of "transitional competence" or "interlanguage" (Corder 1967, 1981; Selinker 1972), the approach changed and the gradual development of the interlanguage (along the interlanguage/restructuring continuum) towards nativelike usage (no matter how it might be construed) became an important item on the research agenda.

It can be argued that a new stage started when — towards the end of the 1960s — Second Language Acquisition (SLA) emerged as a new cross-disciplinary discipline. It drew on language teaching, linguistics, research on child language acquisition and psychology, and gradually established itself in the 1990s as a "normal" discipline with journals of its own, conferences, study and research programmes (Ortega 2009, 2). Recent and up-to-date reviews of SLA are e.g. Abrahamsson 2009; de Bot, Lowie & Verspoor 2005; Doughty & Long 2003; Long 2007; Ortega 2009; Robinson & Ellis 2008; Tomasello 2003 and van Patten & Williams 2007.

SLA research has tended to focus on the early stages of language acquisition (beginning with the seminal morpheme order studies) and to some extent also the intermediate levels. It has been studied what kind of developmental stages there are in grammar, phonology and vocabulary (e.g. Doughty & Long 2003; Ellis 1994). Interest in the research on advanced aspects of SLA (near-nativeness or even nativelikeness) is relatively recent (e.g. Ringbom 1993, 2007; Birdsong 2004, 2007; Abrahamsson & Hyltenstam 2009). A new phase in the largely cognitively dominated SLA is the emergence of a sociocultural approach (e.g. Lantolf & Thorne 2007), which is interested in "legitimate peripheral participation in the activities of communities of practice"[3].

---

[3] *In fact, instead of SLA I would prefer the term "language socialization" as a cover term, as it would recognize the inherent link between language, society and culture (cf. Ochs & Schieffeling 2006).*

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

109

In order to obtain a comprehensive view of the interlanguage continuum it is obviously important to cover the whole domain: all levels from elementary to nativelike proficiency, a reasonable range of typologically different languages, a reasonable range of different purposes and contexts of language use in different communicative cultures, a range of personality types and individual variation (e.g. Dörneyi 2005; Larsen-Freeman & Cameron 2008). The domain of generalization and the validity of generalizations need to be considered and temptations of premature generalizations ought to be resisted.

As advanced L2 proficiency has been relatively neglected I will present a brief review of some recent findings drawing, in particular, on an extensive and note-worthy research programme in progress at the University of Stockholm.

Several terms are used in reference to high-level L2 proficiency: functional bilin-gual, near-native, non-perceivable non-native, native-like (a scale of increasing ability). Since the term "native" is subject to continual dispute, Kenneth Hyltens-tam and his colleagues at the University of Stockholm occasionally use the terms "high-level proficiency in second language use", in Swedish "avancerad an-draspråksanvändning" (AAA). The Stockholm team refers to a *native-like speaker* as someone who, in all respects, uses the language like a native speaker, in spite of the fact that the language in question is not that person's first language. A *near-native* speaker, in turn, is defined as a person who is perceived in normal in-teraction as a native speaker but who can be distinguished from native speakers in some features when his/her language is analysed in greater linguistic detail. An *advanced second language learner/user* refers to a person whose language is close to that of a native speaker but whose non-native usage is perceivable also in normal oral or written interaction (Hyltenstam & Abrahamsson 2003, 571; see also Piller 2002).

The ultimate form of language proficiency, nativelike language use, is an inher-ently problematic and much disputed concept, and a prominent expert on the topic Alan Davies (2003) is doubtful about its relevance and usefulness. It is especially problematic in the case of English, with its several "regional" variants and the phenomena called World Englishes and English as a lingua franca. The concept of "native" speaker has been operationalised using e.g. self-assessment, assessment by native speakers and the verification of de facto native speaker status.

In a large-scale, ambitious and skillfully designed and implemented study (Swedish as L2) Hyltenstam and Abrahamsson (2009) have obtained interesting and important results, which can be summarized as follows:

- Persons perceived in normal conditions as native displayed certain divergences from the native usage when language use was examined closely. Perception as a nativelike user is clearly associated with the (early) age of onset of L2 acquisition (lending support to the Critical Period Hypothesis, CPH).
- Near-native proficiency acquired in adult age is related to high language aptitude.
- None of the persons perceived as native speakers and who had started acquiring Swedish after age 12 or later performed at the native level in demanding language tests. Only very few who had started at 11 or younger passed the tests at the native level.
- Thus, adult acquirers can, in principle, never reach the nativelike level and this is also more uncommon with younger learners than is usually suggested.

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

111

3   PROFICIENCY LEVELS IN LANGUAGE EDUCATION

In addition to SLA research, the development of language proficiency has received increased attention in language education when scales of proficiency have been developed all over the world. The "mother of all proficiency scales" is undoubtedly the Foreign Service Institute's (FSI) scale, which one of the most prominent experts in language education and testing, John B. Carroll, helped to develop. Other scales followed: ACTFL scale, Wilkins's scale (1977, developed in connection with the Council of Europe modern language projects), and above all the CEFR (2001) scales.

As the main interest of this article is the linguistic features of language proficiency, the system proposed at the University of Stockholm by Bartning and Schlyter (2004) for 6 levels of French is of great interest.[4] The system is based on an *oral* learner corpus and, although it has six levels as in the CEFR (2001), it has been developed independently and should not be confused with it.[5] It seems that the project is a pioneering one and thus deserves attention.

The levels are described below in terms of some selected features: type of utterance structure: nominal, verbal non-finite or finite; development of finiteness, subject-verb agreement, temporal and modal systems, negation, noun phrase morphology and subordination.

1) *Le stade initial* (initial stage): eg. nominal utterance structure; utterances with some formulae; bare nouns but also some determiners and non-finite verbs; preverbal negation and some finite verbs forms.

2) *Le stade post-initial* (post-initial stage): eg. both non-finite and finite utterance structure,; polyfunctional base forms (the present for the past and the future, etc); some inflection on verbs and adjectives; paratactic utterance structure but also the emergence of some subordination; irregular verbs in the 3rd person singular/plural appear but in different non-finite and non-targetlike finite forms; past tense appears with stative verbs and some distinction is made between 1st and 2nd person of the present-tense

---

[4] See also http://project2.sol.lu.se/DirektProfil/index.html
[5] The researchers are reported at present to be co-operating with the CEFR French development teams.

verbs; gender marking (if used at these two stages at all) may use one (masculine vs. feminine) as the default or overuse the masculine.

3) *Le stade intermédiare* (intermediate stage): eg. more systematic and regular, with a more or less established finite utterance structure (but still a simple interlanguage system): present tense, *le passé composé* (which may correspond to both the "perfect" and the "preterite") and the future, mostly periphrastic future; non-finite forms less frequent;first cases of the subjunctive; double phrasal negation more or less acquired; non-target-like forms on determiners and adjectives persist in gender agreement.

4) *Le stade avancé bas* (advanced low stage): eg. the typical structures of French grammar emerge: the clitic pronoun before the finite verb, the conditional, the pluperfect and the subjunctive – the last three in isolated cases; more complex forms mainly occur when syntax is not complex, but not always in the correct form; most non-targetlike non-finite forms of the regular verbs have disappeared, but the 3rd person plural form of irregular verbs still not consistent: overuse of the present; overuse of the masculine in agreement of adjectives.

5) *Le stade avancé moyen* (advanced intermediate stage): eg. considerable development of inflectional morphology, but still problems of gender and adjectival agreement; multi-propositional subordination increases; contracted sentences with infinitives and gerunds; in subject-verb agreement there is still some difficulty with non-marked 3rd person plural of the irregular verbs; telic verbs appear in the imperfective and static verbs in the perfect; learners can move on the time axis; the future, the conditional, the pluperfect and the subjunctive are mostly targetlike, with some overuse of the *passé composé* for the pluperfect; still problems with gender agreement on preposed adjectives.

6) *Le stade avancé supérieur* (advanced high stage): eg. inflectional morphology is stabilizing even in multi-propositional utterances; high degree of utterance packaging, ellipsis and integrated propositions; almost native-like use of connectors and of relative and causal clauses; mostly targetlike 3rd person plural present tense of irregular verbs and targetlike use of the sunjunctive.

It may appear somewhat surprising that there has not been similar progress in English charting the linguistic progression in terms of proficiency levels. One reason may be that the definitions of objectives by the Council of Europe from

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

113

the mid 1970s onwards[6] were felt to be sufficient. However, our experience with the DIALANG-project in the late 1990s (and there were similar observations elsewhere) indicated that there was considerable overlap in them, i.e. the progression was not very clear. With the growing use of the CEFR (2001) it has become evident that there is a need for a research-based definition of linguistic progression for particular languages.[7]

In fact, Cambridge ESOL currently co-ordinates a large-scale corpus-based project "Profile English". which aims at providing such a research-based description of learner progression in the mastery of English. Some early results have been presented recently (John Hawkins at the 3rd ALTE conference in 2008; Hendricks & Saville at a conference in Taipei in 2009).

Hendricks and Saville (2009) provide a useful review of some findings obtained so far (I have added some personal comments on the findings):

- Not surprisingly, more frequent properties in the target language are more easily acquired: fewer errors, more of the relevant target language properties learned and earlier acquisition. There is a disproportionate use of frequent items in early L2 English, moving gradually to more nativelike target English pattern of frequency.

  When it is compared how language learners use the most common English verbs (know, see, think, want, mean, get, go, say, come and need) vis-à-vis the British National Corpus (BNC) figures, they are overrepresented (with the exception of "mean"). The distributions come closer to the BNC as the proficiency level progresses from A2 to C2.

- Also as expected, structurally simple properties are more easily acquired: fewer errors, more of the relevant target language properties learned and earlier acquisition. Structural complexity leads to more errors.

  "Verb co-occurrence", the company verbs keep, shows the following cline of complexity:
  NP - V: he went
  NP - V - Part: the boy ran away
  NP - V - NP: she loved her husband

---

[6] In English: Breakthrough, Waystage, Threshold and Vantage. There was also a breakdown of English grammar into six stages by L. G. Alexander et al. (1975) — a work drawing on their very extensive experience in the field.

[7] Now that there is a greater awareness of the importance of the semantic and phonological structure (challenging "syntactocentrism", e.g. Jackendoff, 2007) there is also an opportunity to provide a balanced — and therefore a more useful — account of linguistic progression.

NP - V - Part - NP: he looked up the address
NP - V - NP - PP: he added the flowers to the bouquet
NP - V - NP - NP: she asked him his name
NP - V - S (Wh-move): he asked how she did it
NP - V - PP - S: they admitted to them that they had entered illegally
NP - V - P - Ving - NP: they failed in attempting the climb

There is a clear progression in such constructions from A2 to C2 but no new constructions have, so far, been discovered at the C levels. It would appear that these basic constructions of English have been learned by B2. C-levels may need a subtler kind of analysis. The above corresponds to the frequencies found in the BNC.

- An expected result is also the fact that the fewer items and/or properties to be learned in a given grammatical or lexical domain, the easier the domain is to acquire. Small domains exhibit fewer errors, proportionally more of the L2 properties learned and earlier acquisition. Large domains exhibit the reverse trend. Learning a limited set of (closed-class) morphosyntactic properties (eg. subject person agreement on English verbs, tense inflection on English verbs, singular/plural inflection on English nouns) is easier than learning the large (open-class) set of lexical verbs and nouns in English with all of their semantic and syntactic properties and distinctions. This is something learners of English are very familiar with.

- Also as expected by all with first-hand experience in language teaching, and contrary to claims in the early days of creative constructivism, transfer plays a definite role in L2 acquisition: the more similar L1 and L2 are in some grammatical/lexical domain, the easier the domain is to acquire.[8] Thus, speakers of languages with the definite and indefinite articles find it easier to acquire the article system of English than speakers from source languages without articles (again, this is a familiar phenomenon for Finnish learners of English).

- If English has structural alternatives to choose from, with different degrees of efficiency on different occasions of use, selections will move gradually to English target norms as proficiency increases. Examples of this are eg. presence or absence of relative pronouns, presence or absence of explicit that-complementizer, extraposed vs. unextraposed sentential subjects and infinitivals.

---

[8] *Håkan Ringbom (1993, 2007) has made a substantial contribution to this field.*

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

115

- Overall semantic density of verb meaning and utterance meaning increases from A2 to C2. Many verbs of motion at level A2 do not contain any semantic information besides the motion alone (go, move) while with increased proficiency the proportion of such low-information verbs decreases, being low at level C2.

- The number of prepositional and adverbial phrases carrying spatial information increases with proficiency level. If they are missing, this suggests level A2.

It will be of great interest to read the forthcoming findings of Profile English and to check how they compare with our own Finnish experience as speakers of a typologically quite distinct language. While most of the Profile English findings are likely to apply to the Finnish users of English, there may very well be some features where the progression is somewhat different and there are likely to be some features which persist to cause more difficulties for us than for speakers of other languages.

4 SOME OBSERVATIONS AND ASSUMPTIONS BASED ON PERSONAL EXPERIENCE AS A USER OF ENGLISH AND A RATER OF STUDENT WRITING

Although research is the only reliable source of information, due to the fact that research applies sound principles of disciplined inquiry, and in spite of the fact that projects like English Profile are beginning to deliver interesting results, I assume that an experienced and reflective language teacher has a good idea of what kind of language learners at different levels of proficiency can use. This presumes, however, that the teacher has paid conscious attention to the linguistic features, is interested in them and not only on the satisfactory communication of meaning, and is prepared to revise his/her views on the basis of empirical evidence. On the other hand, I do not assume that teachers fully agree in their perception but that there is in fact some (occasionally even considerable) variation as there is no uniform and systematic framework or systematic training to draw on. Thus it is likely that teachers differ to some extent in what they pay attention to in thinking about and assigning levels.[9] Each teacher and rater is likely to have his/her own "subjective theory" about matters like this.

I do not have such teaching experience but I have extensive experience in marking English essays included in the Matriculation Examination. The essays have been written by 18-19-year-old students: most of them had had 10 years of English, some 5 years and a smaller part 3 years. From the mid 1980s to 2008 I rated more than twenty thousand essays. While I obviously followed the analytical rating criteria (which I had helped to formulate), in which communicativeness is the dominant criterion, I also observed and made occasional notes about the language usage. On the basis of this I have formed a personal view of what characterizes writing at different levels of proficiency. Having participated in several events where rating tendencies have been recorded I also know that I tend to be an "average" rater, belonging neither to the lenient rater group nor the severe one. I believe that although what I will present in the following does not build on a systematic learner corpus research, the "claims" are worthwhile working hypotheses, and I do expect them to be roughly correct.

---

[9] *Stating e.g.: "In my view, a learner/user at level B1 ought to know x,y,z" – even if there is no such performance level description (PLD) available.*

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

117

In this presentation I will limit my attention to vocabulary and grammar in a broad sense. I realize that the focus is limited but I believe that one has to make a start somewhere, and *syntax and semantics are important*, whatever other aspects may be added later on. Leading experts on vocabulary are Paul Meara and Paul Nation. My own PhD dealt with vocabulary learning (Takala 1984). The basic reference in vocabulary testing is Read (2000) and in grammar testing Purpura (2004). Eeva Tuokko (2007) has discussed several aspects of vocabulary and grammar testing in her PhD thesis.

4.1 Level in vocabulary mastery

Although the CEFR (2001) scales for general linguistic range and vocabulary range and control are rather general, they form a useful starting point. They are given below:

Levels	General Linguistic Range (CEFR 2001, 110)
C2	*Can exploit a comprehensive and reliable mastery of a very wide range of language to formulate thoughts precisely, give emphasis, differentiate and eliminate ambiguity. No signs of having to restrict what he/she wants to say.*
C1	*Can select an appropriate formulation from a broad range of language to express him/herself clearly, without having to restrict what he/she wants to say. Can express him/herself clearly and without much sign of having to restrict what he/she wants to say.*
B2	*Has a sufficient range of language to be able to give clear descriptions, express viewpoints and develop arguments without much conspicuous searching for words, using some complex sentence forms to do so.* *Has a sufficient range of language to describe unpredictable situations, explain the main points in an idea or problem with reasonable precision and express thoughts on abstract or cultural topics such as music and films.*
B1	*Has enough language to get by, with sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events, but lexical limitations cause repetition and even difficulty with formulation at times.* *Has a repertoire of basic language which enables him/her to deal with everyday situations with predictable content, though he/she will generally have to compromise the message and search for words.*
A2	*Can produce brief everyday expressions in order to satisfy simple needs of a concrete type: personal details, daily routines, wants and needs, requests for information.* *Can use basic sentence patterns and communicate with memorised phrases, groups of a few words and formulae about themselves and other people, what they do, places, possessions etc.* *Has a limited repertoire of short memorised phrases covering predictable survival situations; frequent breakdowns and misunderstandings occur in non-routine situations.*
A1	*Has a very basic range of simple expressions about personal details and needs of a concrete type.*

Levels	Vocabulary range (CEFR 2001, 112)	Vocabulary control (CEFR 2001, 112)
C2	Has a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms; shows awareness of connotative levels of meaning.	Consistently correct and appropriate use of vocabulary.
C1	Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circum-locutions; little obvious searching for expressions or avoidance strategies. Good command of idiomatic expressions and colloquialisms.	Occasional minor slips, but no significant vocabulary errors
B2	Has a good range of vocabulary for matters connected to his/her field and most general topics. Can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocution.	Lexical accuracy is generally high, though some confusion and incorrect word choice does occur without hindering communication.
B1	Has a sufficient vocabulary to express him/herself with some circumlocutions on most topics pertinent to his/her everyday life such as family, hobbies and interests, work, travel, and current events. Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics.	Shows good control of elementary vocabulary but major errors still occur when expressing more complex thoughts or handling unfamiliar topics and situations.
A2	Has a sufficient vocabulary for the expression of basic communicative needs. Has a sufficient vocabulary for coping with simple survival needs.	Can control a narrow repertoire dealing with concrete everyday needs.
A1	Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.	No descriptor available.

4.2 Vocabulary range

Paul Nation[10], a recognized expert on the learning and teaching of vocabulary, has produced a computer programme, based on the British National Corpus (BNC), which makes it possible to study the vocabulary coverage of texts (including learner texts) in relation to the BNC frequencies. The programme analyses the vocabulary and reports which words belong to the list of first 1000 words, the second list of 1000 words etc., ending up with list 14.

Everyone who has a fair amount of experience in rating learner texts knows that good scripts have an extensive, appropriate and precise vocabulary whereas

---

[10] http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

119

poorer scripts usually have a limited vocabulary. Occasionally their vocabulary may be more extensive but there may be inaccuracies and a lot of spelling mistakes. Here are some authentic examples from writing by 18-19-year old Finnish students with 10 years of English study:

- I have been though that young would be vote personal what kind seemed fine.
- They have not to really imagine themselves although they often given various affection themselves.
- They are sometimes take a part too and succested quite well.
- Common knowledge is wrong what comes to public.
- According to the Christmas time and Valentine's Day man are also worried about the number of bank council.
- Adults how have children and life experiment...
- People are selfishness and heartless.

On the basis of my very extensive rating of matriculation examination compositions I submit that the very best of them (C1.2/C2) may occasionally contain words from the top word lists. At level B1/B2 this is unusual and at level A2/B1 very exceptional. As a rough "educated guess" – or expressed in a more formal manner, as a "working hypothesis" – I suggest that within a range of +/-5% the following obtains:

- Scripts at level A1.1 – A2.1: the vocabulary contains mainly words from the first list of 1000 words.
- Scripts at level A2.2/B1.1: about three fourths (3/4, 75%) of the vocabulary belong to the first list and the rest mainly to the second list.
- Scripts at level B1.2/B2.1: two thirds (2/3, 65%) belong to the first list, 20-25% to the second list and the rest mainly to the third list, with occasional words from the higher lists.
- Scripts at level C1.2/C2: only one third (1/3, 35%) belong to the first list, one fourth (25%) to the second list, 10-15% to the third list and about 5% to the subsequent lists.

The above is a "working hypothesis" but I expect it to be roughly correct. It has to be noted also that it is assumed here that the rubric presupposes relatively demanding exposition or argumentation or sophisticated narration (i.e., the assignment is not a simple pragmatic task such as notes, messages).

I have made a random selection of words from the seven first lists and list 14 and present them below so that the readers can get a better idea of what the vocabulary in the different list is like and thus be able to evaluate my working hypothesis.

Table 1. A selection from Paul Nation's vocabulary lists.

Vocabulary list 1	Vocabulary list 2	Vocabulary list 3	Vocabulary list 4	Vocabulary list 5	Vocabulary list 6	Vocabulary list 7	Vocabulary list 14
able	accident	accuse	abandon	abort	abstain	acclaim	admonish
back	battery	bargain	bankrupt	beard	bale	bait	bequeath
care	calm	capture	carve	captive	canvas	callous	contrite
decide	defend	deserve	dart	dazzle	daunt	defer	desecrate
elect	emotion	echo	enrol	eliminate	emulate	elapse	equanimity
favour	familiar	fatal	feast	fever	flaw	ferment	foible
glass	grateful	glance	gender	furious	grievance	glean	gamut
happen	holy	handy	hatch	humiliate	heed	heady	harangue
improve	impress	idle	imitate	infer	impeccable	impair	immutable
judge	journey	jewel	jealous	jeopardy	judicial	jab	
kind	keen	kidney	knit	knack	keel	kiln	
law	length	layer	launch	loathe	levy	lather	lithe
manage	maintain	mature	merge	mercy	mandatory	mingle	medial
nature	negotiate	nervous	notion	nudge	nag	niggle	nuzzle
oppose	occupy	offend	overlap	oppress	omen	opportune	outwit
parent	patient	panic	penetrate	peck	pertinent	peddle	pallid
question	quote	queue	quiz	quarry	quantify	quaint	quiescent
ready	race	rebel	racial	ration	ratify	rant	roundel
secure	salary	scream	scatter	seize	scrutiny	salvage	secede
tape	taste	tease	tedious	tranquil	throttle	thrift	taut
useful	upset	urge	underlie	undergo	unravel	upkeep	
visit	vary	visual	venture	vicinity	vigil	vouch	(weft)

I assume that most teachers of English agree with me that words in list 14 represent, indeed, high level proficiency.

I believe that in analysing and rating learner scripts, in addition to the communication of the intended meaning, it is useful to analyse at some level of detail how extensive the vocabulary is in learner texts and how appropriate and idiomatic it is. At present, vocabulary probably is, in fact, a more or less explicit criterion in rating learner texts but I suggest that it should be a subject to more explicit attention. For instance, it would be useful to comment on these aspects in benchmarks (e.g. what percentage of words belong to various word lists).

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

121

4.3 Going beyond the word as the lexical unit

One characteristic of English is the *phrasal verbs*, whose idiomatic use can be expected to be one indicator of high-level proficiency. Phrasal verbs contain many of the most common verbs in English: e.g. bring, come, do, get, go, hold, keep, look, make, put, run, set, turn.

Referring to my experience in rating learner scripts in the matriculation examination, I suggest that there is a very clear difference between weak, intermediate and good scripts in terms of how phrasal verbs are used. In the weak scripts they are either non-existent or very limited, they are scarce in intermediate scripts but the good scripts may include several different phrasal verbs (appropriately used).

In the following I present some non-systematic examples, which reflect my perception of the phenomenon.
- Weak scripts (A2.2/B1.1): hardly any phrasal verbs; those that resemble them are often, in fact, concrete expressions: get out, look at, take off, walk in, walk out.
- Intermediate scripts (B1.2/B2.1): e.g. get along with, get rid of, grow up, look forward to, take over.
- Very good scripts (C1.2/C2.1): e.g. break up, break in, bring forth, bring up, cut out for, come across, fall into, go on, go with, grow into, grow up, hold for, keep on, keep up with, look after, look forward to, look on, make up, make use of, put up with, run out of, show off, shut up, take care of, take for (granted), take up, turn into, turn on, turn out, wait for, work for, work on.

Perhaps somewhat unexpectedly, base words are much less common than compound words (e.g. Takala 1984). All languages also have a very large amount of multi-word units: phrases, formulas, patterns, chunks, prefabricate routines to mention only a few terms used (e.g. Pawley & Syder 1983; Fillmore, Kay & O'Connor 1988; Granger 1998). According to Wray (2002) there are dozens of taxonomies to categorize formulas. The traditional syntax-dominated approach to linguistics has treated them as peripheral phenomena and it cannot, in fact, deal with them in a credible and effective manner.

Jackendoff (2007) presents persuasive arguments why syntax cannot be the dominant linguistic component but phonology, syntax and semantics are all autonomous components with more or less close/loose interfaces. In other words, "parallel architecture" is a more promising starting point than the syntactocentrist

approach. According to this view, there is no clear distinction between grammar and lexicon. Lexical units are part of the rules of grammar, somewhat special kinds of rules but rules nonetheless. Unlike the mainstream approach that sees the primary function of language to be thinking (inner speech), Jackendoff and several other linguists consider its main function to be communication of thoughts, i.e. meaning. With evolution, syntax and phonology have developed into more and more effective means for conveying meaning more accurately.[11]

Jackendoff (2007, 56) provides the following set of examples of idiomatic language use (non-canonical utterance types), which cannot easily be dealt with in syntax:

- Into the boat with you!
- How about a cup of tea?
- What, me worry?
- One more beer and I'm leaving.
- The more I read, the less I understand.
- How dare he question our motives!
- Far be it from us to expect any special treatment.

Collocations are an important aspect of the lexicon and their appropriate use can also be taken as an indicator of high-level proficiency. Wray (2002, 63) presents as examples of the collocation dimension: blow a trumpet - blow a fuse - blow your own trumpet - blow the gaff; under the table - under attack - under the microscope - under the weather.

It is perhaps a bit paradoxical (Forsberg 2008) that formulas function as an important communication strategy in early language acquisition (Krashen & Scarcella 1978) and as a basis for subsequent creative language use (e.g. Wong-Fillmore 1976) and at the other end they are indicators of high-level (idiomatic) language use (e.g. Yorio 1989; Ellis 2002; Wray 2002; Schmitt et al. 2004; Forsberg 2008). It is likely that really rich input and continuous use stabilise the appropriate use of formulas. This is a lifelong challenge but also a source of pleasure in language learning: one can always learn new idiomatic expressions and to learn to use them appropriately.

---

[11] *Jackendoff observes that language does have some "paleolexicon", which does not require syntax: hello, yes, oops, ouch.*

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

123

4.4 Syntactic accuracy

The CEFR (2001) offers a scale for grammatical accuracy, which is reproduced below. It appears appropriate but there is a problem referred to in the above: it remains unclear which structures are simple vs. complex and which are easy vs. difficult to acquire/learn.

GRAMMATICAL ACCURACY (CEFR 2001, 114)	
C2	Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).
C1	Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot. Good grammatical control; occasional 'slips' or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect.
B2	Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding. Communicates with reasonable accuracy in familiar contexts; generally good control though with noticeable mother tongue influence. Errors occur, but it is clear what he/she is trying to express.
B1	Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations.
A2	Uses some simple structures correctly, but still systematically makes basic mistakes - for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what he/she is trying to say.
A1	Shows only limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.

In the following I will try to provide some concretization of these proficiency levels by drawing on the scheme by Bartning and Schlyter (2004) for French, on some previous research literature and on personal observations. It is to be noted that the Bartning and Schlyter scheme has been developed independently of the CEFR (2001). As what I will do is a "thought experiment", I will list their levels and the CEFR (2001) levels side by side, but I wish to emphasize that no claim is made about one-to-one correspondence. Readers can judge for themselves how close the links appear to be. The same applies to the samples I have provided: they are not empirically verified but represent my hypotheses. They — like the other samples I give — obviously reflect the influence of Finnish. Learners with other language backgrounds may exhibit somewhat different language properties.

Level	Grammatical accuracy (CEFR 2001)	Linguistic features
C2	*Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).*	Bartning & Schlyter (2004): *Le stade avancé supérieur* (advanced high stage) : eg. inflectional morphology is stabilizing even in multi-propositional utterances; high degree of utterance packaging, ellipsis and integrated propositions; almost nativelike use of connectors and of relative and causal clauses; mostly 3rd person plural present tense of irregular verbs and targetlike use of the subjunctive. *English: some possible problems:* I did it with protest … in one voice … for consequence of … with the view to … he wrote very shortly … It became Mary who took the initiative … She said she had a weak heart… … but it seemed that the accuracy of the figures were not questioned … From here to my cottage are 60 kilometres… Eggs and bacon are my favourite dish … Either he or his sister are sure to help us… This fact is impertinent to the debate …
C1	*Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot.* *Good grammatical control; occasional 'slips' or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect.*	Bartning & Schlyter (2004): *Le stade avancé moyen* (advanced intermediate stage): eg. considerable development of inflectional morphology, but still problems of gender and adjectival agreement; multi-propositional subordination increases; contracted sentences with infinitives and gerunds; in subject-verb agreement there is still some difficulty with non-marked 3rd person plural of the irregular verbs; telic verbs appear in the imperfective and static verbs in the perfect; learner can move on the time axis; the future, the conditional, the pluperfect and the subjunctive are mostly targetlike, with some overuse of the *passé composé* for the pluperfect; still problems with gender agreement on preposed adjectives. *English: some possible problems:* It will be possible the decision to be made …. deserves even larger attention … did not give firm answer to this question… the result was at high degree interpreted in the same way… It should be mentioned in this connection, that this does not allow enough stable estimation of the situation … regard this as something, which was expected, since … is splitted in two parts …. selection of a precise method to carry out the task …. No student, who scored less than 50%, passed the exam… I saw a man the other day, who says you were old friends…
B2	*Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding.* *Communicates with reasonable accuracy in familiar contexts; generally good control though with noticeable mother tongue influence. Errors occur, but it is clear what he/she is trying to express.*	Bartning & Schlyter (2004): *Le stade avancé bas* (advanced low stage): eg. the typical structures of French grammar emerge: the clitic pronoun before the finite verb, the conditional, the pluperfect and the subjunctive - the last three in isolated cases; more complex forms mainly occur when syntext is not complex, but not always in the correct form; most non-targetlike non-finite forms of the regular verbs have disappeared, but the 3rd person plural form of irregular verbs still not consistent: overuse of the present; overuse of the masculine in agreement of adjectives. *English: some possible problems:* We enjoy each others company … I appriciate good conversations … I like to get my voice and my opinions out … Before all I would like to thank you … Neither my elder sister and my elder brother doesn´t live home anymore … It is difficult to point things that I don´t like …. My school was renovated few years ago … I would like to see more world some day … I was interested of ice hockey … I am much exited about the trip … it hadn´t been surprisingly broken … I wish you passed the test tomorrow…

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

125

B1	*Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations.*	Bartning & Schlyter (2004): *Le stade intermédiare* (intermediate stage): eg. more systematic and regular, with a more or less established finite utterance structure (but still a simple interlanguage system): present tense, *le passé composé* (which may correspond to both the "perfect" and the "preterite") and the future, mostly periphrastic future, non-finite forms less frequent, first cases of the subjunctive, double phrasal negation more or less acquired, non-targetlike forms on determiners and adjectives persist in gender agreement. *English: some possible problems:* I usually listen music and look TV … it is hard to say just one good band … I have few hobbys …. in winter I go to snowboarding … I want to hobby with my dog … I need to my hobby only book …. Accually I have not seen them at many weeks … I have a couple foreigner favourites … Some weeks ago a new boy go in to our class … people beleves in macig … … you don´t need anything special equipment … It is annoying if someone looks other people only from the outside and criticize them …
A2	*Uses some simple structures correctly, but still systematically makes basic mistakes - for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what he/she is trying to say.*	Bartning & Schlyter (2004): *Le stade post-initial* (post-initial stage): eg. both non-finite and finite utterance structure, polyfunctional base forms (the present ffor the past and the future, etc), some inflection on verbs and adjectives, paratactic utterance structure but also the emergence of some subordination, irregular verbs in the 3rd person singular/plural but appear in different non-finite and non-targetlike finite forms, past tense appears with stative verbs and some distinction is made between 1st and 2nd person of the present-tense verbs, gender marking (if used at these two stages at all) may use one (masculine vs. feminine) as the default or overuse the masculine. *English: some possible problems*: I´m very interesting to learn.. I have learn basic things … she is in same years than I`m … I`m better speak svenka than English … Lately time I am went riding … I hobby read every day … I be going to lazy in my next holiday … I did not catched any fish
A1	*Shows only limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.*	Bartning & Schlyter (2004): *Le stade initial* (initial stage): eg. nominal utterance structure, utterances with some formulae, bare nouns but also some determiners and non-finite verbs, preverbal negation and some finite verb forms. *English: correct use of formulas and some possible problems:* My name is xxx… I live in xxx… I have wonderful famil … My family belong parents and sister …

Here are some examples of what I take to be C-level use of English grammar.[12] They are authentic learner productions at the age of 18-19 with ten years of English study.

- It has been said that the youth of today has no sense of morals at all and that society as we know it will fall apart when our generation takes charge.
- It is time to understand that also in our own society there is a need to be engaged in the struggle for better human rights.
- I do not underestimate the significance of animal rights organisations, but I believe there is a lot to be done in order to improve human rights first.
- The capability and desire to communicate that led to writing and eventually books has helped lift us to the top of the evolutionary ladder.

---

[12] *It seems to me that these examples also show a clear interaction between cognitive and linguistic levels.*

- Although you've been the most loyal friend I could ever wish for, I haven't always given you the attention or the appreciation you deserve, and for that I apologize sincerely.
- This is very convenient, of course, for when the problem is out of sight, it is also out of mind.
- I suppose that, standing on the verge of adulthood and facing all the possibilities the future has to offer, it's only natural to feel more or less eager to get it all.
- People are prepared and willing to sacrifice themselves in order to make a difference, in order to improve the conditions of living for all mankind, and they may sometimes even succeed.
- The recent years have shown that, contrary to what we have always thought, we perhaps cannot manage without the help of the outside world.
- Only by having personal contacts across racial boundaries are we able to become more informed and less prejudiced.
- Because we can live only as individuals, as drops of water in a sea of anonymity, a certain amount of selfishness is inevitable. Altruism must be honoured as a virtue, not forced as a command.

It would be very useful to be able to draw on corpus-based analyses about the linguistic properties of proficiency levels. Profile English and the Bartning-Schlyter scheme are good examples of the promise of such an approach. While it is possible that this work will not discover many things that experienced language teachers do not know already, it will be of great value as it will provide empirical corroboration of some beliefs, refutation of some other beliefs and yield information about points not attended to.

4.5 Level of orthography/spelling

The CEFR (2001) scale for orthographic control is given below. It appears relevant but it would also benefit from language-specific exemplification.

Anecdotes about how many spelling variants even simple words can exhibit are commonplace.[13] I recall (but cannot locate the source) that 77 different ways to write "Friday" were reported in a Swedish report on assessment.

---

[13] Howlers recorded e.g. in hotel information all over the world are partly due to spelling mistakes.

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

127

Orthographic control (CEFR 2001, 118)	
C2	Writing is orthographically free of error.
C1	Layout, paragraphing and punctuation are consistent and helpful. Spelling is accurate, apart from occasional slips of the pen.
B2	Can produce clearly intelligible continuous writing which follows standard layout and paragraphing conventions. Spelling and punctuation are reasonably accurate but may show signs of mother tongue influence.
B1	Can produce continuous writing which is generally intelligible throughout. Spelling, punctuation and layout are accurate enough to be followed most of the time.
A2	Can copy short sentences on everyday subjects - e.g. directions how to get somewhere. Can write with reasonable phonetic accuracy (but not necessarily fully standard spelling) short words that are in his/her oral vocabulary.
A1	Can copy familiar words and short phrases e.g. simple signs or instructions, names of everyday objects, names of shops and set phrases used regularly. Can spell his/her address, nationality and other personal details.

My personal experience suggests that weak(ish) learner scripts can occasionally have a fairly large vocabulary but there may be considerable problems of accuracy and appropriateness and spelling can be really faulty. In spite of the fact that the message may succeed if the reader makes an effort (and especially if the reader knows some of the source language), communication is uncertain, ineffective and the script deserves a low mark.

Here are a few examples. They are, however, by far not the worst cases I remember having come across.

- On Hospitals haven't enaf good profesional peaple working.
- Without takin over animals natural habitan's.
- It is the tureth that we needed to know.
- Music was so loudly and noisy, that I almost got hedeake.
- ... witch we have built to hold everything in chape.
- Why are the girls better in school and cinder stundets?
- There is happened a big scange of when othe sex meet eatch other.
- Okey, that present have to bee a fanny, but still useless.
- That trip was memoratable.
- Second appeartunety is that it's not chance nothing.

5 DISCUSSION

I have presented some views about the development of language proficiency level descriptions and referred to several R&D projects. I noted that the development of proficiency (progression) has always interested (and worried) language educators. Traditionally language proficiency has been divided into elementary, intermediate and advanced.[14] These levels have been determined on the basis of experience, drawing on whatever research evidence was available, on how complexity was conceptualized, on how frequent linguistic elements were assumed or known to be. An early example of a systematic attempt is Alexander et al. (1975).

A new stage started when, in addition to traditional attention to grammar, there emerged an interest in what people could do with language, for which purposes their language proficiency was sufficient. The fifty-year engagement by the Council of Europe in modernizing language education in Europe (cf. Trim 2007) led - from the 1970s onward - to the publication of description of objectives: in the case of English - Breakthrough, Waystage, Threshold and Vantage, which combined language functions and the related linguistic resources. The next stage was the application of can-do - statements (CEFR) to illustrate language proficiency in six levels, from basic to proficient language use. However, the CEFR (2001) does not present any breakdown of linguistic exponents, as the CEFR was to serve as the reference applicable for *all* languages.

As the "architecture" of language (especially the fundamental property of combinatoriality) inevitably plays an important role in communication (in conveying and negotiating meaning), it is important to have an adequate theory of language. I believe, based on my own experience in language education, that the approach on "parallel architecture" by Jackendoff (2007)[15] is a promising way to view language and language use. This approach proposes that there are three parallel and partly autonomous combinatorial systems, each of which has its own formation rules: phonological formation rules (phonological structures), syntactic formation rules (syntactic structures) and semantic formation rules (semantic

---

[14] In cognitive psychology these are often described in terms of the novice-expert dimension.

[15] And others working on construction grammar, cognitive grammar, head-driven phrase structure grammar, conceptual semantics etc.

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

129

structures). These semi-independent combinatorial systems are linked with systems of interface principles, some very general and some very specialized. The interface principles allow a variety of mappings, including many-to-many mappings ("dirty correspondence"). On this view, semantic structure needs to be formally much richer than surface syntax (cf. also Wray 2002). Syntax has the linear order of phonology but the embedding structure of semantics. Syntax is, however, not dominant: combinatoriality is not based only on syntax.

As mentioned earlier, in Jackendoff's view words are also rules, albeit particular rules. Lexical elements contain also a great number of multi-word units (e.g. phrasal verbs, collocations, formulas). One consequence of this is that in developing language-specific Reference Level Descriptions the lexicon needs to be incorporated as an inherent component. I think that it is important to stress this as there may be a strong temptation to focus on the more easily managed ("simpler") syntax. In this article I have tried to heed this warning.

What corpus-related action is needed in the future work?[16]
- Sufficiently large and representative learner corpuses in different languages (cf. above).
- All learner texts need to be assessed using the CEFR level descriptors (minimum: 2 trained and experienced raters).
- From every proficiency level established in this way, a small but representative sample is to be selected and these to be rated by at least ten other experienced raters. Using a suitable method (eg. Multi-facet Rasch Measurement, MFRM, or some other preferred method) to examine rater behavior, the results can be extrapolated to the whole double-rated text corpus. This subset itself (ten raters minimum) can also function as a useful benchmark sample provided that there is at least 90% agreement on the level.
- The corpus is analysed in a variety of research projects.
- Pedagogical applications are constructed.
- Pre-service and in-service education is arranged for teachers.

The CEFR (2001) has started a new phase in the development of language education even if it is not the "final word" let alone a "Bible". Proficiency-level thinking with related scales, and the elaboration of their linguistic properties, is a topical challenge, which language educators will surely take up with great interest.

---

[16] SLATE and Cefling are examples of projects that are producing important and interesting information. See e.g. http://www.jyu.fi/hum/laitokset/kielet/kesakoulu2009/programme/plenaries

What about the themes that need research? I will present some research questions with an emphasis on language education:

1) What levels are reached in different language programmes in different languages in our country (and in other countries)? How does language background (L1) influence the attainment of levels (Finnish, Swedish, Sami, various immigration-related languages)? How are the levels reflected in the matriculation examination, other examinations and in national assessment results?

2) What language properties (including errors) are typical at different levels? How are there properties related to the L1 and L2?

3) What is the relative role of grammar and vocabulary in the comprehension of spoken and written language? The traditional view has been that vocabulary is more important (cf. the early interest in vocabulary frequency and vocabulary control in textbooks). Recently Shiotsu and Weir (2007) have drawn a reverse conclusion. I am not convinced by their finding. (At the 2009 EALTA Conference Brunfaut reported on a study which also called in question the conclusion; see www.ealta. eu.org.) I suggest that it is also necessary to take into account both the level of proficiency and the cross-linguistic similarity of the studied languages. It is likely, as Ringbom (2007) has suggested, that close language similarity encourages learners/users to transfer grammatical hypotheses, which means that vocabulary plays an important role in such a case. On the other hand, other things being equal, I assume that at levels from A1 to low B1, grammar and vocabulary are roughly equally important but beyond that vocabulary knowledge is a better predictor of comprehension.

4) What is the size of vocabulary at various stages of education and at different levels? In my doctoral dissertation (Takala 1984), I concluded that 30 years ago, in a streamed teaching of English, the students in the advanced stream knew about an average of 1500 English words at the end of the comprehensive school after seven years of English, the intermediate stream students about 1000 words and the basic stream students about 500 words. There was, thus, a great varia-

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

131

tion, which I expect to be true even today[17]. I also discovered that vocabulary-related skills (compounding, derivation) were surprisingly poorly developed. What is the situation like today? How large vocabulary is acquired in other languages with much less out-of-school exposure? Are the vocabulary-related skills equally poorly developed today? I regret to assume so. As far as I can judge, the "boom" in vocabulary R & D (Takala, 1989) appears to have passed in a few years.

I hope that there will be research-based information on such – and related – topics in the near future. Such research would have at least one avid reader.

---

[17] But I expect the vocabulary sizes to be considerably larger, due to the "explosion" of exposure to English in the Finnish society. Today, it is unlikely that there are true beginners when the formal study of English begins: even young children may know some, occasionally, quite a lot of English.

REFERENCES

Abrahamsson, N. 2009. Andraspråksinlärning. Lund: Studentlitteratur.

Abrahamsson, N. & Hyltenstam, K. 2009. Age of acquisition and nativelikeness in a second language: listener perception vs. linguistic scrutiny. Language Learning 59, 3, 249-306.

Alexander, L.G., Stannard Allen, W., Close, R. A. & O'Neill, R. J. 1975. English grammatical structure. A general syllabus for teachers. London: Longman.

Bartning, I. & Schlyter, S. 2004. Itinéraires acquisitionnels et stades de développement en français L2. Journal of French Language Studies, 14, 281-299.

Birdsong, D. 2004. Second language acquisition and ultimate attainment. In A. Davies & C. Elder (Eds.) The Handbook of Applied Linguistics. London: Blackwell, 82-105.

Birdsong, D. 2007. Nativelike pronunciation among late learners of French as a second language. In O.-S. Bohn & M. J. Monro (Eds.) Language experience in second language speech learning. In honor of James Emil Flege. Amsterdam: John Benjamins, 99-116.

Cizek, G. J. & Bunch, M. B. 2007. Standard setting. A guide to establishing and evaluating performance standards on tests. Thousand Oaks: Sage.

CEFR = Common European Framework of Reference for Languages. Learning, teaching, assessment. 2001. Council of Europe. Cambridge: Cambridge University Press.

Corder, S. P. 1967. The significance of learners' errors. International Review of Applied Linguistics 5, 161-169.

Corder, S. P. 1981. Error analysis and interlanguage. Oxford: Oxford University Press.

Davies, A. 2003. The native speaker: Myth and reality. Clevedon, UK: Multilingual Matters.

de Bot, K., Lowie, W. & Verspoor, M. 2005. Second language acquisition. An advanced resource book. London: Routledge.

Doughty, C. J. & Long, M. H. (Eds.) 2003. The handbook of second language acquisition. Oxford: Blackwell.

Dörnyei, Z. 2005. The psychology of the language learner. Individual differences in second language acquisition. Mahwah, N.J: Erlbaum.

Ellis, N. C. (Ed.) 1994. Implicit and explicit learning of languages. New York: Academic Press.

Ellis, N.C. 2002. Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. Studies in Second Language Acquisition, 24, 143-188.

Ellis, N. C. 2008. Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 endstate. In P. Robinson & N. C. Ellis (Eds.) Handbook of cognitive linguistics and second language acquisition. London: Routledge, 372-405.

Fillmore, C. J., Kay, P. & O'Connor, M. 1988. Regularity and idiomaticity in grammatical constructions: The case of "let alone". Language 64, 501-538.

Takala, S. Linguistic Features at Different Levels of Language Proficiency
Some Facts, Assumptions and Suggestions, 103-134

133

Forsberg, F. 2008. La langage préfabriqué – forms, fonctions et frequencies en français parlé L2 et L1. Bern: Peter Lang.

Granger, S. 1998. Prefabricated patterns in advanced EFL writings: collocations and formulae. In A. P. Cowie, (Ed.) Phraseology: theory, analysis and applications. Oxford: Clarendon Press, 145-160.

Hendricks, H. & Saville, N. 2009. Presenting the English profile programme. In 2009 LTTC International Conference on English language Teaching and Testing, 75-86.

Hylstenstam, K. & Abrahamsson, N. (2003) Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.) The Handbook of Second Language Acquisition. London: Blackwell, 539-588.

Jackendoff, J. 2007. Language, Consciousness, Culture. Essays on Mental Structure. Cambridge, Mass.: The MIT Press.

Kaftandjieva, F. 2004. Standard setting. In S. Takala (Ed.) Reference supplement to the preliminary pilot version of the Manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment. Section B. Strasbourg: Council of Europe/Language Policy Division, 1–43.

Kaftandjieva, F. 2005. Linking Curricula and Exams to the CEFR. Paper presented at Cambridge University Press EOI Seminar (Barcelona, March 11, 2005).

Kane, M. 2001. So much remains the same: conceptions and status of validation in standard setting. In G. J. Cizek (Ed.) Setting performance standards: concepts, methods, and perspectives. Mahwah, N.J: Erlbaum, 53-88.

Kelly, L. G. 1969. 25 centuries of language teaching. Rowley, Mass.: Newbury House.

Krashen, S. D. & Scarcella, R. 1978. On routines and patterns in language acquisition and performance. Language Learning 28, 283-300.

Lantolf, J. P. & Thorne, S. L. 2007. Sociocultural theory and second language learning. In B. van Patten & J. Williams (Eds.) Theories in second language acquisition. An introduction. Mahwah, N.J.: Erlbaum, 201-224.

Larsen-Freeman, D. & Cameron, L. 2008. Complex systems and applied linguistics. Oxford: Oxford University Press.

Long, M. H. 2007. Problems in SLA. New York: Erlbaum.

North, B. 2000. The development of a common European framework scale of language proficiency. New York: Peter Lang.

Ochs, E. & Schieffeling, B. 2006. The impact of language socialization on grammatical development. In C. Jourdan, & K. Tuite, (Eds.) Language, Culture, and Society. Cambridge: Cambridge University Press, 168-189.

Ortega, L. 2009. Understanding second language acquisition. London: Hodder Education.

Pawley, A. & Syder, F. 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.) Language and communication. London: Longman, 191-226.

Piller, I. 2002. Passing for a native speaker: Identity and success in second language learning. Journal of Sociolinguistics, 6, 179-206.

Purpura, J. L. 2004. Assessing grammar. Cambridge: Cambridge University Press.

Read, J. 2000. Assessing vocabulary. Cambridge: Cambridge University Press.

Ringbom, H. (Ed.). 1993. Near-natived proficiency in English. English Department Publications 2. Åbo: Åbo Akademi University.

Ringbom, H. 2007. Cross-linguistic similarity in foreign language learning. Clecedon, UK: Multilingual Matters.

Robinson, P. 2005. Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. International Review of Applied Linguistics, 43, 1-32.

Robinson, P. & Ellis, N. C. (Eds.) 2008. Handbook of Cognitive Linguistics and Second Language Acquisition. New York and London: Routledge.

Schmitt, N. (Ed.) 2004. Formulaic sequences: Acquisition, processing and use. Amsterdam: Benjamins.

Selinker, L. 1972. Interlanguage. International Review of Applied Linguistics 10, 209-231.

Shiotsu, T. & Weir, C. 2007. The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. Language Testing 24, 1, 99-128.

Takala, S. 1984. Evaluation of students' knowledge of English vocabulary in the Finnish comprehensive school. Jyväskylän yliopisto. Kasvatustieteiden tutkimuslaitos. Tutkimuksia 350.

Takala, S. (Ed.) 1989. Sanaston opettaminen ja oppiminen. [Vocabulary teaching and learning; in Finnish] Jyväskylän yliopisto. Kasvatustieteiden tutkimuslaitos. Teoriaa ja käytäntöä 44.

Tomasello, M. 2003. Constructing language. A usage-based theory of language acquisition. Cambridge, Mass: Harvard University Press.

Trim, J. L. M. 2007. Modern Languages in the Council of Europe 1954-1997. Strasbourg: Language Policy Division.

Tuokko, E. 2007. Mille tasolle perusopetuksen englannin opiskelussa päästään? Perusopetuksen päättövaiheen kansallisen arvioinnin 1999 Eurooppalaisen viitekehyksen taitotasoihin linkitetyt tulokset. [What level do pupils reach in English at the end of the comprehensive school?; in Finnish] Jyväskylä: Jyväskylä University Printing House.

van Patten, B. & Williams. J. 2007. Theories in second language acquisition: An introduction. Mahwah, N.J.: Erlbaum.

Wong-Fillmore, L. 1976. The second time around: cognitive and social strategies in second language acquisition. PhD thesis, Stanford University.

Wray, A. 2002. Formulaic language and the lexicon. Cambridge: Cambridge University Press.

Yorio, C. 1989. Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam & L. Obler (Eds.) Bilingualism across the lifespan. Cambridge: Cambridge University Press, 55-72.

Zieky, M. J., Perie, M. & Livingston, S. A. 2008. Cutscores: A manual for setting standards on educational and performance tests. Princeton, NJ: Educational Testing Service.