Research Notes

Contents

Editorial Notes

Welcome to issue 33 of *Research Notes*, our quarterly publication reporting on matters relating to research, test development and validation within Cambridge ESOL.

This issue focuses on *English Profile*, a collaborative programme of research, consultation and publication, designed to enhance the learning, teaching and assessment of English worldwide. In this issue we describe how *English Profile* came about, its academic and institutional partners and its three research strands, with contributions from project partners and researchers from the growing number of English Profile Networks.

In the introductory article, Svetlana Kurteš and Nick Saville describe the birth of *English Profile*, a programme rooted in – and building on – the Common European Framework of Reference (CEFR) and other Council of Europe initiatives. They outline the approach to producing Reference Level Descriptions (RLD) begun by Cambridge ESOL and describe current research projects and the various events through which we disseminate research findings. Next, Angeliki Salamoura discusses how research data being collected for *English Profile* can be aligned to the CEFR.

The next three articles describe work underway at Cambridge University where researchers are developing new and extending existing corpora in order to investigate criterial features at each proficiency level within the Corpus Linguistics research strand. Henriëtte Hendriks outlines a research agenda that links language acquisition research and computational approaches to the analysis of learner data, based on an enhanced version of the Cambridge Learner Corpus – Cambridge ESOL's and Cambridge University Press' unique written corpus. Caroline Williams describes the challenges that learner writing pose to an automatic parser, concentrating on English word order errors of Spanish and Chinese speakers. Theodora Alexopoulou then reports on the collection of written materials from classroom settings around the world to form new corpora to supplement the Programme's existing learner data.

We next focus on the Curriculum and Assessment research strand. Anthony Green reports on functional progression in English language teaching materials, based on a survey of documentation including test specifications and coursebooks. Radmila Bodrič investigates the impact of a common European language policy on language teaching in the Serbian context, setting out what is expected of today's teachers.

Two English Profile Network members then consider culture-specific aspects of language which relate to the Language Pedagogy research strand. JoAnne Neff-van Aertselaer compares persuasive texts written in English and Spanish to identify differences and similarities between texts, languages and cultures. Tatiana Larina compares directness, imposition and politeness in English and Russian and describes how they influence intercultural communication.

We finish with a report from the ALTE 2008 conference.

Editorial team for Issue 33: Fiona Barker, Svetlana Kurteš and Kirsty Sylvester, with Stephen McKenna and Chris Lewis.

The English Profile Programme – an overview

SVETLANA KURTEŠ ENGLISH PROFILE CO-ORDINATOR, CAMBRIDGE ESOL
NICK SAVILLE DIRECTOR, RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL

Introduction

English Profile is a collaborative programme of interdisciplinary research, set up to produce Reference Level Descriptions (RLD) for English linked to the general principles and approaches of the Common European Framework of Reference (CEFR; Council of Europe 2001). The RLDs are envisaged as a core set of reference tools for English as a foreign or additional language.

At the start of this issue it is worth outlining a brief history and structure of *English Profile*, as this will give a broader context of the work outlined in the following articles and give some indication of the many parts which make up the whole of this ambitious programme of research and publication. *English Profile*'s approach is rooted in collaboration and innovation; understanding how this project came into being ultimately gives some indication of its future potential as a series of tools for teaching, learning and assessing English.

The task of giving a concise overview of the development of *English Profile* and outlining where it currently stands is not an easy one: the programme is constantly evolving; the parties involved are diverse, and although there are a number of projects already under way, *English Profile* is only just beginning to achieve some of its eagerly anticipated potential.[1]

Having defined its core values and intentions and formed the partnerships and networks to achieve its goals, *English Profile* is currently going through a period of consolidation and is entering a new cycle in its development, shifting from the planning stage to focusing on the achievement of its key aims and delivering concrete results.

The birth of CEFR

While *English Profile* is by definition an extension of the CEFR and proceeds from that Framework, the CEFR itself was, in turn, the culmination of work first proposed in the 1970s (see Trim 2007) and carried out over the following two decades (Council of Europe 2002). The *English Profile* approach is as much rooted in the work which informed the CEFR as it is by the CEFR itself.

The foundation stones of the CEFR can be discovered in work sponsored by the Council of Europe's Modern Languages Project as methods of describing functional language competence in adults: the *Threshold*, *Waystage* and later the *Vantage* levels developed by Professor Jan van Ek and Dr John Trim (1991a, 1991b, 2001).

In the mid 1990s, the Council of Europe began the work of finding a way to collate these different levels into a coherent framework. It is beyond the scope of this short article to detail the development of the CEFR, however, once published, the CEFR's six reference levels became widely accepted. Meanwhile, independently and almost simultaneously, the Association of Language Testers in Europe (ALTE) had developed another scale of levels with the intention of creating a unified framework. This scale became harmonised with the CEFR after 2001 (see Jones 2002).

It is significant that Professor van Ek and Dr Trim, while developing specifications of objectives intended to have wider implications for language learning in Europe, were writing *in English*, and used English as the example in their *Waystage*, *Threshold* and *Vantage* series. This series has been acknowledged as existing RLDs for English at the lower levels and the *English Profile* programme's initial focus on the CEFR's higher 'C' levels recognises that significant work has already been achieved in the past on the lower levels (A2–B2 in particular), indicating an existing gap for a description of the higher levels.

A new approach

Cambridge ESOL's senior representatives, who had been participating in Council of Europe committees on language policy, realised that Cambridge ESOL's experience in research and development for assessment purposes, together with research being carried out within Cambridge University, gave them the expertise and resources to take on the challenge of producing RLDs for English.

The Goethe Institute had already published its RLDs for German, *Profile Deutsche*, (German Profiles) and while that was influential – not least in the choice of name – *English Profile* – the Cambridge ESOL approach was different from the start. A document was drafted in 2004 proposing *English Profile* in outline. It featured a number of key points including:

- an empirical approach
- an agenda rooted in data-driven research
- greater flexibility, accessibility and feedback involving up-to-date technology.

These elements remain at the heart of the Programme today.

The outline approach was presented at Council of Europe meetings on the RLD developments and received encouragement to continue its progress. The British Ambassador to the Council of Europe, via the British Council, also endorsed the project when it began to gather momentum in 2005.

One of the key qualities which had recommended the proposals to the Council of Europe had been the intention to make the Programme an international one, and to avoid

1. We are grateful to Chris Lewis for his contributions to this article.

a purely Anglocentric approach. This was important to reflect the unique position of English as a global language. From the start it was a stated aim of *English Profile* to seek other research partners and to involve stakeholders from within the field of English language learning, teaching and assessment around the world.

Some of the features of *English Profile* which distinguished it from previous work in this field included:

- research based on electronic corpora of learner data, producing results which can be empirically measured and that are not predictable from current language acquisition theories alone.
- incorporating psycholinguistic factors, complexity metrics and frequency measures, in addition to the more traditional linguistic features such as grammatical and lexical components.
- focusing strongly on the impact of different first languages and learning contexts and the effects of language transfer.

The core group

A group of six partners working in related fields was established to get the project off the ground:

- British Council
- Cambridge University Press
- English UK
- University of Cambridge ESOL Examinations
- University of Cambridge Research Centre for English and Applied Linguistics (RCEAL)
- University of Bedfordshire Centre for Research in English Language Learning and Assessment (CRELLA).

From the outset, the diversity of the core group was held to be one of its main strengths, and in the early meetings a series of projects and working groups were set up to reflect three key research strands: Corpus Linguistics, Language Pedagogy, and Curriculum and Assessment[2]. There are articles representing each of these strands in this issue, outlined below.

English Profile Networks

From its outset, *English Profile* was a partnership based on collaboration and with an inclusive agenda. As the Programme has moved from scoping and planning its work into the research projects that will lead to publishable results, the importance of bringing in further partners has grown. The English Profile Network is now becoming an extended base of academics, government advisors and educationalists with an interest in the English language. Members may contribute directly to the development of *English Profile* by providing critical reviews of work in progress, participating in the research programme, promoting *English Profile* projects, disseminating the results and engaging with those who teach and test English using the CEFR.

Because of the international framing of *English Profile* as a programme and also the diversity of the approaches within it, members are able to take part in the networks according to their interests and abilities. Some are now forming regional networks, to focus on matters relating to English language learners in a specific region or with a specific linguistic background. Such regional English Profile Networks are being established in a number of countries and it is hoped these will provide a focus for discussion of key issues relating to the research questions which *English Profile* is investigating and eventually become model platforms for curriculum innovation and reform. While some may choose to become part of the Network based on geographic considerations, others may take a thematic interest in *English Profile* and wish to contribute to one of the research strands according to their academic or technical abilities and resources.

Also, as the projects develop, it is becoming increasingly clear that more data needs to be collected and developed into corpora. (See Dr Theodora Alexopoulou's article on building new corpora, page 15 and Dr Angeliki Salamoura's article on evaluating the CEFR level of new data, page 5). The Network members will play a key role in collecting this data which will be critical to delivering empirically measurable results and conclusions.

Current projects

The Corpus Linguistics Research Team, based in RCEAL and led by Professor John Hawkins and Dr Henriëtte Hendriks, is working on two areas of direct relevance to *English Profile*: 1) A set of criterial features that characterise and distinguish the six levels of the CEFR with respect to English, and 2) The impact of different first languages on performance at each of the levels and their interaction with the criterial features. (See articles by Dr Hendriks and Caroline Williams on pages 7 and 10).

The Language Pedagogy research strand (lead by Dr Svetlana Kurteš) focuses on pedagogically oriented contrastive linguistics, the grammar-pragmatics interface and questions surrounding their didacticisation across cultures and national teaching practices. Plans are well under way to establish an international research group that will look into culture-specific differences in written and spoken academic discourse. (See articles by Professor JoAnne Neff-van Aertselaer and Professor Tatiana Larina on pages 28 and 33 for example).

Cambridge ESOL, together with colleagues from CRELLA – Professor Cyril Weir and Dr Tony Green – are currently working on three longer-term projects that involve closer collaboration with other researchers, both from the current team and the outer network of collaborators. The focus of their research is EFL curricula and assessment (see Dr Green's article on page 19, also Dr Radmila Bodrič's article on page 25).

English Profile: the future

Following the initial launch of the programme of research, there has recently been a period of consolidation, where the work of *English Profile* has shifted emphasis from the

2. These research strands will expand to include other areas as the Programme develops.

setting up of systems and partnerships to now focusing on outcomes. This has involved a restructuring of the team's activities into the following three areas: research and development, setting up a collaborative network, and promotional activities.

Research and development work

Research and development work is gaining momentum, with several research projects either already underway, or in preliminary or preparatory stages. Corpus Linguistics and Language Pedagogy and Assessment Research Teams are expected to continue to collaborate on a number of research questions.

Given the interdisciplinary nature of the research to be undertaken, research teams will engage in parallel and simultaneous investigations on a set of related research questions, observing them from different angles and/or fields of study. Preliminary/interim results, once obtained, will inform further directions of research. In addition, plans are underway to set up new research strands, primarily to focus on issues related to teacher training and other related fields of study.

Setting up and extending a collaborative network

This continues to be one of the priorities for 2008–9. It will be achieved through seminars and other promotional activities that will help the team to identify colleagues and contacts working in the same field who are willing to contribute to *English Profile* by:

- Participating directly in the research programme.
- Becoming local or regional co-ordinators who will promote the Programme, disseminate its results and engage with those who teach and test English using the CEFR and consequently become model platforms for curriculum innovation and reform elsewhere in the region or country.
- Helping the Programme team to define the outcomes they are working towards and thus help ensure that the outcomes of the research programme meet the needs of teachers and learners in their country.
- Contributing data and insights that will enrich the pedagogical tools that are being developed, etc.
- Being a regional focus for international studies on culture-specific differences.

One key event on the calendar of promotional activities will be the launch of *English Profile* in south-eastern Europe at a two-day seminar in Herceg Novi, Montenegro in September 2008.[3]

3. See www.EnglishProfile.org for further information.

Conclusion

Looking beyond the mid-term prospects, the ultimate shape of *English Profile* is now becoming much clearer. It seems at this point unlikely that *English Profile* will take the form of a monolithic publication of one set of RLDs. It would more probably be a series of descriptors covering all levels and communicative/functional domains and involving all specific areas of English. These could form a mosaic – discrete pieces of work, which together create a whole image of the diversity of the English language globally.

Rapid technological change and the increased mobility that globalisation brings will also mean that English will continue to change rapidly, requiring further revision of existing work and new projects to be started. *English Profile*, therefore, has the potential to be not merely a project with a beginning, a middle and a single final outcome, but a permanently ongoing programme, mapping English as it changes, and maintaining a record of those changes for study by future generations.

This issue of *Research Notes* contains an overview of work underway and suggests avenues for future collaboration. At the time of going to press we are looking forward to the sixth *English Profile* seminar to be held in Cambridge in July which promises to be a stimulating event for invited participants from the core group and members of the growing number of English Profile Networks.

References and further reading

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

— (2002) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment Case Studies*, Strasbourg: Council of Europe Publishing.

Jones, N (2002) Relating the ALTE Framework to the Common European Framework of Reference: in Council of Europe, 167–183.

North, B (1995) The Development of a Common Framework Scale of Descriptors of Language Proficiency Based on a Theory of Measurement, *System*, 23, 445–65.

Taylor, L and Jones, N (2006) Cambridge ESOL exams and the Common European Framework of Reference (CEFR), *Research Notes* 24, 2–5.

Trim, J (2007) Presentation to the English Profile Seminar, Cambridge, February 2007, retrieved from www.englishprofile.org/threshold_background.html

Van Ek, J A and Trim, J L M (1991a) *Threshold 1990*, Cambridge: Council of Europe/Cambridge University Press.

— (1991b) *Waystage 1990*, Cambridge: Council of Europe/Cambridge University Press.

— (2001) *Vantage*, Cambridge: Council of Europe/Cambridge University Press.

Erratum

In issue 32 (May 2008) on page 29 Dr Philida Schellekens' affiliation was incorrectly given as the National Research and Development Centre for Adult Literacy and Numeracy, UK. She is in fact an independent consultant. We apologise for any misunderstanding.

Aligning English Profile research data to the CEFR

ANGELIKI SALAMOURA RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL

English Profile and the CEFR

One of the principal aims of the *Common European Framework of Reference for Languages* (CEFR; Council of Europe 2001) is to function as a kind of common language, a 'lingua franca' in the modern languages field that allows language courses, curricula, syllabuses, teaching and learning materials, tests and assessment systems to refer back to a common scale of measurement, thus enhancing comparability and transparency (see also Saville 2004:281). Since the CEFR is not language specific, this aspiration can be achieved to its full extent with the exemplification of the CEFR in individual languages. *English Profile* will contribute to this aim by developing Reference Level Descriptions (RLD) for the English language. The RLDs will detail the features of learner language in English that correspond to the six levels of the CEFR. (For more information on RLDs see Council of Europe 2008a and the *English Profile* website). Critically, the RLDs developed within *English Profile* will draw on *empirical* research data both already in existence (from the Cambridge Learner Corpus) and collected *specifically for English Profile* from a variety of English language learners and learning materials (learner input and output) from around the world.

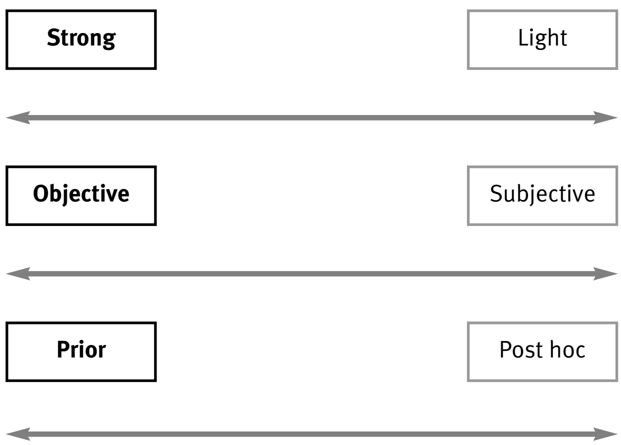
As *English Profile* seeks to illustrate the CEFR levels, any data that will inform this project need to be aligned to the CEFR scale. More importantly, it is imperative that this alignment is done in a systematic, principled way based on good practice and empirical evidence rather than intuition to ensure the validity of this process. The remainder of this article considers some methodological issues of relating research data to the CEFR and outlines *English Profile*'s approach to this activity.

How to align English Profile research data to the CEFR?

English Profile research has thus far drawn heavily on exam data derived from the Cambridge Learner Corpus which contains output from English language learners taking Cambridge ESOL exams.[3] One of the advantages of using exam data which come from tests closely related to the CEFR, such as the Cambridge ESOL examinations (cf. Taylor and Jones 2006), is that one can investigate the language skills of learners whose level of language proficiency has already been measured against the CEFR levels. Currently, plans are under way to enrich the type of input to *English Profile* by collecting a wide range of non-exam samples of learner language that will form a number of further learner corpora: classroom discourse (written and later spoken), naturalistic discourse, academic English discourse and so on (see Alexopoulou's article on page 15). Ellis and Barkhuizen (2005) provide a useful discussion of the theoretical and methodological issues that accompany the collection and analysis of different types of learner language (e.g. experimentally-elicited vs. naturally-occurring samples). In our context, collecting non-exam data (from a variety of sources, learner groups, teaching and learning settings, etc.) poses the following methodological challenges: How can one align research data to the CEFR following a methodology that will allow one to make valid claims about this linkage? What kind of evidence is relevant to demonstrating alignment, and how can it be collected?

The answers to the above questions are certainly not straightforward. The linking of research data to the CEFR is currently uncharted territory. The *Manual for Relating Language Examinations to the Common European Framework of Reference* published by the Council of Europe in 2003 can function as a starting point but it is only partially relevant as it is tailor-made for relating *exams* to the CEFR. In systematising the linking of research data to the CEFR, it would be perhaps best to view alignment, and in particular evidence for alignment to the CEFR, as a continuum rather than as all-or-nothing (Figure 1). As Figure 1 illustrates, sometimes we may have strong evidence for alignment and in some other cases we may have only light evidence due to the nature of the data. Data that derive from exams which are already aligned to the CEFR will undoubtedly carry strong evidence for alignment whereas non-exam data may hold lighter evidence. For instance, when one can obtain only self-assessment statements or teacher assessment in relation to what learners can or cannot do in a second language across the CEFR scale, this evidence will not be as strong as evidence about these learners' CEFR level which originates from a reliable CEFR-aligned exam.

Figure 1: Alignment to the CEFR as a continuum



1. See www.cambridge.org/elt/corpus/learner_corpus.htm for further information.

Another dimension along which evidence for CEFR alignment may vary is objectivity/subjectivity. Evidence built on oral and written performance data (e.g. interviews, written essays or open-ended questions) depends on judgement and is, by definition, relatively subjective. Finally, one could collect evidence for the CEFR alignment of research data prior to collection or post hoc or (ideally) use a combination of both methods. For instance, prior to data collection, one could examine whether the curriculum/syllabus or the textbook(s) which the learners under investigation follow are linked to the CEFR and at which level. That would be a first indication of the CEFR level of the research data to be collected. This evidence can be complemented and enhanced after data collection when, for example, the data collected are assigned a CEFR level by expert judges.

Aligning research data to the CEFR: an example

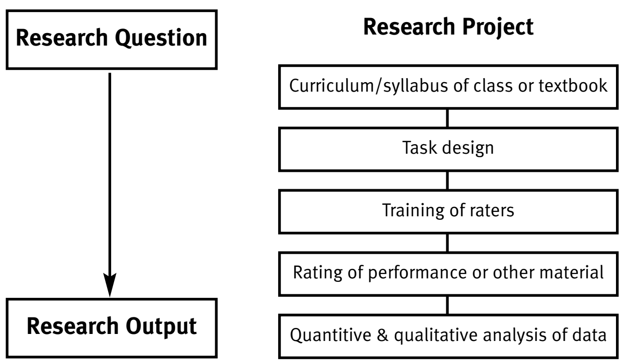
Let us now discuss in more detail an example of research data collection and ways of ensuring alignment to the CEFR. Let us assume that we want to collect written discourse from a classroom setting (e.g. responses to classroom writing tasks). As mentioned previously, before data collection one could find out or analyse the CEFR level of the curriculum/syllabus of the English language course the class follows or the textbook(s) it uses, or both. One could also check if the learners have recently passed an exam or sat a placement test aligned to the CEFR. Another starting point would be to administer assessment using Can-do statements that relate to the Framework, e.g. the CEFR Self-assessment Grid (Table 2, Council of Europe 2001, 26–7,). This assessment can be done by learners themselves (self-assessment) or by teachers who will assess their students' CEFR level. Having, by now, a good indication of the learners' CEFR level, this information can be fed into the design or selection of appropriate classroom writing tasks which should target the learners' CEFR level(s) as closely as possible. The CEFR scales for written production (Council of Europe 2001, 61–5) and the *Manual's* Table 5.8 *Written Assessment Criteria Grid* (Council of Europe 2003, 82) can be a useful starting point at this stage. The CEFR level of the designed task(s) may be further validated by obtaining ratings from expert CEFR raters. Expert CEFR raters can also be recruited to rate the resulting learner written performances after data collection. These raters could be professional writing examiners, item writers, teachers, researchers or other independent markers trained in using the relevant CEFR scales.

As it has probably become evident by now, linking data to the CEFR can be a highly judgemental process. The quality of judgements is, therefore, a decisive factor in this process. In order to ensure the quality of judgements a thorough training of the raters is required. But how can one train people to become expert at applying CEFR scales to learner performances and language tasks? Cambridge ESOL, one of the partners in *English Profile*, was involved in the development of the CEFR in the 1990s, as well as in the piloting of the CEFR linking procedures detailed in the *Manual*. As a result, it has developed expertise around issues of CEFR alignment. Moreover, the alignment of Cambridge ESOL tests to the CEFR is a process integrated into every stage of the design and administration cycle of a test and familiarisation with the CEFR is a necessary part of the induction of test developers in the organisation. Cambridge ESOL has thus built a repository of self-access and face-to-face workshop materials that aim to familiarise inductees with the content of the CEFR scales and to train them in using these scales to rate learner output and tasks.[2] The Cambridge ESOL CEFR training materials build on a variety of *Manual* familiarisation and training activities (see Council of Europe 2003, 25–8; 71–86) but they are also complemented with non-*Manual* mandated activities to suit the organisation's specific needs.[3] The *English Profile* team will be drawing on these materials for CEFR familiarisation and training purposes.

A model for aligning research data to the CEFR

Figure 2 presents a preliminary model for aligning research data to the CEFR which summarises how the *English Profile* team approaches this activity. The alignment is not viewed as an all-or-nothing procedure. It is incorporated in all stages of an *English Profile* research project – from the research question through to the research output. In a research project that involves data collection, CEFR alignment starts by judging the CEFR level of the curriculum/syllabus and/or the textbook(s) of the class in question. It informs the task design and materials selection of the project and subsequently the training of raters who will judge the CEFR level of the collected data. It is further consolidated by the rating of the resulting performances or other material by the trained CEFR raters and, of course,

Figure 2: Preliminary model for aligning research data to the CEFR



2. For more information on a recent Cambridge ESOL CEFR Familiarisation and Training Workshop see: www.cambridgeesol.org/what-we-do/newsroom/2007/cefr_alignment.html
3. Familiarisation activities include assessing one's ability in foreign language(s) using the CEFR Self-assessment Grid highlighting keywords that define salient differences between adjacent levels; and sorting out mixed-up descriptors from a CEFR scale followed by feedback and discussion. Materials used for training raters to apply CEFR scales to learners' performances and language tasks include the Council of Europe's CEFR-calibrated sample performances and tasks (Council of Europe 2008b).

by the qualitative and quantitative analyses of the collected data. CEFR alignment is also viewed as an iterative process – the more empirical evidence one can collect, the stronger the linking claim between data and the CEFR; the more one knows about the CEFR levels in relation to empirical data, the more one can use this knowledge to assess new data or re-assess existing data, thus improving the alignment procedures.

Conclusion

In summary, the *English Profile* team is working towards developing an approach to CEFR alignment that is flexible and adaptable enough to cover all different types of research projects within the Programme; but most importantly, the team is also working towards an approach that will allow them to make valid claims about the linkage of *English Profile* research data to the CEFR levels. The different issues and dimensions of CEFR alignment discussed in this article (cf. Figure 1) are some of the many points that *English Profile* researchers take into account when designing and implementing their studies in order to ensure that *English Profile* outcomes are clearly defined along the CEFR continuum.

References and further reading

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.[4]

— (2003) *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment, Manual: Preliminary Pilot Version*, Strasbourg: Council of Europe Language Policy Division.

— (2008a) Reference Level Descriptions (RLD) for national and regional languages, retrieved from www.coe.int/T/DG4/Linguistic/DNR_EN.asp#P22_1605

— (2008b) Material illustrating the CEFR levels, retrieved from www.coe.int/T/DG4/Portfolio/?L=E&M=/main_pages/illustrationse.html

Ellis, R and Barkhuizen, G (2005) *Analysing Learner Language*, Oxford: Oxford University Press.

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.

Saville, N (2005) An interview with John Trim at 80, *Language Assessment Quarterly* 2/4, 263–288.

Taylor, L and Jones, N (2006) Cambridge ESOL exams and the Common European Framework of Reference, *Research Notes* 24, 2–5.

4. Council of Europe publications can be downloaded from www.coe.int/t/dg4/linguistic/Publications_EN.asp

Presenting the English Profile Programme: in search of criterial features

HENRIËTTE HENDRIKS RESEARCH CENTRE FOR ENGLISH AND APPLIED LINGUISTICS, UNIVERSITY OF CAMBRIDGE

Introduction

In 2005, the six *English Profile* partners – including Cambridge ESOL, Cambridge University Press, RCEAL (University of Cambridge) and CRELLA (University of Bedfordshire) – embarked on a new project regarding L2 learners of English, and the description of their proficiency levels. The *English Profile Project*, as it was soon named, set out to look at the data of the *Cambridge Learner Corpus* (hereafter CLC) in order to look for criterial features in the productions of these learners during Cambridge exams, and to thereby get a better understanding of the various proficiency levels, up to now described only in terms of Can-do statements.

This article proposes to give the reader some background information regarding the reasons for embarking on this project, followed by a short presentation of the research questions as studied by the researchers at the Research Centre for English and Applied Linguistics. It will finish by introducing two articles resulting from the research up to now, and published in this issue of *Research Notes*. The Centre currently has three running research projects connected to the quest for criterial features in the data at two different linguistic levels. The first project, involving Professor John Hawkins, Dr Teresa Parodi and Dr Dora Alexopoulou, looks at sentence level features. The second project, involving Dr Henriëtte Hendriks, looks at discourse level features. A third project looks at eliciting new data complementary to the *CLC*, and is run by Dora Alexopoulou. All three projects can only take place because of the extensive help from Dr Paula Buttery (computational linguistics), Caroline Williams (PhD student) and the tagging and parsing programmes by Professor Ted Briscoe (University of Cambridge Computer Laboratory.

The *CLC* is a large learner corpus of written scripts from Cambridge exams as taken all over the world. The current size of the corpus is around 27 million words and it has some important features very attractive to second language researchers, and not found in any other L2 learner corpora. First of all, the sheer size of the corpus is not easily replicated by any other corpus. Second, it is cross-linguistic in nature, allowing us to compare different pairs of source and target languages. In total, 91 source languages are represented, over 20 of which are represented by large enough samples to be used for quantitative analyses. Moreover, the whole corpus has been tagged and parsed, such that it allows us to test hypotheses in a more sophisticated way than with un-parsed corpora. Finally, the learner corpus provides data in already attested proficiency levels (from A2 to C2 on the Common European Framework of Reference).

Background to the project

In 2001, after a considerable time of deliberation, research and writing, the Council of Europe published the Common European Framework of Reference (hereafter CEFR), a work intended to provide a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc., across Europe. The CEFR was created in the light of the Council's wish to promote awareness of a European identity and thereby to stimulate plurilingualism of the individual as an important part of that identity. CEFR defines plurilingualism as 'the ability to use languages for the purposes of communication and to take part in intercultural interaction, where a person, viewed as a social agent has proficiency, of varying degrees, in several languages and experience of several cultures.' Plurilingualism is promoted in the hope that it will provide linguistic tolerance, and as an essential element of intercultural education (Beacco & Byram 2007,17–8). The European Commission in 1995 proposed the rule of thumb of 'mother tongue plus two', according to which all EU citizens should be proficient in two EU languages besides their own.

As David Little (2007) explains, the CEFR is not language-specific. It describes the communicative functions that learners should be able to perform at different proficiency levels, but does not specify how those functions might be realised in, say, French, German, English or Danish. CEFR's proficiency descriptors are always positive, never referring to what the learner cannot do. They also aim to be definite, clear, brief, and independent. The CEFR was embraced immediately after its publication, but mainly by language testing organisations such as Cambridge Assessment, and the Association of Language Testers in Europe (ALTE) of which Cambridge Assessment is a member. Further intended influence, for example on official curricula in schools across Europe and on curriculum guidelines, has been much slower.

Again immediately after the publication of the CEFR, critique was uttered regarding some issues which were recently presented in a coordinated fashion in *The Modern Language Journal* (2007) by a number of researchers involved or not involved with the birth of CEFR. Questions were raised regarding:

1) The empirical basis of the CEFR.
2) The validity and psychological reality of the six levels of proficiency. In particular, as Hulstijn argues, it is not clear that second language learners necessarily go through all 6 proficiency levels as described by CEFR when acquiring a second language. If indeed this is not the case, then the predictive powers of CEFR are in question (Hulstijn 2007). In similar fashion, it is also not clear if all linguistic features to be acquired follow a linear progression from one proficiency level to another, again calling in question the predictive powers of CEFR (rather, current research mostly suggests that the latter is not the case).
3) The cross-linguistic matching of proficiency levels. Is it indeed possible to state that certain functions indicate a given level of proficiency, regardless of the language in which the function has to be expressed?

A very clear and straightforward answer can be formulated with respect to question number 1: The descriptors in CEFR were developed on the basis of an empirical research project (North 2000; North and Schneider 1998) that used extensive qualitative research with groups of language teachers as informants (North 2007). As a result, the descriptors are carefully empirically calibrated with mathematical values on a common scale and known statistical properties. However, they are not based upon any data available from second language acquisition research. Thus, what is described are teachers' perceptions of language proficiency, not validated descriptions of second language acquisition processes, because, at the time the CEFR was developed, second language acquisition research was not in a position to provide these descriptions (North 2007). The English Profile Programme is, to a certain extent, a reaction to such comments.

Aims of the current project

The aims of the project are to identify criterial features at each of the five proficiency levels represented in the CLC and across language groups[1]. More specifically, the project looks for *linguistic* criterial features, rather than for *functional* criterial features, given that these are already set out in the CEFR. It is hoped that by adding linguistic criterial features to the functional criterial features already available (the *Can-do* statements), measurement of proficiency will become even more accurate, and may allow us to predict certain clusters of linguistic features given a learner of a said proficiency level, or, vice versa, predict on the basis of a cluster of linguistic features what proficiency level the learner is at. The current project only looks at English as a target language, but it is hoped that the methodology used will be valid for other target languages as well. The project moreover specifically addresses some of the critiques of the CEFR as formulated in the past 10 years.

It is felt that in order to appropriately identify linguistic criterial features, one will have to look at individual source-target language pairs. Thus, all language teachers can report specific errors made more prominently by some source-language speakers, and less by other source-language speakers. This begs the question if, for example, a French learner at a given proficiency level will present the exact same learner profile as a Chinese learner at that same proficiency level. They may both be equally good at using the target language for a given range of communicative functions, but they may make source-language specific errors in doing so.

Another important issue regards the progress of learners with respect to the specific linguistic features across proficiency levels. It is to be expected that each individual linguistic feature will improve gradually over time until it reaches a target language level. However, it is not clear how linguistic features in different domains relate to each other, i.e., do tense markers always reach target-language norms

1. Note that A1 level is largely underrepresented in the CLC due to candidates at this level not producing sufficient text for analysis. Cambridge ESOL's exams at A1 level (YLE Movers, for example) do not require candidates to write extended prose, thus making this data inappropriate for the CLC. But see Alexopoulou's article on page 15 for new data collection procedures which should include some lower level texts.

before gender marking, or is this dependent on the features found in the source language, the complexity of markers in the target language, etc.? By studying a large number of features across linguistic domains and at both utterance and discourse level, it is hoped that we may end up showing clusters of linguistic features that predict the development of a next feature, features that specifically trigger the development or allow for the development of the next feature, etc. In doing so, we will be able to provide a much more detailed description of each proficiency level, taking into account specific features of learners of particular source languages.

Findings such as the above will allow us to inform curriculum design, and teaching and testing materials for specific source-target language pairs.

Identifying criterial features

We have identified an initial set of lexical, grammatical, and discourse areas that are promising candidates for criterial feature identification, and these are being tested at the moment. These candidates were chosen on the basis of results from previous research in linguistics and in particular language acquisition studies. Given the research members of the project (linguists, specialists in second language acquisition, and computational linguists), we are able to explore such candidates using different acquisition theories, thereby covering more ground than a regular language acquisition project would normally do. In incorporating all levels of language production, not just the sentence level, we furthermore allow for more explanations of facts, some of them of use for the higher proficiency levels. Some features will end up not being criterial, because they do not show any predictable behaviour in acquisition or across languages, some other features may end up being criterial because the phenomena related to them turn out to be systematic and robust. Given the database of 27 million words, and the fact that the database is manually error-coded and completely tagged and parsed, we have the power of finding such robust phenomena.

Following are just a few of the hypotheses the team are testing with respect to certain linguistic features at the moment.

Lexical choice errors

For such errors – incorrect noun, verb, adjective or adverb – it is initially hypothesised that error rates will decline from A2 to C2 level. Otherwise said, the higher the proficiency level, the fewer (or equal) errors we expect to find. If, however, we find that errors fluctuate across proficiency levels in an unexpected way, lexical choice errors may not be a criterial feature, or the hypothesis may have to be fine-tuned. Another hypothesis related to lexical choice errors proposes that items subject to error will correlate inversely with native speaker frequencies in the *British National Corpus* (hereafter BNC), i.e., the lower the frequencies in the BNC, the more frequent the errors. This hypothesis is based on the idea that it is easier to acquire an item when it is frequent in the input, the BNC being a database representative of the input. A third hypothesis is that errors will vary with the L1 in that genetically distant source languages will exhibit more lexical choice errors than closer source languages.

Article errors

This type of error can be of two kinds. First, articles can be missing where they are obligatory in English, or provided when they should not be. Second, English has two types of articles, the indefinite article *a* and the definite article *the*. Exchanging one for the other often results in an error. This type of error should decline from A2 to C2, i.e., the higher the proficiency level, the fewer missing articles, and error rates should be greater or equal for source languages without articles than for source languages with articles. The latter hypothesis has already been refined based on a specific theory and is studied in a more fine-grained way by Dora Alexopoulou. According to her hypothesis, the acquisition of the English article system is inseparable from the acquisition of specificity. Depending on the marking of specificity in the source language (with article/with other means; with indefinite article but not with definite article), more precise hypotheses can be formulated for the different language pairs.

Discourse appropriateness of spatial expressions

We hypothesise that speakers will follow the source language *locus* and *focus* of spatial information in verbs and other means at lower levels of proficiency and only slowly adapt to the distribution of information as customary in the L2. This will result in different developmental paths across source languages. To exemplify, English native speakers, when seeing a cartoon involving a man pushing a ball so that it rolls across the street, will preferably say:

(a) The man rolls (**Cause** and **Manner** of Motion) the ball across (**Path** of Motion) the street.

French native speakers, when presented with the exact same cartoon will preferably say:

(b) Le garçon traverse (**Path** of Motion) la rue avec un ballon.

and, when in an experimental situation, may say:

(c) Le garçon traverse (**Path** of Motion) la rue en poussant (**Cause** of Motion) le ballon.

It is expected that differences of this kind will have an influence on French learners of English such that they may describe the same scene in English as in:

(d) The man crosses (**Path** of motion) the street pushing the ball.

in which case the information about the manner of displacement of the ball is missing.

Note that the sentence in (d) is grammatically correct, and may well be heard uttered by native speakers of English depending on the context. However, in the given situation, it deviates from native speaker speech. Finding out if and when such deviation starts to occur, or disappears, can give us an indication of the proficiency level of the speakers. It is believed that this type of deviation plays an important role at the B2, C1 and C2 CEFR proficiency levels, and researching them may therefore shed light on the top proficiency levels that are otherwise hard to distinguish on the basis of functional Can-do statements only.

Introduction of the following articles

In the two articles from RCEAL that you will find below, none of the above issues are discussed, but for some first results from those projects, please refer to the paper by Hawkins and Buttery, to appear in the Proceedings of the ALTE 2008 conference.[2] Rather, the articles will present some first results on a project that is being conducted by Caroline Williams, our PhD student on the project. Caroline has been looking at the acquisition of English word-order by Chinese and Spanish learners of the language. While doing so, she looked at this on the one hand from a second language acquisition perspective. On the other hand, however, she looked at the same phenomenon from a computational point of view, checking if non-native writing as found in the CLC poses any problems for the parser and tagger used, given that these were originally developed for native speaker English by Ted Briscoe. Caroline explains that word-order errors are likely to be highly systematic in nature and strongly influenced by the structure of the learner's source language. At the same time, deviations from native speaker word-order have the potential to dramatically affect a parser's ability to retrieve a syntactic representation which accurately reflects the input structure, or the speaker's intended meaning. Results show that word-order errors researched by Caroline indicate that the most frequent errors found are not likely to influence the ability of the parser to accurately retrieve the parts of speech and relations between elements, which is reassuring for its use in the English Profile Programme. Some of the errors of non-native speakers could, however, create distinctive patterns in the parsed output that could possibly predict what the source language of the given speaker is.

The second article, by Dora Alexopoulou, reports some of the work ongoing in the third project. That is, the project that sets out to develop a more elaborate data set, which is not just based on learners' exam scripts, but also on other modes of production and hence communication. One of the very early critiques regarding *English Profile* was its use of the CLC which, of course, has very specific features, as it was designed many years before *English Profile* came into being. Language is produced in a range of contexts besides exam conditions, and many a second language researcher can provide a list of problems arising from using that particular type of data. Also, exam contexts are restricted in terms of the situations they can create for language users to use all of their skills. Hence, the CLC might not be representative of all of the features that second language learners would naturally present, and may at the same time highlight some features more prominently than they would normally occur in the learners' output. Additionally, the data in the CLC are not equally divided over proficiency levels, and across source languages, thereby diminishing the powers of any quantitative analyses. Eliciting new data from all over the world in a controlled fashion is a major endeavour, and Dora will report on the steps taken so far to carry out this data collection exercise. In particular, she will indicate how the current uneven spread of data by proficiency level is already being tackled, and how the project is thinking about getting spoken data that are essentially different from the written exam scripts we already have, even if the new data will still be a long way away from natural speech.

Finally, more results are being produced every day in this vibrant programme, and anyone interested in *English Profile* and its development and findings can find more information on the RCEAL website: www.rceal.cam.ac.uk or the Programme website: www.EnglishProfile.org

2. See page 40 for information.

References and further reading

Beacco, J C and Byram, M (2007) *From linguistic diversity to plurilingual education: Guide for the development of language education policies in Europe*, Strasbourg: Council of Europe, Language Policy Division.

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

Hulstijn, J (2007) The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency, *The Modern Language Journal* 91, 663–7.

Little, D (2007) The common European framework of reference for languages: perspectives on the making of supranational language education policy, *The Modern Language Journal* 91, 645–655.

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

— (2007) The CEFR illustrative descriptor scales, *The Modern Language Journal* 91, 656–9.

North, B, and Schneider, G (1998) Scaling descriptors for language proficiency scales, *Language Testing* 15, 217–262.

Challenges to parsing English text: the language of non-native speakers

CAROLINE WILLIAMS RESEARCH CENTRE FOR ENGLISH AND APPLIED LINGUISTICS, UNIVERSITY OF CAMBRIDGE

Introduction

Crucial to *English Profile* is the use of the Robust Accurate Statistical Parser (Briscoe 2006) to parse the learner data (written texts from the CLC) and extract information about the parts of speech used and their syntactic relations. This parser, however, was developed to work with native English text, text which can be very different from that produced by learners. This article discusses some of the challenges which learner data pose to an automatic parser, focusing in particular on word order errors by speakers whose first language is either Spanish or Chinese. I conclude that on the basis of this work it seems unlikely that these word order errors have a significant impact on automatic parsing, however I show that the output of the parser when faced with such errors may well provide a diagnostic for the first language of the text's author.[1]

Background

Depending on the speaker's proficiency, text by a non-native speaker may differ to a greater or lesser extent from that of a native speaker. At one extreme are ambilingual speakers who write with the proficiency of a native speaker; at the other extreme are beginning learners who might express *I don't like coffee* as *me no like coffee*. For the purposes of this article I focus on the language of learners with a minimum proficiency of B1 in the Common European Framework of Reference (CEFR; Council of Europe 2001). B1 is defined as the level at which a speaker is capable of using language independently, i.e. to construct new utterances, rather than relying solely on memorised phrases.

As studied in other strands of *English Profile* research, the structure of a speaker's second language (L2) is significantly affected by their native language (L1). The existence of a parallel or related structure in the L1 can make the English structure easier or harder to acquire in the first place, depending on the exact nature of the difference, and thus making it more or less likely that learners will make errors in it. Similarly, the existence and frequency of a parallel or related structure in the L1 also affects the choices they make in the L2. For example, since English overwhelmingly requires a Subject Verb (SV) word order whereas Spanish permits both SV and VS depending on context, it seems likely that when speaking Spanish, native English speakers use SV orders with higher frequency than native Spanish speakers. They might, for example, say *Juan llegó* when a native Spanish speaker would say *llegó Juan*. It is crucial to note that this would not be a syntactic error – it is a possible way of expressing the idea that 'John arrived' – but it is dispreferred by native speakers if 'arriving' is not the focus of the utterance (Lozano 2006).

This article reports on preliminary work aimed at establishing some of the properties of English text written by learners whose native language is Spanish or Chinese, because the grammars of Spanish and Chinese are very different with respect to word order. Spanish has a basic SVO order, but this is relatively flexible and there are contexts where other orders are strongly preferred (Butt 2001). Chinese, on the other hand, is a topic-comment language in which the topic is the first element, be it a noun phrase, an adverb, or even a clause (Li 1981). English, of course, is fairly rigidly SVO.

I focus on errors in word order since English has minimal morphology and a relatively fixed word order, i.e. information about a word's part of speech and its relation to other words in the sentence is primarily conveyed through word order. Changes to the word order therefore have the potential to dramatically affect a parser's ability to retrieve a syntactic representation which accurately reflects the input structure, or of the speaker's intended meaning (which their actual syntax may not reflect). Word order errors also seem likely to be highly systematic in nature and strongly influenced by the structure of the speaker's other languages – unlike, say, spelling or lexical errors. Finally, English word order is relatively stable across regional varieties in comparison with aspects of language such as the lexicon.

Methodology

The study uses as its data source a subset of the Cambridge Learner Corpus (CLC). The CLC consists of 27 million words of learner data at CEFR levels A2–C2 (Council of Europe 2001), taken from Cambridge ESOL written examination scripts. These levels cover a progression of competence from very basic written communication skills, to the skills required of a university level student writing in English. Of the subset of the corpus which is available to us, 10 million words have been manually error-coded and corrected (Nicholls 2003) and of this, slightly under 4 million words (11,500 scripts) was assigned to a passing grade, thus ensuring that the scripts are representative of language at the desired proficiency level. The CLC is particularly appropriate for this study since it enables us to study non-native language – and therefore parser behaviour – on samples of language which become progressively more native-like as proficiency increases.

It should be noted, however, that the definition of 'word order error' used in the error coding is in part a subjective judgement. The set of errors includes some orders which are stylistically dispreferred rather than overtly wrong, for example:

> *It was ‹NS type='W'› really a/a really ‹/NS› good lesson.*

The error code appears in ‹ › followed by the learner's actual text, i.e. *really a* which is separated from the corrected version *a really* by the vertical bar.

Furthermore, some of the strings are acceptable strings in their own right, but are judged to be incorrect based on the immediate context, for example:

> *On a sunny day, the brilliant blue sky, the ‹NS type='W'› brick-red/red-brick ‹/NS› roofs and a couple of flying seabirds would make an excellent picture.*

The CLC is marked up with XML and has been separated out into the original version of the corpus, the corrected version of the corpus, and the version in which errors are marked up and corrected, in line with de Mönnink's (2000) suggestions. Both the original and the corrected versions of each text have been tagged and parsed with the RASP (Robust Accurate Statistical Parsing) system (Briscoe 2006) and all three versions are automatically searchable.

From this corpus, all sentences which only include one word order error were extracted for learners whose L1 was Spanish or Chinese at each proficiency level, in order to easily isolate the causes of any parser errors. The number of sentences per language and level is shown in Table 1.

1. This article is based on a paper presented at Computational Linguistics UK 2008, University of Oxford, 11–12 March 2008.

Table 1: The number of sentences used for each language and level combination

	Chinese	Spanish	Total
B1	12	44	56
B2	78	68	146
C1	33	54	87
C2	24	63	87
Total	**147**	**229**	**376**

Unfortunately, as can be seen, it was not possible to obtain equal numbers of sentences for each language or level, and some of the samples are rather small. The results presented in this article should therefore be considered as indicative of possible trends rather than as conclusive findings.

Classifying the errors

In order to examine the challenges posed by different types of word order error to automatic parsers, an error typology was developed. Since we are only working with individual sentences, this typology is based upon a description of the errors in terms of surface word order or the type of structure involved, rather than on an attempt at explaining the nature of the error in terms of the learner's mental rule which generated it.

In this typology the subject, verb and object are treated as forming a 'grid' around which other elements are placed, and errors are described in terms of the constituent which is misplaced with respect to that grid. Thus, for example, *To sum up, I consider myself the right person for that position vacant* is classified as containing an adjective (Adj) error, since the adjective *vacant* is out of place with respect to the noun *position*. Beyond that, where the error involves elements such as an adverb (Adv) and a prepositional phrase (PP), I have treated the least argument-like element as the one which is out of position, as in the following example where the adverb phrase *last week* is treated as out of place: *I went last week to the disco*.

Where a VP and NP were switched, the NP was treated as being misplaced, since the NPs tend to be arguments of the verb and in a sense are therefore dependent upon it. Where the correct order of elements involves a change of part of speech, e.g. *We have received your fax concerning the office renting*, the element which did not change part of speech was labelled as the errorful item, and the sentence was flagged as involving a change. Constituents were identified at the narrowest level which defined the error, therefore *P.S. I'll show you my country around* was treated as a prepositional error, whereas *Report of our department on the overall performance* was treated as a prepositional phrase error.

Apart from these relatively straightforward positioning errors, several types of error appeared to form unitary phenomena and were therefore classified separately, rather than in terms of which element was out of place. Three such error types were those involving inversion (Inv), negation (Neg), or 'of phrases', i.e. those phrases which are either preferably expressed as an 'of phrase', e.g. *volume of exports* rather than *exports volume*, or which could conceivably be expressed that way, but in practice tend to be expressed differently, e.g. *Then there was a closing display of fireworks*. A further type of error which was assigned a separate classification is Compound, since this pertains to an error in the internal structure of a multi-word expression, rather than in the position of the noun or NP in some larger unit, e.g. *But I really wanted a computer home*.

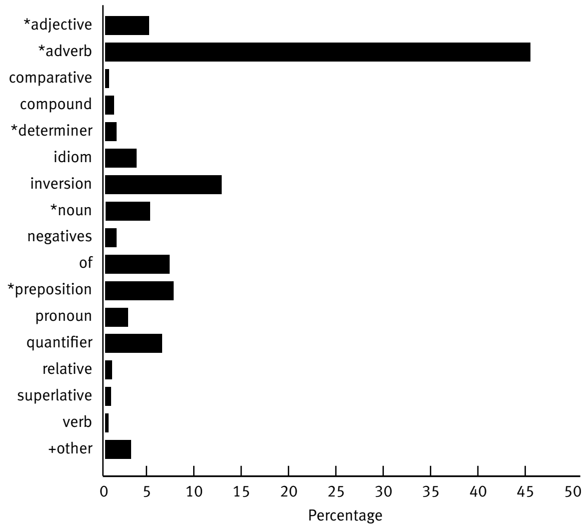
Results

In this section I present the overall pattern of errors together with five other types, namely: compounds, adjective phrase position, adverb position, noun phrase errors and other errors.

Overall pattern of errors

The percentage of each type of error is shown in Figure 1. As shown, by far the biggest category of errors is adverbs, almost three times more than any other group of error. The second largest group of errors occurs with inversion. It is not surprising that these two types of error should be so common in our sample. The rules governing the position of adverbs in English are complex and depend on many factors. It would not be surprising, therefore, if non-native speakers should have failed to master their interactions precisely.

Figure 1: The percentage of each error type in the whole sample[2]



That inversion errors should form such a large category is no surprise either: there is no comparable structure in either Spanish or Chinese which these speakers can draw on. Furthermore, the rules governing the use of inversion can easily be incorrectly identified by the learner: many of the errors in this sample arise from learners automatically inverting after an orthographic form which may be either an interrogative word or a relative pronoun (e.g. *why*, *how*, *where*), possibly on the faulty understanding that it is the form which triggers the inversion rather than its function in the sentence. Both adverbs and inversion contexts therefore offer considerable potential for error, and both types of

2. * refers to a grouping, e.g. errors in the position of the adjective and of the whole adjective phrase.
+ refers to a small number of sentences which were erroneously included due to multiple errors or for which no suitable analysis could be found.

Figure 2: The percentage of each error type in the sample of Spanish errors[2]

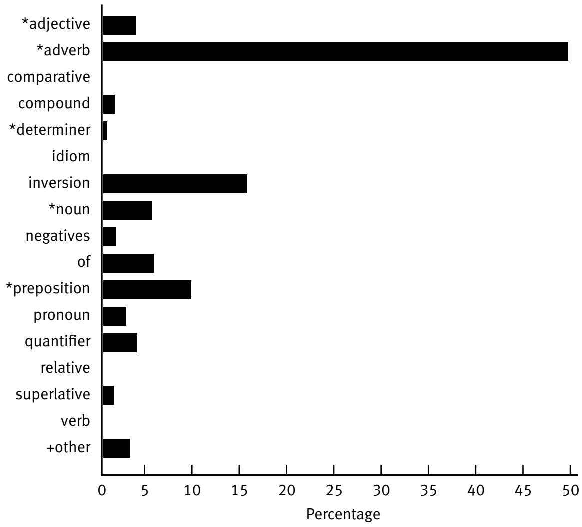
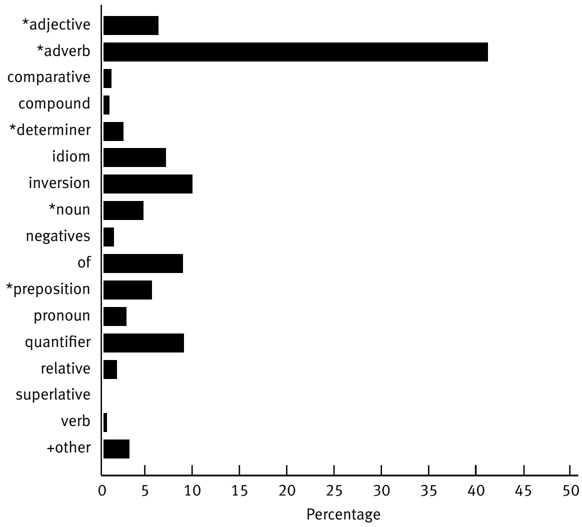


Figure 3: The percentage of each error type in the sample of Chinese errors[2]



construction are frequent in English, such that a learner will regularly have to try and use them. Also frequent were 'of' errors, in this case probably because the choice between *x's y, the y of x* or simply *y x* is relatively lexically-specific in English, and learners may not have had sufficient exposure to these items to learn which structure is appropriate.

Both inversion and adverb errors seem to be more consistently a problem for Spanish speakers than Chinese speakers in our sample (see Figures 2 and 3). This is surprising from a contrastive analysis point of view since the differences between Spanish and Chinese and English do not straightforwardly predict this pattern. One might expect Chinese speakers to have more problems with the positioning of adverbs than Spanish speakers since in Chinese, the system governing the positioning of adverbs operates extremely differently from both Spanish and English, and indeed partly operates at a discourse level rather than in terms of sentence grammar. In terms of inversion, there is no particular reason to expect a significant difference in the extent to which this causes an issue.

Interestingly, increasing proficiency seems to have almost no systematic effect on which of these errors predominates. The one exception is errors involving adverb positioning: as proficiency increases these consistently form an increasingly large percentage of the errors for both Spanish and Chinese. All the other types of error, however, show non-systematic fluctuations in their percentages. This perhaps indicates that adverb phrase errors are the hardest group to overcome, with other types of error diminishing. This increase does not seem to be linked to an increase in adverb use, since the number of adverbs used per hundred words declines between B1 and C2 for Spanish, and increases for Chinese. It may, however, simply be a function of the particular annotation scheme used here, in which annotators seem to have corrected less-preferred positions as well as actively dispreferred positions.

Compounds

One might not expect this type of error to cause trouble for the parser since it is internal to the noun phrase and therefore should not have an impact on the overall structure retrieved. One type of analysis RASP carries out, however, is a grammatical relations (GR) analysis, which represents relations between heads and modifiers. A misordering of elements such as *computer home* instead of *home computer* does have an impact on the GR output since it affects which of the two nouns is identified as the head. In *But I really wanted a <NS type='W'> computer home|home computer</NS>*, RASP identifies *computer* as the head of *computer home* in the uncorrected version, and thus both the DOBJ and DET relations have *computer* as their dependent.[3] This should not have too substantial an impact on actual applications, however, since when querying the corpus, if one needs to retrieve relations at such a specific level (DOBJ), then presumably one would also be retrieving the modifiers.

Adjective phrase position

Initially rather surprising is the fact that Chinese speakers have more problems with the placement of adjectives than Spanish speakers in our sample. In Spanish the adjective follows the noun whereas in Chinese and English it precedes the noun, so one might have expected Spanish learners to have more of a problem with this. On closer examination of the data, however, it became apparent that Spanish and Chinese learners were making different types of error. Some of the errors made by Spanish learners did indeed involve placing an adjective after a noun, but some of them – and most of the errors made by Chinese learners – involved errors in the order of two adjectives, e.g. *Her gray long hair was all over her face*. In practice, it may be that the rule concerning location of adjectives (before the noun) is straightforward and therefore such errors are easily eliminated by Spanish speakers, whereas the rules governing the precise ordering of adjectives are not so simple and therefore continue to occur at all levels.

As one might expect, the latter type of error has little effect on tagging and parsing. The tagging and GR outputs

3. DOBJ is direct object; DET is determiner.

produced by RASP were identical for this particular sentence, regardless of which order the adjectives appeared in. Interestingly, with our example of the former type of error, the adjective is interpreted as a noun modifying the actual noun as if it were a compound, thus assigning the incorrect part of speech to the adjective.

Adverb position

Since the adverb phrases form a rather more diverse group, here we only study individual adverbs such as *again* or *calmly*. One of the most striking observations is that of the 60 adverbs in this sample, only 15 are what may be termed 'productively derived' adverbs, formed by addition of *-ly* from an adjective, e.g. *angrily*. Of these, only 12 are being used as manner adverbs. This difference is also reflected in the token counts, with 17 and 14 out of 157 respectively. That they represent such a small group of errors may be because their syntactic behaviour is fairly uniform in English, with a strong preference for them to appear sentence finally[4]. The remainder of the adverbs are closed class adverbs such as *here*, *very* or *always*. The appropriate position for this latter set of adverbs is less uniform, varying depending on factors such as its semantics (e.g. adverb of time, space etc.) and its subtype within that role (e.g. time adverb of frequency, or of duration), its scope in the sentence, its grammatical function (e.g. adjunct, conjunct etc.) and even its phonological weight (Quirk 1985:500–2). The multitude of relevant factors and the complex interactions between them seem a likely indicator for the continuing prevalence of errors in this type of adverb: there is no simple rule which determines the correct position.

Noun phrase errors

The NP class of errors shows an interesting effect of first language. Firstly, of the 19 NP errors, 12 were made by Spanish speakers. Four of these errors display VS word order, an order which is extremely common in Spanish, but disallowed in English. An example sentence from the corpus is *Last week <NS type='W'> came to London from Madrid, my best friend/my best friend came to London from Madrid </NS>*. These errors are displayed in B1, B2 and C1, indicating that this fundamental word order error occurs even in proficient speakers. Moreover, this error caused RASP to generate inappropriate output. The tagging remains correct, but RASP finds an NCSUBJ (non-clausal subject) relation between *come* and *week* instead of between *come* and *friend*, and thus identifies *last week* to be the subject of the sentence instead of *my best friend*. This may be the correct output based on the input, but it fails to capture the writer's probable meaning. Indeed, this is one of the problems with parsing non-native English: sometimes identification of an error depends on world-knowledge or semantic knowledge; the input is syntactically acceptable and RASP's output is correct with respect to that input. As the input syntax does not reflect the intended meaning, however, applications which depend on that meaning – in this case correct identification of the subject for anaphora resolution – will be hindered.

Other errors

A further three errors reflect Spanish preferences in terms of information packaging, as illustrated in the following example:

That would surely solve most of the described problems.

In the Spanish equivalent, *described* would be an adjective (albeit also one derived from a verb) and would appear in pre-verbal position, unlike in English where it is a verb and it follows the noun. This latter type of error is not restricted to Spanish – there is also one Chinese example – but it is characteristic of the Spanish pattern. In both cases the error may well arise due to uncertainty as to whether the *-ed* form is functioning as a verb or an adjective. It should be noted that there are cases in English in which a verbal participle does function as an adjective, for example in the sentence: *The published do not tell the whole story*. As one might expect given that the pattern does form a legal syntactic string, if not a preferred one in this case, this distinction is correctly identified by RASP. Again the tagging remains correct, but here RASP identifies the type of modifier relation between *described* and *problem* as a non-clausal modifier (NCMOD) as if *described* were an adjective, as opposed to as the predicative relation XMOD. It is possible that a higher than normal frequency of NCMOD relations involving VVN might indicate a text whose writer is a native speaker of Spanish.

Inversion errors include contexts where English would invert the verb and its subject (e.g. *Now, why you don't tell me how your holiday was?*), as well as hyper-correction contexts where learners have erroneously inverted the verb and its subject. (e.g. *Please tell me how can I get it back!*). Omitted inversions form a slightly smaller group – only 21 out of 53 sentences – but this is unlikely to be statistically significant. A minority of these occurred after an adverb such as *only* or *never*, but as might be expected, the overwhelming majority occurred in direct questions. Most occurred in *wh-* questions rather than 'yes–no' questions, but this may be an effect of genre. Similarly, almost all of the 31 hyper-corrections occurred in the context of indirect questions. It is worth noting that these basic errors in the formation of interrogatives continue up to the most advanced level included in our corpus.

Both hyper-corrections and missing inversions cause some problems for the parser. The hyper-correction *Please tell me how can I get it back!* caused the parser to generate an additional CCOMP relation. In contrast, *Did you have a nice holiday or you were with your family?* caused the parser to falsely identify *or* as taking part in an AUX relationship with *do*.

Conclusion

This work has illustrated some of the variation from standard English which occurs in the CLC and shown how the nature and/or quantity of the errors varies depending on the speaker's first language. In the case of noun phrase errors, we presented some evidence of transfer on the part of Spanish speakers when writing in English. The examples examined here seem to have had a relatively insignificant impact on the ability of a parser built for native English to

4. All of the open-class adverbs in our sample modify verbs (Quirk 1985:562).

accurately retrieve the parts of speech and relations between elements, which is reassuring for its use in *English Profile*. Only one of the errors we have looked at seemed to affect RASP's ability to accurately retrieve the correct parts of speech, but although several have affected its GR output, in most cases the error would only show up in queries depending on a detailed analysis of the sentence, rather than relying on its large-scale constituents. Some of the errors made by the non-native speakers, however, seem likely to generate distinctive patterns in the parser output which could be used for automatic error identification in Computer Assisted Language Learning or in automating the error-coding process in the CLC.

References and further reading

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

Butt, J and Benjamin, C (Eds) (2000) (3rd ed) *A new reference grammar of Modern Spanish*, London: Arnold Publishers.

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

de Monnink, I (2000) Parsing a learner corpus, in Mair, C and Hundt, M (Eds) *Corpus linguistics and linguistic theory*, Amsterdam: Rodopi, 81–90.

Li, C and Thompson, S (1981) *Mandarin Chinese: A functional reference grammar*, Berkeley, Los Angeles, London: University of California Press.

Lozano, C (2006) Focus and split-intransitivity: the acquisition of word-order alternations in non-native Spanish, *Second Language Research*, 22/2, 1–43.

Nicholls, D (2003) The Cambridge Learner Corpus – error coding and analysis for lexicography and ELT, in Archer, D, Rayson, P, Wilson, A, and McEnery, T (Eds) *Proceedings of the Corpus Linguistics 2003 conference*, UCREL Technical Paper number 16. Special issue, 572–582.

Quirk, R, Greenbaum, S, Leech, G, and Svartvik, J (1985) *A comprehensive grammar of the English language*, London: Longman.

Building new corpora for English Profile

THEODORA ALEXOPOULOU RESEARCH CENTRE FOR ENGLISH AND APPLIED LINGUISTICS, UNIVERSITY OF CAMBRIDGE

Introduction

This article describes the collection of data from around the world to form new corpora for use by *English Profile* researchers. I describe the contribution that the Cambridge Learner Corpus (CLC) has already made to the Programme and then introduce the main parameters that are informing the design and collection of our new written corpus – soon to be joined by a spoken corpus.

The Cambridge Learner Corpus

The Cambridge Learner Corpus (CLC), a collection of over 100,000 exam scripts written by students taking ESOL exams worldwide has grown into an important resource for the study of learner English, already informing language test development activities (Barker 2006) and various publications and teaching materials for learners of English[1]. The CLC is also of central importance to research carried out within the English Profile Programme. In its original format the CLC includes rich annotations of various linguistic features as well as a very systematic way of error coding. However, for the purposes of *English Profile* research, the CLC has been transformed to a structured set of sentences reflecting the internal structure of phrases composing sentences and the grammatical/dependency relations connecting words and phrases. This transformation was enabled by the RASP Parser (Briscoe *et al*. 2006) and work led by Professor Ted Briscoe at the Computer Lab and Dr Paula Buttery at RCEAL. The parsed CLC enables a wider and more sophisticated range of searches than is possible using the unparsed version. The parsed CLC provides the empirical basis for a research programme investigating the properties of learner grammars at RCEAL, headed by Professor John Hawkins. The central goal of this research is to contribute to the development of Reference Level Descriptions (RLD) that will exemplify the levels of the Common European Framework of Reference (CEFR; Council of Europe 2001) for English. These RLDs will be based on an in-depth analysis of the grammatical properties of learner English, crucially incorporating psycholinguistic metrics of complexity and differentiating learners according to their first languages (L1s), and thus, accounting for any transfer of linguistic features from the first language to English. The ultimate goal is to provide a set of criterial features based on quantified and robust correlations between a given learner stage and the complexity and coverage of learner grammars. CLC therefore provides the empirical basis of this research.

Despite its size and coverage, a systematic analysis of the data properties of the CLC reveals a number of shortcomings in relation to the needs of this specific project which are due to the corpus not being designed for this specific research project (Williams 2007). Firstly, the CLC contains an unbalanced distribution of data across CEFR levels and across different L1s, although the L1 spread is representative of the actual candidature for each specific exam in the corpus. Data from B2 level represent a high

1. See www.cambridge.org/elt/corpus/learner_corpus.htm

proportion of the corpus which is due to its historical development, as candidate scripts at levels B2–C2 (FCE, CAE and CPE) were the first to be added to the corpus in the early 1990s and are also amongst the longest scripts in the corpus. The uneven spread and amount of data across the levels have impeded attempts to build grammar profiles for each level and carry out comparisons between levels. Similarly, the CLC contains a range of discourse genres and topics, all pertinent to written exam tasks, e.g. letters, reports, stories and compositions, which means that other genres, e.g. reviews, instructions or proposals, are underrepresented in the CLC. At lower levels (A2–B1) topics tend to be firmly transactional and descriptive, focusing on the candidate's everyday life while at higher levels (B2–C2) candidates give their opinions on a wide range of topics. Such discrepancies may weaken comparisons of learner grammars between levels, since there is always the question of whether any differences between levels are partially attributable to differences in chosen discourse genres or topics. Finally, the CLC involves language production that is somewhat isolated from the learner and the cultural and educational background in which they operate, thus, depriving research into possible interactions between demographic, cultural and educational parameters and developing learner grammars. Last, but not least, the CLC currently contains only written production, leaving unexplored the properties of learners' speech. In order to address these shortcomings, a data collection exercise is underway with the aim of enriching and complementing existing data in the current CLC by collecting data of specific types, initially written but also spoken.[2]

Collecting a new corpus

The main aim of building a new corpus is to extend the data available to *English Profile* with learner English data from classroom settings around the world, while at the same time collecting information relating to the learners and the educational and sociocultural settings in which they operate. An important feature of this new corpus is that it will consist of language samples produced by learners *on demand* and *for* the corpus.[3] This provides us with a unique opportunity to collect materials exhibiting a wide range of linguistic features, and, therefore, construct a learner corpus balanced for factors that are normally beyond the control of researchers compiling learner corpora from materials produced for another primary purpose (e.g. exam, placement test, university essay etc.). This unique opportunity, is, at the same time, a serious challenge for the design. For most native corpora a wide and diverse sample of existing work (from newspapers and novels to courtroom proceedings) normally guarantees the whole spectrum of linguistic features and vocabulary. However, for the learner corpus we seek to build, the burden of eliciting a wide range of linguistic features and vocabulary falls on the design. As indicated in Sinclair's (1996) definition of a corpus, it is the linguistic criteria used for identifying and organising relevant samples of language that ultimately defines the usefulness of a corpus which is 'a collection of pieces of language, selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language'. Below I identify the main parameters that will guide the development and organisation of materials through the data collection.

Corpus parameters

The new corpus will be organised around two learner-related parameters:

1. the competence/proficiency level of the learner, which will be linked to the 6 CEFR levels
2. the Mother Tongue (or First Language: L1) of the learner.

These two parameters will result in 6 main sub-corpora (corresponding to each CEFR level) multiplied by the number of L1s reflected in the corpus. Data collection will be organised along *text internal parameters* and *text external parameters*, presented in the following sections.

Text internal parameters

The main challenge for the design of materials is to ensure that the compiled corpus will provide us with a balanced sample of L2 language. Two types of text internal parameters, *linguistic* parameters and *contextual* ones will guide the development of materials.

Linguistic parameters

The linguistic parameters relate to grammar, discourse structure, the complexity of L2 production and discourse genres. The rationale here is to develop an inventory of features/types relevant to each parameter and produce tasks that will specifically target the elicitation of those features. One important feature of the approach taken here is its corpus-based nature. The development of tasks will be based on prior corpus-based investigation on the types of texts that in native English tend to naturally favour a given feature or construction.

Grammar

A comprehensive investigation of the properties of learner grammars (at each CEFR level) will need to rely on a wide range of grammatical features. It is important that the designed materials elicit as wide a range of grammatical constructions as possible. To appreciate the importance of the issue consider the acquisition of verbal morphology in English as a concrete example. At first sight, it would seem that design in this domain ought to be trivial, given the fact that every single English sentence contains at least one verb. Just the mere quantity of verbs should guarantee sample representativeness. Not so. To evaluate learner competence in this domain, the first step is to check that learners know how to form the various tenses (present, past, future etc.). So, materials need to elicit the target forms in all tenses. In addition, materials also need to elicit not only declaratives, but also questions and negative

2. In the first instance data collection will focus on written production with the aim of extending to oral production at a later stage.
3. The International Corpus of Learner English (Granger, Dagneaux and Meunier 2003) is an example of a learner corpus involving materials produced on demand for the corpus. However, collection of data is restricted to C levels and consists mainly of essays.

statements, since the paradigm is different in these types of structures with auxiliary selection and word order an important aspect of competence in this domain. In fact, currently the CLC, despite its 27 million words, contains very few questions, making it very hard to properly evaluate whether learners master auxiliary selection and word order and its correlation with tense formation. This is unsurprising, however, given the nature of the tasks that learners are responding to in the CLC.

Similar issues arise with almost every single grammatical phenomenon. It is, therefore, of crucial importance that materials are developed that elicit a wide variety of grammatical constructions and features to ensure adequate samples of each construction/feature for linguistic analysis. Such balanced sampling is important for any corpus. But for a learner corpus it is crucial for one more reason. Learner grammars, in particular at earlier stages, can be incomplete in a number of ways, one of which may be absence of a particular feature or construction. It is important to know that when a feature/construction appears to be absent from learners' grammars at a given stage, this is a true feature of their grammar and not an artefact of the data collection process. Put simply, if materials never target the elicitation of, say, questions or imperatives, we cannot claim that learners cannot form questions.

Discourse structure

Linguistic competence involves knowledge that pertains to the organisation of discourse, putting sentences together and linking syntactic objects in a given sentence to referents and events in the discourse. There is a rich body of investigation in the theoretical literature on this domain – see for example Discourse Representation Theory (Kamp and Reyle 1993), Centering Theory (Grosz *et al.* 1995) and Information Structure (Vallduví 1992). One central question in this domain is the types of noun phrases speakers use to pick up discourse referents, to expand topics or introduce new information to the discourse (Hawkins 1978). By way of illustration, consider the example *Peter invited Bill to the party. He thought* ... Despite the presence of two discourse referents, *Peter* and *Bill*, *he* in the second sentence is normally associated with the subject of the first sentence *Peter*. There is evidence that at early stages learners avoid pronouns and tend to use full noun phrases in place of pronominals (e.g. *Peter* instead of *he*). The question directly relates to whether learners can felicitously alternate pronouns, demonstratives, noun phrases etc. to identify different types of discourse referents.

Complexity

An important feature of the approach taken for the research undertaken at RCEAL and the Computer Lab is the application of psycholinguistic metrics of complexity to the analysis of L2 production, an approach inspired by Hawkins' influential theory of Efficiency and Complexity (Hawkins 2004). Such metrics may involve the complexity of verb-frames, relative clauses, subordination etc. (see Alexopoulou and Keller 2007, Briscoe and Buttery 2007, Korhonen, Krymolowski and Briscoe 2006 among others). It is important that materials are balanced for the expected complexity of the structures elicited so that the developmental path of learners is traced as they move from less complex to more complex structures and the interaction between complexity and learnability of structures is investigated.

Discourse genre

Knowledge of English involves fluency in different genres/modes of communication, e.g. story telling, letter writing, descriptive essay, analytical essay, informal conversations in a familiar setting (family, pub), conversations in a formal setting (job interview) etc. At the same time, certain types of texts favour different types of constructions (e.g. questions or imperatives are unlikely in an analytical essay) and a different set of skills (e.g. developing arguments vs. engaging an interlocutor in a conversation). A balanced sample of discourse genres is essential. As indicated earlier, it is important to know that any differences in learner grammars between stages are a true feature of learner grammars and not attributable to differences between the discourse genres and topics dominating learner production at each stage. The desideratum therefore is a set of materials that, in addition to ensuring a wide range of topics and discourse genres, guarantees homogeneity across levels. This is a significant challenge, for the simple reason that materials have to match students' abilities at each stage; put simply, it makes no sense to ask a beginner to write an analytical essay (even if this beginner is a lawyer or journalist). In addition, teaching practices and materials pace the introduction of different types of written work as students progress through the CEFR levels.

Contextual parameters

Below I identify two further parameters that will play an important role in the development of materials, linkage of materials and production to CEFR levels and the naturalness of classroom-based L2 production. The two parameters are included for completeness, however, a proper discussion of the issues raised in these two areas goes beyond the scope of the present article.

Linking materials and L2 production to CEFR levels

The problem of linking the corpus to CEFR levels has two dimensions. Firstly – as already happens with all Cambridge ESOL exams – materials need to be linked to the six CEFR levels prior to the data collection, so that what is asked from learners matches the capabilities of their level. Second, production data in themselves need to be evaluated and linked to CEFR levels. Linking the materials to CEFR levels is of paramount importance since the ultimate goal is to provide criterial features distinguishing the levels. A strategy on this is being developed in collaboration with Dr Angeliki Salamoura and colleagues at Cambridge ESOL (see Salamoura's article on page 5).

Naturalness of classroom-based L2 speech

Two issues arise in relation to the naturalness of classroom-based L2 production. First, the linguistic parameters guiding the development and organisation of materials need to be incorporated in tasks that match the socio-cognitive reality of classrooms as well as the life experiences of subjects. Second, in order to properly evaluate the nature of learner grammars, it is important to compare them with native grammars.

One of the key research objectives of the Corpus Linguistics team at RCEAL is to evaluate learner grammars by specifically contrasting features of learner grammars against features of native English as they emerge in large corpora such as the BNC. One potential issue for this comparison is that the BNC does not contain the types of classroom-based written work we will be collecting. Thus, interactions between discourse genre/topics and linguistic features might skew comparisons between BNC and classroom-based learner data. To address this issue, we are currently exploring the possibility of building a control corpus with data from native speakers of English matching the profile of the learners (e.g. for age and educational background) completing the same tasks.

Text external parameters

While the focus of a given corpus lies, naturally, on production data, information on the educational context and the learners themselves can be valuable for researchers interested in the effect of various educational and sociolinguistic variables as well as in the developmental path of specific sets of learners. Information will be collected along two main parameters:

i) the educational context

ii) the learners.

Educational context: country and teaching institutions

Different practices are relevant in different countries. While in some countries the teaching of English is primarily done in mainstream state schools, in other countries English is taught in specialised (private) language schools. Information on the general educational context in a given country is normally widely available, so we will only focus on collecting information about the specific institutions that provide us with data. For example, the place of English in the curriculum, the status of English teachers within a given institution, the motivation of students etc.

Input: teachers, teaching materials

Information on the input that students receive is of vital importance for research on educational practices. Information will be collected on:

- Teachers: education, motivation, familiarity with the CEFR levels, native speakers or not etc.
- Teaching materials: course materials, test materials, how are they used in classroom, variety of English taught – British vs. American – curricula, syllabuses.
- Intensity of input: frequency of classes, immersion in an English speaking environment, availability of English on TV/cinema etc.

Learners

Information on the learners who produce the data is indispensable in analysing and contrasting written responses from different sets of learners (e.g. 15 year olds in Spain vs. 15 year olds from bilingual schools in Argentina). Demographic information will be collected relating to the age of the learner (at the time of data collection, when they started learning English), their education, their motivation, etc.

Importantly, information on learners will in effect allow us to exploit the longitudinal nature of some of the collected data. At least in principle, it will be possible to follow data of a particular class in a given school over 2–3 years and compare with another set of learners for a similar period of time. While at the moment, it is hard to evaluate whether, in practice, it will be possible to follow up a whole class over several years, it is certainly a fact that such longitudinal data would make this corpus unique among current learner corpora.

Last, but not least, learners will be matched to a CEFR level in accordance with procedures developed for evaluation of their level prior and after the data collection.

Logistical aspects

Data collection will rely on partnerships with educational organisations worldwide. We have just completed a pilot study at secondary schools in Lombardy, Italy organised and led by Cambridge ESOL and local Ministry staff. We have had an enthusiastic response from local schools and a very interesting first set of data which we are currently studying for the development of our database.

Data collection will eventually be managed through the *English Profile* website; we expect to have a first version of this up and running by the end of 2008.

Conclusion

I have laid out a roadmap to what is a very exciting and, at the same time, ambitious and challenging enterprise. I have concentrated on issues relating to the content and design of the materials to be collected, abstracting away from practical and technical issues. This is a starting base which will hopefully grow and expand through collaboration and exchange of ideas with our partners worldwide.

References and further reading

Alexopoulou, T and Keller, F (2007) Locality, cyclicity and resumption: at the interface between the grammar and the human sentence processor, *Language* 83/1, 110–160.

Barker, F (2006) Corpora and language assessment: trends and prospects, *Research Notes* 26, 2–4.

Briscoe, E and Buttery, P (2007) The influence of prosody and ambiguity on English relativisation strategies, paper at *Interdisciplinary Approaches to Relative Clauses*, Cambridge, September 2007.

Briscoe, E, Caroll, J and Watson, R (2006) The second release of the RASP system, in *Proceedings of the COLING/ACL interactive presentation session*, Sydney, Australia.

Granger, S, Dagneauz, E and Meunier, F (Eds) (2003) *The International Corpus of Learner English, Handbook and CD-ROM*, Version 1.1, Louvain-la-Neuve: Presses Universitaires de Louvain.

Grosz, B J, Joshi, A K and Weinstein, S (1995) Centering: a framework for modelling the local coherence of discourse, *Computational Linguistics* 21/2, 202–225.

Hawkins, J (1978) *Definiteness and Indefiniteness*, London: Croom Helm.

— (2004) *Efficiency and Complexity in Grammars*, Oxford: Oxford University Press.

Korhonen, A, Krymolowski, Y and Briscoe, E (2006) A large subcategorisation lexicon for natural language processing applications, in *Proceedings of the 5th International conference on Language Resources and Evaluation*, Genova, Italy.

Sinclair, J M (1996) *Preliminary recommendations on corpus typology*, EAGLES Document TCWG-CTYP/P, retrieved from www.ilc.pi.cnr.it/EAGLES.

Vallduví, E (1992) *The Informational Component*, New York: Garland.

Williams, C (2007) *The Cambridge Learner Corpus for Researchers on the English Profile Project*, Version 2, internal English Profile report.

English Profile: functional progression in materials for ELT

ANTHONY GREEN CENTRE FOR RESEARCH IN ENGLISH LANGUAGE LEARNING AND ASSESSMENT, UNIVERSITY OF BEDFORDSHIRE

Introduction

The Common European Framework of Reference for Languages: learning, teaching, assessment (CEFR; Council of Europe 2001) is intended to provide common terms for users including learners, teachers, publishers and language testers to communicate about language learning and levels of language ability across borders, target languages and educational sectors. It offers scales that describe in general terms what language learners can typically do with a language at different levels of ability and raises key issues for the user to consider. The levels are presented as three broad categories A, B and C. Finer distinctions are made within each of these as between A1 and A2 and, for certain purposes, plus levels, such as B1+ or B2+, have also been described. The CEFR has been widely adopted around the world as a means of helping users to set targets for learning, teaching and assessment. However, because it is not specific to any one language, it cannot provide sufficiently detailed specifications of the grammar, vocabulary and other features of a particular language associated with each level.

The Council of Europe envisages that more detailed Reference Level Descriptions (RLD) will be required to expand on and exemplify the CEFR scales for specific languages to inform the planning of language programmes and for assessment. The Council of Europe has provided a guide for the development of these RLDs. According to this guide, the purpose of RLDs is 'to describe or transpose the [CEFR] descriptors that characterise the competences of users/learners at a given level in terms of linguistic material specific to that language and considered necessary for the implementation of those competences' (Council of Europe 2005). Already RLDs have been produced for German (Glaboniat *et al.* 2003) and Spanish (Instituto Cervantes 2006) with work underway for a number of other languages.

English Profile and the CEFR

English Profile is a long-term collaborative research programme to produce RLDs for English as a foreign or additional language. Building on the pioneering work, predating the CEFR, to specify learning target specifications for English by van Ek and Trim known collectively as the Threshold series and including A1 – *Breakthrough*, (Trim forthcoming) A2 – *Waystage* (van Ek and Trim 1998) B1 – *Threshold* (van Ek and Trim 1998) and B2 and above – *Vantage* (van Ek and Trim 2001), these are intended to provide an integrated, cumulative description of the communicative tasks learners can be expected to perform in English at each level of the CEFR. The project outcomes should assist English language teaching, learning and assessment; guiding curriculum development, teacher training, self-directed learning and language test development.

English Profile currently embraces three project strands:

- Language Pedagogy – focusing on a range of issues surrounding English language learning and teaching across cultures
- Curriculum and Assessment – focusing on how language skills develop, both in terms of learners' knowledge and how learners are able to use the language when taking tests
- Corpus Linguistics – working with computational linguists in Cambridge to investigate the language which learners actually produce at each level as represented in the Cambridge Learner Corpus.

This article reports on preliminary work undertaken at the University of Bedfordshire focusing on curricula and materials, with particular attention to the higher levels (B2–C2). The intention is to explore how the highest levels of the CEFR (C1 and C2) may be distinguished from the levels below. Initially, following the language model employed in the CEFR, our focus is on *communicative functions*, defined in the *Threshold* specifications (van Ek and Trim 1998) as, 'What people do by means of language… things such as describing, enquiring, denying, thanking, apologising, expressing feelings, etc.'

Bearing in mind the advice of the Council of Europe *Guide for the Production of RLD* (Council of Europe 2005) that their success will depend partly on the degree of consensus that they are able to marshal among language professionals we chose to employ a research methodology that would allow us to explore current beliefs among these professionals about higher levels of English language ability. Starting from the perspective of language functions and using the broadest distinctions between A, B and C, we have been gathering materials from a range of English language coursebook publishers, programme providers and testing bodies around the world to gain insights into how higher levels of English proficiency are portrayed and how these might be distinguished from lower levels.

In future phases of the project, the materials collected must also be approached from the perspective of notions and, ultimately, of grammatical or lexical exponents to provide a more complete picture, through the lens(es) of the CEFR, of the ways in which communicative language use is conceived in these materials. The claims about the language that typifies each level abstracted from the materials must then be tested against the experience of educators and evidence of the language actually produced by learners available from the Cambridge Learner Corpus and other sources.

This process, it is anticipated, will allow us, ultimately, to arrive at a thorough description of communicative English language use that employs the organising principles of the CEFR itself to bridge the divide between the framework and the multiple and varied practice(s) of specifically *English* language learning, teaching and assessment.

Communicative functions in the CEFR

In the Breakthrough, Waystage, Threshold and Vantage specifications, communicative functions are offered as learning targets. As learners at all levels may need or want to do similar things with language, the lists of functions provided are very similar across levels. For this reason it is suggested in the draft *Guide* that in principle, when developing RLDs:

> 'The typology and internal classification of the [functions should] remain unchanged from one level to another for the same language, the denomination of the same [function] remains unchanged.' (Council of Europe 2005, 8)

However, the Vantage specification makes it clear that more functions do emerge at higher levels, allowing for, 'a more sensitive sub-categorisation of functions' than at Threshold. The CEFR suggests qualitative change in the functions that characterise learner language between the A and B levels and the C level. The A and B levels appear to offer an expanding repertoire of communicative functionality reflected in expanding access to contexts for language use. Broadly:

- A1 is said to be characterised by 'a very finite rehearsed, lexically organised repertoire of situation-specific phrases' for expressing personal information
- A2 by the expansion of 'social functions'
- B1 by 'the ability to maintain interaction and get across what you want to, in a range of contexts' and by increasing flexibility
- B2 is said to involve 'argument, effective social discourse and… language awareness' with conversation management, discourse skills and negotiation appearing at the B2+ level.

This suggests that where the A1 learner is restricted in the contexts within which they can use language, the B2+ learner is capable of participating quite effectively in the public sphere in expressing points of view and putting forward arguments. The C level, on the other hand, appears to represent not so much an increase in functionality and in the range of contexts for use as increasing ease and subtlety of communication. At C1 there is, 'a deepening awareness of access to a broad range of language, which allows fluent, spontaneous communication' and, at C2 a level of, 'precision, appropriateness and ease' in using the language. It is apparent in the CEFR that passing from B2 to the C level should enable the learner to access higher education, professional fields of employment and the literary culture associated with a language.

Data collection

To explore the nature of this consensus, we surveyed such metadocuments as test specifications, proficiency scales, syllabuses, and coursebook outlines (and their support material in all media) at B2, C1, C2 levels including English language curricula in various countries to cover Europe, South America, Asia, the Middle East and the UK.

The very diversity of the materials collected introduces a challenging complexity to the task of relating them. Our review of the literature pointed to the eclectic nature of pedagogic materials and the variety of ways in which they are informed by conceptions of communicative functions. We found similar variation in the materials we collected. As the schemes vary so much in how they conceive of and treat functions and in the detail they provide, it was not always a straightforward matter to transfer these to our database. Following discussion, lists of Can-do statements were also considered in our analysis on the grounds that many these include references to language functions as we conceive of them in our literature review and appeared either to complement or to substitute for lists of functions in the schemes we analysed.

In summary, our sources included lists of functions and Can-do statements extracted from:

- Publications disseminated by educational institutions available for download on the world wide web such as test specifications and handbooks, proficiency scales and support materials for textbooks
- Publications sourced through *English Profile* partner organisations such as the British Council and Cambridge University Press
- Bestselling international textbooks and related support materials identified via sales figures supplied by the Bournemouth English Book Centre (a specialist supplier of ELT materials)
- Curriculum and syllabus documents sourced by contacts expressing interest in *English Profile* or attending seminars introducing the Programme
- Materials sourced from educational contexts in which the project team members had previous experience.

We developed a database that would allow us to organise, store, and retrieve the lists of functions and Can-do statements taken from each of the schemes included in the study. Because the database is designed to record the occurrence of functions across the schemes, it facilitates comparisons between these. This allows us to search for commonalities within the C level across contexts and to identify how the C level might differ from B2.

Further comparisons could be made between the Council of Europe categories (Wilkins 1976) that formed the basis for our own definition of language functions and the functions described for each of the schemes analysed.

In the first instance a process of interpretation was needed to relate the functions and Can-dos we had listed to the Council of Europe levels. Linking to a framework such as the CEFR is challenging; the procedures for linking examinations to the framework are the subject of an extensive manual and reference supplement produced by the Council of Europe which reveals how complex a task it is to provide substantive evidence of a relationship between a test and the CEFR. Less work has been done by the Council of Europe to support the linking of learner portfolios, textbooks and curricula to the CEFR, but there is no reason to suppose that relating these would be any more straightforward.

In relating the materials we collected to the CEFR for the purposes of our analysis, guidance was sought from the relevant literature, from the materials themselves (most of which claimed some link to the CEFR, with or without supporting evidence) and from external sources that had attempted to map schemes to each other or to other external frameworks. We do not intend to suggest that these linking procedures meet the necessary standards to support the claimed relationships and the links we draw should not be taken as anything more than speculative. At this stage, however, our focus is on the descriptions used rather than the performance they describe and we believe that the approach we have taken to linking the materials we have collected can help us in our limited aim of outlining possible differences between learners at the B2 and C levels.

As noted above, the expansion of functional categories in Breakthrough, Waystage, Threshold and Vantage and the progression found in the CEFR suggest that the C level may introduce new functions that do not appear at B2. For the purposes of analysis, we sought a comprehensive set of functions that could highlight the differences between levels and that might highlight areas that the Threshold series categories would not easily encompass. Following our review of the literature on communicative functions in the CEFR, we chose to use Wilkins' (1976) list of functions as a basis for the comparison between levels. The Wilkins categories are organised into three levels of detail. The six major categories (argument, emotional relations, judgement and evaluation, personal emotions, rational enquiry and exposition, suasion) are subdivided and each sub-category has a list of associated micro-functions. Although Wilkins (1976) does not claim that these lists are complete or exhaustive, they do provide the most comprehensive reference list of functional descriptors available with close to 300 functions listed.

Analysis

We collected over 300 statements in the form of functional descriptions or Can-do statements covering the B2 and C levels. Our analysis of these statements was undertaken in two stages. The first involved comparing the proportion of function words found in Wilkins (1976) that also appeared in our database at the B2 and C levels. The second involved an exploration, using corpus analysis tools, of the contexts in which the functional descriptors occurred.

In the first stage, using our reference list of function words from Wilkins (1976), we analysed the sets of functions and Can-do statements from the documents we had collected to establish which function words occurred at each level. This provided a broad indication of which functions might be more closely associated with the C levels.

In the second stage, we refined our exploration of the materials, carrying out a keyword analysis to identify which functions were particularly associated with each of the two levels and which of these were common to both. In identifying keywords, we employed *KeyWords Extractor*, a programme in the *Compleat Lexical Tutor* online lexical analysis suite (www.lextutor.ca), which carries out comparisons between wordlists and a reference corpus of general language. The programme 'determines the defining lexis in a specialised corpus, by comparing frequency per word to frequency in a reference corpus' (Cobb 2000). The programme computes a *keyness factor* representing the number of times more frequent the word in the focal text is than in the default reference corpus (the Brown corpus). By highlighting the specific vocabulary that characterises a text, the keyword analysis provides an indication of topic. In the case of our lists of materials it offered a means of identifying which words were salient to each level. Having extracted lists of keywords describing language functions for each level, we then compared the contexts in which the keywords common to both occurred at the B2 and C levels. To do this, we used the programme *WordSmith Tools* (Scott 2004) to generate concordance lines indicating how the contexts for these keywords differed across levels.

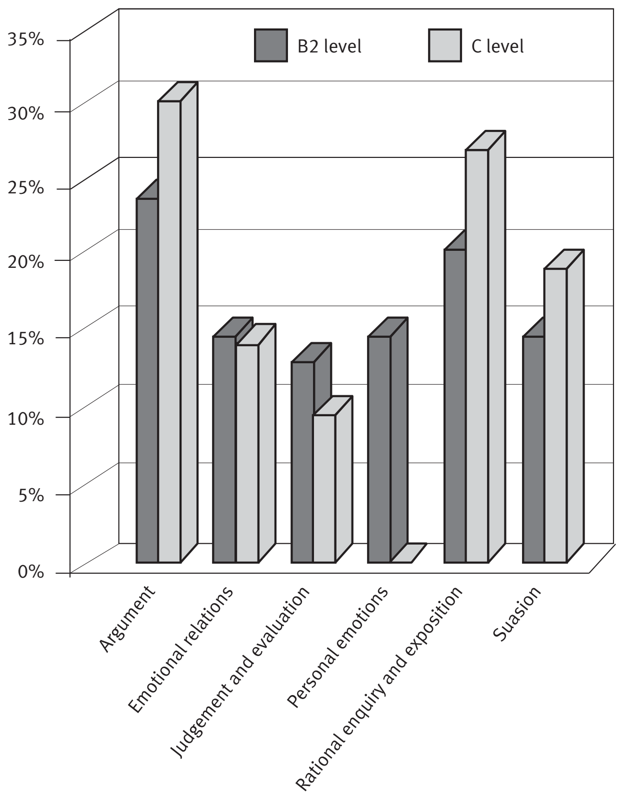
Results

Of the two hundred and fifty-eight micro-functions listed by Wilkins (1976) seventy-four (just over 28%) appeared on either the list of functions or of the Can-do statements that we compiled. There is limited agreement between the two lists with thirty-two of these words appearing both as Can-dos and as functions. Focusing on the words that appeared only at the C level, there was no direct agreement between the lists of words from Wilkins (1976) found in our lists of functions and those found in our list of Can-dos.

Figure 1 shows the proportion of the Wilkins' (1976) function words associated with each major category that were found at each level. Sixteen (50%) of the 34 functions listed by Wilkins under *rational enquiry and evaluation* appear here at either the C or B2 level or at both, compared with just nine (15%) of the 62 listed under *judgement and evaluation*. Among the micro-functions in this category that were not included here were; *excuse, pardon, conciliate, forgive, approve, commend, applaud, accuse, condemn* and *deplore*.

As the number of statements varies by level, it is appropriate to compare the proportion rather than the raw number of functions in each category for the B2 and C levels. A comparison of Wilkins (1976) functional categories (Figure 1) across functions and Can-do statements suggests that at the C level we see an increase in *argument, suasion* and *rational enquiry and exposition*. The proportion of functions is similar at the B2 and C levels for *emotional relations* while *judgement and evaluation* and, particularly, *personal emotions* seem to become less salient at the higher level in the materials we analysed.

Figure 1: Proportion of function words in Wilkins' (1976) categories found at the B2 and C level in the materials database



The following of Wilkins' (1976) functions occur only at the C level in the lists of functions. Wilkins' superordinate categories of communicative function are given in parenthesis.

- acknowledgement (emotional relations)
- disagree; inform (argument)
- justification; proposition (rational enquiry and exposition)
- recommend; threaten (suasion).

Among the Can-do statements, the following occur only at the C level:

- advocate (argument)
- assess (judgement and evaluation)
- demonstration; illustration (rational enquiry and exposition).

Clearly it would be wrong to infer from this comparison that none of the functions in these lists can occur at the B2 level. Taking the example of *inform*, synonyms such as *tell* or *give information* are found at B2 and below and so the choice of the word *inform* at the C level may suggest rather a subtle change of perspective. However, taken together, the pattern of function words emerging at the C level is suggestive of a shift in focus and points to *rational enquiry and exposition*, *argument* and *suasion* as being of particular relevance to the C level. It is notable that a similar conclusion was reached by Shaw and Weir (2007) in their retrospective analysis of writing examinations at these levels.

In the second stage of the analysis we carried out a keyword analysis of the materials. We first set out to identify keywords appearing at each level (B2 and C) that had no counterpart at the other level. We found that a large number of keywords were shared between the two levels with *texts, topics, contexts, discourse, vocabulary, lexical, grammatical, punctuation description, narration, formal* and *appropriate* among those with the highest keyness factors for both lists. In fact, all of the keywords at B2 level with a keyness factor greater than 100 either appeared among the C level keywords or were members of the same word family as a C level keyword.

Selecting words with a keyness factor of 100 or more, the following words were identified as unique to the C level with no closely related words (those in the same word family) appearing as a keyword at B2 level: *coherent, colloquial, confidently, conveys, critically, demonstrates, edit, evaluate, finer, genre, integrating, interprets, proficiency, slang, structurally, structured, subtleties.*

Of these, *judgement and evaluation* is one of Wilkins (1976) major categories, *demonstration* and *interpretation* are listed under *rational enquiry and exposition*. *Structuring discourse* is one of the *Vantage* categories and also implies *editing* (although this is not explicitly mentioned in *Vantage*). The appearance of such words as *colloquial, slang, genre* and *subtleties* point to the importance of the sociolinguistic dimension of competence at this level while *confidently, critically, edit* and *finer* suggest accuracy and a degree of metalinguistic, pragmatic awareness and critical distance. In terms of the 'functional competence' described in the Bachman and Palmer (1996) model, the C level appears to open up opportunities for learning through language (heuristic function) and creativity (imaginative function) that are not available at the B levels.

Where the comparison of functions by category described above indicates salient areas of difference between B2 and C, these keywords could be considered to define the shared ground between them. On the other hand, the contexts within which each of these words appeared at the different levels provided a further indication of the distinctions being drawn.

Having identified the keywords for the analysis, we used *WordSmith Tools* to generate concordance tables for each. The concordances showed the contexts in which each keyword appeared at the B2 and C levels and allowed us to compare these, further highlighting relevant differences in the expected performance of learners across levels.

The keywords include the following functions from Wilkins (1976): *argue, confirm, describe, hypothesis, infer, predict, report.*

In addition, we also considered the following keywords, some of which (*compare*, *express*) also implied a language

Figure 2: Example of Concordance Lines for ‘argue’ at the C Level

write a well structured detailed	ARGUMENT	emphasising important points
follow a complex	ARGUMENT	even when it is not clearly structured
develop an	ARGUMENT	further
understand in detail an	ARGUMENT	in a discussion programme
present a clear smoothly flowing description or	ARGUMENT	in a style appropriate to the context

function, others suggesting notions (*emotion*, *information*) or social factors (*context*, *audience*): *attitude, audience, compare, context, emotion, event, express, fact, idea, information, instruction, intend, issues, knowledge, opinion, purpose, question, viewpoint.*

In the following paragraphs we describe the contexts in which one of these words, *argue* (and its derivatives *arguing, argues, argument, arguments, argumentative*), appears at the B2 and C level and consider the implications for differentiating between B2 and C level learners.

Figure 2 shows a few concordance lines for *argument* taken from C level descriptors. Using these concordance lines, we built up comparisons between the contexts for keywords that appear at each level. Figure 3 shows in the form of a diagram the contexts for *argument* and its derivatives at each level.

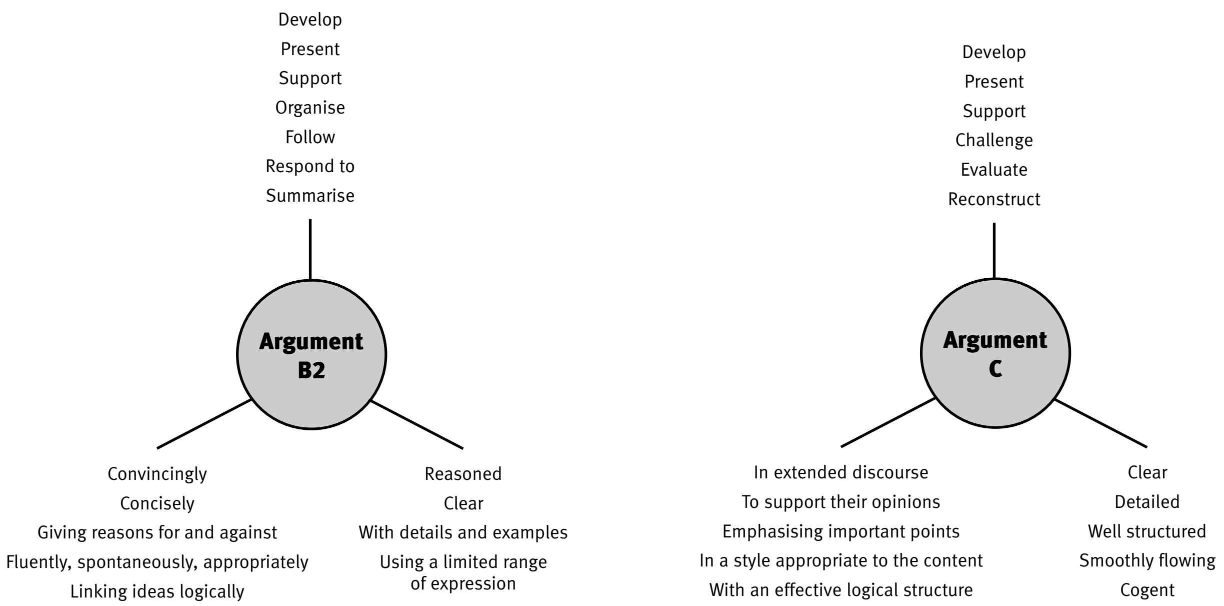
Tracing the occurrence of *argue* in the materials with illustrative examples, at the B2 level, learners are said to be able to present an argument, developing it with ‘reasons for and against’ (Sabanci University Syllabus) or details and examples and organising their essay length written work or talks by ‘linking ideas logically’ (face2face Upper Intermediate). However, this level of skill may depend on a familiar topic and involve a ‘limited range of expression’ (face2face Upper Intermediate). At C level, these constraints are removed and the learner is expected to be able to use a ‘wide range of complex grammatical structures’ (City & Guilds International ESOL Expert Level), shaping their production to generate effective arguments that highlight important points and use a ‘style appropriate to the context’ (New Headway Advanced). At this level arguments are not only coherent, but well or consistently supported and developed. However, even at the C1 level, learners may attempt to avoid argument tasks ‘by resorting to simplification through the use of description or narration in place of argument or hypothesis’ (ACTFL Proficiency Guidelines Advanced High). Receptively, learners gain the ability to follow subtle and inexplicit arguments including those found in a serious newspaper or discussion programme; at B2 arguments may need to be ‘clearly signalled’ (New Cutting Edge Intermediate) and the topic familiar.

Conclusion

The analysis has suggested that, in spite of the diversity of the materials we collected, it is possible to abstract a broad set of claims about the distinctive features of C level performance that could be refined, enhanced and recast as a coherent set of communicative functions. These claims could then be tested against both the perceptions of educators and the analyses of learner language towards the development of RLDs for English.

Our analysis suggests that such function sets as arguing, criticising and evaluating; editing, refining and structuring; defining, defending and justifying; abstracting, interpreting and integrating are salient at the C level, together with increasing sensitivity to the relationship between choices

Figure 3: Contexts for *argument* and its derivatives at B2 and C levels



of linguistic form and contexts for use. Although the Threshold series functional categories of *imparting and seeking information, expressing and finding out attitudes, socializing, deciding on and managing the course of action, assuring and repairing communication and structuring discourse* may be capable of encompassing the range of functions identified at the C level in the materials we have analysed, this categorisation does not, in itself, seem to discriminate clearly the heuristic, imaginative and pragmatic opportunities (Bachman and Palmer 1996) that the C level appears to open up for the learner. It may be that additional categories of function are required or that a different approach to presenting them should be found to describe the C level to better capture the progression that is suggested.

A criticism of the CEFR has been that there is not enough in the illustrative scales that relates to young learners (at lower secondary or below) or migrant learners of languages. The schemes we have collected to date do include some intended for migrant learners of English (the UK Adult ESOL core curriculum and Canadian Language Benchmarks 2000), and so might help to address the second of these gaps, but we did not find material that was explicitly addressed to learners younger than 16. As is apparent from our analysis, complex language of the kind associated with the C levels is most often used to address complex subjects and questions have been raised about whether it is as much linguistic as cognitive complexity that is required to carry out tasks at the higher levels; cognitive complexity that younger learners cannot be expected to have acquired.

In relation to this issue, the distinction made in the CEFR between interaction and production in spoken and written language appears particularly important to the C level. In productive language use, which North (2007) reminds us is associated with cognitive academic language proficiency (CALP; Cummins 1979) and formal, planned uses of language, accuracy, complexity and precision are of central importance. In interpersonal communication skills it is idiomaticity, fluency and sensitivity to social context that come to the fore. This will need to be reflected in the proposed RLDs.

Domain, and learners' ability to manage the pragmatic implications of the personal, professional, educational and public contexts, appear to be central to the C level and will have implications for how linguistic exponents of the specified functional and notional categories can be presented. In addition to structural exponents, there is a need to exemplify the kinds of sophistication and subtlety with which C level learners are said to be able to select appropriate forms from a repertoire of options.

Can-do statements at the C level of the CEFR have proved the most difficult to scale (North 2000) and this raises the possibility that the Can-do approach is better suited to the clear-cut tasks described at the lower levels. It is certainly more straightforward to make the constrained judgement about whether a learner is able to 'introduce him/herself and others and… ask and answer questions about personal details' than to decide whether they 'can use language flexibly and effectively for social, academic and professional purposes' with all that this assumes about the range of contexts within which the learner uses language, and the interpretation required on the part of the judge in interpreting such phrases as 'flexibly' and 'effectively'. It is an often-voiced criticism of outcomes-based assessment schemes that as tasks become more complex and distinctions between success or failure more nuanced, the value of the Can-do approach may begin to break down.

Building on this study it should now be possible to draft a set of functional categories that define the salient features of the C levels with respect to communicative functions. As discussed above, these categories in themselves do not provide sufficient detail to define the C level. It would however also be helpful to carry out a more detailed investigation of some of the material we have included in this study not only at the level of the metadocuments used here, but also analysing the content of the textbooks and test materials to which they relate. This would help us to cast more light on how the functions described are operationalised by course and test providers. These descriptions must then be:

- integrated with notions, situational contexts and themes
- tested against and contextualised by grammatical exponents using empirical evidence collected from naturalistic settings, the classroom and the examination hall
- tested and calibrated against the perceptions of language educators in contexts where the *English Profile* may be used
- revised on the basis of feedback and matched to finer-grained descriptions of level (C1, C2).

Achieving a scaleable set of descriptors for the English C level RLDs will clearly be a challenging task for the *English Profile* consortium and will require a flexible approach to the current CEFR scales. This article represents a preliminary and necessarily tentative step, but one that will now enable us to begin the empirical work, informed by the perspective of the educators who use the CEFR, of specifying and refining the description of the C levels for English.

References and further reading

Bachman, L F and Palmer, A S (1996) *Language Testing in Practice*, Oxford: Oxford University Press.

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

— (2003) *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment, Manual: Preliminary Pilot Version*, Strasbourg: Council of Europe Language Policy Division.

— (2005) *Guide for the Production of RLD*, Version 2, retrieved from www.coe.int/t/dg4/linguistic.

Cummins, J (1979) Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters, *Working Papers on Bilingualism*, 19, 121–129.

Glaboniat, M, Müller, M, Rusch, P Schmitz, H, and Wertenschlag, L (2005) *Profile Deutsch*, Berlin: Langenscheidt.

Instituto Cervantes (2006) *Plan curricular del Instituto Cervantes*, Madrid: Instituto Cervantes/Biblioteca Nueva.

North, B (2000) *The Development of a Common Framework Scale of Language Proficiency*, New York: Peter Lang.

— (2007) The CEFR Illustrative Descriptor Scales, *The Modern Language Journal*, 91/4, 656–9.

Scott, M (2004) *WordSmith Tools*, Version 4, Oxford: Oxford University Press.

Shaw, S and Weir, C J (2007) *Examining Writing in a Second Language*, Studies in Language Testing Volume 26, Cambridge: Cambridge University Press/Cambridge ESOL.

Trim J L M (forthcoming) *A1 Breakthrough Level of the Common European Framework of Reference*, commissioned by the Council of Europe.

Van Ek, J A and Trim, J L M (1998) *Threshold 1990* (2nd ed.), Cambridge: Council of Europe/Cambridge University Press.

— (1998) *Waystage 1990* (2nd ed.), Cambridge: Council of Europe/Cambridge University Press.

— (2001) *Vantage*, Cambridge: Council of Europe/Cambridge University Press.

Wilkins, D (1976) *Notional Syllabuses*, Oxford: Oxford University Press.

Language pedagogy in an era of standards

RADMILA BODRIČ UNIVERSITY OF NOVI SAD, SERBIA

Introduction

This article investigates the underlying trends of a common European education policy in language teaching, learning and assessment. It also focuses on how the teacher's role has changed in the light of these new tendencies. Finally, the article discusses how and to what extent this truly European dimension in language education has been embraced in the Serbian educational context.

Foreign language learning in Europe

The reasons why people want to learn foreign languages are many and varied. Generally, they revolve around the need to communicate. This need to communicate with one another across languages is, of course, an old and obvious one. In the changing economic and political context of Europe, this need for communication and cross-cultural understanding is more urgent than ever. Many long-established institutions have undergone a profound process of change leading towards the promotion of international understanding, tolerance and cultural diversity. One such institution is the Council of Europe, an intergovernmental organisation whose primary goal is to promote the unity of the continent, to protect human rights, pluralist democracy and the rule of law. Many Council of Europe projects in the field of modern languages have directed their efforts towards deepening mutual understanding among citizens in Europe.

During the last decade of the 20th century, enormous changes occurred in most European countries, which provided a powerful incentive for innovation and reforms of education systems. A number of these will be cited for the purposes of this article. They are as follows:

1. A strong process of integration setting up a uniform European education policy, which, in turn, has brought about standards ensuring easier international comparability of education systems, levels of achievement, and educational but also professional or employment mobility.
2. A tendency towards high-quality and efficient education, thereby creating standards for self-assessment and evaluation of all participants and segments in the process of education.
3. Ever-increasing expectations from schools to support the individuality of students and prepare them for future challenges through the adequate transfer of knowledge, skills, and responsible action in the classroom community. By doing so, schools enable students to grow beyond the boundaries of their own cultures.

A new paradigm for the new age

A new paradigm has been created for the new age, but as Kohonen (2007:17) notes 'changes of the magnitude of paradigmatic shifts…do not take place overnight. They are inevitably a function of time, conscious effort and explicit concrete support in any profession'. Indeed, change is not an event, it is a process, and the process of change (realigning educational practices to the requirements/standards of the new age) must understandably be an ongoing one and it must be implemented effectively. In 2001, the European Year of languages, the Council of Europe's Language Policy Division launched two fundamental documents which have influenced and still continue to exert a powerful influence on the developments in foreign language education across Europe and beyond. These documents are the *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* and the *European Language Portfolio* (the CEFR and ELP respectively in further text). The CEFR reference levels were developed independently of any specific language. The CEFR serves as a benchmark against which sets of reference level descriptors for any national or regional language can be compiled. Its creation is an ongoing process for it is constantly being updated and revised (Council of Europe 2001).

Since the methodological implications of the CEFR are many, it is extensively used by course designers, textbook writers, testers, teachers and teacher trainers – it is of interest to all who are directly involved in language teaching and testing. Its main aims include encouraging practitioners in the field of language to reflect on their practice, particularly in respect of practical language learning needs, objective setting and keeping track of learner progress. It owes its great popularity to the availability of a transparent and independent common scale of reference which ensures the international comparability of certificates and other materials.

Unlike the CEFR which is an instrument for professionals in language education, the ELP, which complements it, is an instrument for the learner. It has been introduced by the Council of Europe in order to help learners set clear language learning priorities, monitor their progress, and continuously assess the results of their learning. The ELP enables learners to record their language learning achievements and their experience of using other languages and meeting other cultures. In this way the ELP gives learners a say in their education, develops their self-awareness; it raises awareness of individual language competences and enhances intercultural awareness and promotes life-long learning of languages, in addition to encouraging non-judgmental self-assessment (see Schneider and Lenz 2001).

The concrete benefits different groups of people gain from implementing the CEFR and the ELP are as follows (Bodrič 2006):

Language learners:

- get involved more actively in the learning process, enabling them to set clear and precise learning goals
- get a clear idea of where they are and where they are headed on the way to achieving their language learning goals
- are able to present a standardised set of skills which are recognised by employers, educational and other institutions throughout Europe and beyond.

Teachers and schools:

- can help learners gain more autonomy
- gain a basis for a unified language and language teaching policy
- fulfill one of the prerequisites to official recognition.

Teacher trainers:

- can train teachers to integrate the CEFR levels in their teaching
- gain a comprehensive and standardised set of guidelines to pass on during pre- or in-service teacher training.

Employers:

- gain an accurate insight into an applicant's language and intercultural skills and experience
- are able to set up language training courses within the workplace using the ELP.

These two crucial reference documents along with the developments in theories of foreign/second language acquisition, in teaching methods, psycholinguistics, neurolinguistics, and other related disciplines have prepared the ground for reconsidering some long-standing pedagogic routines. The new age has imposed a changed role or profile of the foreign language teacher and learner alike. So what is expected of a twenty-first century teacher and what are the standards s/he should conform to?

The foreign language teacher: the changing role in the classroom

The language teacher of today needs to help "mould" a learner or citizen in linguistically and culturally diverse Europe, into a well-balanced, knowledgeable, reflective thinker, an inquirer and a communicator, that is to say, a true language learner and user who is ready to self-invest in their language education. To do this successfully, the teacher ought to possess professional skills, background knowledge, personal qualities and the urge to invest in and commit to becoming a life-long learner in the multidisciplinary fields of foreign language acquisition. In order to teach towards the demands of the twenty-first century, teachers need to undertake a range of activities, the key ones are described below.

Firstly, teachers need to grant language learners maximum access to learning and maximise the outcomes of the learning process. They can do this through providing a wide range of resources and encouraging learners to work out a variety of solutions to problem-solving situations, thereby fostering creativity, higher-order thinking, effective communication through collaboration, development of interpersonal skills and personal and social responsibility.

Secondly, teachers need to involve themselves more in curricula reviews by using information obtained from formative and summative assessment. In this way teachers can reasonably influence curriculum content and performance standards. Teachers' assessment practices are an important part of, and reference for, researchers in the validation of standards (see McKay 2005). If different approaches to teaching and assessment are to take place, they are, understandably, most likely to succeed if they come from the teachers themselves.

Teachers also need to utilise efficient, transparent assessment and evaluation procedures in order to fully maximise learning outcomes. In this respect, the CEFR and ELP are of paramount importance for they have brought much-needed clarification in the teachers' understanding of language learning proficiency levels, thereby helping to establish a unified language teaching policy throughout Europe and beyond. This is discussed below.

Finally, teachers need to continuously reflect on their teaching and make necessary changes based on feedback gained from consultations with students, results of assessment and evaluation processes, and their personal research, thus helping themselves reconstruct their experience and make a shift in pedagogic routines.

These are but some of the requirements of the multi-faceted role of teachers, but by far the most powerful influence on foreign/second language education which has established itself as a benchmark against which to measure the effectiveness of language instruction and learner performance has been the use of the CEFR.

The CEFR and foreign language teaching and testing

The CEFR has ushered in a new era in foreign language teaching and testing through extensive standard-setting and test equating. Given that national language qualifications vary in their standards and defined levels of language proficiency, the implementation of the CEFR as a benchmark has greatly contributed to widely-recognised modern language certification across Europe and worldwide. There exist a number of internationally recognised tests, aligned with the CEFR, which give an accurate and true reflection of candidates' communicative competency across the four language skills, all of them being real life skills. Many applied linguists and educators mention the 'washback' effect arguing that if important decisions are related to test results, then teachers will teach to the test by changing their teaching methods to target the specific areas tested. On the other hand, the 'washback' of these exams has also had a positive effect, inasmuch as it has led to a spontaneous and widespread adoption of CEFR standards on the part of many schools preparing candidates for international examinations that wish to remain competitive. There are dangers present in any large-scale attempt at standardisation, however, one must not lose sight of the fact that the results of standardised testing can be used to show what areas educators, and not just students, do well in and what areas are less developed, thus largely influencing the curriculum in general. Many European countries today are increasingly introducing language tests for migration and citizenship purposes, thereby providing evidence of attained proficiency in the official language of the country (Saville 2006:2). Saville (ibid:3) notes that:

> 'Whereas high-quality tests and appropriate uses can facilitate the process of language learning and integration of test takers within a multicultural society, poor-quality tests or inappropriate uses can endanger integration and may lead to social exclusion.'

Although there is general consensus that there is no "perfect" test for any given purpose, it is rigorous analysis, that is to say research, validation, constant monitoring and updating that make any given test highly applicable and reliable for decision-making purposes such as employment, education, etc.

The role of the CEFR is so powerful that CEFR levels of competence, treated as required performance standards, form the basis for the curricula of the corresponding language courses or examinations. For instance, the new Curriculum and Syllabus at the English Department of the Faculty of Philosophy (University of Novi Sad, Serbia), where I teach, aims to bring all students of English as a foreign language to the Council of Europe C2 Level (ALTE Level 5 – 'Mastery'). Therefore, in order to make the studies more efficient, a proposal has been made that students enrolling in the programme should already be excellent at B1 Level (ALTE Level 2) or at least average at B2 level (ALTE Level 3). The entrance examination speaking test materials are designed accordingly, to test the limits of the candidate's ability through a series of speaking tasks of increasing difficulty up to the threshold of Level C1 (i.e. to the upper band of Level B2). Experience indicates that some candidates at the Entrance Examination are already able to achieve, and in some cases exceed, this level of communication. The Serbian Ministry of Education has adopted the CEFR as the basis for language teaching and learning throughout primary, secondary and university education. In the same vein, national foreign language curricula at all levels are being rewritten following the specifications of the CEFR for it has been recognised that its implementation can raise the standards of L2 learning, teaching and assessment and facilitate the recognition of language qualifications from a wide range of sources. To illustrate, the Centre for Serbian as a foreign language, established at the Department of Serbian Language and Linguistics, Faculty of Philosophy, University of Novi Sad has done much to encourage the adoption of the CEFR. The syllabus has been designed to relate learner performance to the CEFR levels. The CEFR is thus used as a benchmark for designing courses, textbooks and tests respectively. Moreover, the Threshold level for Serbian as a foreign language has been completed but is still unpublished. Work on the remaining reference level descriptors for Serbian is in progress.

It must be duly noted that private foreign language schools in Serbia are at the forefront of foreign language teaching, learning and assessment. By organising tuition in small groups of five to ten learners, by being flexible and progressive, they are the first to adopt the latest developments in foreign language teaching. By far the greatest penetration of international examinations into the Serbian market has been through private language schools. Awareness is raised in teaching and management staff of the importance of international standardised examinations and of the benefits that schools obtain from promoting and incorporating them. Therefore, it is to these schools that we mostly owe the current level of standardisation and the good pass rates in examinations. A number of them have fully integrated and promoted European quality assurance standards for language teaching and testing provision. As with state-run schools, the quality of their services varies, but those at the forefront offer high-quality education and adhere closely to European standards. In addition, the professional development of foreign language teachers focuses on teaching and assessing according to the aims of the CEFR and its levels.

The CEFR is 'often a misconstrued and therefore a widely maligned document' (Dicker 2007:27). Specifically, there have been and still are those who believe that the CEFR is unnecessary and that it creates a great deal of needless "box-ticking" for both teachers and students. They also claim that it depersonalises something which, by its very nature, is very personal: the process of language learning and the teacher-student relationship. In addition, they claim that it is based on and therefore promotes mistrust in the teacher and their abilities both to teach and to assess.

Unfortunately, these are the dangers which are inherently present in any large-scale attempt at standardisation. Whether the impact of introducing the CEFR and the ELP will be as thoroughly dehumanising as their critics say or whether it will make language learning and teaching much easier and more efficient as their advocates state, remains to be seen. Should the CEFR and ELP prove to be worthy, they shall indeed usher in a new era in language teaching or, if not, be consigned to the past along with the other unsuccessful attempts at bringing that much-needed order to language and language learning. As a teacher and researcher, I am secretly hoping that this will, indeed, be the Holy Grail of language teaching and that we will be able to integrate it into our work without undue loss of individuality.

Conclusion

The massive project of uniting the people of Europe has led to the creation of a new educational scheme which presupposes plurilingual and pluricultural models of education, the creation and alignment of educational standards, levels of achievement in the educational process and the like. Given that foreign languages have assumed a high position in the curriculum, that they are being taught at an increasingly early stage (Eurydice 2005) and that students have a chance to learn at least two foreign languages during compulsory education, much has been done to encourage a longer-term productive change and a more context-sensitive approach to second/foreign language learning, teaching and assessment.

References and further reading

Bodrič, R (2006) The Common European Framework of Reference for Languages and European Language Portfolio – languages without frontiers, *Proceedings from the International Congress "Applied Linguistics Today"*, Belgrade: 28–30 May 2006.

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

Dicker, C L (2007) Whatch'a say? The brass tack realities underlying the CEFR, in Rasulić, K and Trbojević, I (Eds) *ELSII 75 Proceedings Vol. 2/ International Conference to Mark the 75th Anniversary of the English Department, English Language and Literature Studies: Interfaces and Integrations (2004; Belgrade)*, Faculty of Philology, University of Belgrade, 27–39.

Eurydice (2005) Key Data on Teaching Languages at School in Europe, Brussels: Eurydice European Unit.

Kohonen, V (2007) Developing Foreign Language Education through Transformative Teacher Growth, *Humanising Language Teaching, 9/1*, retrieved from www.hltmag.co.uk/jan07/index.html

McKay, P (2005) Research into the assessment of school-age language learners, *Annual Review of Applied Linguistics* 25, 243–263.

Saville, N (2006) Language testing for migration and citizenship, *Research Notes* 25, 2–4.

Schneider, G and Lenz, P (2001) *European Language Portfolio: Guide for Developers*, Lern- und Forschungszentrum Fremdsprachen University of Fribourg/CH, retrieved from www.coe.int/T/DG4/Portfolio/documents intro/Eguide.pdf

Arguing in English and Spanish: a corpus study of stance

JOANNE NEFF-VAN AERTSELAER UNIVERSIDAD COMPLUTENSE, MADRID

Introduction

This article examines the role of interpersonal and textual metadiscourse in the construction of persuasive texts written by English- and Spanish-speaking writers in order to contrast, quantitatively and qualitatively, articles written by editorialists in both English and Spanish. The pragmatic-rhetorical metadiscourse strategies used are those listed in Dafouz's (2007) taxonomy, which is loosely based on Crismore, Markannen and Steffensen's (1993) but with modifications in order to accommodate the particular characteristics of the languages contrasted. In the analysis presented here, the textual metadiscourse markers (used principally to organise the text for the readers) comprise logical markers, sequencers and glosses, while the interpersonal metadiscourse markers (used primarily to engage the reader in the argumentation) include hedges, certainty markers and attitude markers.

This study is part of a larger project (see Neff *et al.* 2006) which aims to describe Spanish EFL students' pragmatic competency, for which it is first necessary to provide base-line information on preferred and dispreferred rhetorical features in the texts of expert writers, in this case, those of English and those of Spanish editorialists.

Indicators of cultural differences

There are a number of studies, some from psychology (questionnaires and interviews) and some from the examination and/or comparison of conversation in Spanish with conversation in English that claim that Spanish (peninsular) conceptual structures may differ from those of other cultures. For instance, studies from social psychology (Fischer, Manstead and Rodríguez 1999) suggest that Spanish culture is honour-based, as compared to individualistic-based, which reflects values of Dutch culture. Important values in an honour culture 'centre around social connectedness, and should be the ones in which one's personal worth in relation to others, especially one's family, is apparent' (Fischer, Manstead and Rodríguez 1999:154). In more individualistic cultures, core values centre around the competent and independent self. In Hernández–Flores' (1999:40) discussion of politieness strategies in Spanish colloquial conversations, she proposes that an important component of Spaniards' sociocultural identity is the concept of affiliation, which is linked to *confianza*, or a closeness or sense of familiarity. Thus, one may behave as if one had a close relationship to a group in order to affirm his or her right to be a member.

Such underlying conceptual structures may be related to the findings by Ballesteros (1999) of differences in the use of positive- or negative-face strategies in speech acts carried out by Spanish or British speakers. In his study of requests and commands in Spanish and English conversations, Ballesteros shows that the Spanish speakers clearly preferred to use positive politeness in these speech acts, while the British speakers used negative politeness strategies.

These positive politeness strategies seem to be related to the use of *we* in spoken and written discourse in Spanish. Spanish grammar books discuss the use of *we* in formal academic, political, and journalistic writing as referring only to one speaker or writer.[1] Matte Bon (1999:266) refers to this as a politeness strategy, but in the common-sense meaning of the word, e.g., not connected to Brown and Levinson's theory. However, as the studies of Hernández-Flores (1999) and Fischer, Manstead and Rodríguez (1999) show, more than signaling the courtesy of the speaker, the use of *we* has to do with a politeness strategy of inclusion, that is, attending to the concept of *affiliation*. In the case of written text (and it seems in conversational practice as well, cf. Ballesteros 1999), the alignment between writer and reader (DuBois 2007:7) or the pragmatic-interactional configuration established by both the forms of address and the metadiscourse markers, examined in the present study and displayed in Table 1, seems to be somewhat different in Spanish (peninsular) and English.

Table 1: Social distance – authority scale

	Social distance (solidarity factor)	Power relationships (authority)
Spanish	—	—
English	+	—

The social relationships, based on perceived rights and duties (Leech 1983:126[2]), are important in establishing the status and thus, the stance of participants. The present study of metadiscourse suggests that the social relationships between writer-reader are perceived differently in Spanish (peninsular) and English, in terms of social distance. While the Anglo editorialists use a greater number of hedges, the Spanish writers use a greater number of certainty markers, a finding corroborated by other studies of newspaper editorials (Marín-Arrese *et al*. 2004). In general, the fewer hedges and more certainty markers displayed in the Spanish texts mean less indirectness and suggest that the Spanish writers' perception of their rights and obligations toward the reader is different from that of the Anglo discourse community.

1. Brown and Levinson (1987:127) refer to several types of "we"; the inclusive "we", as including the speaker and the hearer, is identified with positive politeness. Hyland (2001:223) points out that there is a growing preference for the use of "I" over "we" in academic writing in the hard sciences.
2. In their well-known study of forms of address, Brown and Gilman (1960) use the term "solidarity factor", which, following Leech (1983), I refer to here as "social distance". Although forms of address have been changing somewhat radically and complexly since WWII in many different European countries (Clyne *et al.* 2003), one can still find indices of differences in social distances between the Anglo and Spanish discourse patterns. More recent usages may depend on many different factors: age, political affiliation, educational status, etc.

Metadiscourse as indicators of cultural differences

The issue of cultural differences, as reflected in academic writing, has long been controversial, ever since Kaplan (1966) put forth his much debated proposition of "cultural thought patterns" in academic English as Second Language (ESL) writing. Although Kaplan has been criticised for his hypotheses, especially those related to "thought patterns" of groups of writers, a myriad of newer works continue to produce evidence which supports the supposition that language groups do indeed differ in their interactional patterns in academic writing.

Both Régent (1985), for differences between French and Anglo academic writing, and Clyne (1987), for differences between German and Anglo academic texts, have convincingly demonstrated the existence of cultural discrepancies as reflected in organisational patterns. Both studies deal mainly with preferences in the layout of information for readers, although Clyne also discusses rhetorical aspects, such as how much digressive argumentation is permitted in German academic writing as compared to the Anglo discourse community.

Instead of studying academic texts, the analysis presented here focused on (signed) opinion articles, a subclass of persuasive texts (van Dijk 1988) whose final aim is to convince the readership by means of logical (i.e. textual) as well as affective (i.e. interpersonal) strategies. Opinion articles are eminently subjective, since they reflect the writer's personal stance, but at the same time they need to appear to be established on objective bases in order to be accepted by a wider discourse community. Both types of metadiscourse, textual and interpersonal, contribute to the construction of a persuasive text through the use of balancing mechanisms. While logical markers, sequencers and code glosses bring cohesion and coherence to the text, hedges, certainty and attitude markers achieve persuasion through writer-reader identification and personal involvement.

In this study, the concept of metadiscourse is used as a means to identify possible differences or similarities between the texts, languages and cultures compared. Metadiscourse is defined here 'as the linguistic resources used to organise a discourse or the writer's stance towards either its content or the reader' (Hyland and Tse 2004:157). This notion includes a heterogeneous array of cohesive and interpersonal features which help relate a text to its context by assisting readers in organising and interpreting the text in a way preferred by the writer and with regard to the dominant rhetorical conventions of the discourse community. A variety of metadiscourse taxonomies have been proposed (Beauvais 1989; Crismore, Markkanen and Steffensen 1993; Hyland 2005; Vande Kopple 1985, 2002), which adopt different approaches and propose different metadiscourse categories. Generally speaking, most of these classifications adopt a functional perspective (with the exception of Beauvais' model) and distinguish between textual and interpersonal dimensions. While the textual dimension refers to the explicit resources used by the writer to organise the text and guide the reader through it, the interpersonal dimension is essentially interactive and evaluative and expresses the tenor of the discourse. More recently, Hyland (2005:45) has replaced the terms *textual* and *interpersonal* with Thompson's (2001) *interactive* and *interactional* since Hyland views all metadiscourse as referring 'to interactions between the writer and the reader'.

While I agree with the interactive function of both textual and interpersonal markers, the taxonomy followed here still follows the traditional terminology and adopts a functional perspective; however, I have included a second level of analysis to respond to the differences identified in the pair of languages compared (English and Spanish). Previous contrastive research (Dafouz 2003) has shown that existing taxonomies needed to be modified to accommodate the meanings expressed in other types of texts not previously analysed, in this case, newspaper opinion articles. The taxonomy used in this analysis differs mainly in the sublevels identified in the initial macro-categories devised by Crismore *et al.* (1993), since initially it was found that at a functional level Spanish and English texts seemed to coincide in the number and frequency of metadiscourse categories used. However, a closer examination revealed that the types of markers used varied considerably. The next section will develop in detail the types of metadiscourse categories surveyed in this study.

Method

The corpora of expert writers compared in this study come from the English-Spanish Contrastive Corpus (ESCC) of argumentative texts (signed editorials) dealing with international affairs, economy and the European Union (Marín and Neff 2001), 113,475 words in English and 151,011 words in Spanish.[3]

The analysis for the different metadiscourse categories was based on Dafouz's 2007 taxonomy. WordSmith Tools (Scott 1999) was used to find the most frequent items, although many tokens had to be analysed manually to avoid ambiguities in the classification. For example, in the case of logical connectors, it was necessary to eliminate from the concordance lines all the cases in which *and* or *but* were not used as inter-clausal connectors.

Within the textual metadiscourse categories (text organisers), this study examines only three textual metadiscourse markers (logical markers, sequencers and code glosses), since in a previous analysis (Dafouz 2003) with a smaller corpus, these three showed statistically significant differences in the professional writing of English and Spanish editorialists. Logical markers refer to cohesive elements (principally connectors) which help readers interpret pragmatic connections between ideas by signaling additive, contrastive and resultative relations in the writer's argumentation. These markers are subdivided into four types:

- additives – *and, moreover, in addition, furthermore, also*
- adversatives – *but, however, nevertheless, yet*
- consecutives – *consequently, finally, thus, therefore*
- conclusive markers – *in conclusion, concluding, to conclude, to sum up/summing up.*

3. For the future contrastive EFL study, data from these corpora will be compared with two other corpora: LOCNESS, (149,790 words) essays written in English by American college students, and SPICLE (194,845 words), essays written by 3rd- and 4th-year EFL Spanish students.

Sequencers mark particular positions in a series and help guide the reader along the lines of the argumentation. In this study the following items were analysed: *first, second, third, in the first place, on the one hand, on the other hand, last, next, then*. The third type of textual markers examined were code glosses, which explain, rephrase or exemplify textual material. These markers were divided into two types:

- reformulators – *in other words, that is*
- exemplifiers – *for example, for instance, such as.*

As in the analysis of the textual constituents, three major interpersonal markers were studied: hedges, certainty markers and attitude markers. Hedges express partial commitment to the truth-value of the text in the form of epistemic verbs (*may, might, seem, could*), probability adverbs (*maybe, perhaps*) and other epistemic expressions involving adjectives (*it is likely, it is possible*). Certainty markers, on the other hand, emphasise the force or the writer's full commitment to the certainty of the message (*clearly, undoubtedly, obviously*), while attitudinal markers express the writer's stance to the propositional content in the message and the readership, by using deontic verbs (*must, have to*), attitudinal adverbs (*surprisingly, unfortunately*) and other attitudinal expressions (*it is necessary, remarkable*). Table 2 summarises the metadiscourse categories analysed and offers some examples.

For the study of the data, the statistical differences were calculated by norming the results per 10,000 words and comparing the results of the two groups by using *Keywords Tool* in order to calculate the chi-square test of significance.

Table 2: Metadiscourse categories analysed

TEXTUAL MARKERS	Examples
Logical markers [additives, adversative, consecutive and conclusive markers]	*Furthermore, but, however, yet, finally, therefore, to sum up,* etc.
Sequencers [reformulators and exemplifiers]	*First, second, in the first place, on the one hand/other hand,* etc.
Glosses	*For example, for instance, in other words,* etc.
INTERPERSONAL MARKERS	
Hedges [epistemic verbs; probability adverbs/adjectives]	Epistemic verbs (*can, could, might, may*), *perhaps, it is likely,* etc.
Certainty markers [adverbs/adjectives]	*It is clear/clearly, it is obvious/obviously, it is certain/certainly,* etc.
Attitude markers [Deonic modals; attitudinal adverbs/adjectives]	Deontic modals (*have to, must, should*), attitudinal adverbs, (*unfortunately, surprisingly*) and attitudinal expressions (*it is necessary,* etc.)

Results: Preferred and dispreferred features in argumentative texts

The analysis, generally speaking, shows that the major differences between the Spanish and the English expert writers involve the use of both textual metadiscourse –

in particular, the use of logical connectors – and interpersonal markers

Results for textual markers

As can be seen in Table 3, the Spanish professional writers used more additive connectors, a total of 2620 versus the 1407 tokens present in the texts of their English-speaking counterparts. A closer examination showed that the Spanish group concentrated most of these markers in the form of additive linkers (1868 tokens), leaving adversatives, consecutives and conclusives in second place. The English group divided the use of logical markers into additives (824) and adversatives (549).

Table 3: Textural markers preferred in English and Spanish argumentative texts

Total TEXTUAL Markers	Logical		Sequencers		Glosses	
	Raw fig.	10,000W	Raw fig.	10,000W	Raw fig.	10,000W
English	1407	123	22	1.9	44	3.9
Spanish	2620	262	67	4.5	44	2.9
Differences		P>.005		P>.03		*P>.06

These results corroborate earlier studies (Dafouz 2003; Mauranen 1993; Neff *et al*. 2004) which proposed that differences in the use of additive and adversative markers may be due to the way cultures construct argumentative texts. Mauranen (1993:236) argued that while Finnish writers present their claims in the form of a conclusion, Anglo-Americans present them in the form of a result. In other words, Finns build the argumentation using a progressive strategy that entails moving forward in the presentation of ideas and adding evidence to the original claim. Anglo-American writing, on the other hand, exhibits a retrogressive strategy, which requires reconstructing the argumentation and presenting different sides of the argumentation to reach a plausible result. In the case of the Spanish editorial writing, the findings suggest that these authors may prefer to add justification to the original idea but continually develop the argumentation in the same direction (i.e. progressively) via additive markers. In contrast, the English-speaking discourse community makes a higher use of adversative markers in order to build arguments which contrast the pros and cons of an opinion.

As for code glosses, the findings reveal that both groups of professionals used them very sparsely (coinciding in 44 tokens, a non-significant difference, marked with an asterisk), a result which may respond to rhetorical conventions. Perhaps due to space constraints, editorialists do not seem to use these markers to organise their information nor to exemplify or clarify content. Thus linguistic economy may well function as a crucial criterion in the writing of opinion articles.

Results for interpersonal markers

The findings for the interpersonal markers, displayed in Table 4, reveal that two categories showed a significant difference: hedges and certainty markers. Both sets of writers have hedges as the greatest number of interpersonal markers, followed by attitude markers and lastly, certainty expressions. The findings revealed a significant difference (P<.007) between the number of hedges used by the Spanish editorialists and those used by the English-speaking editorialists, although caution must be exercised here, as some of the hedging of Spanish expert writers may be found in dependant subjunctive clauses. However, this finding does seem to coincide with a feature which other English-Spanish researchers (Williams 2005) have found for expert writing (biomedical articles) in Spanish: the “forcefulness” of stating opinions as compared with the negative politeness strategies used in the texts of expert Anglo-American writers.

Table 4: Interpersonal markers preferred in English and Spanish argumentative texts

Total INTER-PERSONAL Markers	Hedges		Certainty		Attitude	
	Raw fig.	10,000W	Raw fig.	10,000W	Raw fig.	10,000W
English	543	49.3	81	7.1	321	24.2
Spanish	406	36.4	122	8.2	388	25.9
Differences		P>.007		P>.05		*P>.9

In regard to the certainty markers, the Spanish editorialists used significantly more markers than did the English editorialists (P>.05). At a more delicate level of analysis, it can be observed that English experts prefer the adverbial forms (*certainly*, 18 tokens; *surely*, 15 and *clearly*, 11). The Spanish experts prefer more forceful evaluative adjectives and adverbials, such as *sin duda* [no doubt], 39 tokens; *es cierto* [it is true/certain], 16 tokens; and, *ciertamente* [certainly/truly], 13 tokens.

Although there is not a significant difference between the two groups as far as the number of tokens in attitude markers, again a closer examination does reveal some differences as to which particular markers are chosen. Spanish authors prefer the use of *deber* [must/should], with 237 tokens, while Anglo-American authors prefer *should*, with 153 tokens as compared to 83 tokens of *must*.[4]

Conclusion

The findings presented in this study of texts written by Spanish (peninsular) and Anglo editorialists suggest that there may be at least two major differences involving the stance adopted by each discourse community. The analysis of the textual metadiscourse markers: logical (additives, adversative, consecutive and conclusive markers), sequencers and glosses (reformulators and exemplifiers), showed that the Spanish texts had a significantly greater number of additive discourse markers than the English texts. The results for the Spanish writers may reflect Spanish rhetorical conventions, which seem to favour a

4. Previous studies (Neff 2006) have suggested that, as *deber* translates for both *must* and *should*, it appears that the Spanish EFL writers’ overuse of *must* is a reasonably good indication of transfer from the L1.

progressive argumentation strategy, building up evidence of the same type, clause by clause, and hence, the need for many additive markers. English argumentation strategies appear to prefer the initial establishment of the major premise(s) and then the posterior development of a balanced consideration of the pros and cons, and, thus, these texts show a more symmetric use additive and adversative textual markers. Concerning the category of glosses, which showed no significant differences between the two groups of editorialists, it seems that expert writers do not allow themselves this possibility of clarification, perhaps because editorial space is limited.

In the case of interpersonal metadiscourse (hedges, certainty markers and attitude markers), both the Spanish and English texts showed similar numbers in one of the three categories, that is, attitude markers. The two groups showed significant differences, however, in the use of hedges ($P<.007$) and of certainty markers ($P<.05$). A closer examination of the certainty markers revealed that English experts prefer the adverbial forms (*certainly, surely, clearly*). Although the Spanish experts use these forms as well as adjectival constructions (*es cierto* [it is true/certain]), they seem to favor more forceful evaluative adjectives and adverbials, such as *sin duda* [no doubt]; *es cierto* [it is true]; and *ciertamente* [certainly/truly]. For the attitude markers, although the numbers used by the two groups reveal no significant differences, the Spanish expert writers appear to favour assertive markers, such as *debe*, *hay que*, *tiene que*, all meaning 'must/have to'. Interestingly, the Spanish experts also use these devices as topic introducers, such as *debemos recordar el tema de la financiación de los partidos* [we should/must remember the topic of the funding of political parties]. Again, this result seems to point to different overall argumentation strategies, but not to divergent genre-driven strategies. The significant differences in the categories of hedges and certainty markers do, however, seem to suggest that the Spanish writers' conceptualisation of their reading public is different from that of the Anglo-American discourse community. This finding has been at least partially corroborated by translation research (Williams 2005) which has shown that Spanish research articles tend to present claims more forcefully than is the tendency in Anglo academic discourse.

By and large, the findings drawn from this study hold important implications for cross-linguistic comparisons and genre characterisation and also for the teaching of literacy in the EFL/ESL learning contexts.

References and further reading

Ballesteros, F (1999) *La cortesía verbal: análisis pragmático lingüístico de las exhortaciones impositivas en inglés y español: el ruego en el mandato*, Madrid: Universidad Complutense de Madrid.

Beauvais, P (1989) A speech act theory of metadiscourse, *Written Communication*, 6, 11–31.

Biber, D and Finegan, E (1989) Styles of stance in English: lexical and grammatical marking of evidentiality and affect, *Text* 9/1, 94–124.

Brown, R and Gilman, A (1960) The pronouns of power and solidarity, *American Anthropologist*, 4/6, 24–39.

Brown, P and Levinson, S (1987) (2 Ed) *Politeness*, Cambridge: Cambridge University Press.

Clyne, M (1987) Discourse structures and discourse expectations: Implications for Anglo-German academic communication in English, in Smith, L (Ed.), *Discourse across Cultures. Strategies in World Englishes*, New York: Prentice Hall International, 73–83.

Clyne, M, Kretzenbacher, H, Norrby, C and Warren, J (2003) Address in some western European languages, Moskovsky, C (Ed.), *Proceedings of the 2003 Conference of the Australian Linguistic Society*, retrieved from www.als.asn.au

Crismore, A, Markannen, R and Steffensen, M (1993) Metadiscourse in persuasive writing. A study of texts written by American and Finnish University students, *Written Communication* 10/1, 39–71.

Dafouz, E (2003) Metadiscourse revisited: a contrastive study of persuasive writing in professional discourse, *Estudios Ingleses de la Universidad Complutense* volume 11, 29–52.

Dafouz-Milne, Emma (2007) The pragmatic role of textual and interpersonal metadiscourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse, *Journal of Pragmatics*, 40/1, 95–113.

Du Bois, J (2007) The stance triangle, in Englebretson, R (Ed.), *Stancetaking in Discourse: Subjectivity in Interaction*, Amsterdam: John Benjamins, 139–182.

Fischer, A, Manstead, A and Rodríguez, P (1999) The role of honour-related vs. individualistic values in conceptualising pride, shame, and anger: Spanish and Dutch cultural stereotypes, *Cognition and Emotion* 13/2, 147–179.

Hernández–Flores, N (1999) Politeness ideology in Spanish colloquial conversation: The case of advice, *Pragmatics* 9/1, 37–49.

Hyland, K (2001) Humble servants of the discipline? Self-mention in research articles, *English for Specific Purposes* 20, 207–226.

Hyland, K (2005) *Metadiscourse. Exploring interaction in writing*, London: Continuum.

Hyland, K and Tse, P (2004) Metadiscourse in Academic Writing: A Reappraisal, *Applied Linguistics* 25/2: 156–177.

Kaplan, R (1966) Cultural thought patterns in intercultural education, *Language Learning* 16, 1–20.

Leech, G (1983) *The Principles of Pragmatics*, London: Longman.

Marín-Arrese, J, Hidalgo, L and Plaza, S (2004) Evidential, epistemic and deontic modality in English and Spanish: The expression of writer stance in newspaper discourse, in Facchinetti, R and Palmer, F (Eds) *English Modality in Perspective*, Berlin: Peter Lang, 121–139.

Marín, J and Neff, J (2001) Contrastive analysis of evidentiality in English and Spanish: corpus studies of argumentative texts written by native and non-native writers, paper presented at *II International Contrastive Linguistics Conference*, Santiago de Compostela (November).

Mauranen, A (1993) *Cultural Differences in Academic Rhetoric*, Frankfurt-am-Main: Peter Lang.

Neff, J, Ballesteros, F, Dafouz, E, Díez, M, Martínez, F, Prieto, R and Rica, J P (2004) Formulating Writer Stance: A Contrastive Study of EFL Learner Corpora, in Connor, U and Upton, T (Eds) *Applied Corpus Linguistics: A Multidisciplinary Perspective*, Amsterdam/Philadelphia: Rodopi, 73–90.

— (2006) A Decade of English-Spanish Contrastive Research on Written Discourse: The MAD and SPICLE Corpora, in Carretero, M, *et al.* (Eds) *A Pleasure of Life in Words. A Festschrift for Angela Downing,* Volume II, Madrid: Servicio de Publicaciones de la Universidad Complutense de Madrid, 199–218.

Régent, O (1985) A comparative approach to the learning of specialized written discourse, in Riley, P (Ed.), *Discourse and Learning*, London: Longman, 105–120.

Scott, M (1999) *WordSmith Tools*, Version 3, Oxford: Oxford University Press.

Thompson, G (2001) Interaction in Academic writing: learning to argue with the reader, *Applied Linguistics*, 22/1, 58–78.

van Dijk, T (1988) *News as Discourse*, Hillsdale, NJ: Lawrence Erlbaum.

Vande Kopple, W (1985) Some exploratory discourse on Metadiscourse, *College Composition and Communication* 36, 82–93.

— (2002) Metadiscourse, discourse and issues in composition and rhetoric, in Barton, E and Stygall, G (Eds), *Discourse Studies in Composition*, Cresskill, NJ: Hampton Press, 91–114.

Williams, I (2005) Thematic items referring to research and researchers in the discussion section of Spanish biomedical articles and English-Spanish translations, *Babel* 51/2,124–160.

Directness, imposition and politeness in English and Russian

TATIANA LARINA PEOPLES' FRIENDSHIP UNIVERSITY OF RUSSIA, MOSCOW

Introduction

Success in intercultural communication depends greatly on the understanding of communicative intentions of interlocutors and the pragmatic meaning of their utterances. Numerous problems in communication stem from the fact that people do not only speak different languages, but use them in different ways according to specific social and linguistic norms, values, and socio-cultural conventions. The problem is accentuated by the fact that such notions as *politeness* and *imposition* vary across cultures. Consequently, culture-specific communicative strategies dictate a different choice of language means applied in identical situations.

The goal of this article is to demonstrate the difference in directness vs. indirectness, perception of imposition and politeness in Russian and English communicative cultures, in particular, in Face Threatening Acts (FTA) and their influence on ethnic communicative styles. I will analyse sociocultural differences and values to show that in intercultural communication politeness does not always require indirectness while directness does not always presuppose impoliteness.

The intercultural aspect of politeness

As numerous cross-cultural studies have shown, politeness despite its universal character is a culture-specific phenomenon (Blum-Kulka 1992; Blum-Kulka, House and Kasper 1989; Hickey and Stewart 2005; Pizziconi 2003; Sifianou 1992; Watts 2003; Watts, Ide and Enlich 1992; Wierzbicka 1985, 1991, 1987 and others). Differences in politeness systems reflect differences in social relationships and are determined by culture-specific values. Blum-Kulka notes that 'systems of politeness manifest a culturally filtered interpretation of interaction' (1992:270).

People from different cultures do not always share ideas as to what is polite and what is not. The same verbal or non-verbal act being polite in one culture may be percieved as inappropriate or even rude in another culture. Watts (2003:14) claims that even the lexemes *polite* and *politeness* may vary in meaning and connotations associated with them. The results of my analysis of understanding politeness by Russian and English speakers (Larina 2004) corroborate this statement.

Being polite in another culture requires skills in using culture-specific strategies that might differ from those in your own culture and consequently require modification of verbal and non-verbal behaviour. Analysing how English is used by Russian speakers, Thomas notes that 'while having an excellent 'linguistic command' of the language, they often unwittingly appear discourteous or domineering' (Thomas 1984:227). She has well-grounded reasons for this claim. Russians are often perceived by westerners, especially by the British, as impolite people since they often sound over-assertive, argumentative, and even aggressive. They may ask private questions, they like to give advice and may interfere into conversation, they prefer to express their opinions as well as their communicative intentions directly, feel free to use imperatives and so on. But from the Russian point of view such conduct in many contexts can be quite acceptable and is not considered as impoliteness. On the other hand English politeness is not always assessed by Russians in a positive way. English people are often perceived by Russians as over-polite (which is not so good), ceremonious, and distant.

To understand differences in communicative behaviour which lead to such negative stereotyping we have to undertake an analysis of social relationships and cultural values. As Wierzbicka (1985:145) claims, linguistic differences are due to 'aspects of culture much deeper than mere norms of politeness' and are associated with cultural differences. The understanding of cultural differences which influence communicative behaviour is a necessary part of intercultural communicative competence.

Politeness is tied up with the most basic principles of sociocultural organisation and interpersonal relationships within social groups and should be viewed in the context of Social distance and Power distance which are considered the main dimensions of cultures (see Hofstede 1980, 1991). Social distance (D) and Power distance (P) usually go together: more individualist cultures are characterised by a lower P distance index, those which are more collectivist, have a higher P distance index. In these terms English and

Russian cultures maintain the following differences: English culture is individualist with low Power distance, while Russian culture is collectivist with a higher Power distance. In other words, the scale of social distance (D) (horizontal relations) is longer in English culture since the scale of power distance (P), reflecting the vertical hierarchical relations is longer in the Russian system than in the English one.

In different cultures the notion of 'distance' also varies. As Wierzbicka notes, 'in Anglo-Saxon culture distance is a positive cultural value, associated with respect for autonomy of the individual. By contrast, in Polish it is associated with hostility and alienation' (Wierzbicka 1985: 156). The same could be said about Russian culture where distance is often perceived as indifference. English proverbs emphasizing the value of distance (such as '*A hedge between keeps friendship green*'/'*He travels the fastest who travels alone*'/'*Come seldom, come welcome*' and others) sound rather odd to Russians as they value solidarity and closeness. Russian proverbs claim that '*It is better to have 100 friends than 100 roubles' ('Ne imey sto rubley, a imey sto druzey')/'Even death could be nice while you are among people' ('Na miru y smert' krasna')/'Without a friend one is an orphan, while having a friend one is a member of the family' ('Bez druga sirota, s drugom semyanin')*.

Social distance in English culture becomes prominent in the zone of privacy, which envelops every person irrespective of their age or status. It is one of the most important cultural values which regulates social relationships. Paxman (1999) calls it 'one of the defining characteristics of the English', 'one of the country's informing principles' and claims that 'the importance of privacy informs the entire organisation of the country, from the assumptions on which laws are based, to the buildings in which the English live' (Paxman ibid.117–8). Moreover, it is vivid in communication and explains a lot of peculiarities of English verbal and non-verbal behaviour.

'Privacy' is an essential English notion. A truly comparable word for it doesn't exist in other European languages (French, Italian, Spanish, or Polish). Neither does one find it in Russian. In different contexts it can be translated in different ways. As a cultural concept it can be interpreted as personal autonomy. The notice *Private* (*No admittance*) on doors is translated into Russian as *Strangers are forbidden to enter (Postoronnim vhod vospreschion)*. Privacy can be viewed as a personal zone where nobody is allowed to intrude.

The value of privacy in English culture encourages people to follow strict norms to protect their right from imposition, to demonstrate respect for independence of every individual. Meanwhile the lack of it in Russian culture and the high level of solidarity let people behave in a more direct way, being less formal and vigilant in guarding their own personal space than one is in English. This cultural difference explains a lot of characteristics peculiar to both politeness systems, as well as to their communicative styles.

Directness versus indirectness and politeness

To protect privacy, English communicative culture, which is perceived as essentially avoidance-based, negatively oriented (Stewart 2005:117) has the most conventionalised set of linguistic strategies for Face Threatening Act (FTA) redress, which are aimed at minimising the imposition on the hearer. One of the main strategies of independence is 'be conventionally indirect' (Brown & Levinson 1987). In all Speech Acts (SA) with pragmatic meaning *I want you to do it* the English make special effort to diminish and soften their imposition and show their respect for other people's autonomy (privacy). It does not matter whether the H is obliged to comply with the S (*Could I have the menu?*), whether the action is of benefit to the speaker (*I am just wondering if you could possibly help me with this*) or to the hearer (*Would you like to come to my birthday party?* or *Why don't you go and see the doctor?*).

Brown and Levinson (1987:142) claim that the indirect speech acts are universal and are probably constructed in essentially similar ways in all languages. Intercultural comparisons disprove the claim about the universality of indirectness. Watts (2003:15) writes that non-Russian commentators of the social behaviour of Russians note a high degree of unmitigated directness in various SA types, which contradicts the English tendency towards showing distance, reserve and formality. Comparative analyses of English and Russian communicative behaviour in such speech acts as Request, Advice, Invitation, and even Command (Instruction) show that though some of the negative politeness strategies introduced by Brown and Levinson do exist in Russian communication, they are much less important and frequent as in English (see Larina 2003, 2006). Moreover, in many situations where they are conventional in the English context, in Russian they might be perceived as inappropriate.

For example they are not typical for the situations, which due to the power of the Speaker or other contextual characteristics do not offer any contextual or pragmatic option for the Hearer but to perform the action. The English requests *Can I ask you to write down your answers?* (a teacher to their pupils)/*Could you please come to my office for a moment?* (a boss to his secretary)/*Would you mind moving your car please? It is parked illegally* (a police officer to a driver) perplex and amuse Russian speakers since indirectness is not used in those situations in Russian. In such contexts they would express their communicative intention directly using imperative: *Zapishite, pozhaluysta vashi otvety (Put down your answers, please)/Zaydite, pozhaluysta, ko mne v cabinet (Come to my office please)/Perestav'te vashu mashinu. Parkovka zdes' zapreschena (Move your car. Parking is forbidden here)*. Interestingly in the last situation even the modifier *please* can be omitted as the policeman possesses some power over the hearer who broke the rules, giving them the right to be direct.

In intercultural communication differences in strategies can cause communicative problems concerning not only the level of politeness/impoliteness but can lead to wrong interpretation of communicative intentions of the speaker. It might occur when a foreign speaker understands the form of the utterance and its linguistic meaning but is unaware of its pragmatic meaning. Giving an example of a pragmalinguistic failure, Thomas makes an interesting observation: the utterance *X, would you like to read?* which in an English classroom would be a highly conventionalised polite request/directive to do so, in a Russian classroom often elicited the response *no*, *I wouldn't* from students who had no intention of being rude, but who genuinely thought that their preferences were being consulted (Thomas 1983:101). Such situations when the ignorance of pragmatic differences causes misunderstandings and provokes communicative breakdowns and even conflicts are not rare. A Russian cosmonaut during his training in the US had a conflict with his English teacher who was not happy with his behaviour and wanted to refuse to teach him. When the interpreter tried to sort out the problem, the Russian assured him that he did not mean to upset his teacher and added: *'Actually she is a bit strange. Yesterday, for example, she asked me whether I would like to write the exercise. I just said 'No, I'd prefer to do it orally' but she suddenly got offended'*. The reason for their mutual misunderstanding lies in strategic differences. As in the Russian context students are not given options but are told directly what they are supposed to do the Russian cosmonaut failed to understand the pragmatic meaning of the teacher's utterance which did not concern his willingness but was a directive to write the exercise. The teacher was not aware of the fact that her request could have been interpreted as a question and her student did not mean to be rude.

Similar misunderstandings can occur at the workplace. An English businessman complained about his Russian secretary being very slow and inefficient. She needed several reminders to do each job. But, actually, her sluggishness is not so surprising since she was told *Could you possibly type these letters?* or *Could I possibly trouble you to take a moment to do it?* Russian speakers do not perceive such utterances as instructions to do something immediately but rather as very soft requests they can attend to when they have time available. They infer that taking care of such requests can be postponed.

Misunderstandings of this sort occur because in Russian communication indirect questions are not considered as a norm if the Hearer (H) is supposed to carry out the act and has no other option. In such situations Russian speakers fail to grasp the pragmatic meaning of the utterances as they differ from their literal meanings. In the Russian context the situations discussed above require direct requests which sound natural and quite polite: *Chitayte, pozhaluysta (Read, please)*/*Napishite, pozhaluytsa, eto uprazhneniye (Write this exercise, please)*/*Napechatayte, pozhaluysta, ety pis'ma (Type these letters, please)*. These examples confirm Thomas' observation that 'polite usage in Russian permits many more direct imperatives than does English'. She also points out that 'transferred into English, such direct imperatives seem 'brusque and discourteous' (Thomas 1983:102).

When the Hearer has a pragmatic option whether to do the act or not it can be expressed in Russian both by direct (imperative) and indirect (question) utterances: *Day mne, pozhaluysta, tu knigu (Give me that book, please)* or *Ne mog by ty mne dat' tu knigu? (Could you give me that book?)*. The only difference is that the Russian question is negative and the word please is excessive in the question. Nevertheless as numerous data reveal imperative sentences are preferable in Russian communication in Requests. An imperative modified by '*please*' is the most frequent form employed to make a request and it does not sound as demanding and imposing as in English. Indirect utterances are also possible but, as they sound more formal and distant, they are considered to be more appropriate for a high register of communication and mostly appear on the formal level or in the situations with 'a high cost' of the request . Traditionally they are hardly ever used in interactions between equals (friends, students); rather they are used by those who have more power (parents talking to children, bosses addressing their subordinates, etc.).

In various situations it is conventional and acceptable for Russians to approach the H directly. Directness in this case does not mean impoliteness. Note that in order to be more polite Russian speakers may intensify their request instead of softening it. Literarily such utterances can be rendered as *Be so kind, give me that book please (Bud' dobr, day mne tu knigu, pozhaluysta)* or even *Be so kind, give me that book please. I am asking you very much (Bud' dobr, day mne tu knigu, pozhaluysta. Ya tebia ochen' proshu)*. This example shows that while an explicit performative sounds too intrusive in English, and English speakers tend to avoid it, as 'in many circumstances it seems to imply an unequal power relationship or a particular set of rights on the part of the Speaker' (Thomas 1995:48), Russians on the contrary intensify their requests by adding very much (ochen'). In Russian communication explicit performatives are not perceived as intrusive and it is quite appropriate to say *I ask you…, I invite you…, I advise you…* and so on.

The English language has a more elaborate system of requestive utterances in comparison with Russian. For formulating requests it offers a set of models with numerous hedges, modifiers and downtoners which dissociate the H from the act and minimise the imposition: *Do you think you could possibly do X, please?*/*I was just wondering whether you could possibly do X* or *I was thinking maybe you wouldn't mind doing X*. Such formulas are not used in Russian as Russian speakers never minimise their imposition to such an extent, they are perceived as too elaborate, ambiguous and over-polite.

However, the main differences between the two languages in this respect emerge not so much in the set of formulas but in the choice of linguistic forms and strategies for formulating a request. An attempt to reduce the imposing nature of a request is much more typical of English communication than of Russian. Russians prefer more structurally direct requests than the English, who tend to choose structurally indirect constructions and make a bigger effort to minimise their imposition. A higher level of Power distance typical of Russian culture in asymmetrical contexts on the one hand and a higher level of solidarity on the other hand allow the Russian speakers to be more direct and demanding. Such speech behavior is taken for granted and at the same time does not sound impolite. This fact proves that in intercultural aspect directness does not always mean impoliteness and politeness does not always require indirectness.

Directness versus indirectness and imposition

While analysing directness versus indirectness it is necessary to be aware of the fact that this characteristic refers to the form of the utterance, its structure and language. However, there is no direct link between directness of the form and directness of its pragmatic meaning; in other words, between directness and imposition. On the pragmatic level, direct utterances (those which are formed by imperatives) do not always sound imposing. They may be softened by intonation, and context which includes numerous extra-linguistic features, such as the interlocutors' age, social position, the relationship between them, the place of interaction and other factors which reduce assertiveness. Above all, directness is determined by the conventions of the culture.

In intercultural communication it is also important not to consider indirectness as the main means of softening the imposition. In languages other than English various linguistic (lexical and grammatical) devices serve this purpose. In Russian culture, for example, an imperative modified by *please ('pozhaluysta')* as has been mentioned does not sound as demanding and imposing as in English. In Russian *please* seems to have a stronger pragmatic meaning than in English and it easily converts directives into requestives. Russian imperative modified by *please* is perceived as a polite request and is acceptable at all the levels of politeness.

Another important means of softening the imperative are forms of address. They play an important part in interpersonal communication because they contain information about a given language, indicate the social status of the interlocutors and the type of relations between them, in particular the degree of intimacy, social and status distance etc.

The Russian language offers a greater variety of forms of address than English. Firstly, it has the T/V distinction. The polite *vy (vous)* pronoun and verb forms used with it also modify the imperative and make it sound more polite. Secondly, among various forms of address the Russian language uses a patronymic name which follows the first name *(Vladimir Ivanovich/Maria Vladimirovna)*. Such combined forms of address (given name and patronymic) sound less formal and distant than honorific names *(Gospodin Nikitin/Gospozha Nikitina* which correspond to *Mr./Mrs. Nikitin)* but more formal and respectful than a bare first name *(Vladimir/Maria)*. Besides, there are a lot of diminutive names *(Volodia/Voloden'ka/Masha/Mashen'ka etc.)* used while addressing children, close friends, and family members. They are also acceptable at workplace when talking to peers, inferiors and younger colleagues. Thus it is considered appropriate for a boss to use an imperative utterance addressing when their middle-aged secretary *Maria Vladimirovna, napechatayte, pozhaluysta, eti pis'ma* (literal translation – *Maria Vladimirovna, type those letters please)*. Addressing a younger secretary, the manager might say *Mashen'ka* (diminutive of *Maria*), *napechatayte, pozhaluysta, eti pis'ma (Mashen'ka, type those letters please)*. Despite the use of the imperative verb, these utterances do not sound as commanding in Russian as in English. They are perceived as a softened command or even as a polite request.

Not only proper names but also many Russian nouns have diminutive forms. They express intimacy and affection and when used in imperative sentences, they reduce the imposition and make the utterance sound very soft: *Dochen'ka, prinesi mne, pozhaluysta, stakanchik vodichki*. On the pragmatic level this utterance, which is translated literally as *Daughter* [diminutive], *bring me a glass* [diminutive] *of water* [diminutive] *please* is not so direct and imposing as in English and can be compared with such an English model as *Honey, could you bring me a glass of water, please*.

Besides the linguistic and pragmatic differences in the structure and perception of imperative utterances we have discussed earlier, there is another important distinction between Russian and English impositive acts. Due to cultural differences such SAs as Request, Offer, Invitation, Advice and even Criticism and Reprimand are not as face-threatening and imposing in Russian as they are in British culture. This has to be taken into account when communicative intentions of Russian speakers are interpreted.

One of the reasons why Russians are perceived as imposing and demanding is their willingness to give advice to anyone, including strangers, even without being asked for it. Children seem to be the most frequent recipients of such 'wisdom'. They are recommended not to read on public transport since this is bad for the eyes. On a frosty day they are instructed by passers-by to put on caps that should protect them from catching a cold. It is common for shoppers to give advice to each other as to which brands are preferable. Giving advice to clients is considered one of the professional duties of shop assistants. Russian people receiving advice do not feel that it is a violation of their privacy, rather they interpret it as an act of goodwill and a sign of friendliness, and solidarity. Advice is perceived as an involvement speech act rather than a face-threatening act in Russian communication.

Offers and invitations are other vivid examples of how differently imposition is viewed in English and Russian communicative cultures. An invitation urges the invitee to accept rather than leaving them a comfortable way of rejecting it. As a result such English invitations as *It would be nice to have tea together, but I am sure you are very busy* (an example of English politeness strategies given by Scollon and Scollon 2001:51) or *I was wondering if you would like to come over to me for a meal this Saturday evening. I know it's fairly short notice. So please don't worry if you have other plans* (from real communication) would sound rather impolite and even offensive to a Russian speaker. Since giving options in these situations is inappropriate and could be interpreted as evidence of the Speaker's insincerity, rather than a demonstration of their respect for the Hearer, an indirect invitation may trigger a negative response. Note that Russian interlocutors while inviting would rather intensify their pressure on the H than give them options and soften the imposition. Thus instead of saying *It would be nice to have tea together, but I am sure you are very busy*, which sounds more than strange to Russian speakers, they could say *Let's go out for coffee.*

Stop working. Relax. It's time to have a break. In Russian context direct invitations would be more appropriate as the Speaker is expected to demonstrate a sincere desire to see the Hearer.

Scholars note that there are some instances when the imperative is polite in English, but these are rather limited. The imperative may be used when it is implied that the H will benefit from the suggested action, and the S is confident that the H will enjoy the action. It is used, for example, in well-wishing (*Have a lovely day/Enjoy your week-end/Take care*) or when a host invites their guest to enter and have a seat, or encourages them to have some more food or drink (*Come in/Sit down/Have some more wine/Have another chocolate* etc.). In these situations the S can afford to put pressure on the H. Leech points out that this indicates the S's sincerity in having the H accept the offer or invitation (Leech 1983:109).

Interestingly, this is the main reason why Russians prefer imperatives when inviting. The more pressing the invitation, the more sincere it sounds (*U menia den' rozhdeniya v voskresenye. Ty obiazatel'no dolzhna pridty – I am having a birthday party next Sunday. You must come by all means*). Being so demanding the Speaker demonstrates his great desire to see the Hearer, they can offer options related to the time of the encounter but not to the fact of coming: *Ne mozhesh pridty v 7, prihodi pozhe. Budem tebia zhdat' (If you can't come at 7, come later. We'll be waiting for you)* or *Prihodi, kogda smozhesh (Come when you can)* etc. When encouraging a guest to have some more food or drink, Russian communicators are also more pressing, perhaps because hospitality and modesty are among the most important cultural values.

Thus the notion of imposition is also culture-specific and it does not always mean impoliteness. Whether the Speaker is willing to disguise or soften the imposition by offering a semantic (formal) option or whether they find it unnecessary or even inappropriate, depends on the culture.

Summary

When we participate in intercultural communication we should bear in mind that directness and imposition do not necessarily imply impoliteness. Russian culture is a case in point. In many contexts direct utterances are socially appropriate and preferable. Besides indirectness, the Russian language has a wide repertoire of linguistic means to soften imposition. The speakers' choice of directness instead of indirectness can be triggered by the asymmetry of socio-cultural relationships (Social and Power Distance), differences in cultural values reflected in the way people use the language in their interaction, and by differences in politeness systems. Although politeness is a universal category, it is a culture-specific phenomenon and has different means of expression.

The comparative analysis conducted on the basis of ethnographic observations, questionnaires, and interviews shows that following the principle of non-interference, English speakers tend to avoid direct utterances in all SAs with the pragmatic meaning 'I want you to do that'. Brought up in an individualist culture, English respondents value privacy, individual autonomy, and their cultural norms require a more distant system of behaviour. SAs that can be generalised by the formula 'I want you to do that' are considered face-threatening. Elaborate negative strategies and preference for conventional indirectness are typical not only on formal occasions, but also prevail in everyday encounters. English communicators tend to use indirect formulas with semantic options in all SAs, including directives, where no functional options are available.

In a collectivist Russian culture it is uncommon to be as vigilant in guarding one's personal space as one is in English. People are more available to each other, which implies reduced social distance. Therefore, negative strategies are less typical of Russian communication; treating the Hearer in a direct way is sociably acceptable and in some situations is a preferred strategy. In Russian, when the S wants the H to do something, they express this wish more directly. The general closeness of interpersonal relations allows that. It is due to the collectivist nature of cultural values that the structure of such speech acts as Invitation, Advice, and Offer can hardly be considered as imposition, rather as positive acts of involvement. Request is also a less face-threatening act in Russian than in English.

Another reason for the observed cultural differences is Power distance which is greater in Russian communication. In asymmetrical situations (parents – children, teachers – pupils etc.) this allows speakers having power and rights to be more direct. In English communication superiors treat subordinates as equals, thus emphasising the cultural value of equality.

Communicative strategies dictate the choice of language means. Imperative utterances which are widely used in Russian communication do not indicate speakers' impoliteness. Polite usage in Russian does not exclude direct imperatives while in English it does.

Conclusion

Culture-specific politeness strategies and linguistic means used for their realisation shape distinctive communicative styles. My data challenge the assertion that English communicative style is direct (Gudykunst & Ting-Toomey 1990). I suggest that in interpersonal communication English style is indirect and implicit compared with Russian. It can be called Hearer-oriented and form-oriented. English speakers put the main emphasis on the form of the utterance and on softening the imposition, while Russian interlocutors are more concerned with the meaning than with the form. They express their intention in a more direct way. The Russian style of interpersonal communication is more direct and explicit compared with English and can be called message-oriented.

These differences confirm that British communicative culture tends to be avoidance-based and distance oriented. They also prove that Russian culture is involvement-based and solidarity oriented. These findings will contribute to the wider Language Pedagogy research strand within English Profile which will focus – amongst other topics – on the acquisition and development of intercultural competence of L2 English learners.

References and further reading

Aijmer, K (1996) *Conversational Routines in English: Convention and Creativity*, London, New York: Longman.

Blum-Kulka, S (1992) The metapragmatics of politeness in Israeli society, in Watts, R J, Ide, S and Enlich, K (Eds), 255–280.

Blum-Kulka, S, House, J and Kasper, G (Eds)(1989) *Cross-cultural pragmatics: Requests and apologies*, Norwood, NJ: Ablex.

Brown, P and Levinson, S D (1987) *Politeness: Some universals in language usage*, Cambridge: Cambridge University Press.

Gudykunst, W and Ting-Toomey, S (1990) *Culture and Interpersonal Communication*, California: Sage Publications.

Hickey, L and Stewart, M (Eds)(2005) *Politeness in Europe*, Clevedon, UK: Multilingual Matters Ltd.

Hofstede, G H (1980) *Culture's Consequences: International Differences in Work-Related Values*, Beverly Hills CA: Sage Publications.

— (1991) *Cultures and Organizations: Software of the mind*, London: McGraw-Hill Book Company (UK) Limited.

Kasper, G (1990) Linguistic Politeness: Current Research Issues, *Journal of Pragmatics* 14, 193–218.

Larina, T V (2003) Kategoriya vezhlivosti v angliyskoy y russkoy kommunikativnyh kul'turah, Moskva: RUDN.

— (2004) Koncept vezhlivost' v kommunikativnom soznaniyi russkih y anglichan, in Szypielewicz, L (Ed.) *Człoviek Świadomość Komunikacja Internet*, Warszawa, 284–293.

— (2006) *Directness vs. Indirectness in Russian and English Communicative Cultures*. General and Theoretical Papers, Essen: LAUD 2006. Paper No. 647, Universitat Duisburg-Essen.

Leech, G N (1983) *Principles of Pragmatics*, London: Longman.

Leech, G and Svartvik, J (1994)(2 ed) *Communicative Grammar of English*, London, New York: Longman.

Paxman, J (1999) *The English: A Portrait of a People*, London: Penguin Group.

Pizziconi, B (2003) Re-examining politeness, face and the Japanese language, *Journal of Pragmatics*, 35, 1471–1506.

Scollon, R and Scollon, S (2001)(2 ed) *Intercultural Communication: A Discourse Approach*, Oxford: Blackwell Publishers Ltd.

Sifianou, M (1992) *Politeness Phenomena in England and Greece*, Oxford: Oxford University Press.

Stewart, M (2005) Politeness in Britain:'It's Only a Suggestion', in Hickey, L and Stewart, M (Eds), 116–129.

Thomas, J (1983) Cross-cultural pragmatic failure, *Applied Linguistics*, 4, 91–112.

— (1984) Cross-Cultural Discourse as 'Unequal Encounter': Towards Pragmatic Analysis, *Applied Linguistics* 5/3, 226–235.

— (1995) *Meaning in Interaction: an Introduction to Pragmatics*, London, New York: Longman.

Tsui, A B M (1994) *English Conversation*, Oxford: Oxford University Press.

Watts, R J (1992) Linguistic politeness and politic verbal behaviour: reconsidering claims for universality, in Watts, R J, Ide, S and Enlich, K (Eds), 43–69.

— (2003) *Politeness*, Cambridge: Cambridge University Press.

Watts, R J, Ide, S and Enlich, K (Eds) (1992) *Politeness in Language: studies in its History, Theory and Practice*, Berlin, New York: Mouton de Gruyter.

Wierzbicka, A (1985) Different Cultures, different languages, different speech acts, *Journal of Pragmatics*, 9, 145–161.

— (1991) *Cross-cultural Pragmatics: The Semantics of Human Interaction*, Berlin: Mouton de Gruyter.

— (1997) *Understanding cultures through their key words: English, Russian, Polish and Japanese*, New York: Oxford University Press.

ALTE 2008 conference report

Many Cambridge Assessment staff were involved in the multi-lingual ALTE 3rd International Conference 2008 which we hosted in Cambridge on 10–12 April. The conference theme, *The Social and Educational Impact of Language Assessment*, bridged the world of language assessment and educational, social, cultural and economic environments and contexts. Cambridge ESOL staff and consultants led many sessions, including papers, posters, keynote sessions and a plenary by Dr Lynda Taylor (Cambridge ESOL consultant) on *Setting language standards for teaching and assessment: a matter of principle, politics, or prejudice?*

The opening plenary: *Using Learner Language from Corpora to Profile Levels of Proficiency (CEFR) – Insights from the English Profile Project*, was given by Professor John Hawkins (University of Cambridge & University of California, Davis) with Dr Paula Buttery (University of Cambridge) and described the findings made by English Profile researchers to date (see Hendriks' article on page 7). Professor Micheline Chalhoub-Deville (University of North Carolina, Greensboro) in her plenary spoke about *Standards-based assessment in the USA: Social and Educational Impact* whilst Professor Tim McNamara (The University of Melbourne) spoke about *Recognizing the Other: Language assessment, immigration and citizenship*, a topic of current concern in language testing and assessment. Dr Brian North (Eurocentres, Switzerland) discussed *The educational and social impact of the CEFR in Europe and beyond: a preliminary overview* whilst in the final plenary Professor James Purpura (Teachers College, Columbia University, New York) discussed *The Impact of Language Assessment on the Individual*. Videos of the plenaries and other selected sessions are available on the ALTE website: www.ALTE.org

Cambridge ESOL staff (from the Research and Validation and other Groups) contributed a number of presentations and took part in a keynote session and a forum on *Language Testing, Migration and Social Inclusion* (see the online Programme for full details). A selection of the contributions made by Cambridge ESOL staff are summarised below; these show the range of language assessment products, teaching awards and other activities that we are involved in.

Using corpora for language assessment: trends and prospects

Fiona Barker considered the recent development of the use of corpora for language assessment and the related impacts on test providers, takers and users. Computerised collections of written and spoken texts (corpora) have existed for around fifty years and have created many opportunities for research and application within diverse fields of language study including theoretical and applied linguistics. Within the language testing community, corpora have been acknowledged as having the potential for informing language assessment since the 1990s and throughout the last decade significant progress has been made in building and using corpora for language assessment.

This paper provided an overview of key native-user corpora of English (e.g. BNC, COBUILD) and their application in related fields such as language teaching and lexicography, before describing the development of corpora for language assessment, notably collections of learner writing and speech (e.g. Cambridge Learner Corpus) and domain-specific corpora, including collections of academic speech. Major contributions to the field were noted, along with the challenges faced by language testers when building and seeking new ways in which to use corpus resources. Fiona considered what Cambridge ESOL is currently doing with corpora, including work underway within English Profile as reported in this issue. The paper concluded with a number of future directions for using corpora in language testing which aim to build on what has already been achieved in this area.

The ALTE Validation Unit and the diffusion of good testing practice

Michael Corrigan presented on the work of the ALTE Validation Unit which involves four language test providers (CAPLE of the University of Lisbon, CVCL of the Università per Stranieri di Perugia, the Goethe-Institut and the Language Centre, National University of Ireland, Maynooth) who seek to develop their quality assurance capacity with psychometric analysis, and Cambridge ESOL as host organisation. Although three of the four organisations have had rigorous quality assurance procedures in place for many years, the integration of quantitative techniques into their procedures is something new and challenging. Good testing practice in the title refers to the inclusion of quantitative analysis in test production and administration.

Qualitative analysis can be used in many aspects of test production and administration. It is therefore important to ensure that workable and understandable procedures are developed so that those conducting test production and administration processes have no difficulty in making use of the statistical analysis done by others. An example of the development of such procedures connected with the use of an item bank was given. These procedures needed to take account of the technology used for the procedure, the staff operating the procedure, the requirements of the test and the way in which different information would be used in decisions in the procedures.

Another important aim of this work is to develop the capability of the staff of the organisation through quantitative feedback on the items produced by the system. Scalar items (where a response which is not totally right can receive a partial credit – the available score might therefore be 0 – 1 – 2) were discussed as an example. Scores obtained on scalar items are expected to be sequentially more difficult for a candidate to obtain, so weak candidates on the whole get 0, strong candidates get 2 and those in-between get 1. If scores are not sequentially ordered in this way, or one of the categories is under-used, this can be discovered by statistical analysis. Feedback on items which are not performing as well as expected can allow those involved in item development to investigate and edit items as well as equipping them to avoid such problems in the future.

Language proficiency and testing for migration purposes: What are the practical implications?

Sacha DeVelle presented a paper on language testing for migration purposes. The impact of large-scale immigration movements in recent years has highlighted the need to assess language ability for migration purposes. The role of fairness is of utmost importance when evaluating English ability, given that quality and fairness are paramount features in high-stakes English proficiency tests.

Recent sweeping changes to immigration procedures in the United Kingdom, Australia and Canada have raised a number of questions about the feasibility of an exemption list for native English speakers and testing language ability for L2 English speakers. These questions were discussed within the context of 2007 immigration trends for the IELTS test. The broader implications for using IELTS as a “fit for purpose” test for migration purposes were also addressed in light of test quality and fairness. Finally, the presentation highlighted the important need for mediation between policy makers and language testing experts when deciding on what tests fit who, and the corresponding testing implications.

On related topics, Szilvia Papp presented a paper entitled *The language requirements of the UK test for citizenship: Is it a valid test?* and Nick Saville took part in the *Language Testing, Migration and Social Inclusion Forum* which was moderated by Lynda Taylor.

Assessment for teaching: Cambridge ESOL's CLIL exam

TKT: Content and Language Integrated Learning (CLIL) is a new test for teachers designed and developed by Cambridge ESOL as an additional module to TKT, which will be released in October 2008. CLIL candidates are expected to be practising teachers (both native and NNS) who are engaged in the teaching of subject content through the medium of English. They may be working in bilingual contexts, teaching students for the International Baccalaureate, or working in English-as-Additional-Language contexts in the UK.

Mick Ashton and Evelina Galaczi gave a presentation which gave an overview of the development stage of the CLIL exam and focused on the CLIL construct, its operationalisation in the context of the exam, and the challenges of addressing universal aspects of CLIL practice in the test tasks while avoiding issues relating to subject specificity.

The presenters further reported findings (both quantitative and qualitative) from a trial which took place in December 2007–February 2008 in 11 locations around the world prior to the exam's launch, and involved 259 trial participants. The sample of trial participants was chosen so that it resembled key background characteristics of the anticipated CLIL candidature. The statistical analysis of the test items and the qualitative feedback from the candidates and exam administrators indicated that the trial was successful, with the majority of items performing well, and the trial participants showing an overall positive attitude to the test.

Implementation of an Italian item bank for large-scale test construction

Danilo Rini (CVCL, Università per Stranieri di Perugia) and Michael Corrigan (Cambridge ESOL) presented on the implementation of an item bank which is part of CVCL's continued programme of improving their quality assurance provision. CVCL provides tests of Italian as a Foreign Language and the ALTE Validation Unit is collaborating by providing psychometric analysis and advice in this item banking project. The use of item banks for standardised assessment is becoming more common but the implementation of such a resource is not an easy task. The presentation described the steps taken in respect of the CVCL bank:

i) the development of the pretesting programme to populate the bank and the design of the item bank itself,
ii) the plans for future use of the bank and quality assurance measures to be taken.

The presentation also reported on the success of the most recent pretesting sessions, with more than 70 per cent of items being banked as a result.

Item banks are of great value to organisations providing many forms of the same test. An item bank can be populated with many items of known characteristics, such as difficulty, discrimination and area of the specifications tested. New forms of a test can then be constructed by searching the bank for appropriate items. Test construction using items of known characteristics means that it is far easier to ensure that a new test form is equivalent to an old one.

It was decided that pretesting for this bank would be done by asking existing CVCL test centres to arrange for students of Italian as a Foreign Language to sit these tests. Most of these students are likely to be candidates of a CVCL test in the near future so are likely to have been studying relevant material, be familiar with the test and be motivated to do well. To encourage their participation, the structure and graphical layout of the papers were made as similar as possible to live test papers and participants were promised some feedback on their performance. As it was felt important to make the pretest authentic, anchor items were included in the body of the test, rather than as a separate, additional test. Anchor items are items of known difficulty and their function is to calibrate the new items so that the difficulty of each item in the bank is relative to that of all the rest, whichever pretest it was included in.

Logistical issues involved finding centres willing to do pretesting which have the resources to administer the tests as if they were live tests. Once a nucleus of such centres was established, a pretesting programme was agreed, whereby participants for regular pretests would be found by the centres, they would administer the tests and then return the papers to CVCL. The development of items, the construction, printing and dispatch of papers, the analysis of results and the return of feedback on examinee performance were then all timetabled to fit in with the pretest dates.

The development of the item bank comprised both the development of the user interface and the pattern of workflow that the bank would serve. The principal stakeholders for these developments were CVCL staff who construct items and test papers. The item bank user interface therefore had to be intuitive and adaptable. The pattern of workflow included a number of editing stages and pre- and post-pretesting stages. For this reason the item bank is, in fact, really a number of banks, each storing a set of items at the same stage in the item and test paper construction process.

When fully populated, the item bank will be used for the construction of live tests. It will then be possible for users to select items by difficulty and other characteristics. They will ensure that, in the final test, the range of difficulties and the mean difficulty of all items is appropriate, as is the coverage of the specified content and item types. After live tests constructed from the bank are used, certain quality checks can be made to ensure that the system is producing appropriate tests and adjustments can be made if necessary. One such check involves a comparison of the difficulty of an item found in a pretest with that found in the live test; whole live tests can also be compared with past live tests in a similar fashion and equating adjustments made.

Continuing the ALTE 2008 theme

The next two issues of *Research Notes* will consider the impact of language testing and assessment in a wide range of contexts, continuing the ALTE 2008 conference theme. In addition, a volume to appear in the *Studies in Language Testing* series will feature selected papers and plenaries from this event. This is currently being collated and will be published in 2010; further details will be announced in a future issue of *Research Notes*.