Table of Contents

From the Editors

ADAPTING TO THE CEFR IN ELT: EVALUATION AND ASSESSMENT OF LANGUAGE COMPETENCE

This issue includes articles concerning different approaches to tackle the impact of the Common European Framework of Reference (CEFR) on English Language Teaching (ELT) and assessment. Special attention is paid to the assessment of language competence in a higher education context. The use and impact of evaluation and assessment results is of paramount importance when these results are a prerequisite, for instance, to start postgraduate studies or to be able to ask for a job where a specific level is required.

In this context, validity and reliability of standard tests and their alignment to the CEFR levels and placement tests in higher education institutions are key concepts within the European framework. Due to the fact that the English language has an important impact as a tool for communication worldwide, most research in defining the different CEFR levels and evaluating their description (and how this description is understood and applied) refer to the English language. But the application of such levels to other European languages is also relevant particularly if we want to establish a clear language reference that may be understood and shared among languages. Thus, we may talk about the comparison of educational achievements across countries and in different parts of the same country as well as among languages.

Another important issue within the CEFR context is the modes in which assessment and tests are carried out, depending on the institutional media. Multimodal means of assessment, online assessment or in situ assessment may also determine the way levels are measured and accessibility for test takers.

In the first article presented in this volume **Pikabea, Lukas and Figueras** survey the different models that have been used in order to certify Basque language levels

according to the European framework, and the number of itineraries a test taker may choose to be able to obtain a certified level. In order to do so, they combine a number of tools in their study including technical qualitative analysis, interviews with people responsible for the management of test administration within an institution, and questionnaire design to gather data, among other procedures. The importance of their exhaustive research is to establish a framework for the validation of the existing accreditation systems for Basque that accounts for an analysis of whether the specifications for fluency in the language that allowed test alignment with the CEFR were properly followed in the different accreditation institutions. Finally, they put forth a proposal for adaptations where necessary and monitoring of such adaptations.

Papageorgiou deals with the issue of how different assessment tools around the world are aligned with the CEFR levels, that is, (1) how assessments are brought into alignment with other existing standards and frameworks and (2) how assessment results are interpreted when compared to another assessment frame. Papageorgiou also identifies those areas that still need refinement in relation to the CERF levels, such as why and how these levels are selected in policy making or the fact that two tests are assessed as belonging to a same CERF level do not necessarily have the same content or level of test difficulty. The relevance of developing adequate alignment tools and theories cannot be emphasized strongly enough.

In his article, *Measuring the impact of CLIL on language skills: a CEFR-based approach for Higher Education,* **Jiménez-Muñoz** discusses the difficulties faced by CLIL instructors when applying CEFR criteria. Issues like the lack of English language level on the part of the students that get enrolled in a CLIL classroom at university level; the need of developing specific skills to teach through a second or foreign language; the lack of financial means to accomplish all instructional purposes, both language and content ones; and the way instructors overcome these problems, are evaluated. The relevance of this study lies in the analysis of results based on tools that aim at evaluating student progress after a CLIL experience in a way that fits the university time schedules, adjusting to university terms and their timing.

In the fourth article of the volume, **Beinhoff** deals with the relevance of developing a European framework for language evaluation focusing on the representation of speech development -particularly perceptive skills- in the CEFR level descriptions. According to this author, vagueness is a feature of the speech-related CEFR sections and related level descriptions so that assumptions made therein -the supposed linear progression between levels amongst them- have not been sufficiently tested yet. By presenting an exploratory study on speech perception in language learners this paper investigates what kind of influence listeners' levels of proficiency in the second language and their L1 backgrounds have when perceiving intelligibility. The results break new ground by identifying that proficiency levels and L1 background do (although not always) influence intelligibility and partially confirm the idea of a linear progression as proposed in the CEFR.

The volume also includes a final article entitled "Motivation and constraints of illocution in the lexical constructional model: the case of the *Aux NP* construction", in which its author, **Del Campo**, addresses the motivation and constraints of illocutionary meaning production. By analysing the realization procedures of the *Aux NP* construction in relation to their potential to exploit the semantic base of requestive acts, the author explores how our knowledge of illocution is understood in terms of high-level situational models which are activated to produce speech act meaning, and the way such operations motivate the conventionalized value of linguistic expressions. As a result, always within the framework of the Lexical Constructional Model (LCM), a comprehensive understanding of the constructional nature of illocutionary meaning on the basis of naturally occurring data is provided.

Finally, **Annemieke Meijer** reviews the volume *English-Medium Instruction at Universities*, authored by Doiz, Lasagabaster and Sierra. The reviewer presents this five-part volume as "an interesting and timely addition to the growing literature on the use of English as the language of instruction at universities in non-native contexts", in which a varied picture of current issues and practices is provided by means of contributions from eighteen authors from many and diverse countries. The selection, even though a bit arbitrary somehow, is highly interesting not only because of their

diversity but also because of the unexpected connections established, all contributions adding thus to the overall picture.

Mª Carmen Campoy-Cubillo
Nuria Edo-Marzá
Editors
Universitat Jaume I, Spain

Methodology for the process of validating qualifications in the Basque language and their adaptation to the Common European Framework of Reference for Languages (CEFR)

Iñaki Pikabea Torrano
i.pikabea@ehu.es
José Francisco Lukas Mujika
jf.lukas@ehu.es
Universidad del País Vasco, Spain
Neus Figueras Casanovas
nfiguera@xtec.cat
Generalitat de Catalunya, Spain

ABSTRACT

In the Basque Country there are currently more than 140,000 people with accredited qualifications that certify their knowledge of Euskara (Basque Language). These qualifications have been issued by a number of official bodies dependent on the Basque Government, but had not been validated against each other until recently.

In order to validate these qualifications and adapt them to the Common European Framework of Reference for Languages, a group of experts were commissioned by the Basque Government to undertake a study conducive to this end. This paper outlines the methodology and conclusions of that study.

Keywords*: methodology, evaluation**, qualifications**, Euskara (the Basque language), the European Framework of Reference, validations.*

I. INTRODUCTION

In the Basque Country (a region comprising part of the north of Spain and the south-west of France) there are currently over 140,000 people who possess one or several of the more than 20 different qualifications that accredit knowledge of the Basque language (Euskara) at different levels. These qualifications are required to be able to work as government administration personnel (teaching staff, healthcare employees, etc.).

Since 1998 there have been demands from the general public for a single system and a common approach to accrediting competence in the Basque language. In 2006, those responsible for such affairs at the Vice-Ministry for Language Policy of the Basque

Government contacted the authors of this paper and stressed the need to have an accreditation system for knowledge of the Basque language which was capable of agglutinating the different models of such knowledge and/or the routes that can be taken to certify it, while employing a unified approach based on the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001, 2003, 2004, 2009).

It was also understood that the work undertaken in European bodies such as ALTE (Association of Language Testers in Europe) or EALTA (European Association for Language Testing and Assessment) should be incorporated, as should the opinions of recognized experts from the field of language assessment (Alderson, 2006; Cizek & Bunch, 2007; Downing & Haladyna, 2006; Verdía, Conde, De Samblanc & Cassany, 2002; Zieky & Livingston, 2008). In a first phase, the commission of experts drew up a report (Figueras, Pikabea & Lukas, 2008) describing the work done and its results, and also including recommendations for improvement of the different tests. An outcome of this process was the approval by the Basque Government of the framework for the validation of qualifications and accredited certificates of knowledge in Basque and their adaptation to the CEFR. This framework was incorporated into a number of legislative decrees, culminating in the decree[1] which defined the validations and drew parallels between the various qualifications and certificates of knowledge of the Basque language (Pikabea, Lukas, Figueras, 2009).

Through this system of accreditations and validations practically all the qualifications in Euskara (a total of 22, with some having been granted since 1975) have become officially recognized within the CEFR. In this way, the validations between the different qualifications and certificates are recognized, as is their relationship with the four levels of the European Framework. The equivalences of the qualifications and certificates from IVAP[2], HABE, Osakidetza, Ertzaintza, the Basque Government's Education, Universities and Research Department, the Royal Academy of the Basque Language (*Euskaltzaindia*), the Department of Education of the Government of Navarre and the *Escuelas Oficiales de Idiomas* (Government-run Language Schools) were established with the B-1, B-2, C-1 and C-2 levels of the European Framework.

Three years later, the Basque Government once again contacted the authors of this paper and asked for a second report. The different systems of evaluation had been updated by the various bodies involved after the publication of the above mentioned decree, and it was necessary to assess to what extent the changes responded to the recommendations of the 2008 Report. This second report was finalized in 2011 (Figueras, Pikabea, Lukas, 2011). In this article the research procedures followed in the first and second reports are briefly outlined, together with a series of suggestions put forward on the basis of the different analyses and observations. The authors have adopted a general approach, as detailed results and data for the different qualifications and organizations need to remain confidential.

II. OBJECTIVES AND LIMITATIONS

The objectives proposed were the following:

1. Analysis of current systems for the accreditation of knowledge of the Basque language amongst adults.

2. Establishment of a common set of specifications for fluency in the language that would enable the objective comparison of levels, tests and the degree of benefit drawn from the various systems for certification amongst adults.

3. Proposal for adaptations considered necessary, taking the CEFR criteria and levels as a basis.

4. If possible, the design of a table of equivalences for the various certifications of knowledge of Basque amongst adults.

5. Monitoring of adaptations, 4 years on.

III. METHODOLOGY

III.1. First phase (2006-2008)

Based on the Council of Europe's recommendations for harmonizing exams with the CEFR levels as described in the Manual for Relating Examinations to the CEFR

(familiarization, specification, standardization of judgments, empirical validation), the steps and calendar to be followed were set out and a protocol of actions drawn up. These were then forwarded to all the bodies participating in the project, with the request to provide the researchers – within a period of one month – with:

1. a certain number of test forms (corresponding to the last three years)

2. the completed questionnaires designed to this end

3. a certain number of candidates' written scripts and oral performances, following set requirements

That is, on the one hand, protocols or questionnaires were drawn up for the analysis of the examinations and, on the other, expert judges were selected and used to assess the test tasks and the candidates' scripts and performances in relation to the CEFR levels.

The following procedures were followed:

Questionnaire analysis. The Questionnaire was developed based on the Forms included in the 2003 Council of Europe Manual and included questions on: 1) general description of the exam; 2) drafting and administration of the tests; 3) correction and scoring; 4) analysis of data and a later review; 5) qualitative-technical analysis of the tests; and 6) estimated global level of each examination in CEFR terms.

System of judges. Judges were selected from amongst expert teachers, teacher trainers and item writers in the Basque country, and the procedures followed in their training and standardization were also those recommended in the 2003 Council of Europe Manual, including: 1) familiarization of the judges (a total of 30) with the CEFR descriptors; 2) judges' analysis and evaluation of the samples of oral and written production of the examinees (a total of 39 written comprehension tasks, 16 oral comprehension tasks, 36 pieces of written production from the students and another 34 items of oral production from the students).

The data thus gathered were examined thoroughly and, as a conclusion, a report was drawn up in 2008. Table 1 reflects the possible comparability of qualifications, as well as the adaptations needed to conform to the CEFR levels.

To this end, the following criteria were used:

- Analysis of test data provided by their devisers (Lukas, 1998)

- Qualitative-technical analysis of tests

- Analysis of the tests and CEFR level assignments by the judges

When required, a scale of grey tones was used to indicate, beside each qualification, the areas in which the qualification in question had to be adjusted in order to have a comprehensive, integrated equivalence with the CEFR[3] levels.

Abbreviations:

> **wc**: written comprehension   **we:** written expression
>
> **oc:** oral comprehension   **oe**: oral expression

Interpretation and positions of colours:
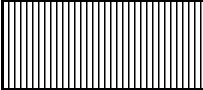
Same column	Equivalence in general terms
	Absence of specific test
	Does not reach required level
	Exceeds the required level

Table 1 shows a graphical summary of the conclusions drawn from research in the 2008 report. This summary was considered extremely useful by examination bodies, as it showed where adjustments were necessary, and extremely helpful by government administration bodies, as it helped decision-making.

Numbers 1-5 in the Organization column refer to the five exam-providing institutions in the Basque country, their names not being given for the sake of confidentiality.

Table 1. Summary of equivalences to the CEFR (2008 Report).

CEFR LEVELS										
ORGANISATION	**A1**	**A2**	**B1**		**B2**		**C1**		**C2**	
1	wc	wc	**Certificate A**	wc	**Cer B**	wc	**Cer C**	wc	**Cer D**	wc
	we	we		we		we		we		we
	oc	oc		oc		oc		oc		oc
	oe	oe		oe		oe		oe		oe
2	wc	wc	**Certificate A**	wc	**Cer B**	wc	**Cer C**	wc	**Cer D**	wc
	we	we		we		we		we		we
	oc	oc		oc		oc		oc		oc
	oe	oe		oe		oe		oe		oe
3	wc	wc		wc	**Cer A**	wc	**Cer B**	wc		wc
	we	we		we		we		we		we
	oc	oc		oc		oc		oc		oc
	oe	oe		oe		oe		oe		oe
4	wc	wc	**Certificate A**	wc	**Cer B**	wc	**Cer C**	wc		wc
	we	we		we		we		we		we
	oc	oc		oc		oc		oc		oc
	oe	oe		oe		oe		oe		oe
5	wc	wc	**Certificate A**	wc	**Cer B**	wc	**Cer C**	wc		wc
	we	we		we		we		we		we
	oc	oc		oc		oc		oc		oc
	oe	oe		oe		oe		oe		oe

III.2. Second phase (2011)

A number of years after these proposals for validated qualifications were first put forward (2008), the Department of Language Policy of the Basque Government commissioned a review and the monitoring of the changes and improvements made by each body responsible for setting examinations in the Basque language.

In the process monitored for this research, a protocol of activities was drawn up again and then remitted to all those bodies participating in the project, the certifying bodies or organizations being requested to ensure that such examinations for the validation of qualifications should:

- provide the same number of tests (corresponding to recent examination announcements) for each of the levels examined

- complete the questionnaires designed with this aim

- designate people with a thorough knowledge of the tests to be interviewed by the researchers

Technical qualitative analysis of the tests

Apart from the EALTA Recommendations, other points taken into account for the technical analysis of the tests were, on the one hand, the usual criteria for item development and, on the other, the recommendations made in the 2008 report.

Monitoring of suggested recommendations

The recommendations made in the Report of the Commission of Experts on the Validation of Basque Language Qualifications and their Adaptation to CEFR (2008) were reviewed to determine whether or not each recommendation had been taken into account.

Questionnaire analysis

The questionnaires sent to examination providers were aimed at collecting information about five different aspects, namely:

1. Drawing up the tests

2. Medium-term adjustments

3. Completing the tests

4. Technical analysis of the tests

5. Commentaries

For each affirmative response some evidence was required to accredit its veracity. And each negative response also required an explanation.

Interviews with representatives of the board

The information gathered and analyzed, both through questionnaires constructed *ad hoc* as well as through the technical qualitative analysis of the tests, was completed with interviews with the experts and/or those responsible for the various examinations of the

institutions involved in the research. In all, 10 people from five institutions were interviewed.

Figure 1 shows the different sources of information employed in order to reach the conclusions drawn in this report.
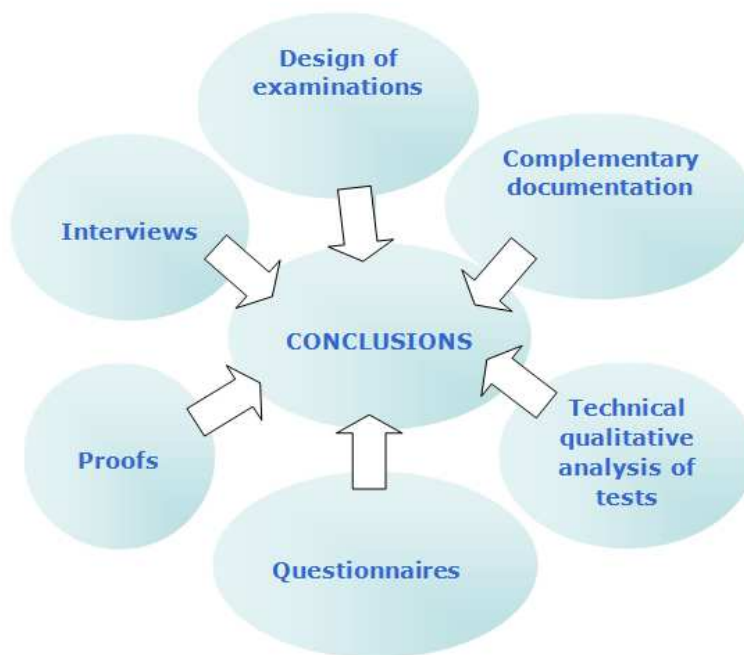

Figure 1. Triangulation of the different sources of information.

IV. RESULTS

The results and conclusions were included in two reports (2008 and 2011).

Team of evaluators

The teams of evaluators are very heterogeneous. Evaluation is considered to be a speciality that is sufficiently demanding and painstaking to have stable dedicated teams of evaluators in each institution.

Designs and protocols

A clear improvement and greater professionalization in test development and analysis are observed. In most cases this was already the starting point. A very high level was seen in the measurement tools that were analyzed and in the processes and documentation supporting them.

Comparison of requirements

The comparison of requirements for qualifications which, theoretically, accredit the same CEFR level was not reviewed. This would be a very important experimental study which would make it possible to identify possible imbalances between qualifications at the same level issued by different departments.

Validation of qualification at CEFR level

No institution claims to have experimental data that guarantee that its qualifications do, in fact, correspond to CEFR levels, whereas they all state that they follow the Framework descriptors strictly. However, no field research has been carried out to allow an informed comparison of evaluation processes for other languages in Spain or in other countries.

Qualifications at C2 level

An analysis of the two existing qualifications that currently accredit level C2 reveals a different conception of the evaluation process for this level. While one particular body may present a concept similar to that used at previous levels but at a greater level of difficulty, the other body may interpret the certification as the final part of an ongoing evaluation process, where the work done during the course (a prerequisite to be able to sit the exam) is taken into consideration. This provides candidates with two ways of being evaluated, with different requirements.

Unnecessary evaluation processes

It has been observed that in some cases candidates have to pass tests which, in reality, do not provide data of interest. This happens in certain evaluations where the candidate, before completing the examination that will provide him/her with an accreditation, is required to sit a pre-test, the results of which are only used to decide whether the candidate is ready to sit the examination or not.

Variety of examination formats

The designs of the evaluations that were analyzed are complete and their content fulfils the criteria for validity. However, it has been observed that, in some cases, the training

given to the students for the acquisition of Basque is conditioned by the type of exam. As a consequence, it becomes too restricted, as it is too dependent on the exam format. This is why some bodies propose a greater variety of tests in their examinations, thereby avoiding an over-repetitive format.

V. CONCLUSIONS AND IMPLICATIONS

The proposal for the validation of qualifications in Basque and for their adaptation to the CEFR levels also included several suggestions to improve the design and uses of the tests. The suggestions were addressed to the management or administrative bodies and referred to topics or lacunas needing short- or medium-term consideration with the goal of enhancing the accreditation processes. Amongst the most notable suggestions, the following are worth mentioning:

Open/restricted exam sessions

Some bodies have both open and restricted examination sessions, but there is no empirical evidence about the level of competence of the candidates, whether they are comparable or not. This is a task that will have to be addressed in the future.

Examples of classified oral/written productions

What the authors propose is the establishment of a classified collection of samples, to act as a reference for all accrediting organizations, given that this will reduce the risk of there being significant differences amongst the sample collections of different bodies. Thus, just one collection of reference samples will help to obviate there being different levels of requirements from one organization to another at any one CEFR level.

Inter-institutional Technical Commission

The report proposed the creation of an inter-institutional technical commission to debate and agree on those technical aspects that are of concern to the evaluation processes, with the goal of avoiding overlaps (exam calendars, etc.) or disparity of criteria.

Team of evaluators

Test development is a sophisticated operation. In many organizations teams of item writers have a short existence, and in others item development is commissioned to

outside contractors. In order to guarantee the necessary coordination and coherence, it is recommended that each institution take total control of the process, including the design of the tests, in order to foster higher quality.

Qualifications at C2 level

It is recommended that the two institutions certifying C2 and currently using totally different procedures reach an agreement as to which formula is the fairest and most appropriate.

Accreditation bodies

Concern arises from not knowing which organizations will, in the future, be officially allowed to carry out the evaluation and accreditation of candidates' knowledge and acquisition of the Basque language, as well as the possible consequences thereof.

Unnecessary evaluation processes

In section 4 of this paper we have mentioned that some institutions require the candidate to sit a pre-test prior to the final official test. The data available show that the results obtained in the oral and written comprehension sections of the pre-test that the candidates take in their study centers match those obtained in the final examination. In view of this evidence, we strongly recommend that the accrediting institutions focus on the assessment of productive skills (oral and written production), thus avoiding unnecessary duplication of costs incurred by the Administration.

Request for advice

While the authors were drawing up their reports, certain departments requested advice on evaluation topics that would help clear up doubts that might arise. At an administrative level, it would be useful to have a team of experts who could also respond to this type of internal requests.

Evaluation of oral comprehension

In general, the models for the evaluation of oral production are apparently similar across institutions: the examinee listens to an audio accompanied by items that have to be answered correctly, but what format of audio text should the candidate be provided with? Should it be a monologue, a conversation, a clip from a real interview, a sequence

from a TV programme, a standard language production… or what? Unification of criteria within each organization or body and, if possible, on an inter-institutional level is important.

Statistical analysis of the tests

Pilot testing needs to be an indispensable condition prior to the final application of a certificate test. It is equally important to analyze how the test items worked after their final application. Besides justifying the technical quality of the tests employed, these analyses will be used to create the bank of calibrated items to guarantee the quality of future examinations. Finally, the bodies that administer a preliminary test should analyze its predictive validity, as well as its reliability.

Establishment of standards or cut scores

It would also be useful to set down some procedure for establishing such scoring methods, although this may involve modifying current legislation in some cases.

Linguistic normalization and accreditation

The reports also raise the question of a possible clash between the goals of language normalization in the Basque Country and the system of accreditation. Possible contradictions would have to be detected and, in such cases, decisions deemed opportune would need to be taken.

Resources and criteria for disabled candidates

The drawing up of a general, single norm of compulsory compliance is recommended, and this would have to comply with and respond to the content of State legislation corresponding to that of the International Convention on the Rights of Persons with Disabilities.

Notes

[1] Decree 297/2010 in the Official Gazette of the Government of the Basque Country.

[2] IVAP: Basque Institute for Public Administration; HABE: Institute for the Acquisition of and Literacy in the Basque Language by Adults; Osakidetza: the Basque National Health Service, Ertzaintza: the Basque Police Force.

[3] This example serves as a rough guide, the names of the institutions and the qualifications awarded having been omitted. Not all institutions issue qualifications at all CEFR levels.

REFERENCES

Alderson, J.C. et al. 2006. Analyzing tests of reading and listening in relation to the CEF: the experience of the Dutch CEF construct project. *Language Assessment Quarterly.* 3(1) 3-30.

Cizek, G. and Bunch, M. 2007. *Standard Setting. A Guide to Establishing and Evaluating Performance Standards on Tests.* Sage Publications

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

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

Council of Europe. 2004. *Reference Supplement to the Preliminary Pilot Version of the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Strasbourg: Language

Council of Europe. 2009. *Manual for relating examinations to the CEFR.* Available at: http://www.coe.int/t/dg4/linguistic/manuel1_en.asp (last accessed 22-11-2011)

Downing, S.M and Haladyna, T.M. (eds.). 2006. *Handbook of test development.* London: Lawrence Erlbaum Associates.

Figueras, N., Pikabea, I. and Lukas; J.F. 2008. *Informe de la Comisión de Expertos sobre convalidación de titulaciones de euskara y adecuación al MCER.* Gobierno Vasco: Viceconsejería de Política Lingüística (unpublished).

Figueras, N., Pikabea, I. and Lukas, J.F. 2011. *Informe de seguimiento de la Comisión de Expertos sobre convalidación de titulaciones de euskara y adecuación al MCER.* Gobierno Vasco: Viceconsejería de Política Lingüística (unpublished).

Lukas, J.F. 1998. *Análisis de ítems y de tests con ITEMAN.* Leioa: Servicio de Publicaciones de la Universidad del País Vasco.

Pikabea, I., Lukas, J.F and Figueras, N. 2009. "Proceso de comparabilidad de titulaciones de euskara y adecuación al Marco Común Europeo de Referencia (MCER)". *Revista de Investigación Educativa.* 27(2) 469-486.

Verdía, E., Conde, G., De Samblanc and Cassany, D. 2002. "Marco común europeo de referencia y Portfolio de las lenguas.", *Mosaico*, 9, monográfico. Revista para la Promoción y Apoyo a la Enseñanza del Español.

Zieky, M. P. and Livingston, S. 2008. *Cut scores: A Manual For Setting Standards Of Performance On Educational And Occupational Tests.* ETS.

Issues in aligning assessments with the Common European Framework of Reference

Spiros Papageorgiou
spapageorgiou@ets.or
Educational Testing Service, MI, USA

ABSTRACT

One of the main aims of the Common European Framework of Reference is to help providers and users of assessments "describe the levels of proficiency required by existing standards, tests and examinations in order to facilitate comparisons between different systems of qualifications" (Council of Europe 2001: 21). Providers of language assessments both inside and outside Europe follow various methodologies to align their assessments with the CEFR levels, as several case studies show (Figueras and Noijons 2009; Martyniuk 2010). This paper discusses the use of the CEFR in the field of language assessment, focusing in particular on issues related to alignment. The paper presents the types of validity evidence that should be accumulated to support an alignment claim and concludes with directions for future research in order to further enhance our understanding of using the CEFR for the design of assessments and the interpretation of assessment results.

Keywords: *Alignment, cut scores, Common European Framework of Reference.*

I. INTRODUCTION

The publication of the Common European Framework of Reference (CEFR) in 2001 has been recognized as the "most significant recent event on the language education scene in Europe" (Alderson 2005b: 275). The main purpose of the CEFR is to provide a common basis for the elaboration of language syllabuses, examinations, and textbooks by describing in a comprehensive way what language learners have to learn to do in order to use a language effectively for communication (Council of Europe 2001: 1). The language proficiency levels and their language performance descriptors are central to the CEFR's descriptive scheme of language use, as noted by Little (2006: 169). They serve one of the main aims of the Council of Europe as described in Chapter 3 of the CEFR volume, that is: "to help partners to describe the levels of proficiency required by existing standards, tests and examinations in order to facilitate comparisons between

different systems of qualifications" (Council of Europe 2001: 21). Such comparability of language qualifications in Europe was difficult to achieve prior to the CEFR because of the plethora of diverse educational systems and traditions. Alderson (2007: 660) pointed out that "the six main levels of the CEFR have become a common currency in language education, and curricula, syllabuses, textbooks, teacher training courses, not only examinations, claim to be related to the CEFR".

Nowadays, providers of language assessments, both inside and outside Europe, follow various methodologies to align their assessments with the CEFR levels, as reported in several case studies in two edited volumes (Figueras and Noijons 2009; Martyniuk 2010). The most common approach to bring tests into alignment with the CEFR is the one recommended in the *Manual* published by the Council of Europe (2009). The approach consists of two main stages: content alignment and setting of cut scores. The main purpose of this paper is to discuss the use of the CEFR in the field of language assessment, with a particular focus on issues related to the alignment of assessments with the CEFR. Before discussing alignment issues, however, it is important to first consider the work that led to the development of the CEFR and its levels, which is presented in the next section.

II. THE DEVELOPMENT OF THE CEFR AND ITS LANGUAGE PROFICIENCY SCALES

The Council of Europe (not be confused with the European Union) is the continent's oldest political organization, founded in 1949. Its general aim is to foster common democratic principles among its 47 members. The Council of Europe has been active in the area of languages for more than forty years with two complementary bodies: the Language Policy Division in Strasbourg, France, and the European Centre for Modern languages in Graz, Austria.

In order to promote plurilingualism and pluriculturalism among European citizens, the Council of Europe published a number of documents in the 1970s that have been influential in second language teaching, such as the notional-functional syllabus by Wilkins, which describes what a learner communicates through language (Wilkins 1976), and three ascending levels describing language achievement: Waystage (Van Ek

and Trim 1991), Threshold (Van Ek and Trim 1998) and Vantage (Van Ek and Trim 2001). In 1991, at an intergovernmental symposium in Switzerland, the development of a common framework for learning, teaching and assessment was deemed desirable in order to:

- promote and facilitate cooperation among educational institutions in different countries;
- provide a sound basis for the mutual recognition of language qualifications;
- assist learners, teachers, course designers, examining bodies and educational administrators in situating and coordinating their efforts;

(Council of Europe 2001: 5)

The authoring group produced an initial version in 1996 and the final version of the CEFR was published after feedback and consultation in 2001, the European Year of Languages, in English and French. Since then, the CEFR volume has been freely available online on the Council of Europe website (www.coe.int/portfolio) in more than 30 languages. These include non-European languages, such as Arabic and Japanese, revealing the strong interest in the document world-wide, not only within Europe.

Although the CEFR contains a rich description of the language learning process, it is widely accepted that the CEFR language proficiency scales are the best known part of the 2001 volume (Little 2006). The proficiency scales of the CEFR have gained popularity because they offer a comprehensive description of the objectives that learners can expect to achieve at different levels of language proficiency. They describe language activities and competences at six main levels: A1 (the lowest) through A2, B1, B2, C1 to C2 (the highest). Borderline levels are further elaborated using a 'plus' between A2+ (between A2 and B1), B1+ (between B1 and B2) and B2+ (between B2 and C1). The scales comprise statements called 'descriptors', which are always phrased positively, as they are intended to motivate learners by describing what they can do when they use the language, rather than what they cannot do (Council of Europe 2001: 205). The performance descriptors of the CEFR are designed following an action-oriented approach: language users are seen as members of a society who have tasks to accomplish, including those that are not language-related (Council of Europe 2001: 9).

Because of the action-oriented emphasis, the descriptors are also frequently referred to as "can-do statements".

The scales and descriptors in the 2001 edition of the CEFR were primarily developed during a large research project in Switzerland (North 2000; North and Schneider 1998). The project applied a variety of qualitative and quantitative methods for the initial analysis and collection of more than 2000 language descriptors used in proficiency scales around the world, the consequent selection and refinement of 1000 of these descriptors and, finally, the placement of the descriptors at different proficiency levels that subsequently formed the CEFR levels (see also Appendix A and Appendix B in Council of Europe, 2001). A number of studies and research projects such as the DIALANG project (Alderson 2005a; Alderson and Huhta 2005) have shown that the descriptors can be consistently replicated in a range of contexts, thus offering validity evidence for the use of those descriptors across a variety of contexts.

Language assessment is specifically discussed in Chapter 9 of the CEFR, which serves as a useful introduction to important notions and principles in the field. Fundamental terms such as validity and reliability are explained, and different types of assessment are described (e.g., formative versus summative assessment; norm-reference testing versus criterion-referencing testing). The next section focuses on the process of aligning assessments with the CEFR, which has been the topic of many studies in the field of language assessment.

III. THE PROCESS OF ALIGNING ASSESSMENTS WITH THE CEFR

The CEFR has been extremely influential in the field of language assessment, as evidenced by the 2005 special issue of the Language Testing journal on language assessment in Europe (Alderson 2005b) and the various alignment studies in the two edited volumes mentioned above (Figueras and Noijons 2009; Martyniuk 2010). The demand for alignment of assessments with various external standards has increased not only in Europe, but worldwide, because of education reforms which push for accountability, including close monitoring of students' progress and use of standardized tests (Deville and Chalhoub-Deville 2011). In the United States, the No Child Left Behind Act and more recently the Common Core State Standards, an initiative

supported by most states in the United States to describe the skills and abilities expected by students at each grade level, have further raised the demand to bring assessments into alignment with frameworks and standards.

The *Manual* published by the Council of Europe (2009) offers a recommended set of procedures for aligning tests with the CEFR, which consists of two main stages: content alignment and setting of cut scores. For content alignment, the Specification chapter of the *Manual* suggests forms to be completed for each language skill. These forms contain several questions regarding the extent to which the content of an assessment covers communicative language activities, contexts, text types and other aspects of language ability described in the CEFR. Thus, the completed forms constitute a claim of content coverage in relation to the CEFR. The second stage involves the setting of minimum scores on the test that would indicate that a test-taker has demonstrated the performance expected at that CEFR level (Standardization Training and Benchmarking chapter and Standard Setting Procedures chapter). These minimum scores (cut scores) are established following a well-researched process in the educational measurement literature called "standard setting" (Cizek and Bunch 2007). During standard setting, a panel of expert judges (often called "panelists") is required, under the guidance of one or more meeting facilitators, to make judgments on which examination providers will base their final cut score decisions. Statistical information about test items and the distribution of scores might also be used to help panelists with their judgment task. A fairly common practice in standard setting meetings is that more than one round of judgments is implemented (Hambleton 2001; Plake 2008). Between rounds, the panel discusses individual judgments, receives the statistical information about items and scores and repeats the judgments. Even though the panel will offer a recommended cut score, the decision on whether to accept this score rests with the examination provider. In this sense, standard setting is in fact a procedure for recommending cut scores for implementation by the provider of the test (Cizek and Bunch 2007; Tannenbaum and Katz 2013). Procedures for validating the recommended cut scores are also presented in the *Manual*.

IV. EXPLORING THE QUALITY OF ALIGNMENT WITH THE CEFR

In the field of educational measurement, alignment typically refers to the extent to which the content of an assessment covers the skills and abilities described in an external framework and standards. Such exploration of content coverage is an integral part of the Specification chapter in the *Manual*. Webb (2007) proposed a process to evaluate the alignment of assessments with content standards based on four criteria:

- Categorical Occurrence, which addresses the issue of whether a test covers the content discussed in the standard.

- Depth-of-Knowledge (DOK) Consistency, which addresses the extent to which an assessment is as cognitively challenging for test-takers as one would expect, given the description of what students are expected to know and be able to do in the standard.

- Range of Knowledge Correspondence, which deals with the extent to which the breadth of knowledge in the assessment corresponds to the expected one in the standard.

- Balance of Representation, which addresses the extent to which specific knowledge is given more or less emphasis in the assessment compared to the standard.

Although the *Manual* (Council of Europe 2009) does not provide criteria similar to the ones by Webb (Webb 2007) for evaluating alignment of test content with the CEFR, it could be argued that the various forms that should be completed during the Specification stage do cover the above criteria to some extent.

Content alignment as described in both the *Manual* (Council of Europe 2009) and Webb (2007) requires the use of human judgment. This use of human judgment is a central issue in the process of setting cut scores (Zieky and Perie 2006: 7). As Kantarcioglu and Papageorgiou (2011) noted, judgments are not only involved during the planning of a standard setting meeting, for example, when a standard setting method is chosen, but in every step of the decision-making process of setting cut scores, that is: deciding on the number of levels with which to classify test-takers, selecting and training panelists, and scheduling the activities in the standard setting meeting. Despite this reliance on

judgments, the standard setting meeting and its outcomes can be evaluated based on several criteria typically grouped into three categories (Council of Europe 2001; Hambleton and Pitoniak 2006; Kane 1994):

- Procedural validity, examining whether the procedures followed were practical, implemented properly, that feedback given to the judges was effective, and that sufficient documentation has been compiled.

- Internal validity, addressing issues of accuracy and consistency of the standard setting results.

- External validation, by collecting evidence from independent sources which support the outcome of the standard setting meeting.

The *Manual* presents in detail how (mostly) quantitative data under these three categories should be collected and analyzed to support the proposed cut scores in relation to the CEFR levels. Studies reporting on the alignment of assessments with the CEFR routinely employ such quantitative techniques to provide validity evidence for the setting of cut scores in relation to the CEFR levels. For example, the alignment of the reading and listening scores of the Michigan English Test with the CEFR levels (Papageorgiou 2010b) involved examination of both intrajudge and interjudge consistency, such as standard error of judgment, agreement coefficient, and Kappa indices as part of the internal validation of the cut score. In another study, Kantarcioglu et al. (2010) applied the many-facet Rasch model (Linacre 1994) to explore the judges' agreement in setting cut scores for the Certificate of Proficiency in English of Bilkent University to the CEFR levels.

A qualitative approach to investigating the judges' decision-making process when setting cut scores to the CEFR was employed by Papageorgiou (2010a). The study investigated the factors reportedly affecting the panelists' decision to set a cut score and the problems faced when setting cut scores in relation to the CEFR. The panelists' group discussions were analyzed based on a coding scheme built both inductively, that is, drawing codes from the actual data, and deductively, that is, drawing codes from existing theory, such as qualitative research into participants' experiences in standard setting (Buckendahl 2005). The findings of the study suggest that decision-making might be affected by factors irrelevant to the description of expected performance in the

CEFR, such as panelists' personal expectations and experiences, which might threaten the validity of the cut score. The study also found that the CEFR might be useful for defining learning objectives, but is not sufficiently specified for the purpose of setting cut scores.

To conclude, research approaches that evaluate alignment with the CEFR include both quantitative and qualitative techniques and there are a growing number of relevant studies employing both. However, future research still needs to address several issues regarding CEFR alignment, as discussed in the next section.

V. FUTURE RESEARCH IN CEFR ALIGNMENT

When aligning test scores with the CEFR, an important decision to be made is whether a score that demonstrates sufficient performance on the assessment also indicates sufficient performance in relation to the CEFR. This is particularly the case for language assessments reporting results in the form of a pass/fail result. A pass/fail result is usually the case with licensure examinations, intended for professionals such as doctors or pilots, who need to pass the exam in order to practice their profession. Language assessments might also report pass/fail results, typically accompanied by a certificate which documents that a test-taker performed satisfactorily on the assessment. If the content of this assessment has been aligned with a specific level on the CEFR, then the implication is that all test-takers with a "pass" certificate should be at the intended CEFR level. Therefore, two decisions need to be made regarding the use of the scores from such an assessment: first, whether a score indicates that a test-taker has passed the assessment, and second, whether this "pass" score indicates that the targeted CEFR level has been achieved (see Council of Europe 2009: 58). More research is needed to understand the relationship between these two cut score decisions, which for now remains unclear.

Aligning assessments with the CEFR has important implications for policy-making. There has been considerable criticism of the uses of the CEFR as a policy document (McNamara 2006; McNamara and Roever 2006), in particular when it comes to immigration. According to Alderson (2007: 260), an unintended consequence of the adoption of the CEFR as a tool by policy-makers is that these officials have no

understanding of the nature of language learning, yet they impose requirements for language proficiency without any consideration as to whether such levels are achievable. For example, language tests are extensively used as gatekeepers for immigration purposes (Shohamy and McNamara 2009) based on language requirements defined in terms of CEFR level (see for example requirements by the UK Border Agency at http://www.ukba.homeoffice.gov.uk). However, the rationale behind the selection of a given CEFR level for a specific purpose such as immigration is not always clear. Therefore, more research is needed in local contexts to identify reasonable language requirements for specific language uses in order to inform policy-making.

Another important implication of CEFR alignment for learners, teachers, and score users is the interpretation of results from different assessments that claim alignment with the same CEFR level. These different assessments should not be interpreted as equivalent in terms of difficulty or content coverage (Council of Europe 2009: 90). Achieving CEFR Level B1 on a general proficiency test intended for young learners and a test of English for Specific Purposes (ESP) does not mean that the scores on these two tests have the same meaning because the intended test purpose, test content, and test-taking population are notably different. One way to provide more accurate information about assessment results is to provide empirically-derived, test-specific performance levels and descriptors designed for a given assessment, for example by adopting a scale anchoring methodology (Garcia Gomez, Noah, Schedl, Wright, and Yolkut 2007). Such levels and descriptors can be provided in addition to information about CEFR alignment.

VI. CONCLUSION

As discussed in this paper, the CEFR and in particular its language proficiency scales and descriptors might offer language teachers, learners and users of assessments an opportunity to better understand the meaning of the results of these assessments (Kane 2012). However, alignment with the CEFR should not be considered a substitute to ongoing procedures for validation (Fulcher 2004). The *Manual* strongly emphasizes that a prerequisite for any effort to achieve alignment with the CEFR is that an assessment be of high quality, otherwise alignment is "a wasted enterprise" (Council of Europe

2009: 90). For example, if an assessment is not reliable, setting a minimum score on this assessment to indicate adequate performance at a given CEFR level will not be particularly meaningful. Moreover, it should also be pointed out that the theoretical underpinnings of the CEFR remain weak (Alderson 2007) and that its language proficiency scales are primarily a taxonomy that makes sense to practitioners, rather than empirically validated descriptions of the language learning process (North and Schneider 1998: 242-243). Moreover researchers have noted several problems with the use of the CEFR for designing test specifications (Alderson et al. 2006). Therefore, content alignment of an assessment with the CEFR cannot provide sufficient evidence of content validity or substitute various language learning theories that should be considered when designing an assessment.

Alignment with the CEFR might not be straightforward because, by design, the description of what learners are expected to do in the CEFR is under-specified to allow for a wider application across a variety of contexts. Unfortunately, this intended under-specification might also mean that alignment of assessments for specific groups of test-takers may be particularly challenging, for example, in the case of assessments for young learners (Hasselgreen 2005). Despite these issues, it could be argued that alignment of assessments with the CEFR remains an important area of inquiry in the field of language assessment because it has the potential to raise awareness of important assessment issues, for example in contexts where local tests are developed (Kantarcioglu, et al. 2010).

Acknowledgements

REFERENCES

Alderson, J. C. 2005a. *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.

Alderson, J. C. 2005b. Editorial. *Language Testing, 22*(3), 257-260.

Alderson, J. C. 2007. The CEFR and the need for more research. *The Modern Language Journal, 91*(4), 659-663.

Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., and Tardieu, C. 2006. Analysing tests of reading and listening in relation to the Common European Framework of Reference: The experience of the Dutch CEFR Construct Project. *Language Assessment Quarterly, 3*(1), 3–30.

Alderson, J. C., and Huhta, A. 2005. The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing, 22*(3), 301-320.

Buckendahl, C. W. 2005. Guest editor's introduction: Qualitative inquiries of participants' experiences with standard setting. *Applied Measurement in Education, 18*(3), 219-221.

Cizek, G. J., and Bunch, M. 2007. *Standard setting: A guide to establishing and evaluating performance standards on tests*. London: Sage Publications.

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

Council of Europe. 2009. Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment. A Manual. Retrieved 15/02/2009, from http://www.coe.int/t/dg4/linguistic/manuel1_en.asp

Deville, C., and Chalhoub-Deville, M. 2011. Accountability-assessment under No Child Left Behind: Agenda, practice, and future. *Language Testing, 28*(3), 307-321.

Figueras, N., and Noijons, J. (Eds.). 2009. *Linking to the CEFR levels: Research perspectives*. Arnhem: CITO.

Fulcher, G. 2004. Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly, 1*(4), 253-266.

Garcia Gomez, P., Noah, A., Schedl, M., Wright, C., and Yolkut, A. 2007. Proficiency descriptors based on a scale-anchoring study of the new TOEFL iBT reading test. *Language Testing, 24*(3), 417-444.

Hambleton, R. K. 2001. Setting performance standards on educational assessments and criteria for evaluating the process. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 89 -116). Mahwah, N.J.: Lawrence Erlbaum Associates.

Hambleton, R. K., and Pitoniak, M. J. 2006. Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 433-470). Westport, CT: Praeger Publishers.

Hasselgreen, A. 2005. Assessing the language of young learners. *Language Testing, 22*(3), 337-354.

Kane, M. 1994. Validating the Performance Standards Associated with Passing Scores. *Review of Educational Research, 64*(3), 425-461.

Kane, M. 2012. Validating score interpretations and uses. *Language Testing, 29*(1), 3-17.

Kantarcioglu, E., and Papageorgiou, S. 2011. Benchmarking and standards in language tests. In B. O' Sullivan (Ed.), *Language Testing: Theories and Practices* (pp. 94-110). Basingstoke: Palgrave.

Kantarcioglu, E., Thomas, C., O'Dwyer, J., and O' Sullivan, B. 2010. Benchmarking a high-stakes proficiency exam: the COPE linking project. In W. Martyniuk (Ed.), *Relating language examinations to the Common European Framework of Reference for Languages: Case studies and reflections on the use of the Council of Europe's draft manual* (pp. 102-116).

Linacre, J. M. 1994. *Many-facet Rasch measurement* (2nd ed.). Chicago: MESA Press.

Little, D. 2006. The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact. *Language Teaching, 39*(3), 167-190.

Martyniuk, W. (Ed.). 2010. *Relating language examinations to the Common European Framework of Reference for Languages: Case studies and reflections on the use of the Council of Europe's draft manual*. Cambridge: Cambridge University Press.

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

North, B., and Schneider, G. 1998. Scaling descriptors for language proficiency scales. *Language Testing, 15*(2), 217-262.

Papageorgiou, S. 2010a. Investigating the decision-making process of standard setting participants. *Language Testing, 27*(2), 261-282.

Papageorgiou, S. 2010b. *Setting cut scores on the Common European Framework of Reference for the Michigan English Test.* (Technical Report). Ann Arbor:

University of Michigan. Retrieved 09/09/2011, from http://www.cambridgemichigan.org/sites/default/files/resources/MET_Standard Setting.pdf.

Plake, B. S. 2008. Standard setters: Stand up and take a stand! *Educational measurement: Issues and Practice, 27*(1), 3-9.

Shohamy, E., and McNamara, T. 2009. Language tests for citizenship, immigration, and asylum. *Language Assessment Quarterly, 6*(1), 1-5.

Tannenbaum, R. J., and Katz, I. R. 2013. Standard setting. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology: Vol 3. Testing and assessment in school psychology and education* (pp. 455-477). Washington, DC: American Psychological Association.

Van Ek, J. A., and Trim, J. L. M. 1991. *Waystage 1990*. Cambridge: Cambridge University Press.

Van Ek, J. A., and Trim, J. L. M. 1998. *Threshold 1990*. Cambridge: Cambridge University Press.

Van Ek, J. A., and Trim, J. L. M. 2001. *Vantage*. Cambridge: Cambridge University Press.

Webb, N. L. 2007. Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education, 20*(1), 7-25.

Wilkins, D. A. 1976. *Notional syllabuses*. Oxford: Oxford University Press.

Zieky, M. J., and Perie, M. 2006. A primer on setting cutscores on tests of educational achievement. Retrieved 08/08/2011, from http://www.ets.org/Media/Research/pdf/Cut_Scores_Primer.pdf

Measuring the impact of CLIL on language skills: a CEFR-based approach for Higher Education

Antonio José Jiménez-Muñoz
jimenezantonio@uniovi.es
Universidad de Oviedo, Spain

ABSTRACT

Advocates of CLIL (Content and Language Integrated Learning) have extolled the virtues of this approach to fostering both content and language alike. However, the generalised and varied implementation of EMI (English as a Medium of Instruction) in universities worldwide has led many lecturers to question these claims. This paper presents a CEFR (Common European Framework)-based model for measuring the impact of EMI at the tertiary level, the aim being to provide further evidence of the progress made in language-learning in modules taught in a foreign language. Using questionnaires based on the revised and refined CEFR descriptors from EAQUALS (Evaluation and Accreditation of Quality in Language Services), students answer an initial self-assessment survey about their background and language skills, which is controlled by a final questionnaire targeting their perceived progress throughout their instruction. Designed from the ground up with the CEFR as a backdrop, this method can be easily tallied with objective assessment to uncover data about students' linguistic performance in CLIL contexts.

Keywords: CLIL, CEFR, EMI, educational research, plurilingualism

I. CLIL AS A CONTESTED APPROACH

Most universities around the globe now offer full or partial degrees taught through a foreign language. English has long been the language of science, but these degrees have made English "the language of higher education in Europe" (Coleman 2006: 1). Part programmatic development, part explicit competition in a wider tertiary environment, English has become the de-facto language for academic discourse and those refusing to provide English-taught modules endanger their global scientific visibility (Alexander 2006). Particularly in Europe, there is a strong consensus on the methodological approach to be used when a content module or degree is taught through a language other than the students' mother tongue (often in English as the Medium of Instruction, or EMI). Content and Language Integrated Learning (CLIL) "is part of mainstream school education in the great majority of countries at primary and secondary levels" (Eurydice

2006: 13) and, more recently, it has also become a major move towards multilingualism at the tertiary level (Fortanet-Gómez 2013). Unlike immersion programmes, multilingualism and CLIL assume that the role of language for the participants need not be transparent for either lecturer or student (Lagasabaster and Sierra 2010) and devise interventions to foster linguistic skills alongside content.

Rather than being a strict method, CLIL is "essentially methodological" (Marsh 2008: 244). Its theorists claim it fosters a flexible, inclusive approach which can be applied through many specific methodologies, since both content and language are integrated. Its advocates stress that by "integrating language and subject teaching, various forms of educational success can be achieved where classrooms comprise learners with diverse levels of linguistic competence" (Marsh 2006: 3). They believe it creates an "innovative fusion of non-language subject with and through a foreign language" (Coyle, Hood and Marsh 2010: 1) and, as it focuses mainly on explaining meaning and not language per se, it allows for "implicit and incidental learning" in "naturalistic situations" (Marsh 2002: 72). CLIL reportedly enables one "to learn as you use and use as you learn" (Marsh 2002: 66) rather than learning language on its own or as separate from content.

While the introduction of CLIL/EMI opens a window for the revision of instructional strategies in many disciplines, a considerable number of participants have observed the problems it adds to their programmes of study. Language level is often at the core of this criticism; content lecturers have complained about the lack of sophistication in their students' use of English (Erling and Hilgendorf 2006: 284) and questioned their own abilities to tackle linguistic issues (Airey 2013: 64), particularly in the case of non-native English (Coyle 2008: 105-106). The seemingly implicit need for the "watering down and simplification of content to make it comprehensible" (Costa and Coleman 2010: 13) questions the quality of learning in EMI contexts. Some have thus called for the "limiting effect" of a foreign language as a medium of instruction on students' final performance to be accepted as something unavoidable (Clegg 2001: 210).

Prima facie, CLIL as theory and its implementation as practice seem to be at odds. Opponents, however, often overlook the fact that provisions made for vital structural readjustments affecting budget, staff development and educational design (Mellion 2006) may not always have been well designed (Airey 2011: 43-45). There is clearly a

need for more development programmes, and advances in teacher-training (Marsh, Mehisto, Wolff and Frigols-Martin 2012) may help design better programmes that target the specific skills needed to teach through a second language (Ball and Lindsay 2013). Lecturers nevertheless attempt to overcome these linguistic hurdles with a number of mediations, techniques and approaches, but the success of these actions is often left unexplored, analysed subjectively or tightly linked to results. A closer assessment of both the effectiveness of those practices and their connection to the expected learning outcomes is essential to guarantee quality. For the learner, it is not a case of either-or: students need both content and language for future studies and employment (Rienties, Brouwer and Lygo-Baker 2013). Until conclusive proof is found, the controversy over the usefulness of CLIL and EMI to promote excellence in both content and language learning will continue. Our study seeks to begin to address the lack of empirical evidence by establishing an evaluation framework that could be readily applicable to various CLIL implementations in order to provide proof towards their effect on the English of learners.

II. THE NEED FOR EVIDENCE-BASED RESEARCH

There is a strong need to provide evidence for the hypotheses surrounding CLIL. The hopeful advantage of teaching "two for the price of one" and the "added value of CLIL" both need to be supported (Bonnet 2012: 66) to avoid the "risk of becoming a 'buzz word' without evidence-based research" (Hunter and Parchoma 2012). Unless this research is carried out to assess the quality of language-learning under CLIL programmes, they run the risk of becoming undervalued despite their obvious merits. Most criticism is focused on the idea that A2-B1 students fresh from high school will find it impossible to cope with the linguistic demands of academic tasks; as a student progresses towards C1-level, this perception of English as limiting students' academic performance becomes gradually unsustainable. However, several reasonable questions remain: how much does a student progress in his or her language proficiency with no specific language instruction? How successful are particular instruction tactics to teach the language and motivate students towards linguistic subtleties? How aware of their linguistic progress are students and how do they qualify it?

There has been comparatively little hands-on research assessing the effectiveness of CLIL implementations, and most of what has been conducted has been mainly restricted to primary and secondary education (Thomas and Collier 2002; Barnett, Yarosz, Thomas, Jung and Blanco 2007; Cobb, Vega and Kronauge 2009; Lindholm-Leary 2011). Some studies, such as Kirkgöz (2005, 2009), Vizconde (2006) and Londo (2012), have centred upon student motivation and attitudes towards the use of EMI in higher education, while others have linked particular skills to academic results (Morrison, Merrick, Higgs and Le Métais 2005; Rienties, Beausaert, Grohnert, Niemantsverdriet and Kommers 2012). However, there is very little empirical research on students' actual performance which is data-driven (Llinares and Whittaker 2010). Gradually, studies focused on particular skills (Ruiz de Zarobe 2010, Navés 2011) and the effect of language level on learning content (Aguilar and Muñoz 2013) are appearing, but more solid and standardised evidence is needed for the language learning taking place under CLIL/EMI conditions, particularly in higher education contexts.

III. A CEFR-BASED RESEARCH METHOD

The multiplicity of different linguistic skills displayed by students in class requires conflating these skills into a framework which makes students' progress in reading, writing, listening and speaking understandable, referable and translatable. With its detailed use of descriptors, we have a perfect backdrop in the CEFR for the interpretation of particular learner skills. Thus, thinking inversely, those tools which aim to assess that linguistic performance would benefit from observing the CEFR from the ground up, rather than having their outcomes transferred to CEFR descriptors and levels. Our approach designs a framework which grades student abilities and interprets them into CEFR-coded language, which is in turn easier for linguists to demonstrate against students' actual oral and written performance. It also helps when it comes to tracking learner performance over time, so that a student would progress or regress among different levels in that particular skill during their period of EMI instruction. Consequently, it may prove a valuable tool to show both qualitative and quantitative evidence for the actual linguistic learning taking place in CLIL/EMI programmes.

The aim of our study, then, is to offer an initial but solid framework for such evaluation based upon the CEFR levels and descriptors as revised and refined by EAQUALS (Evaluation and Accreditation of Quality in Language Services). These expanded, new levels and descriptors are more readily usable by university students as they are more varied and their wording is less technical than the original CEFR descriptors developed in the 1990s. The new EAQUALS design also helps identify a total of eleven levels instead of the original six, which adds granularity and refinement, thereby making these descriptors particularly useful for shorter periods of reduced direct instruction, such as university semesters.

III.1. Design phase

The implementation of our framework has three major stages: design, data-collection and evaluation. In the design phase, content lecturers allocate CEFR-compliant skills and sub-skills for students to perform, as specified in the module syllabus, and their attached tasks. These particular skills are then individually mapped onto their respective CEFR level, which offers a detailed multi-level grid of the expected level from students. Doing so has several inherent diagnostic advantages, such as detecting potential problems, designing strategies to promote excellence or setting an egress CEFR level, as well as realising the linguistic complexity of those academic tasks set by lecturers. Often overlooked in L1 contexts, this effect in module design might hinder students' academic performance on purely linguistic grounds. These sub-skills are the major part of a questionnaire reflecting actual student ability as per CEFR levels both in each grouped skill and globally. Our survey adds more details about students' background and profile to enable subject modelling and prototyping.

III.2. Data collection

First, students must complete an initial questionnaire (at the beginning of the semester, module or course) detailing their attitude towards studies, intrinsic motivation and self-perception of their English-language skills. The language part of the questionnaire is standardised according to CEFR/EAQUALS descriptors (see references below), but students are not given any indication about the level each descriptor belongs to. They

are required to evaluate their confidence in doing a number of linguistic tasks. In this way, students are in fact grading themselves in the skills and sub-skills identified as being required for their university core modules and tasks. Additionally, it allows quantitative fine-graining, in five degrees, of their confidence in their performance in each skill, and fallibility control by means of items of the same skill/level in the same questionnaire. A second questionnaire is presented towards the end of the research period. When data has been compiled and analysed, this method allows student perceptions to be verified through an analysis of the oral or written work they have produced in their modules, which in turn enables language to be prototyped and linked to students' academic achievement.

III.2.1 Initial survey

Please note that this survey is a tailored questionnaire, specifically designed for first-year students in an EMI Economics degree programme at a Spanish university and, as such, it could serve as a proof-of-concept, to be expanded and adapted to other contexts. Students have a minimum A2 level of English in all four skills, although their background and language abilities vary greatly. The survey is computerised, nominal and all items are compulsory. Items 1-3 are demographic. Items 4-15 enquire about the English-related educational background of the student to enable modelling/prototyping. Items 16-19 respond to those skills needed in students' particular degree programmes and in this case vary from B1 to C1 levels, since a full A1-C2 range is possible but unmanageable. The full CEFR range is kept, however, for a more general evaluation of control items 20-27. CEFR level references, in square brackets, are stated for the purposes of this paper, but are not present in the actual student survey.

Table 1. English Level Self-Assessment: Initial Survey

1. Please state your sex.	• Male • Female
2. Which year were you born?	(open numeric answer)
3. Which is your first language?	• Spanish • French • Italian

	• German • An Asian language • Another Romance language • Another Germanic language • A Slavic language • Other
4. Please state whether you can use other languages and, if so, how well, using 1 (= not spoken) to 5 (fluent).	• Spanish (1-5) • French (1-5) • Italian (1-5) • German (1-5) • An Asian language (1-5) • Another Romance language (1-5) • Another Germanic language (1-5) • A Slavic language (1-5) • Other – please specify (1-5)
5. How many years have you been learning English?	(open numeric answer)
6. Which of these aspects do you find particularly difficult about learning English? (You can select more than one)	• Grammar structures • Vocabulary • Pronunciation • Understanding what I read • Understanding native speakers • Understanding other foreigners when they speak English • Writing long, formal texts • Writing short, informal texts • Speaking English with native speakers • Speaking English with other foreigners
7. Which of these aspects do you find easier about learning English? (You can select more than one)	• Grammar structures • Vocabulary • Pronunciation • Understanding what I read • Understanding native speakers • Understanding other foreigners when they speak English • Writing long, formal texts • Writing short, informal texts • Speaking English with native speakers • Speaking English with other foreigners
8. Which of these do you regularly do in English, if any? (You can select more than one item)	• Listening to music • Reading books or magazines • Reading websites • Speaking to other people • Talking on the phone/webcam • Writing letters or e-mails • Using social networks, text-based chat or forums • Watching online videos, films or TV programmes • Other – please specify
9. Have you ever been to an English-speaking country, lived, studied or worked with people whose first language was English, so you had to use English for communication? Please state the total amount of time these experiences lasted for.	• Never • 1 month • 1-3 months • 4-6 months • 6-12 months • More than 1 year • More than 2 years • More than 3 years

	• 4 years or more
10. Please state your general attitude towards these points, using 1 (very negative) to 5 (very positive).	• English language (1-5) • English lessons (1-5) • English teachers (1-5) • British or American culture (1-5) • The culture of other English-speaking countries (1-5) • British or American native speakers (1-5) • Speakers from other English-speaking countries (1-5) • Speaking in English with foreigners (1-5)
11. How would you describe your general level of English?	• Basic or Elementary • Pre-intermediate • Intermediate • Upper-intermediate • Advanced • Proficient / Bilingual
12. What are your motivations for studying/improving your English? (You can select more than one)	• Travelling around the world • Meeting people from different countries • Having more opportunities in the future • Knowing more about the language and culture • Finding a job in the future • Using the language as a tool to work in my country • Using the language as a tool to work abroad • Selling goods to other countries • Being socially respected • Other – please specify
13. What was your approximate university access test grade in the English paper? Please use a comma for decimals. Example: 8,2	(open numeric answer)
14. Have you taken any extra English courses, apart from those at school?	• Never • About one year • About 2-3 years • About 4-5 years • About 6-7 years • More than 7 years
15. Which is the highest English certificate you have obtained?	• No certificate, just high-school / Official Language School Basic level / Cambridge KET with merit or PET / Trinity 3 or 4 (CEFR A2 , or similar certificate) • Official Language School Intermediate level / Cambridge PET with merit or FCE grade D / Trinity 4 or 5 (CEFR B1, or similar certificate) • Official Language School Advanced level / Cambridge PET with distinction, FCE grade B or C, CAE grade D / Trinity 6, 7 or 8 (CEFR B2, or similar certificate) • Official Language School C1 / Cambridge FCE grade A, CAE grade B or C, CPE grade D / Trinity 10 or 11 (CEFR C1, or similar certificate) • Cambridge CAE grade A, CPE grade A, B or C / Trinity 12 (CEFR C2, a similar certificate, or native speaker)
16. Can you do the following? Answer using 1 (No) to 5 (Yes) to express how confident you feel in each of these skills.	• I can understand in detail texts within my field of interest or the area of my academic or professional speciality. [B2.2] • I can understand the motives behind the characters' actions and their consequences for the development of the plot in a narrative or play. [B2.2] • I can rapidly grasp the content and the significance of news,

	articles and reports on topics connected with my interests or my job, and decide if a closer reading is worthwhile. [B2.1]
	• I can read and understand articles and reports on current problems in which the writers' express specific attitudes and points of view (e.g. arts reviews, political commentary). [B2.1]
	• I can take a quick look through simple, factual texts in magazines, brochures or on a website, and identify whether they contain information that might be of practical use to me. [B1.2]
	• I can identify the main conclusions in clearly written texts which argue a point of view. [B1.2]
	• I can follow clear, written instructions (for example for a game, the use of a cosmetic or medicine or when using a piece of electronic equipment for the first time). [B1.1]
	• I can find out and pass on straightforward factual information. [B1.1]
17. Can you do the following? Answer using 1 (No) to 5 (Yes) to express how confident you feel in each of these skills.	• I can use a variety of linking words efficiently to mark the relationships between ideas clearly. [B2.2]
	• I can develop an argument systematically in a composition or report, emphasising decisive points and including supporting details. [B2.2]
	• I can present a topical issue in a critical manner and weigh up the advantages and disadvantages of various options. [B2.1]
	• I can write clear, detailed descriptions on a variety of subjects related to my field of interest. [B2.1]
	• I can communicate with reasonable accuracy and can correct mistakes if they have led to misunderstandings. [B2.1]
	• I can summarise non-routine information on familiar subjects from various sources and present it to others with some confidence. [B1.2]
	• I can develop an argument in writing well enough to be followed without difficulty most of the time. [B1.2]
	• I can write short, comprehensible connected texts on familiar subjects. [B1.1]
18. Can you do the following? Answer using 1 (No) to 5 (Yes) to express how confident you feel in each of these skills.	• I can identify speaker viewpoints and attitudes as well as the information content in recordings on social, professional or academic subjects. [B2.2]
	• I can understand the main ideas of complex speech delivered in a standard dialect, including technical discussions in my field of specialisation. [B2.1]
	• I can understand in detail what is said to me in standard spoken language even in a noisy environment. [B2.1]
	• I can follow the essentials of lectures, talks and reports and other forms of academic or professional presentation in my field. [B2.1]
	• I can understand announcements and messages on concrete and abstract topics spoken in standard dialect at normal speed. [B2.1]
	• I can understand straightforward information about everyday, study- or work-related topics, identifying both general messages and specific details, provided people speak clearly in a familiar accent. [B1.2]
	• I can follow a lecture or talk within my own field, provided the subject matter is familiar and the presentation straightforward and clearly structured. [B1.2]

Language Value 6 (1), 28-50	• I can understand the information in audio material on topics that interest me provided it is spoken clearly in a standard accent. [B1.2] • I can follow a lot of what is said around me, when people speak clearly and without using idioms and special expressions. [B1.2] • I can understand the main points of clear standard speech on familiar, everyday subjects, provided there is an opportunity to get repetition or clarification sometimes. [B1.1]
19. Can you do the following? Answer using 1 (No) to 5 (Yes) to express how confident you feel in each of these skills.	• I can keep up with a discussion and express my ideas and opinions clearly, precisely and convincingly even in formal meetings. [B2.2] • If I don't know a word or expression I can find another way of saying what I mean. [B2.2] • I can express exactly what I want to, focusing on both what I say and how I say it. [B2.2] • I can take an active part in extended conversation on most general topics. [B2.1] • I can evaluate advantages and disadvantages, and participate in reaching a decision. [B2.1] • I can account for and sustain my opinions in discussion by providing relevant explanations, arguments and comments. [B2.1] • I can give clear, detailed descriptions on a wide range of subjects related to my fields of interest. [B2.1] • I can speculate about causes, consequences, hypothetical situations. [B2.1] • I can develop a clear, coherent argument, linking ideas logically and expanding and supporting my points with appropriate examples. [B2.1] • I can use standard phrases like "That's a difficult question to answer" to gain time and keep the turn while formulating what to say. [B2.1] • I can generally correct slips and errors if I become aware of them or if they have led to misunderstandings. [B2.1] • I have sufficient vocabulary and can vary formulation when expressing myself on matters connected to my field and on most general topics. [B2.1] • I can produce stretches of language with a fairly even tempo; although I can be hesitant as I search for expressions, there are few noticeably long pauses. [B2.1] • I can explain why something is a problem and comment on what other people think. [B1.2] • I can develop an argument well enough to be followed without difficulty most of the time. [B1.2] • I can sum up what has been said about something we are discussing. [B1.2] • I can start, maintain and close simple face-to-face conversation on topics that are familiar or of personal interest. [B1.1] • I can agree and disagree politely. [B1.1] • I can briefly explain and justify my points of view. [B1.1] • I can give a short prepared presentation, without practising word for word, and answer clear questions. [B1.1] • I can ask someone to clarify or elaborate what they have just said. [B1.1]

20. Select the statement which best represents your general ability in reading English.	• I can understand any kind of text, including those written in a very colloquial style and containing many idiomatic expressions or slang. [C2] • I can understand in detail lengthy, complex texts, whether or not they relate to my area of speciality. [C1.2] • I can understand in detail a wide range of lengthy, complex texts likely to be encountered in social, professional or academic life, though I may want time to reread them. [C1.1] • I have a broad active reading vocabulary, which means I can read with a large degree of independence, adapting style and speed of reading to different texts and purposes. [B2.2] • I can read with a large degree of independence, using dictionaries and other reference sources selectively when necessary. [B2.1] • I can understand the main points in straightforward texts on subjects of personal or professional interest. [B1.2] • I can understand the main points in straightforward factual texts on subjects of personal or professional interest well enough to talk about them afterwards. [B1.1] • I can understand short, simple texts on familiar subjects, which consist of high-frequency, everyday or job-related language. [A2.2] • I can understand short, simple texts containing familiar vocabulary including international words. [A2.1] • I can read very short, simple texts very slowly by understanding familiar names, words and basic phrases. [A1.2] • I can recognise names, words and phrases I know and use them to understand very simple sentences if there are pictures. [A1.1]
21. Select the statement which best represents your general ability when writing in English.	• I can produce written work that shows good organisational structure, with an understanding of the style and content appropriate to the task. I can produce text which is proof-read and laid out in accordance with relevant conventions. [C2] • I can write well-structured texts which show a high degree of grammatical correctness and vary my vocabulary and style according to the addressee, the kind of text and the topic. [C1.2] • I can write clear, well-structured texts on complex topics in an appropriate style with good grammatical control. [C1.1] • I can use a range of language to express abstract ideas as well as topical subjects, correcting most of my mistakes in the process. [B2.2] • I can write at length about topical issues, even though complex concepts may be oversimplified, and can correct many of my mistakes in the process. [B2.1] • I can write about a variety of familiar subjects well enough for others to follow my story or argument. [B1.2] • I can write short, comprehensible connected texts on familiar subjects. [B1.1] • I can write about my everyday life in simple sentences (people, places, job, school, family, hobbies, etc.). [A2.2] • I can write about myself using simple language. For example: information about my family, school, job, hobbies, etc. [A2.1] • I can write simple sentences about myself, for example,

	where I live and what I do. [A1.2] • I can write about myself and where I live, using short, simple phrases. [A1.1]
22. Select the statement which best represents your general ability to understand spoken in English.	• I can understand any kind of spoken language, whether live or broadcast, delivered at fast speed, even in a noisy environment. I can appreciate irony and sarcasm and draw appropriate conclusions about their use. [C2] • I can understand a wide range of idiomatic expressions and colloquialisms, appreciating shifts in style and register. [C1.2] • I can understand enough to follow extended speech on abstract and complex topics of academic or vocational relevance. [C1.1] • I can understand standard spoken language, live or broadcast, even in a noisy environment. [B2.2] • I can understand the main ideas of complex speech on concrete and abstract topics delivered in a standard dialect, including technical discussions in my field of specialisation. [B2.1] • I can understand straightforward information about everyday, study- or work-related topics, identifying both general messages and specific details, provided people speak clearly in a familiar accent. [B1.2] • I can understand the main points of clear standard speech on familiar, everyday subjects, provided there is an opportunity to get repetition or clarification sometimes. [B1.1] • I can understand enough of what people say to be able to meet immediate needs, provided people speak slowly and clearly. [A2.2] • I can understand simple information and questions about family, people, homes, work and hobbies. [A2.1] • I can understand people if they speak very slowly and clearly about simple everyday topics. I can understand people describing objects and possessions (e.g. colour and size). [A1.2] • I can understand simple words and phrases, like "excuse me", "sorry", "thank you", etc. I can understand the days of the week and months of the year. I can understand times and dates. I can understand numbers and prices. [A1.1]
23. Which of these is closest to what you can do when understanding spoken English?	• I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly. [A1] • I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements. [A2] • I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear. [B1] • I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of

	films in standard dialect. [B2]
	• I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort. [C1]
	• I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent. [C2]
24. Which of these is closest to what you can do when understanding texts in English?	• I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues. [A1]
	• I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables, and I can understand short simple personal letters. [A2]
	• I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters. [B1]
	• I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose. [B2]
	• I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field. [C1]
	• I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works. [C2]
25. Which of these is closest to what you can do when talking to others in English?	• I can interact in a simple way provided the other person is prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say. I can ask and answer simple questions in areas of immediate need or on very familiar topics. [A1]
	• I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even though I can't usually understand enough to keep the conversation going myself. [A2]
	• I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events). [B1]
	• I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible. I can take an active part in discussion in familiar contexts, accounting for and sustaining my views. [B2]
	• I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social and professional purposes. I can formulate ideas and opinions with precision and relate my contribution skilfully to those of other speakers. [C1]
	• I can take part effortlessly in any conversation or discussion

	and have a good familiarity with idiomatic expressions and colloquialisms. I can express myself fluently and convey finer shades of meaning precisely. If I do have a problem, I can backtrack and restructure around the difficulty so smoothly that other people are hardly aware of it. [C2]
26. Which of these is closest to what you can do when speaking on your own in English?	• I can use simple phrases and sentences to describe where I live and people I know. [A1] • I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job. [A2] • I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions. [B1] • I can present clear, detailed descriptions on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options. [B2] • I can present clear, detailed descriptions of complex subjects integrating sub-themes, developing particular points and rounding off with an appropriate conclusion. [C1] • I can present a clear, smoothly-flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points. [C2]
27. Which of these is closest to what you can do when writing in English?	• I can write a short, simple postcard, for example, sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form. [A1] • I can write short, simple notes and messages relating to matters in areas of immediate needs. I can write a very simple personal letter, for example thanking someone for something. [A2] • I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions. [B1] • I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences. [B2] • I can express myself in clear, well-structured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select style appropriate to the reader in mind. [C1] • I can write clear, smoothly-flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works. [C2]

III.2.2 Final survey

This survey builds heavily upon the initial questionnaire. For this reason, items 31-42, being identical to 16-27, are not reproduced here.

Table 2. English Level Self-Assessment: Final Survey.

28. After receiving lectures and performing tasks in another language, how has your English level changed? What skills have you improved the most? Write a short comment about your ideas.	(Open answer)
29. How much do you consider you have improved your language skills after the module? 1= not at all; 5= I have improved a lot	• Grammar structures (1-5) • Vocabulary (1-5) • Pronunciation (1-5) • Understanding what I read (1-5) • Understanding native speakers (1-5) • Understanding other foreigners when they speak English (1-5) • Writing long, formal texts (1-5) • Writing short, informal texts (1-5) • Speaking English with native speakers (1-5) • Speaking English with other foreigners (1-5)
30. How would you describe your general level of English at present?	• Basic or Elementary • Pre-intermediate • Intermediate • Upper-intermediate • Advanced • Proficient / Bilingual
…	

III.3. Data evaluation

Finally, in the evaluation stage, quantitative data from these channels above are collated and analysed in order to trace the impact of EMI on students' self-perceived evolution. When analysing data following our method, two strategies are observed: control and scoring. Student data are controlled by three groups of items per skill, so that their level is not severely influenced by a single choice. Student responses are weighted and, using a software package for statistical analysis, greater variance in student choices for descriptors at the same level is deleveraged. Student responses are scored and distributed across the CEFR levels as follows:

Table 3. CEFR level default scores.

Level	Default Score
C2	100
C1.2	90
C1.1	80
B2.2	70
B2.1	60
B1.2	50
B1.1	40
A2.2	30
A2.1	20
A1.2	10
A1.1	5

Items which show module-specific sub-skills (16-19) are analysed in a specific way: for each item, each of the self-assessed options (1-5) is multiplied by its default level score shown in Table 3 and the mean of the multiple choice values is adjusted to the value of the maximum potential score (5 x 100 C2 level score). Choices in items 20-27 are given the value of the default score for their level. The total for the skill is calculated using the weighted mean of the values in each category, unless adjusted by the statistical software, and interpreted as the level for that particular skill as referred to above. As an indication, we calculate the global level of the student as the mean of the four skills.

Table 4. Outline of the analysis scoring.

Skill	Item (weight)	Score	Total
Reading	16 (50%)	Adjusted mean (choice 1 x level score, choice 2 x level score , choice 3 x level score …)	Weighted mean
	20 (25%)	Choice x level score	
	24 (25%)	Choice x level score	
Writing	17 (50%)	Adjusted mean (choice 1 x level score, choice 2 x level score , choice 3 x level score …)	Weighted mean
	21 (25%)	Choice x level score	
	27 (25%)	Choice x level score	
Listening	18 (50%)	Adjusted mean (choice 1 x level score, choice 2 x level score , choice 3 x level score …)	Weighted mean
	22 (25%)	Choice x level score	
	23 (25%)	Choice x level score	
Speaking	19 (50%)	Adjusted mean (choice 1 x level score, choice 2 x level score , choice 3 x level score …)	Weighted mean
	25 (25%)	Choice x level score	
	26 (25%)	Choice x level score	
Overall	Only as an indication of student's overall level		Mean

III.3.1 Sample analysis

The sample below is from Julia A., a first-year 19-year-old female student in an Economics degree programme at a state-funded Spanish University. Her L1 is Spanish, with 12 years of English instruction. Julia finds pronunciation and speaking to foreigners particularly difficult, and thinks that writing is her strongest skill. She regularly reads webpages and forums in English, but she does not often listen to or speak English. Julia has been to an English-speaking country for less than three months altogether, and she shows a positive attitude towards the English language and cultures. She has received extra-curricular English lessons for about two years, and she has not obtained any kind of language certificate. Julia describes her level of English as intermediate, and expects to use English in the future to improve her employability.

Table 5. Initial survey (24 Sept 2012).

Skill	Item	Score	Total
Reading	16	Option 1, B2.2: 4 x 70 = 280 Option 2, B2.2: 3 x 70 = 210 Option 3, B2.1: 5 x 60 = 300 … Mean: 263.8 Adjusted mean: 26380 / 500 = 52.8	58.9 [B1.2]
	20	Option 5, B2.1: 60	
	24	Option 4, B2.2: 70	
Writing	17	Option 1, B2.2: 4 x 70 = 280 Option 2, B2.2: 4 x 70 = 280 Option 3, B2.1: 5 x 60 = 300 … Mean: 272.3 Adjusted mean: 27230 / 500 = 54.5	64.8 [B2.1]
	21	Option 3, C1.1: 80	
	27	Option 4, B2.2: 70	
Listening	18	Option 1, B2.2: 3 x 70 = 210 Option 2, B2.1: 4 x 60 = 240 Option 3, B2.1: 5 x 60 = 300 … Mean: 245.8 Adjusted mean: 24580 / 500 = 49.2	49.6 [B1.1]
	22	Option 6, B1.2: 50	
	23	Option 6, B1.2: 50	
Speaking	19	Option 1, B2.2: 3 x 70 = 210 Option 2, B2.2: 4 x 70 = 280 Option 3, B2.2: 3 x 60 = 180 … Mean: 196.8 Adjusted mean: 19680 / 500 = 39,4	44.7 [B1.1]
	25	Option 3, B1.2: 50	
	26	Option 3, B1.2: 50	
Overall		Only as an indication of student global level	54.5 [B1.2]

Table 6. Final survey (12 July 2013).

Skill	Item	Score	Total
Reading	16	Option 1, B2.2: 4 x 70 = 280 Option 2, B2.2: 4 x 70 = 280 Option 3, B2.1: 5 x 60 = 300 … Mean: 283.2 Adjusted mean: 28320 / 500 = 56.6	63.3 [B2.1]
	20	Option 4, B2.2: 70	
	24	Option 4, B2.2: 70	
Writing	17	Option 1, B2.2: 5 x 70 = 350 Option 2, B2.2: 4 x 70 = 280 Option 3, B2.1: 5 x 60 = 300 … Mean: 298.7 Adjusted mean: 29870 / 500 = 59.7	69.9 [B2.2]
	21	Option 3, C1.1: 80	
	27	Option 4, B1.1: 80	
Listening	18	Option 1, B2.2: 4 x 70 = 280 Option 2, B2.1: 4 x 60 = 240 Option 3, B2.1: 5 x 60 = 300 … Mean: 266.2 Adjusted mean: 26620 / 500 = 53.2	61.6 [B2.1]
	22	Option 4, B2.2: 70	
	23	Option 6, B2.2: 70	
Speaking	19	Option 1, B2.2: 4 x 70 = 280 Option 2, B2.2: 5 x 70 = 350 Option 3, B2.2: 4 x 60 = 240 … Mean: 231.2 Adjusted mean: 23120 / 500 = 46.2	53.1 [B1.2]
	25	Option 4, B2.1: 60	
	26	Option 4, B2.1: 60	
Overall		Only as an indication of student global level	62 [B2.1]

Table 6. Skills variance (from 24 September 2012 to 12 July 2013)

Skill	Initial	Final	Variance	Level Variance
Reading	58.9 [B1.2]	63.3 [B2.1]	+4.4	+ 0.5
Writing	64.8 [B2.1]	69.9 [B2.2]	+5.1	+0.5
Listening	49.6 [B1.1]	61.6 [B2.1]	+12	+1
Speaking	44.7 [B1.1]	53.1 [B1.2]	+8.4	+0.5
Overall	54.5 [B1.2]	62 [B2.1]	+7.5	+0.5

IV. PROVIDING EVIDENCE: FURTHER STEPS

For the particular case of Julia, her self-assessment reveals a significant improvement in all skills despite receiving no language lessons. Before readily attributing that improvement to EMI, further and deeper work needs to be carried out to assess the objective extent of that influence and its social triggers (Llinares and Morton 2012). Here, we have contrasted this possible influence of EMI against three recorded task-based observations to evaluate student performance by analysing discourse and written work. Some variation was found between their own self-assessment and their actual performance in many of the sub-skills included in the questionnaire (Hernandez-Nanclares and Jimenez-Munoz 2014). The objective assessment we present here would need to be performed on a substantial cohort at a number of institutions over a period of time to give further evidence supporting the CLIL hypotheses. The research method we present here is flexible and subject to refinement and adaptation for that purpose. It could contribute to the considerably small amount of quantitative research conducted on the impact of language-based interventions on language competence and academic performance within CLIL contexts. Furthermore, modelling of student profiles would enable preventive interventions to be designed. Such interventions are particularly well suited to the context of plurilingual higher education, where it is essential to attest the quality of teaching and learning, and to ascertain cross-disciplinary best practices.

REFERENCES

Aguilar, M. and Muñoz, C. 2013. "The effect of proficiency on CLIL benefits in engineering students in Spain". *International Journal of Applied Linguistics*, forthcoming. 1 March 2014 <http://onlinelibrary.wiley.com/doi/10.1111/ijal.12006/pdf>

Airey, J. 2012. "'I don't teach language'. The linguistic attitudes of physics lecturers in Sweden". In Smit, U. and E. Dafouz (Eds.) Integrating Content and Language in Higher Education. Gaining Insights into English-Medium Instruction at European Universities. Special Issue of AILA Review 25, 64-79.

Alexander, J. R. 2006. "International programmes in the German-speaking world and englishization: a critical analysis". In Wilkinson, R. and Zegers, V. (Eds.)

Realizing Content and Language Integration in Higher Education. Maastricht: Universitaire Pers, 77-95.

Ball, P. and Lindsay, D. 2013. "Language demands and support for English-medium instruction in tertiary education: learning from a specific context". In Doiz, A., D. Lasagabaster and J.M. Sierra (Eds.) *English-medium instruction at universities: global challenges*. Bristol: Multilingual Matters, 44–64.

Barnett, W. S. Yarosz, D. J. Thomas, J. Jung, K. and Blanco, D. (2007). "Two-way and monolingual English immersion in preschool education: an experimental comparison". *Early Childhood Research Quarterly* 22, 227-293.

Bonnet, A. 2012. "Towards an evidence base for CLIL: how to integrate qualitative and quantitative as well as process, product and participant perspectives in CLIL research". *International CLIL Research Journal* 1 (4), 65-78.

Clegg, J. 2001. "Can English-medium education work? Why we should be honest about its failures and courageous about change". In Trewby R. and S. Fitchat (Eds.) *Language and Development in Southern Africa*. Windhoek: Macmillan, 210-226.

Cobb, B., Vega, D., and Kronauge, C. 2009. "Effects of an elementary dual language immersion school program on junior high school achievement" In Hough, D. L. (Ed.) *Middle grades research: Exemplary studies linking theory to practice*. Charlotte, NC: Information Age Publishing, 27-47.

Coleman, J. 2006. "English-medium teaching in European higher education". *Language Teaching* 39 (1), 1-14.

Costa, F. and Coleman, J. A. 2010. "Integrating Content and Language in higher education in Italy: ongoing research'. *International CLIL Research Journal* 1 (3), 19-29.

Coyle, D. 2008. "CLIL - a pedagogical approach from the European perspective". In Van Deusen-scholl, N. and N. Hornberger (Eds.) *Encyclopaedia of Language and Education 2nd Edition Vol. 4*. New York: Springer, 1200-1214.

Coyle, D. Hood, P. and Marsh, D. 2010. *CLIL: Content and Language Integrated Learning*. Cambridge: Cambridge University Press.

Erling, E. and Hilgendorf, S. 2006. "Language policies in the context of German higher education". *Language Policy* 5, 267-292.

Fortanet-Gómez, I. 2013. *CLIL in Higher Education. Towards a Multilingual*

Language Policy. Bristol: Multilingual Matters.

Hernandez-Nanclares, N. and Jimenez-Munoz, A. 2014. "English as a Medium of Instruction: research project in bilingual economic education". *International Review of Economics Education* 15, in print.

Hunter, M. and Parchoma, P. 2012. "Content and Language Integrated Learning: shifting boundaries and terrain mapping". In Hodgson, V., C. Jones, M. de Laat, D. McConnell, T. Ryberg and P. Sloep (Eds.) *Proceedings of the eighth International Conference on Networked Learning, 2-4 April, 2012*. Maastricht: Maastricht School of Management, 514-521.

Kirkgöz, Y. 2005. "Motivation and student perception of studying in an English-medium university". *Journal of Language and Linguistic Studies* 1 (1), 101-123.

Kirkgöz, Y. 2009. "Students' and lecturers' perceptions of the effectiveness of foreign language instruction in an English-medium university in Turkey". *Teaching in Higher Education* 14 (1), 81-93.

Lasagabaster D. and Sierra, J. M. 2010. "Immersion and CLIL in English: more differences than similarities". *ELT Journal* 64, 376-395.

Lindholm-Leary, K. J. 2011. "Achievement and language proficiency of Latino students in dual language programmes: native English speakers, fluent English/previous ELLs, and current ELLs". *Journal of Multilingual & Multicultural Development* 32(6), 531-545.

Llinares, A. and Whittaker, R. 2010. "Writing and speaking in the history class: data from CLIL and first language contexts". In Dalton-Puffer, C., T. Nikula, & U. Smit (Eds.) *Language use and language learning in CLIL classrooms*. Amsterdam: John Benjamins, 125-144.

Llinares, A., and Morton, T. 2012. "Social perspectives on interaction and language learning in CLIL classrooms". In Alcón-Soler, E. and M.P. Safont-Jorda (Eds.) *Discourse and learning across L2 instructional contexts*. Amsterdam: Rodopi, 105-131.

Llinares, A., Morton, T. and Whittaker, R. 2012. *The roles of language in CLIL.* Cambridge, UK: Cambridge University Press.

Londo, M. 2012. *Optimising EMI in higher education through CLIL: case study at Vienna*. Vienna: Lap Lamberg.

Marsh, D. 2002. *CLIL/EMILE, The European Dimension: Actions, Trends and Foresight Potential*. Jyväskylä: University of Jyväskylä.

Marsh, D. 2006. **"**English as a medium of instruction in the new global linguistic order: Global characteristics, Local Consequences". In Stewart, S., Olearski, J and D. Thompson (Eds.). *Proceedings of the Second Annual Conference for Middle East Teachers of Science, Mathematics and Computing*. Abu Dhabi: METSMaC, 29-38.

Marsh, D. Mehisto, P. Wolff, D. and Frigols, M. J. 2012. *European framework for CLIL teacher education: a framework for the professional development of CLIL teachers*. Graz: European Centre for Modern Languages.

Mellion, M. J. 2006. "The Challenge of Changing Tongues in Business University Education". In Wilkinson R. and V. Zegers (Eds). *Realizing Content and Language Integration in Higher Education*. Maastricht: Universitaire Pers, 212-227.

Morrison, J., Merrick, B., Higgs, S., and Le Métais, J. 2005. "Researching the performance of international students in the UK". *Studies in Higher Education* 30 (3), 327 - 337.

Navés, T. 2011. "The promising benefits of integrating content and language for EFL writing and overall EFL proficiency". In Ruiz de Zarobe, Y., J. M. Sierra and F. Gallardo del Puerto (Eds.) *Content and Foreign Language Integrated Learning. Contributions to Multilingualism in European Contexts*. Bern: Peter Lang, 75-103.

Rienties, B. Brouwer, N., and Lygo-Baker, S. 2013. "The effects of online professional development on higher education teachers' beliefs and intentions towards learning facilitation and technology". *Teaching and Teacher Education* 29, 122-131.

Ruiz de Zarobe, Y. 2010. "Written production and CLIL: An Empirical Study". In Dalton Puffer, C., T. Nikula, and U. Smit (Eds.) *Language Use and Language Learning in CLIL classrooms*. Philadelphia: John Benjamins, 191-212.

Thomas, W. P. and Collier, V. P. 2002. *A national study of school effectiveness for language minority students' long-term academic achievement*. Santa Cruz, CA: University of California at Santa Cruz.

Vizconde, C. J. 2006. "Attitudes of student teachers towards the use of English as language of instruction for science and mathematics in the Philippines". *The Linguistics Journal*, 1 (3), 7-33.

The Common European Framework of Reference for Languages: a European framework for foreign language speech development?

Bettina Beinhoff
bettina.beinhoff@anglia.ac.uk
Anglia Ruskin University Cambridge, United Kingdom

ABSTRACT

This paper explores the representation of speech development and particularly perceptive skills in the CEFR level descriptions. The speech-related CEFR sections and related level descriptions are vague and none of the assumptions made therein (such as the supposed linear progression between levels) have been sufficiently tested yet. This paper presents an exploratory study on speech perception in language learners at different levels of proficiency and from different first language (L1) backgrounds (Spanish and German). The study is based on transcriptions in response to short narratives, and investigates what kind of influence listeners' levels of proficiency in the second language (in this case English) and their L1 backgrounds have on how intelligibility is perceived. The results suggest that proficiency levels and L1 background do indeed influence intelligibility (though not always as anticipated) and partially confirm the idea of a linear progression as proposed in the CEFR.

Keywords: *Language assessment, second language acquisition, speech perception, intelligibility, pronunciation, non-native speech*

I. INTRODUCTION

Ever since its publication in 2001, the Common European Framework of Reference for Language (CEFR, Council of Europe 2001) has received much attention from a variety of professional communities, such as language assessors, language teachers and researchers in the area of second and foreign language acquisition. Each of these communities has made significant contributions to the discussion and it is probably fair to say that in this way, the CEFR has stimulated a truly interdisciplinary debate.

The CEFR was introduced to serve as a 'common basis for the elaboration of […] curriculum guidelines […] across Europe' (Council of Europe 2001: 1). Its central part is a description of proficiency levels of learners' skills with the aim of facilitating

comparisons in language teaching and assessment across Europe. After presenting a brief overview of the main areas of research dealing with the CEFR, this paper will focus on phonological competence which is one of six communicative language competences (together with lexical, grammatical, semantic, orthographic and orthoepic competence) that define learners' levels of proficiency within the CEFR. In the following I will be mainly concerned with the representation of speech perception in the CEFR and I will argue that although much work still needs to be done in order to develop the CEFR to a stage where it can be fully used in language teaching and assessment, the CEFR has the potential to be an incredibly useful tool for bringing more structure into the description of 'learner Englishes' and it can provide an alternative to the increasingly criticised use of native speaker norms in second language (L2) and foreign language education. This discussion will be followed by an explorative study which takes 'intelligibility' as a means of testing whether differences between learners at different CEFR proficiency levels actually exist.

Throughout the debate surrounding the CEFR, many perceived shortcomings of the framework were raised across a wide range of areas, such as lack of detail in its descriptor scales which restricts the CEFR's applicability in language testing (Weir 2005) and specifically in test development (Alderson et al. 2006), the political and social agenda underlying the CEFR and the way this affects language testing (Fulcher 2004) and possible 'abuses' of the CEFR in testing migrant's language abilities (Krumm 2007). Nevertheless, the CEFR has been noted to make valuable contributions, including providing a general framework for language testing which builds on familiar concepts (the notion of 'beginner', 'intermediate' and 'advanced' skills) to develop new approaches (e.g. Little 2007), its potential for improving curriculum design (Westhoff 2007) and to spark discussion about the applicability and comparability of language tests across platforms (e.g. Tannenbaum and Wylie 2008), across national borders (e.g. Phakiti and Roever 2011) and across languages (e.g. Breton et al. 2010).

While most of these debates were mainly centred on language testing and assessment, further research looked at specific areas of language use and development, such as discourse structure, where for example Evison (2013) investigated turn-taking in academic spoken discourse and McCarthy (2010) studied fluency and its connection to turn-taking mechanisms such as turn-openings and turn-closings. Most attention,

however, has been devoted to issues concerning lexical and grammatical development within the CEFR with studies based on several of the 'bigger' European languages; e.g. English (Milton 2010), French (Kusseling and Lonsdale 2013) and German (Hancke and Meurers 2013). Moreover, some studies attempted to link criterial features of the CEFR to second language acquisition theory (see Hawkins and Filipović 2012 and Hawkins and Buttery 2010 for a discussion of the development of grammatical features within the CEFR with regard to Universal Grammar and cognitive theories), however, none of them dealt with phonological development.

II. SPEECH DEVELOPMENT IN THE CEFR

Compared to discourse structures and lexical and grammatical development, phonological development has received little attention in research and discussions on the CEFR so far. This is problematic because more general findings from studies in grammatical and lexical development or discourse structure may not be applicable to phonological development because it is known to diverge from other competences and does not follow the same development rates (Flege and Bohn 1989). In very extreme cases the differences between a learner's level of phonological competence and the level of competence in other areas of development can lead otherwise highly proficient second or foreign language learners to speak with an accent that is hardly intelligible (cf. the 'Joseph Conrad Phenomenon' in Scovel 1981). This discrepancy is not addressed in the CEFR proficiency level descriptions or in CEFR-related research publications.

In addition, CEFR level descriptions provide no detail on speech production and perception. While all CEFR level descriptions are intentionally vague, mostly as a matter of practicality and for historical reasons (Trim 2010), the descriptions for speech production and perception are particularly limited. There is some information on 'phonological control' which refers to speech production and mainly focuses on accentedness, intelligibility and overall fluency (see table 1).

Table 1: Requirements for phonological control (Council of Europe 2001: 117)

Note: The labels A1-C2 refer to the proficiency level of the learner, where A1 stands for beginner, A2 for waystage or elementary, B1 for threshold or intermediate, B2 for upper intermediate, C1 for advanced and C2 for mastery level.

	PHONOLOGICAL CONTROL
C2	As C1
C1	Can vary intonation and place sentence stress correctly in order to express finer shades of meaning.
B2	Has acquired a clear, natural, pronunciation and intonation.
B1	Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur.
A2	Pronunciation is generally clear enough to be understood despite a noticeable foreign accent, but conversational partners will need to ask for repetition from time to time.
A1	Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by native speakers used to dealing with speakers of his/her language group.

Perceptive skills, however, are not described as part of phonological control, nor are they described in any detail anywhere else in the CEFR. Rather, speech perception is mentioned as part of other skills; for example under "General phonetic awareness and skills" which combines production and perception:

> "Many learners, particularly mature students, will find their ability to pronounce new languages facilitated by: an ability to distinguish and produce unfamiliar sounds and prosodic patterns; an ability to perceive and catenate unfamiliar sound sequences; an ability, as a listener, to resolve (i.e. divide into distinct and significant parts) a continuous stream of sound into a meaningful structured string of phonological elements; an understanding/mastery of the processes of sound perception and production applicable to new language learning. These general phonetic skills are distinct from the ability to pronounce a particular language." (Council of Europe 2001: 107)

Teasing out the features which relate specifically to perception, the above quote differentiates between the ability to

(a) distinguish unfamiliar sounds and prosodic patterns,

(b) perceive unfamiliar sound sequences,

(c) turn a continuous stream of speech into meaningful input and

(d) understand or master the processes required in learning how to perceive (and presumably understand) a new language.

These perceptive skills are not linked to any specific proficiency level whereas more general issues of speech perception are part of "Listening comprehension". However, much of the listening comprehension level descriptions refer to the discourse level and lexical features, including content and register of speech, and are otherwise rather vague.

II.1. Linear progression

The information given in the CEFR at each proficiency level suggests a linear progression from very basic skills and knowledge at level A1 and A2 through the intermediate stages B1 and B2 to the advanced stages C1 and C2. As Hulstijn (2007) quite rightly pointed out, it is by no means clear whether this progression from one level to the next corresponds to the learners' actual development. Widely discussed theories in L2 speech acquisition, such as the Speech Learning Model (Flege 1995) and the Native Language Magnet Model (Kuhl 1993) propose specific ways in which learners progress in acquiring the sound system of a new language. For example, in the Speech Learning Model, Flege (1995) suggests that it is necessary to create a category for a new speech sound in order to be able to successfully identify it and distinguish it from similar first language (L1) and L2 sounds. Similarly, Kuhl's Native Language Magnet Model (1993) suggests that new speech sounds are perceived with reference to already known speech sounds (typically from the learner's L1). Only with repeated exposure to these new sounds does the learner (implicitly or explicitly) perceive these new sounds as meaningful new speech sounds. This indicates that speech perception and speech production are closely linked, and, in fact, research suggests that these two levels interact (e.g. Lacabex et al. 2008). This suggests that research into L2 speech production

and perception could potentially add much-needed detail to the CEFR level descriptions.

However, so far, research conducted with reference to L2 speech acquisition models has not looked at the specific development of L2 learners or include L2 learners at different levels of proficiency in line with the CEFR. Rather, the focus was mostly on learners from different first language backgrounds (e.g. Anderson-Hsieh et al. 1992, Munro et al. 2006). A notable exception is Hulstijn et al.'s (2011) study which looked at pronunciation skills to determine the connection between speaking proficiency and linguistic competences within the CEFR for learners of Dutch as an L2. This study found that pronunciation skills are an important factor in determining a learner's proficiency level, together with knowledge of vocabulary and grammar in combination with processing speed. This particular study looked at language production, but given the importance of perceptive skills in the L2 acquisition process, it is crucial to establish whether L2 speech perception is a distinguishing factor between CEFR proficiency levels, and if it is, to investigate what the differences between the proficiency levels are.

II.2. Native speaker norms and non-native speaker performance

As a brief additional note, I would like to add a few comments on a potential area of use for the CEFR which has not received much attention in academic debate as yet. Variability in learners' (non-native) speech perception is usually studied in terms of how it varies from native speaker (NS) abilities and to what extent it follows NS norms. This, however, fails to address the fact that for many, if not for most, non-native speakers, NS proficiency in the perception and production of their second/foreign language is unattainable. In addition, there has been much debate about the suitability of NS norms for non-native speakers, especially with regard to the ever increasing international and intercultural use of the English language (cf. e.g. Canagarajah and Wurr 2011, Cook 1999, He and Zhang 2010, Timmis 2002).

This debate reveals a dichotomy between those proposing 'nativeness' (which in itself is difficult to define) as a central pronunciation target and those who regard intelligibility to a wide native and non-native speaker audience as more important (Levis 2005). This need for more suitable norms and aims for language learners has

been considered by Fitzpatrick and Racine (2013) in their recent study on using L1 performance profiles as an alternative tool for L2 assessment, albeit without explicit reference to the CEFR.

This lack of alternative norms and factors is particularly an issue in speech development where, so far, most research included well-established factors such as 'Age of Learning' and 'Length of Stay in Target Language Country' (cf. Piske et al. 2001), which are not necessarily relevant for the majority of language learners. Especially with regard to the worldwide use of English, many learners – in Europe and beyond – tend to learn English in their first language environments and may not necessarily travel to a target language country; and if they do, they may not stay in the country for extended periods of time. Thus, norms and factors which were developed in contexts where language learners lived in the L2 environment for extended periods of time may not be suitable for most learners. In fact, factors such as 'age' have been shown to have a very different impact for learners in L1 environments as compared to those in L2 contexts (see Lecumberri and Gallardo 2003).

Thus, a more appropriate way of studying and assessing non-native speech would consider the learners' progress in relation to their own development and that of language learners from comparable backgrounds. In this way, the reference framework which the CEFR provides could potentially be used to track and assess learner development based on individual progress, rather than native speaker norms. Furthermore, specifying CEFR proficiency levels for non-native speech development would provide more detailed and appropriate descriptions of learner language than what is currently available (see e.g. the descriptions of different learner Englishes in Swan and Smith 2001). Descriptions of language learners' typical speech patterns (if these really exist) at any given proficiency level could therefore also contribute greatly to fairer assessment and realistic learning aims and targets.

III. INTELLIGIBILITY AND CEFR PROFICIENCY LEVELS: AIMS OF THE STUDY

The above discussion reveals that it is by no means clear whether the learner progression as suggested in the CEFR corresponds to the learners' actual development. In addition, the information given on speech perception in the CEFR is almost non-existent and does not give any information on what learners 'can do' at any particular level. The following study is a very first attempt at shedding some light on

(a)   whether there are any differences in speech perception between learners at different CEFR proficiency levels, and

(b)   if there are differences, to reveal any general trends of what might constitute these differences.

Intelligibility is one of the main concerns in the CEFR level descriptions for 'Phonological control' and it also features strongly under 'General phonetic awareness and skills'. In addition, further research into issues of intelligibility is crucial as it is still not fully determined what pronunciation features contribute to intelligibility (cf. Harding 2013). Therefore, this study will use 'intelligibility' to address the main research aim by investigating if learners of English from different proficiency levels (B1 and C2) can understand given speech samples equally well. It is possible that 'intelligibility' may follow a linear progression as it is necessary to understand before it is possible to engage in communication. Therefore, it seems likely that there may be differences between learners at B1 and C2 level. Another aim is to find out what it is that inhibits intelligibility and whether it is different for learners at B1 and C2 level of proficiency.

Given that the L1 is considered to strongly influence speech perception, this study will also investigate if learners of English from two different L1s (German and Spanish) show differences in how they perceive L2 speech. This is quite likely and is implicitly included in the CEFR section on perception where distinguishing unfamiliar sounds and perceiving unfamiliar sound sequences are central concerns (German and Spanish are phonologically very different and may therefore lead to different problems with intelligibility).

Finally, to account for the global use of English and the growing number of learners and non-native speakers (NNS) of English which – at least theoretically – makes NNS to NNS communication much more likely than NNS to NS communication, this study

aims to investigate intelligibility of NNS English. This study uses Spanish NNS accents in English as the speech stimuli because it is one of the most widely distributed NNS accents in English. Previous research is unclear about whether sharing the same L1 as the speaker is beneficial in understanding L2 communication, such as Spanish learners of English understanding Spanish-accented English (a detailed discussion is beyond the scope of this paper, but see e.g. Major et al. 2002 and Munro et al. 2006), however, it is possible that familiar sounds and patterns from L1 Spanish may be easier to understand for Spanish L1 speakers compared to, for example, German L1 speakers, because Spanish speakers are more familiar with such structures.

IV. METHODOLOGY AND PROCEDURE

In this study, 20 German and Spanish learners of English (from here on called "listeners") responded to recorded speech samples spoken by four Spanish NNS of English (from here on called "speakers"). The listeners were German and Spanish learners of English at different CEFR proficiency levels: ten German learners of English; five at proficiency level B1 (i.e. threshold level) and five at proficiency level C2 (i.e. proficient user) and ten Spanish learners of English; five at proficiency level B1 and five at proficiency level C2. In addition, five English NS took part in the study as a control group. The proficiency levels B1 and C2 were chose because the likelihood to obtain distinctive results was deemed relatively high.

All participants were students in Cambridge, UK, where the English NS and the C2 listeners studied a variety of subjects at the local universities and the B1 listeners were students at residential language schools. All listeners grew up monolingual with German, Spanish or English as their L1; the German and Spanish listeners had learned English at schools in their L1 environments before coming to Cambridge. Their proficiency levels were established through proficiency tests they had taken up to two months previously (for the B1 listeners) and up to one year previously (for the C2 listeners, who had been living in the UK since then). The Spanish and German listeners had stayed in English-speaking countries between three weeks and six months (for level B1) and between five months and seven years (for level C2). The age at which they started learning English was similar across all proficiency levels. The Spanish listeners

started learning English when they were between 6 and 9 years old and the German listeners when they were between 8 and 13 years old.

The speech samples were recordings of four Spanish NNS of English with varying degrees of influence from their L1 Spanish in their English pronunciation. All of these speakers were female. The speech samples were recorded in a sound-treated room with a digital recorder. The speakers were asked to describe three different picture stories and to talk about themselves without revealing any personal information. From these recordings, short utterances were isolated using the speech processing software Audacity (version 2.0.3). These utterances varied in length between five and 15 words. For the experiment, eight speech files per speaker were selected from these utterances (i.e. 32 speech stimuli in total).

The speech samples contain a great range of variation that is generally found in Spanish NNS accents of English (cf. Coe 2001), such as

- Variation in vowel length and vowel quality

- Variation in specific consonants (e.g. pronunciation of English /z/ towards /s/ or /□/; /b/ towards /v/ or /β/; /r/ towards [□] and [r]; /h/ towards /x/)

- Regular sentence rhythm, which often leads to vowels being produced as full vowels (instead of the weak forms) in unstressed syllables

- Strong devoicing of final consonants

- Narrower pitch range, leading to what is often perceived as a 'flat' intonation

The four speakers in this study show all of the above types of variation but differ in the extent to which this variation surfaces in their accents. In an auditory analysis, the accent of Speaker 1 showed a greater amount of the above features than any of the other three accents. The accent of Speaker 4 showed relatively few of these features while the accents of Speaker 2 and Speaker 3 ranked in-between the two.

Speech samples were played in silent conditions over headphones using a Praat (Boersma and Weenink, 2013, version 5.3.42) script on a laptop. The samples were arranged in two different randomisation patterns to address possible influences of

adjacent speech samples on the responses and also to avoid samples of the same speaker appearing in direct succession.

The listeners were asked to transcribe exactly what they heard. They could listen to each sample only once and had to write down what they heard immediately after each file was played. The task was self-paced to give sufficient time for the transcriptions. Prior to the actual task the listeners did a short familiarisation session with sentences which were not used in the study. In addition, the listeners filled out a questionnaire which asked for additional information on the participants' background and familiarity with languages other than their L1 and familiarity with NS and NNS accents of English. As expected, all Spanish listeners were more familiar with Spanish-accented English than the German and English listeners.

V. RESULTS

The 32 speech stimuli were transcribed by all 25 listeners resulting in 800 transcribed sentences. These were coded against a transcription of the speech stimuli (which were carefully transcribed by the researcher and checked by a colleague). For each transcription the number of correctly transcribed words was counted; spelling mistakes were not considered to be incorrect transcriptions in cases where the listener had clearly identified the intended word. Intelligibility scores were obtained by calculating the percentage of correctly transcribed words in each stimulus (following Derwing and Munro 1997); mean scores for each participant group were calculated for each sentence and across all stimuli of each speaker. The mean scores for stimulus transcriptions ranged from 16% (lowest) to 100%, where listeners at B1 level usually achieved the lowest value (in stimulus 1 from Speaker 1, Spanish B1 listeners scored 24% as the lowest value and German B1 listeners 16%) while the lowest value for listeners at C2 level was 24% for the German listeners and 43% for the Spanish listeners. Each group managed to transcribe a number of stimuli 100% correct. Across speakers, the lowest value for correct transcriptions was 47% (for the Spanish B1 group's transcriptions of Speaker 1's utterances) and the highest was 94% (Spanish C2 group's transcriptions of Speaker 3's utterances).
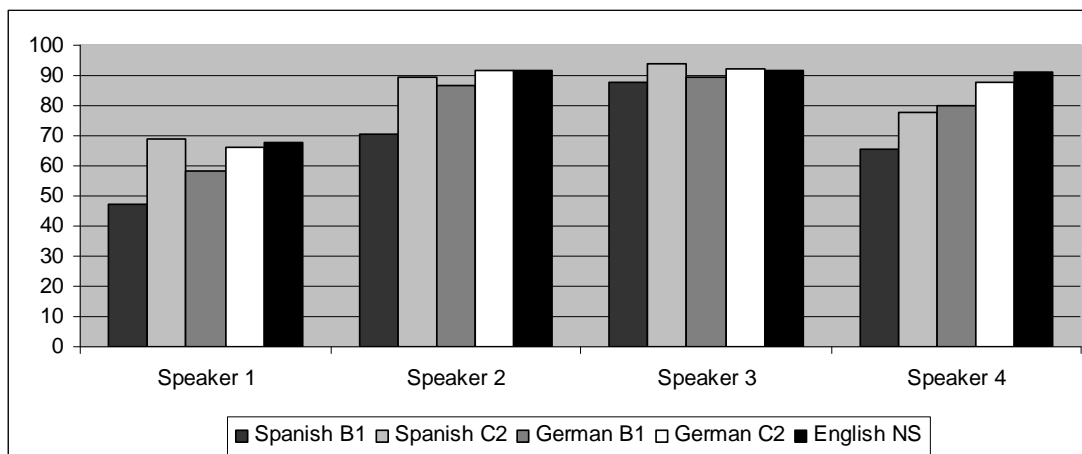
Figure 1: Mean percentages of correct transcriptions across Speakers.

The percentage of correctly transcribed words is lowest for Speaker 1 (47% correct transcriptions for the Spanish B1 group) and highest for Speaker 3 (94% for the Spanish C2 group). All participant groups seemed to have understood Speaker 3 very well, the above values are almost at ceiling level for this speaker across all groups (between 88% and 94%; see figure 1). The main differences between listener groups are evident in the results for Speaker 1, 2 and 4, where in general listeners at B1 level achieved fewer correct transcriptions than the corresponding listeners at C2 level, and for Speaker 4 the results of the German B1 listeners are level with those of the Spanish C2 listeners. According to the results, German listeners made more correct transcriptions of the utterances of Speaker 2 and 4 than the corresponding Spanish listeners; for Speaker 1, the Spanish C2 level listeners made more correct transcriptions than the German C2 listeners, but the Spanish B1 listeners transcribed fewer words correctly than the German B1 listeners. The number of correct transcriptions of the English NS is always very similar to that of the German C2 group. However, this general trend is only partially confirmed by a one-way ANOVA (with the factor 'listener-group' and the dependent factor 'speaker') and a post-hoc Tukey test, which reveal significant differences between the following pairs (see table 2).

Table 2: Significant differences in correct transcriptions (non-significant results are not reported).

	Speaker 1	Speaker 2	Speaker 3	Speaker 4
Spanish B1	Spanish C2 (p = 0.017) German C2 (p = 0.044) English (p = 0.024)	Spanish C2 (p = 0.01) German B1 (p = 0.03) German C2 (p = 0.00) English (p = 0.00)	Spanish C2 (p = 0.09)	German B1 (p = 0.05) German C2 (p = 0.002) English (p = 0.001)
Spanish C2	Spanish B1 (p = 0.017)	Spanish B1 (p = 0.01)	Spanish B1 (p = 0.09)	None
German B1	None	Spanish B1 (p = 0.03)	None	Spanish B1 (p = 0.05)
German C2	Spanish B1 (p = 0.044)	Spanish B1 (p = 0.00)	None	Spanish B1 (p = 0.002)
English	Spanish B1 (p = 0.024)	Spanish B1 (p = 0.00)	None	Spanish B1 (p = 0.001)

The results listed in table 2 indicate that the only statistically significant differences occur between the Spanish B1 group and other groups, which suggests that the other groups transcribed the stimuli equally correctly. A closer look at the transcriptions reveals that the lower number of correct transcriptions for the Spanish B1 listeners is influenced by a comparatively high number of blanket statements such as 'I did not understand a word'. In quite a few cases listeners from this group seemed to have not understood the utterance at all or they may have given up because the utterance appeared to be too difficult to understand. This did not happen as much with the German B1 listeners who by and large at least managed to understand a few words, although in this group there are also some cases where no word has been transcribed or where all transcribed words were incorrect.

It was not always possible to determine what lead to misunderstandings of certain words or whole utterances. However, there are a number of items which give some idea of what may have been the main issues involved. For example the word "catch" in the utterance "cat trying to <u>catch</u> a fish" (Speaker 1, stimulus 3) was misunderstood to mean "cut" by eight listeners (two Spanish B1, 2 Spanish C2, one German B1 and three German C2 listeners) or "cat" by three listeners (one Spanish B1, one Spanish C2 and one German B1 listener). This indicates that the final voiceless affricate was not perceived as such by these listeners, and indeed, the speaker placed more emphasis on the initial part of this affricate, though the final fricative is still audible. It also appears that in this case the vowel was perceived as a more open vowel by those who

understood "cut". In fact, in this example many listeners opted to transcribe a word which would actually make sense in this context; i.e. "trying to *cut* a fish" makes more semantic sense than "trying to *cat* a fish". Interestingly, only one German listener (from the C2 group) was able to identify the word correctly and all English NS transcribed "catch" correctly. All other listeners either transcribed "cat" or "cut" or did not attempt to transcribe the word.

Another word which stands out as difficult to perceive and transcribe is the word "beaver" in the utterance "the <u>beaver</u> got a fish in his hands" (Speaker 2, stimulus 1). All listeners were able to correctly transcribe this sentence except for the word "beaver" which was transcribed as "river" by four listeners (one Spanish B1, two German B1 and one English NS listener) and as "weaver" by one German C2 listener. Other transcriptions were nonsense entries (such as "beaber" or "viver", usually with an added question mark); five listeners correctly identified the word "beaver" (two Spanish C2, two German C2 and one English NS listener). All the other listeners did not attempt to transcribe this word. Misunderstandings were apparently caused by the speaker's pronunciation of the /b/ sounds which varied towards a bilabial fricative, which explains why some listeners perceived it as a /v/ or /w/. Again, some listeners opted for the closest known word (such as "river" and "weaver") in their transcriptions. An almost identical pattern emerged in sentence 1 of speaker 3 "the white cat is looking at the <u>beaver</u>", where all listeners were able to transcribe the sentence correctly but for the final word, which was transcribed as "river", "viver", "weaver" or similar.

Further combinations of factors caused misunderstandings, most notably final devoicing in combination with variation in vowel quality. In stimulus 2, speaker 2 "he suddenly <u>hugged</u> the hedgehog", the word "hugged" is transcribed as "hack" by eight listeners (three Spanish C2, one German B1, four English NS listeners) and as "hacked" by three listeners (one Spanish C2 and two German C2 listeners); one German C2 listener was undecided between "hack" and "hug", six listeners transcribed the word correctly (two Spanish B1, one German B1, two German C2 and one English NS listener). In this utterance, the speaker devoiced the final consonant cluster /gd/ and produced the vowel slightly more fronted.

Another noteworthy case is stimulus 8 by speaker 2 "and then I was <u>living</u> for three years", where the word "living" was transcribed as "leaving" by three German B1 and three German C2 listeners. All other listeners transcribed this word correctly, indicating that the variation in vowel length in this item was only problematic for the German listeners. However, in another instance of variation in vowel length (in stimulus 4 of speaker 3 "the cat is <u>hitting</u> the beehive") eight listeners transcribed "hitting" as "heating" (two Spanish B1, two Spanish C2, two German B1, one German C2 and one English NS listener). In addition, six listeners transcribed this word as "hid in" (one Spanish C2, one German B1, one German C2 and three English NS) which is due to the speaker's voicing of the intervocalic consonant. Another instance of vowel variation which lead to misunderstandings was in stimulus 8, speaker 4 "the box was full of <u>bees</u> and they start chasing the hedgehog". Here the word "bees" was transcribed as "beers" by seven listeners (two Spanish B1, one Spanish C2, two German C2 and two English NS listeners) and as "bears" by four listeners (two Spanish C2, one German B1 and one German C2 listener). One listener (German C2) was undecided between these two options and wrote down both, four listeners (Spanish C2, German B1, German C2 and English NS) transcribed the correct word, all other listeners did not attempt to transcribe this word.

VI. DISCUSSION

The main aim of this study was to investigate whether there are any differences between learners at different CEFR proficiency levels in order to determine whether the proficiency levels as set out in the CEFR correspond to actual differences in the performance of language learners. There are uncertainties about the validity of assumed learner progression from one level to the next (cf. Hulstijn 2007) especially with regard to speech perception and production. The results of the present study, however, reveal that learners of English at B1 level had in general more difficulties understanding the speech stimuli than learners of English at C2 level, which was especially visible in the results for speaker 1, 2 and 4. Interestingly, these differences were only statistically significant for the Spanish listener groups and not for the German groups (a finding which is in itself interesting and should be investigated further). Nevertheless, these

results confirm that there is indeed a difference between these two proficiency levels with regard to speech perception at least in some L1 groups. Considering that the development of second/foreign language speech production and perception are considered to be closely interconnected (cf. Lacabex, García Lecumberri and Cooke 2008), this result confirms that some of Hulstijn et al.'s (2011) findings on pronunciation skills may also be applicable for speech perception in some contexts.

In the transcriptions, some items stood out since many participants had difficulties understanding them. These items were analysed more closely to find general trends of what might cause intelligibility issues between learners at different proficiency levels. The main finding here is that many Spanish B1 listeners did not transcribe a particular stimulus because they did not understand a single word of the stimulus (as many of these listeners later reported or noted on their transcription sheets). It is also likely that they may have perceived the utterance as too difficult to understand and thus did not attempt to transcribe anything, or 'gave up'. It is quite possible that these listeners may have been able to transcribe at least some part of these stimuli had they been given the opportunity to listen to them again. After all, 'repetition' of speech is included in level A2 of 'phonological control' (i.e. speech production; Council of Europe 2001) and though it does not appear in the CEFR section on "General phonetic awareness and skills" there is no reason why it should not be part of level descriptions for speech perception. There were no further findings regarding a possible influence of specific patterns on the intelligibility between proficiency levels, however, this does not necessarily mean that there is none. Bearing in mind that the speech stimuli for the present study were based on natural speech in order to account for the kind of accents these learners are likely to encounter in everyday life, it is possible that speech stimuli which are controlled for specific features may reveal perceptual differences. However, based on the present study, it is not possible to show any general trends of what might constitute differences in speech perception and intelligibility between learners at B1 and C2 proficiency level other than that Spanish B1 listeners were more likely not to understand and transcribe entire stimuli.

The transcriptions and especially those of the difficult items were further analysed to reveal any differences between the Spanish and German learners of English. Based on the CEFR's account of "General phonetic awareness and skills" (Council of Europe

2001) it is tempting to predict that the Spanish listeners – who are more familiar with Spanish accents than the German listeners, as revealed by the background questionnaire – would generally be better able to understand the Spanish accents because these accents may include more familiar sounds and prosodic patterns. And indeed, in one stimulus the word "living" was misunderstood to mean "leaving" by most German speakers but not by the Spanish listeners. This may indicate that Spanish listeners are more likely to be able to process variation in vowel length as compared to German listeners (who would not normally have this feature in their accent). However, a different stimulus containing the same feature only in a different word ("hitting") was misunderstood to mean "heating" by both Spanish and German listeners. Clearly this aspect would be worth further investigation; based on the present results, however, it appears that sharing the same L1 was not an advantage in understanding the speakers. In fact, for the stimuli spoken by speakers 2 and 4, the German listeners made fewer incorrect transcriptions than the Spanish listeners. However, for speaker 1, the Spanish C2 listeners transcribed more items correctly as compared to the German C2 group, while the Spanish B1 group made fewer incorrect transcriptions than the German B1 group.

VII. CONCLUSION

This paper was made up of two parts; in the first part it presented a brief overview of previous research into the Common European Framework of Reference for Languages (CEFR; Council of Europe 2001) and gave an account of the representation of speech development in the CEFR. In the CEFR, speech production is covered by the level descriptions of "phonological control" which focuses mainly on intelligibility and accentedness of the speaker as perceived by native speakers of the language (cf. Council of Europe 2001: 117) and is otherwise rather vague. Speech perception, however, is not included in any detail and only some possible features concerning speech perception are included in the section on "General phonetic awareness and skills" (ibid.: 107). This was followed by a review of previous research into speech perception which pointed out the main areas that would benefit from further investigation in relation to the CEFR; i.e. the supposed linear progression between proficiency levels as described in the CEFR,

but not confirmed by research findings, and the close connection between speech perception and production, which is included to a limited extent in the CEFR section on "General phonetic awareness and skills". In addition, this paper briefly explored the potential of the CEFR to act as an alternative way of describing and evaluating learner speech without necessarily having to resort to native speaker norms and targets. Given that there is a rising interest in such an alternative (cf. Fitzpatrick and Racine 2013) and considering that native speaker norms are not always suitable in language learning contexts (cf. e.g. Cook 1999, He and Zhang 2010) it is well worth exploring this possibility.

The second part of this paper was devoted to an exploratory study of speech perception by learners of English at two different CEFR proficiency levels, with ten German learners of English (five at proficiency level B1 and five at proficiency level C2) and ten Spanish learners of English (five at proficiency level B1 and five at proficiency level C2). In addition, five English native speakers took part as a control group. These participants listened to speech samples based on recordings of four Spanish NNS of English with varying degrees of influence from their L1 Spanish in their English pronunciation. The participants were asked to transcribe each stimulus exactly as they heard it. The transcriptions were then analysed for the number of correctly transcribed words and for any specific sounds and patterns which may have been difficult to understand.

The study shows that some of the assumptions made in the CEFR are valid and correspond to learners' differences at the two levels of proficiency; for example, listeners at C2 level consistently transcribed more sentences correctly than the corresponding B1 level group (though this was only statistically significant for the Spanish listeners). While this study only looked at these two levels, which are quite different in terms of the learners' proficiency in the language, the results show that it would be worth looking at differences between adjacent levels, e.g. C1 and C2, which would also confirm whether the assumption that C1 and C2 are the same (as depicted in "Phonological control") holds in speech perception.

While this study did not find any particular patterns regarding what was difficult to understand for B1 vs. C2 learners and for German vs. Spanish learners, it must be born

in mind that the stimuli were naturalistic and geared towards assessing overall intelligibility. However, there were some items which proved difficult for many listeners and these may be a suitable starting point for further investigation. In general, what this study shows is that intelligibility is a suitable measure with which to investigate differences in the perception of accents between speakers at different proficiency levels.

In this paper, I hoped to show that it is crucial to conduct research into the representation of speech development in the CEFR, for the benefit of learners, language practitioners, researchers and the further development of the CEFR. The results from the study, though limited, have revealed that there are indeed several aspects which are worthy of further investigation in the area of speech perception. It is hoped that the beginning wave of research into the CEFR and especially its proficiency level descriptions will encourage further research within this framework in all areas of language acquisition research.

Notes

[1] The author would like to thank two anonymous reviewers for their invaluable and inspiring comments on an earlier abstract of this paper.

REFERENCES

Alderson, J.C., Figueras, N., Kuijper, H., Nold, G., Takala, S., and Tardieu, C. 2006. "Analysing tests of reading and listening in relation to the common European framework of reference: The experience of the Dutch CEFR construct project". *Language Assessment Quarterly: An International Journal*, 3 (1), 3-30.

Anderson Hsieh, J., Johnson, R., and Koehler, K. 1992. "The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure". *Language Learning*, 42 (4), 529-555.

Audacity software, Version 2.0.3 (August 2013). Open Source Software http://audacity.sourceforge.net/

Breton, G., Bolli, G.G., and Perlmann-Balme, M. 2010. "All different all equal? Towards cross-language benchmarking using samples of oral production in French, German and Italian". *ForumSprache*, 2010 (4), 5-19.

Canagarajah, A., and Wurr, A. 2011. "Multilingual communication and language acquisition: New research directions". *The Reading Matrix*, 11 (1), 1-15.

Coe, N. 2001. "Speakers or Spanish and Catalan". In Swan, M. and M. Smith (Eds.) *Learner English*. Cambridge: Cambridge University Press, 90-112.

Cook, V. 1999. "Going beyond the native speaker in language teaching". *TESOL Quarterly*, 33 (2), 185-209.

Council of Europe 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. 09 August 2011 <http://www.coe.int/t/dg4/linguistic/cadre_en.asp>

Derwing, T.M. and Munro, M.J. 1997. "Accent, intelligibility, and comprehensibility". *Studies in Second Language Acquisition*, 19 (1), 1-16.

Evison, J. 2013. "A corpus linguistic analysis of turn-openings in spoken academic discourse: Understanding discursive specialisation". *English Profile Journal,* 3, e4.

Fitzpatrick, T. and Racine, J.P. 2013. "Using L1 performance profiles as an alternative to native speaker norms". Unpublished poster presented at BAAL 2013, September 2013.

Flege, J.E. and Bohn, O.S. 1989. "An instrumental study of vowel reduction and stress placement in Spanish-accented English". *Studies in Second Language Acquisition* 11, 35-62.

Flege, J.E. 1995. "Second language speech learning. Theory, findings, and problems". In Strange, W. (Ed.) *Speech Perception and Linguistic Experience*. Baltimore: York Press, 233-277.

Fulcher, G. 2004. "Deluded by artifices? The common European framework and harmonization". *Language Assessment Quarterly: An International Journal*, 1 (4), 253-266.

Hancke, J., and Meurers, D. 2013. "Exploring CEFR classification for German based on rich linguistic modelling". *Learner Corpus Research Conference*, September 2013, Book of Abstracts, 54-56.

Harding, L. 2013. "Pronunciation assessment". In Chapelle, C.A. (Ed.) *The Encyclopedia of Applied Linguistics*. Oxford: Blackwell, 4708-4713.

Hawkins, J.A. and Buttery, P. 2010. "Criterial features in learner corpora: Theory and illustrations". *English Profile Journal*, 1, e5.

Hawkins, J.A. and Filipović, L. 2012. *Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework* (Vol. 1). Cambridge University Press.

He, D. and Zhang, Q. 2010. "Native speaker norms and China English: From the perspective of learners and teachers in China". *TESOL Quarterly*, 44 (4), 769-789.

Hulstijn, J.H. 2007. "The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency". *The Modern Language Journal*, 91 (4), 663-667.

Hulstijn, J.H., Schoonen, R., de Jong, N.H., Steinel, M.P. and Florijn, A. 2012. "Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of Reference for Languages (CEFR)". *Language Testing*, 29 (2), 203-221.

Krumm, H.J. 2007. "Profiles instead of levels: The CEFR and its (ab) uses in the context of migration". *The Modern Language Journal*, 91 (4), 667-669.

Kuhl, P.K. 1993. "Innate predispositions and the effects of experience in speech perception: the native language magnet theory". In de Boysson-Bardies, B., S. De Schonen, P. Jusczyk, P. McNeilage and J. Morton (Eds.) *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. Dordrecht: Kluwer, 259-272.

Kusseling, F. and Lonsdale, D. 2013. "A corpus-based assessment of French CEFR lexical content". *Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 69 (4), 436-461.

Lacabex, E.G., García Lecumberri, M.L. and Cooke, M. 2008. "Identification of the contrast full vowel-schwa: training effects and generalization to a new perceptual context". *Ilha do Desterro* 55, 173-196.

Lecumberri, M.G. and Gallardo, F. 2003. "English FL sounds in school learners of different ages". *Age and the Acquisition of English as a Foreign Language*, 115-35.

Levis, J.M. 2005. "Changing contexts and shifting paradigms in pronunciation teaching". *TESOL Quarterly* 39 (3), 369-377.

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 (4), 645-655.

Major, R.C., Fitzmaurice, S.F., Bunta, F. and Balasubramanian, C. 2002. "The effects of non-native accents on listening comprehension: Implications for ESL assessment". *TESOL Quarterly*, 36 (2), 173-190.

McCarthy, M. 2010. "Spoken fluency revisited". *English Profile Journal*, 1, e4.

Milton, J. 2010. "The development of vocabulary breadth across the CEFR levels". In Bartning, I., M. Martin and I. Vedder (Eds.) *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research*, 211-232. 28 November 2013 < http://eurosla.org/monographs/EM01/EM01tot.pdf >

Munro, M.J., Derwing, T.M. and Morton, S.L. 2006. "The mutual intelligibility of L2 speech". *Studies in Second Language Acquisition* 28, 111-131.

Phakiti, A. and Roever, C. 2011. "Current issues and trends in language assessment in Australia and New Zealand". *Language Assessment Quarterly*, 8 (2), 103-107.

Piske, T., MacKay, I.R.A. and Flege, J.E. 2001. "Factors affecting degree of foreign accent in an L2: a review". *Journal of Phonetics* 29, 191-215.

Praat – Doing Phonetics by Computer. Version 5.3.42 (March 2013). Paul Boersma and David Weenink. University of Amsterdam <http://www.fon.hum.uva.nl/praat/>

Scovel, T. 1981. "The effects of neurological age on nonprimary language acquisition". In Andersen, W. (Ed.) *New Dimensions in Second Language Acquisition Research*. Rowley, MA.: Newbury House, 33-42.

Swan, M. and Smith, B. 2001. *Learner English*. Cambridge: Cambridge University Press.

Tannenbaum, R.J. and Wylie, E.C. 2008. Linking English-language test scores onto the Common European Framework of Reference: An application of standard-setting methodology. Princeton, NJ: Educational Testing Service. 28 November 2013 <http://origin-www.ets.org/Media/Research/pdf/RR-08-34.pdf>

Timmis, I. 2002. "Native-speaker norms and International English: a classroom view". *ELT Journal*, 56 (3), 240-249.

Trim, J.L.M. 2010. "The Modern Languages Programme of the Council of Europe as a background to the English Profile Programme". *English Profile Journal*, 1, e2.

Weir, C.J. 2005. "Limitations of the Common European Framework for developing comparable examinations and tests". *Language Testing*, 22 (3), 281-300.

Westhoff, G. 2007. "Challenges and opportunities of the CEFR for reimagining foreign language pedagogy". *The Modern Language Journal*, 91 (4), 676-679.

Motivation and constraints of illocution in the lexical constructional model: the case of the *Aux NP* construction

Nuria del Campo Martínez
nuria.delcampo@hotmail.es
Universidad de La Rioja, Spain

ABSTRACT

This article addresses the motivation and constraints of illocutionary meaning production. Within the framework of the *Lexical Constructional Model* (LCM), I explore how our knowledge of illocution is understood in terms of high-level situational models which are activated to produce speech act meaning and the way such operations motivate the conventionalized value of linguistic expressions. In so doing, I analyze the realization procedures of the *Aux NP* construction in relation to their potential to exploit the semantic base of requestive acts. I will study the most conventional linguistic realizations of the construction and explore the way in which such realizations are used to produce a requestive meaning. The resulting account provides a comprehensive understanding of the constructional nature of illocutionary meaning on the basis of naturally occurring data.

Keywords*: Illocution, cognitive models, conceptual metonymy, conventionalization, idiomatic construction, requestive speech acts, Lexical Constructional Model.*

I. INTRODUCTION

The existence of conventional speech acts was first dealt with by Searle (1975) early in the development of speech act theory. While pragmaticists have generally neglected the conventionalization of illocution (Leech 1983; Sperber and Wilson, 1995, *inter alia*), the systemic-functional approach (Halliday 1994; Halliday and Matthiessen 2004) and Dik's (1989, 1997) functional account have devoted a great effort to formalize the value of conventional speech acts. In general terms, functional grammar theories have argued for sentence types as codified carriers of basic illocutions, the remaining resulting from derivation processes or from the language options. Alternatively, the cognitive linguistic approach has accounted for illocution in terms of metonymically grounded inferential schemas which become conventionalized through usage (Pérez 2001; Pérez and Ruiz de Mendoza 2002; Panther and Thornburg 2003; Stefanowitsch, 2003; Ruiz de Mendoza and Baicchi 2007; Brdar-Szabó 2009). Conventional illocutions have been discussed as

constructions (i.e. form-meaning pairings, like those described by Lakoff 1987; and Goldberg 1995, 2006) that have entrenched speech act values.

On the grounds of the observations on the constructional nature of speech acts, the *Lexical Constructional Model* (LCM) (Ruiz de Mendoza and Mairal 2008, 2011; Mairal and Ruiz de Mendoza 2009; Butler 2009) has incorporated illocution as part of a meaning construction system. The LCM, which draws insights from functional models of language, Cognitive Linguistics and constructionist approaches, especially from the work by Goldberg (1995, 2006), is concerned with the connections between syntax and all aspects of meaning construction, positing four levels of representation: level 1 deals with lexical and constructional argument structure, level 2 with implicated meaning captured by low-level models, level 3 with conventionalized illocutionary meaning and level 4 with discourse aspects, including cohesion and coherence phenomena. Each of the levels is either subsumed into a higher configuration or acts as a cue for the activation of relevant conceptual structure that yields an implicit meaning derivation. The integration of lower-level structures into higher-level ones is regulated by two cognitive processes, constructional subsumption and cued inferencing. Constructional subsumption is the constrained incorporation of lower level structures into higher level configurations. Cued inferencing is a form of linguistically guided interpretation based on cognitive operations such as metaphor, metonymy, reinforcement and mitigation, among others. The LCM aims at the highest possible degree of explanatory adequacy, insofar as it avoids the proliferation of analytical categories. Instead, it assumes that all levels of linguistic description and explanation may make use of the same or at least comparable cognitive processes. This assumption is termed the *equipollence hypothesis*, which has enabled the model to achieve a high degree of regularity and parsimony in the study of meaning construction. Several linguistic processes have been attested to be pervasive in different levels of meaning construction, such as lexical-constructional integration, subsumption, metaphor and metonymy and inferential activity.

The illocutionary component of the LCM treats constructions as form-meaning pairings like other kinds of construction. What distinguishes illocutionary constructions from the others is the idiomatic nature of the linguistic form and the situational generic grounding. Constructions with an illocutionary meaning have also been dealt with at the

layer of argument structure, as in the case of the manipulative subjective-transitive construction (e.g. *I want you out by lunchtime*) studied by Ruiz de Mendoza and Gonzálvez (2010). Even though the LCM has not provided an inventory of illocutionary constructions, its explanatory apparatus is consistent with the descriptions developed by Ruiz de Mendoza and Baicchi (2007). In their approach, illocutionary constructions are discussed in terms of the metonymic activation of high-level scenarios in application of a number of socio-cultural conventions stipulated within a description labeled the *Cost-Benefit Cognitive Model*. This article develops the illocutionary layer of the LCM by analyzing how cognitive models are exploited by speakers to produce speech act meaning and the way such operations motivate the conventionalized illocutionary value of linguistic expressions. In so doing, it analyzes the cognitive grounding of the *Aux NP* requestive construction and its various realization procedures. On the basis of the LCM notion of situational meaning, this work formulates a generic structure for requestive acts and examines the reasoning schemas behind the different lexico-grammatical resources used for their expression.

The understanding of illocution in terms of the constructional realizations that activate pieces of knowledge makes necessary to provide a refined description of the cognitive model types involved and of all the mechanisms that take part in meaning derivation. This is not only for the LCM account but also for other cognitively-oriented theories where illocutionary expression is considered realizational of semantic structures. This will be made apparent by a brief revision of the shortcomings presented by cognitive approaches to illocution. Then it will be shown how these shortcomings are overcome within the constructionist perspective of the LCM, which has been preliminary outlined by Ruiz de Mendoza and Baicchi (2007). It will be further explained the cognitive model types underlying implicit meaning derivation and how the activation of high-level scenarios yields illocutionary acts which may become conventionalized. The analytical tools proposed by the LCM will be used to study the conventional and non-conventional realizations of the *Aux NP* construction and the way such realizations produce requestive illocutions, giving evidence of their explanatory adequacy.

II. THE COGNITION OF ILLOCUTION

Within the framework of Cognitive Linguistics, Panther and Thornburg (1998: 756) have addressed illocution by pointing to the problems that the lack of consideration of the cognitive mechanisms has caused in inferential approaches, which are, first, the fact that, even though illocutionary interpretation is based on inference, speakers can grasp the indirect force of a speech act effortlessly (e.g. the request value of *Could you pass me the salt?*); and second, that they ignore the inference mechanisms involved in the interpretation of illocution as well as their cognitive grounding. In order to overcome these two shortcomings, Panther and Thornburg (1998, 2004) propose that our knowledge of illocutionary meaning is organized in the form of scenarios, which are conceptual constructs of meaning representation abstracted away from prototypical situations where people attempt to get their needs satisfied through expressions of different kinds. Illocutionary scenarios are stored in long-term memory and can be accessed metonymically by activating relevant parts in them. For example, indirect requests such as *Can you open the window?*, *Will you shut the door?* and *Do you have hot coffee?* activate pre-conditions for the performance of a request, which are the addressee's ability and willingness to help, and his possession of the required object. The activation of these pre-conditions affords access to the whole speech act category of requesting.[1]

The key elements that make Panther and Thornburg's proposal interesting from a cognitive perspective are storage in long-term memory and metonymic instantiation. Their formulation has been revised, however, due to the lack of consideration of socio-cultural variables that affect inferencing. These variables are listed by Ruiz de Mendoza and Baicchi (2007: 103) as the following: (i) the power relationship between speakers, (ii) the degree of optionality conveyed, (iii) the degree of politeness, (iv) the degree of cost-benefit, (v) the degree of prototypicality, (vi) the semantic motivation of different kinds of indirect speech acts, and (vii) the cognitive grounding of illocutions. In Ruiz de Mendoza and Baicchi's proposal, following preliminary work by Pérez and Ruiz de Mendoza (2002), socio-cultural variables of this kind are captured by cognitive models that combine with scenarios and form what they call high-level situational models.

High-level models are constructed on the basis of generalizations over cases of everyday interaction where people attempt to satisfy or report their needs. Everyday interaction is captured by low-level situational models, which consist in life scenarios such as taking a taxi, going to the dentist, teaching a class, and the like (see Ruiz de Mendoza, 2007, for a thorough description of cognitive model types). The activation of low-level scenarios produces implicated meaning. An example is provided by *I waved down a taxi*, where the waving sign implies that the speaker got into the taxi, he asked the driver to take him to the destination, and that he arrived safely. The implicature is obtained through the metonymic access to one relevant part of a low-level model about taking a taxi. The abstraction over the common structure shared by low-level models allows us to construct higher-level representations. For instance, from our observation of people begging in a wide range of contexts, we derive generic structure which makes up the high-level model of begging and allows us to interpret each specific instance.[2] In contrast to low-level models, high-level models capture a number of socio-cultural generalizations that carry different types of pragmatic information like optionality, politeness and cost-benefit variables. These variables derive from a single description called the *Cost-Benefit Cognitive Model*, which is defined by Ruiz de Mendoza and Baicchi (2007) as a high-level model based on the concept of mutual manifestness proposed by Sperber and Wilson (1995). The *Cost-Benefit Cognitive Model* captures the relevant socio-cultural information of high-level scenarios associated to illocutionary meaning. Let us reproduce Ruiz de Mendoza and Baicchi's (2007: 111) formulation of the *Cost-Benefit Cognitive Model* in order to explain how it underlies the construing of illocutionary meaning:

(a)   If it is manifest to A that a particular state of affairs is not beneficial to B, and if A has the capacity to change that state of affairs, then A should do so.

(b)   If it is manifest to A that a potential state of affairs is not beneficial to B, then A is not expected to bring it about.

(c)   If it is manifest to A that a potential state of affairs is beneficial to B, then A is expected to bring it about provided he has the capacity to do so.

(d)   If it is manifest to A that it is not manifest to B that a potential state of affairs is (regarded as) beneficial for A, A is expected to make this manifest to B.

(e)   If it is manifest to A that it is not manifest to B that a potential state of affairs is beneficial for B, A is expected to make this manifest to B.

(f)   If it is manifest to A that a state of affairs is beneficial to B and B has brought it about, A should feel pleased about it and make this feeling manifest to B.

(g)   If it is manifest to B that A has changed a state of affairs to B's benefit, B should feel grateful about A's action and make this feeling manifest to A.

(h)   If it is manifest to A that A has not acted as directed by parts (a), (b), and (c) of the 'cost-benefit' model, A should feel regretful about this situation and make this feeling manifest to B.

(i)   If it is manifest to B that A has not acted as directed by parts (a), (b), and (c) of the 'cost-benefit' model and A has made his regret manifest to B, B should feel forgiveness for A's inaction and make it manifest to A.

(j)   If it is manifest to A and B that a particular state of affairs is not beneficial to B but A has no power to change it to B's benefit, A should still feel sympathy for B over the non-beneficial state of affairs and make this manifest to B.

(k)   If it is manifest to A that A is responsible for a certain state of affairs to be to A's benefit, A may feel proud about this situation and make it manifest to B.

Ruiz de Mendoza and Baicchi place the *Cost-Benefit Cognitive Model* at the core of the derivation of illocutionary meaning. In their view, the activation of relevant parts of the model creates an inferential path which can become conventionalized. A case in point is the *Can You VP?* sequence for requests, which was originally intended as a way of reminding the addressee to help if it was within his range of abilities. This value was obtained through the activation of part (c) of the *Cost-Benefit Cognitive Model*. This convention structures the high-level model of requests and shapes their definitional parameters (i.e. optionality and politeness). The repeated use of the *Can You VP?* expression in request contexts conventionalized their meaning to the extent that it ended up yielding a default illocutionary value. Conventional forms of this kind have constructional status, that is, they are the formal part of form-meaning pairings conveying an illocutionary act. By contrast, those expressions which are unable to supply relevant points of access to the convention that shapes the conceptual

representation of a speech act category require inferential activity to produce illocutionary meaning, which is regulated through metonymic access to high-level models, much in the same way as Panther and Thornburg (1998, 2004) have claimed. The theoretical implications of constructional conventionalization in terms of the application of socio-cultural norms are approached at a later stage in this article.

III. ILLOCUTIONARY CONSTRUCTIONS

Since Searle (1975) acknowledged that certain linguistic forms became conventionally accepted for the performance of an indirect illocutionary force, the research on the issue has received a great deal of attention. One of the most important contributions to the subject has been carried out by functional grammar theories (Dik 1989, 1997; Halliday and Matthiessen 2004), which have been largely criticized for overgrammatizalizing illocutionary phenomena which could be accounted for within the domain of pragmatics (see Leech 1983: 56; Butler 1996: 66, for criticism in this direction). Nevertheless, although the emphasis placed within functional approaches on the grammatical side of language lacks of consideration of inferential reasoning, this position has managed to incorporate into grammar a number of relevant illocutionary distinctions which had been assigned to pragmatics. The development of a constructional approach like the one put forward in this study comes closer to the work by other functionalists like Risselada (1993), who disagrees with the idea that grammatical mood codifies basic speech act types, given the wide variety of illocutionary meanings that each mood option has, and rather suggests assigning a certain illocutionary value to each sentence type and counting them as reference points. Risselada's (1993: 74) approach to illocution is based on the assumption that the illocutionary force of speech act types is expressed by means of combinations of the linguistic properties that reflect the characteristic features of the speech act involved. In its most explicit form, an utterance expresses all the essential features of a speech act category. Implicit utterances, by contrast, are due to pragmatic variables such as power or politeness or to the fact that the shared background knowledge provides speakers with the necessary information to derive their illocutionary value.[3] Even though Risselada does not explicitly talk about constructions in her account, her proposal covertly points to a constructional view of illocution. Her

pairings of formal properties of speech acts with illocutionary meaning cannot be regarded otherwise. This theory is in line with cognitively-oriented approaches such as the one put forward by Ruiz de Mendoza (1999), later developed in Ruiz de Mendoza and Baicchi's (2007) work. Ruiz de Mendoza's notion of *specialization of function*, like Risselada's *degree of explicitness*, defines the ability of a given expression to activate a higher or lower number of meaning conditions of a speech act type.

Following Risselada's and Ruiz de Mendoza's insights and working within Cognitive Linguistics, Pérez (2001), puts forward an approach to illocution according to which the meaning conditions of speech act categories were paired with the linguistic means through (i.e. realization procedures) which they were communicated.[4] The notion of illocutionary construction posited by Pérez refines Risselada's and Ruiz de Mendoza's work in two aspects. In the first place, Pérez extends the concept to include linguistic properties such as sentence type, grammatical resources, lexical elements and suprasegmental features. This is quite an advantage, since they serve to increase the level of specialization of an expression to convey an illocutionary force. The type of illocutionary construction put forward in the present study also captures the array of properties proposed by Pérez. The second refinement is that the semantic makeup of illocutionary constructions is accounted for in terms of propositional ICMs specifying the meaning conditions of a speech act category. In this way, Pérez views constructions as pairings of form and function, where form consists in realization procedures capable of activating the semantic variables of an ICM. The higher the number of variables that are activated by a realization procedure, the more prototypical the realization is for the expression of an illocutionary act. However, Pérez does not refer to these realizations as constructions with fixed and modifiable elements.

In contrast to Pérez, Ruiz de Mendoza and Baicchi (2007) do recognize a constructional status for those formulations with instantiation potential for the corresponding scenario. As has been explained above, these authors contend that expressions which become entrenched as inferential shortcuts acquire a constructional character. The *Can You VP?* construction mentioned before is a case in point. Ruiz de Mendoza and Baicchi's proposal regards constructions as conventionalized linguistic forms whose capacity to activate parts of a scenario becomes conventionalized. Such a conception of the term is

similar to the one put forward by Pérez, although differs from the latter in two essential aspects. The first difference concerns the semantic structure of the construction, structured in the form of cultural high-level models. As has been explained, high-level models are conceptual representations of abstract knowledge of illocutionary meaning. The formulation of high-level models to account for illocution seeks to capture the multi-faced amount of information that speakers possess during communication. Later it will be shown that the description of illocutionary acts in terms of high-level models attains a greater degree of explanatory adequacy. The second difference has to do with the formal composition of illocutionary constructions. Ruiz de Mendoza and Baicchi consider all the realization procedures specified by Pérez (e.g. grammatical resources, lexical items, intonation, etc.), with the difference that they are arranged into stable structural configurations. A description of illocutionary acts in terms of Ruiz de Mendoza and Baicchi's constructions is very attractive for the following reasons: (i) it accounts for the motivation of form from meaning, and (ii) it makes it possible to build into grammar a wide range of illocutionary values.

The LCM elaborates on the view of illocution that has been proposed by Ruiz de Mendoza and Baicchi. Illocutionary constructions are conventionalized linguistic forms whose capacity to activate parts of a high-level scenario becomes conventionalized. The LCM approach, however, refines their account by placing a stronger emphasis on the constructional composition of illocution and regard illocutionary constructions as form-meaning pairings made up of fixed (*Can You* in *Can You VP?*) and modifiable (*VP* in *Can You VP?*) elements. The fixed elements cannot be changed without altering the meaning implications conveyed and the variable elements can be parametrized in a constrained way. Constructions may also incorporate further elements with a wide range of meaning implications (e.g. the adverb *please* or beneficiary indicators in the case of requests). Illocutionary constructions may also have to a degree of variation in their form with a consequent variation in their meaning. Sequences like *Could You VP?* and *Do You Think You Could VP?* are variations of the *Can You VP?* construction. The meaning variation of these constructional variants is associated with degrees of indirectness and politeness as well as differences in register. Because of this, the LCM accounts for constructions that have elements in common in terms of family resemblance relationships (Ruiz de Mendoza and Gonzálvez 2010).[5] A case in point is

the *Aux NP* construction, which can be realized in many different ways to produce requestive speech acts. The constructional realizations of the *Aux NP* form are analyzed to explore their grounding in the conventions of the *Cost-Benefit Cognitive Model* and the interplay between linguistic structures instantiating requests and their conceptual motivation. Throughout the analysis I provide evidence in support of the LCM approach to illocution in terms of high-level scenarios and conventional constructions.

IV. REQUESTIVE SPEECH ACTS

Requestive speech acts ask other people to act in the way we want them to. Requestive illocutions ranges over many diverse acts like asking, ordering or begging. Before we go into the differences among these values, it should be noted that they are all included by Ruiz de Mendoza and Baicchi (2007) within a broad category of illocutionary acts that instruct the addressee to act to the speaker's benefit. Let us consider the conventions of the *Cost-Benefit Cognitive Model* that structure the cognitive grounding of requestive acts:

> *If it is manifest to A that a particular state of affairs is not beneficial to B, and if A has the capacity to change that state of affairs, then A should do so.*

> *If it is manifest to A that a potential state of affairs is not beneficial to B, then A is not expected to bring it about.*

Even though requestive illocutions are understood against the same socio-cultural background (that we have to satisfy other people's needs), they are distinct in nature. We should first differentiate ordering from requesting and then requesting from begging. What distinguishes ordering from requesting has to do with the ratings of the power variable. In orders, speakers hold a position of authority over their addressees. Because of this authority, the speaker who utters an order works under the expectation that the addressee will carry out the action. The addressee's lack of optionality to decide upon the realization of the action triggers off the required response. This is not the case with requests, which are performed by speakers who do not have any kind of authority over their addressees. However, this does not mean that the addressee's optionality is unconstrained, since his choice is restrained by the conventions that bind him to help the

speaker if it is within his range of abilities. The power component that makes orders different from requests has led Ruiz de Mendoza and Baicchi to address these categories as distinct illocutions. By contrast, requesting is considered within the same category as begging, in spite of presenting important differences. Unlike requests, in begging the speaker believes that the addressee is not desirous to give him what he wants and adopts a submissive role to obtain the addressee's compliance. This distinction is manifested through different constructional realizations. While requests tend to use mitigators or beneficiary indicators, beggings use repetitions and exclamations. Nevertheless, acts of requesting and begging display the same cost-benefit ratings and are considered within the same category. In keeping with Ruiz de Mendoza and Baicchi's account, this study of requestive speech acts agrees in differentiating orders from requests based on the power variable and also in considering begging as a special form of requesting.

The present analysis will only consider the illocutionary acts contained within the category of requesting. Although the *Aux NP* form can be found in the performance of orders, as will be shown in next section, the meaning conditions of the construction are directly tied to the semantics of requesting and needs to be approached in relation to requests. To see how the various linguistic realizations of the construction express requestive values, it is necessary to define the meaning conditions that make up the generic structure of this illocutionary category. Hence I will put forward a high-level scenario for requests by generalizing over the features of requesting scenarios grounded in the two conventions of the *Cost-Benefit Cognitive Model*. We derive the generic structure of requests from everyday situations where we attempt to get our needs satisfied by other people.[6] Two possible low-level scenarios for requesting encompass a situation in which a person in a needful situation makes somebody else aware of his ability to help and a situation in which a person is asking for help while pretending he is not in need. These low-level scenarios have elements in common upon which the high-level scenario may be constructed. This generic structure captures the semantics of the act of requesting:

(a)   A person is in need of something.

(b)   The person makes somebody else aware of the need.

(c)   The person makes this other person aware of his ability to help.

(d)   The person appeals to the addressee's willingness.

(e)   The addressee may be persuaded to help.

The realizational resources for this scenario may be exemplified in the following utterances:

(1)   I am thirsty.

(2)   Maybe I could have a glass of water.

(3)   Can you give me a glass of water?

(4)   Would you give me a glass of water?

(5)   You will give me a glass of water, won't you?

The above realizations instantiate relevant parts of the scenario formulated for requests. Utterances (1) and (2) point to the manifestness of the needful situation in which the speaker finds himself. Utterances (3) and (4) address the addressee's ability and willingness to satisfy the speaker's need respectively. These examples are instances of the *Aux NP* construction parametrizing the meaning value with different degrees of mitigation. To finish with, utterance (5) spells out that the addressee should be willing to help in compliance with socio-cultural conventions. As will be shown in the next section, the use of various realization procedures in requests displays peculiarities in meaning that reveal different forms of construing a shared conceptual representation.

V. THE *AUX NP* CONSTRUCTION

The *Aux NP* construction is probably the most conventional form for the performance of requests. The formal part of this construction consists of an auxiliary plus a second person subject and a variable verb. The high-level scenario for requests constitutes the semantic base of the construction. This scenario is a manifestation of the conventions of the *Cost-Benefit Cognitive Model* according to which speakers should be willing to help others if it is within their range of abilities. The requestive meaning of the construction was originally derived by means of an inferential schema giving access to these conventions, and has become conventionalized through usage. Let us see how this meaning value is parametrized through various realization procedures.

V.1. The *can you* realization

The different realization procedures of the construction are primarily related to the auxiliary verb, as the subject pronoun is almost invariable.[7] The auxiliary is realized by a modal verb.[8] Modal verbs capture the relations between participants and the realization of the state of affairs in which they are involved (Dik 1989: 205). They include distinctions related to ability and willingness and also to the obligation or permission imposed on participants. One of the most recurring modals used is the form *can*, mostly due to the fact that the parameter of the addressee's ability is relevant to requests. In application of the conventions of the *Cost-Benefit Cognitive Model*, the *can you* form asks the addressee about his capacity to do something for the speaker. Asking the addressee about his capacity to act makes him aware that he is indeed able to carry out the action and reminds him that he is culturally bound to act if he has the ability to do so. In most contexts, this realization gives easy access to the high-level scenario, which is then applied to the specific situation. However, there may be cases where this procedure does not fit to be used as a request. By way of illustration, consider the following examples:[9]

(6)   Can you *see* into the future? (Google Books)

(7)   Can you *smell* the flowers? (Coca)

(8)   Can you *drive* a truck? (Google Books)

(9)   Can you *speak* German? (Bnc)

Utterances above are cases of the construction that function as mere questions.[10] This is due to the parametrization of the variable verb, which needs to be realized by an action-controlling denoting action involving some kind of benefit to the speaker in order to yield a request interpretation. In (6) and (7), the verb denotes a non-controllable activity, which is incompatible with the nature of requesting. In (8) and (9), the verb designates a controllable action but there is no indication of the potential benefit to be obtained by the speaker. These utterances could only be interpreted as requests in marked contexts where it is clear that the speaker is interested in getting the action carried out and that the performance of the action involves some benefit to the speaker. The fact that the

action is beneficial to the speaker is generally made explicit through the use of a beneficiary indicator. The instances of the construction featuring this characteristic convey an easy request value:

(10) Can you bring *me* my purse? (Google Books)

(11) Can you get *me* a drink? (Coca)

(12) Can you write down a recipe *for me*? (Google Books)

A similar effect is achieved through the use of mitigating devices, which have the function of softening the directive force of the request or of urging the addressee to act in the way described:

(13) Can you *please* give me a second? (Coca)

(14) Can you *kindly* open the door? (Google Books)

The request interpretation is coded here by the interpersonal adverbs *please* and *kindly*, whose function is that of increasing the degree of politeness. The mitigation brought about by resources of this kind is motivated by the need to soften the impact of the request by increasing the degree of the addressee's optionality. Granting someone with optionality is regarded as a sign of politeness in our social system and optionality and politeness are thus closely intertwined. Even higher degrees of politeness can be achieved through the use of other mitigating strategies, like the replacement of *can* for *could*. Past modals increase the indirectness of requests, thereby offering the addressee a greater degree of optionality to comply with the speaker's wishes (see Taylor, 1995, and Pérez, 2001, for an explanation of the mitigation of past modals in cognitive terms).[11] The following examples illustrate this:

(15) Could you pass me the sugar? (Google Books)

(16) Could you complete the questionnaire for me? (Google Books)

Utterances (15) and (16) above display the highest degree of specialization as realization procedures for requests. First, because the past form of the modal does not only point to the addressee's ability to carry out the action but also to his willingness by giving him optionality. This activates one further variable of the scenario: mitigation. Thanks to the mitigating properties of past modal verbs, these two examples manage to

make this aspect of requests explicit. And second, because the speaker's interest in getting the action carried out is conveyed through beneficiary indicators (i.e. *me* and *for me*). The instantiation of these parts of the high-level scenario makes the interpretation of these utterances as instances of requesting straightforward. The degree of mitigation conveyed by the past form of *could* can be further increased with the addition of the adverb *please*. There are occasions on which higher degrees of mitigation are required in the performance of a request. Consider situations in which the cost of the requested action is significant, as in (17), or in which the context of the utterance is formal, as in (18):

(17)　Could you *please* hurry home and watch the children for me? (Google Books)

(18)　Could you *please* bring me a cup of hot coffee? (Coca)

The diverse mitigation strategies found in (17) and (18) give rise to subtle formal realizations which, by activating a higher number of variables of the scenario, constitute even more specialized procedures.

V.2. The *will you* realization

Another common way of parametrizing the auxiliary verb of the *Aux NP* form is through the use of the modal *will* appealing to his willingness to act to the speaker's benefit. Through application of the conventions of the *Cost-Benefit Cognitive Model*, the addressee should be willing to perform an action to the speaker's benefit. The *will you* form enquires about the addressee's willingness to act. In unmarked contexts, this realization procedure yields a preferred conventional request interpretation, but it may function as a question:

(19)　Will you find true love? (Google Books)

(20)　Will you ever go back to the world of business? (Coca)

Likewise, this type of realization could be used to perform different speech acts like advising and offering. This is so because the modal *will* is affected in various ways depending on the conditions that apply in each particular interaction. The following are some examples of such a situation:

(21)   Will you buy that car? It's such a beauty. (Coca)

(22)   Will you eat some more cake? (Google Books)

For this realization procedure to produce a request reading, the specified action has to be beneficial for the speaker. This information can be clear from the context or made explicit through beneficiary indicators. Its explicitation obviously results in more codified instances of requesting. Observe how the manifestness of this part of the high-level scenario increases the degree of specialization to the extent that it is not possible to interpret utterances as instances of a different speech act:

(23)   Will you lend *me* money? (Coca)

(24)   Will you buy *me* a pencil set for Christmas? (Bnc)

The impact of the resulting request can be mitigated through the use of *please*. This adverb generally indicates that the speaker seeks a benefit from the realization of the action, but it may occasionally have the opposite effect. In some cases, the adverb may produce forceful demands by implying that the addressee should have acted as required without being told to do so. This use of *please* is reinforced by an imposing falling intonation. This type of intonation is often used by people who have some kind of authority over their addressees. Compare the different uses of the adverb in (25) and (26) below:

(25)   Will you hold the door open for me, *please*? (Google Books)

(26)   Will you *please* bring me my back my bag? (Coca)

As was the case with the previous type of realization, the request meaning can be further specified by means of a past form. It has already been explained that the past tense displays a mitigation that seems appropriate for the politeness that is expected in the performance of requests. By increasing the addressee's optionality, the use of the form *would* softens the force of the act and points with increasing certainty to a request interpretation:

(27)   Would you drive me to the station? (Bnc)

(28)   Would you give me a hand with the washing up? (Google Books)

Because of its instantiation potential for the mitigation that is proper of requests, the *would you* sequence represents a highly conventionalized procedure. Needless to say that the combination of this type of realization with the adverb *please* results in even more polite and thus adapted instances of requesting as in the following examples:

(29)   Would you please tell me where the library is? (Google Books)

(30)   Would you please pass the steak sauce? (Coca)

As may become apparent by (29) and (30), the use of these resources increases the degree of politeness of this realization that fits best in formal contexts where there is a distant relationship between participants.

V.3. Negated modals

The use of negated modals is another type of realization procedure of the construction under scrutiny, although its request meaning is less explicit than in the previous cases. The reasoning schema behind this realization affords metonymic access to the parts of the scenario where the speaker appeals either to the addressee's ability or willingness to comply, but the negated form of the modal presupposes the addressee's refusal, which gives rise to unmitigated requests marked by their impoliteness. Let us see how this meaning is conveyed through the negated form of *can* in the examples:

(31)   Can't you behave properly? (Coca)

(32)   Can't you wipe your feet on the rug? (Google Books)

Through application of the conventions of the *Cost-Benefit Cognitive Model*, the addressee should have acted as required without being asked to do so. Since in normal circumstances, the speaker would expect that the addressee has the ability to act, he enquires about any unexpected inability on the part of the addressee to carry out the action. In unmarked contexts, this realization procedure has a strong power to activate the directive scenario, particularly because it makes explicit the speaker's expectation that the addressee has the ability to perform the action. The request interpretation of the construction can be cancelled out uttered in a marked context where the addressee is not abided to do anything about the situation described (cf. *Can't you hear the whistle blowing?*).

We may find a related realization procedure making use of the negated form of *will*. In this case, the verb used does not assume the addressee's inability but rather his unwillingness to comply with the speaker's wish. See how this type of realization gives rise to a request:

(33)   Won't you sit quiet? (Coca)

(34)   Won't you close the window? (Google Books)

In (33) and (34), the addressee has not carried out the required action counter to expectations and the speaker enquires about any unexpected unwillingness on this part. This realization procedure produces a request interpretation by reminding the addressee that he is abided to act by socio-cultural conventions. The sequence can be nonetheless function as a question (cf. *Won't you buy clothes online anymore?*) in contexts where the addressee is not expected to act. Realizations with negated modals can be performed as well by means of the imperative sentence type. Take the case of the following examples:

(35)   Calm down, can't you? (Coca)

(36)   Hurry up, won't you? (Google Books)

In contrast to interrogative-based realizations, the use of imperative sentences indicates irritation on the part of a speaker who is urging the addressee to act. The resulting request is thus more forceful and the optionality of the addressee is notably reduced.

V.4. Conditional forms

Conditional forms are recognized as a conventional pragmatic mitigator of directive values (see Dancygier and Sweetser, 2005 and Fauconnier, 1985, among others). In the case of requests, the use of the conditional tense is meant to distance the addressee from the required action. This opens up the degree of addressee's optionality, which reduces the force of the act by increasing the indirectness of the request. The most common ways of using a conditional in the construction are the following:

(37)   Would you mind if I use your bathroom? (Coca)

(38)   Would you mind handing me that book over there? (Google Books)

Utterance (37) is a request for permission. These differ from prototypical cases of requesting in that both the speaker and the addressee are expected to perform the action: the speaker will carry out the action that the addressee will grant permission. Requests for permission are therefore conditional. The action will be carried out only if the addressee gives his consent. This conditional character finds an adapted vehicle for expression in this realization procedure. This case differs from the one observed in (38). Utterance (38) exemplifies a request that makes use of the conditional appealing to the addressee's willingness to comply. In application of the conventions of the *Cost-Benefit Cognitive Model*, we are expected to do our best to help others and, at the same time, they expect not to be put to too great an effort in that respect. The conditional form tells the addressee that if the carrying out of the action is too costly for him, he can choose not to do it. The same realization can be used with the opposite meaning, that is, reducing the addressee's freedom by reminding him he should act as required in compliance with the principles of interaction:

(39)   Would you mind not smoking? (Coca)

In (39), the addressee is treated as if he had not realized that he is acting in a way that is negative for the speaker. The conditional form appeals to his willingness to stop the negative action in compliance with socio-cultural conventions. The resulting act is forceful and impolite. In addition to these conventional realization procedures, there are others that accommodate along a prototypical cline. Consider:

(40)   Would you be so kind as to bump up the temperature in here by a degree or two? (Coca)

(41)   Would you be so kind as to water my plants while I'm away? (Google Books)

The previous type of realization mitigated the act of requesting by increasing the addressee's optionality in relation to the cost-benefit variable. The realization procedure illustrated in (40) and (41) above, the mitigation is upgraded in relation to the politeness parameter. By enquiring about the addressee's willingness, the speaker is in fact reminding the addressee that he is bound by conventions to act if it is within his range of abilities. When the required action is presented as seeking a benefit for the speaker, the conditional softens the force of the act and functions as a mitigating device. By contrast, when the action is presented as an alternative of something negative being

done by the addressee, the conditional is used to force the addressee to consider the underlying conventions, thereby rendering an impolite act:

(42)   Would you be so kind as to remove your feet from the table? (Google Books)

In the example, the speaker treats the addressee as if he were not observing the conventions of politeness, giving rise to an impolite request that forces the addressee to act as required.

V.5. Summary of realization procedures of the *Aux NP* construction

Table 1 below provides a non-exhaustive description of the meaning conditions of the high-level scenario for requests and the ways in which they attain linguistic expression through the various realizations of the *Aux NP* construction.

Table 1. Realization procedures of the *Aux NP* construction

Request scenario	Realization procedures
Speaker's need	Beneficiary indicators (*for me*)
Speaker's willingness	Conditional forms, beneficiary indicators (*for me*)
Addressee's ability	Can you…? Could you…? Can't you…?
Addressee's willingness	Will you…? Would you…? Won't you…?
Cost-benefit ratings	Would you mind…? Would you be so kind…?
Optionality	Past modals (*could*, *would*), use of *please*, conditional forms, beneficiary indicators (*for me*)
Mitigation	Past modals (*could*, *would*), conditional forms, interpersonal adverbs (*please*, *kindly*), mild intonation and stress

VI. CONCLUSION

The present work is a case study of the constructional composition of illocutionary meaning within the LCM. The type of illocutionary constructions postulated here pair

the semantic makeup of speech acts with the constructional realizations through which they are communicated. The formal composition of constructions includes properties such as sentence type, grammatical elements, lexical properties and suprasegmental features. The meaning conditions defined in the high-level scenario include semantic variables and pragmatic features like power, politeness, optionality and cost-benefit variables. Such variables are culture-specific and their realization is related to the context of situation of each interactional exchange. High-level scenarios provide the base of a vast number of illocutionary constructions for a speech act type. The different meaning conditions of high-level scenarios are activated through diverse linguistic resources, giving rise to constructions with different degrees of codification. The higher the degree of codification of a construction, the easier it is to grasp the intended meaning and the more specialized the construction is. Conversely, if a construction is implicit but still attains important levels of effectiveness by giving access to relevant parts of a scenario, it is likely to be conventionalized for a specific illocutionary value. The process whereby constructions become conventionalized is constrained by socio-cultural conventions of the kind postulated within *Cost-Benefit Cognitive Model*. The interpretation of non-conventional constructions requires the use of inference and relies on the realization of variable elements and contextual information or shared background knowledge.

This study examines the applicability of the analytical tools put forward by the LCM to account for the various realization procedures of the *Aux NP* construction in relation to their potential to activate the semantic base of requesting speech acts. Once described the high-level scenario for the category of requesting, I have identified the different ways in which the realizations of the *Aux NP* construction provide the addressee with access to the relevant parts of the scenario. The formal composition of the construction has proved both realizational of lexico-grammatical devices and conventionally associated with them.

The analysis carried out has provided evidence in support of the LCM approach to illocution. However, the results suggest that further research on the subject is still needed. It would be advisable to develop the description of the conventions of the *Cost-Benefit Cognitive Model* in order to account for the distinctions among the various

realization procedures of different requestive acts and also to study the relationship between form and meaning among other constructions performing requests.

Notes

[1] In later work, Panther (2005) has gone further and referred to metonymy as an inference schema rather than a substitution relation or a reference point phenomenon, as has been defended by many cognitive linguists (Langacker 1993; Kövecses and Radden 1998, *inter alia*). Specifically, Panther has argued that metonymies provide natural inference schemas which are regularly used by speakers in meaning interpretation. The role of metonymy as an inference schema has been supported by later research carried out by Ruiz de Mendoza and Baicchi (2007), who identify metonymy at the base of illocutionary derivation. The present proposal adheres to such a conception of metonymy in order to account for the illocutionary meaning that derives from the activation of scenarios and which later on becomes entrenched (in Langacker's terms, 1999: 105) through a conventionalization process.

[2] In the LCM, interactional knowledge is structured in the form of situational cognitive models, to be differentiated from non-situational models. Situational cognitive models capture the interaction among entities within a specific time and place. Non-situational cognitive models include variables which are not dependent on time and place. Cognitive operations on non-situational models regulate inferred meaning at the core grammar level, yielding conversion processes and constructional alternations. Operations like metaphor and metonymy on situational models guide pragmatic inferencing (implicature derivation, illocutionary meaning and discourse connections).

[3] Risselada's (1993) definition of *explicit* and *implicit* speech acts is equivalent to the traditional distinction between *codified* and *inferred* speech acts. The degree of explicitness or codification is in both cases determined by the number of meaning conditions of the speech act under consideration which are instantiated by the linguistic form.

[4] The term *realization procedures* was first introduced by Ruiz de Mendoza and Otal (1997) to define the options offered by the linguistic system for the realization of a communicative strategy. In later work by Pérez (2001) and Ruiz de Mendoza and Baicchi (2007), this notion is used referring to entrenched lexico-grammatical devices with instantiation potential with respect to cognitive models. In the present work, realizational procedures which have become conventionalized are regarded as constructions in their own right.

[5] The notion of *family resemblance* was originally propounded by Wittgenstein (1978) to make reference to those categories whose members do not share a set of common attributes but rather display a network of similarities.

[6] For similar descriptions of requesting from a constructionist perspective, see Pérez (1996, 2001) and Ruiz de Mendoza and Baicchi (2007).

[7] It is possible, however, to find contexts in which the subject pronoun does not point to the addressee (cf. *Will he stop making noise?*). Instances of this type represent implicit requests to the addressee to get a third person to carry out the action. Except for these cases, the realization of the construction involves a second person subject (i.e. *you*).

[8] A useful accounts of modal verbs in terms of force dynamics from a cognitive perspective can be found in Talmy (1988).

[9] The description of the realization procedures of the *Aux NP* construction results from the analysis of a corpus of one hundred and sixty-five instances of the construction. The data upon which the study is based has been drawn from the original editions of the British National Corpus (BNC), the Corpus of Contemporary American English (COCA), WebCorp and Google Books.

[10] To resolve the ambiguity of the *can you* form, the LCM postulates two different constructions. One is the polar interrogative construction, which is interpreted as a question about ability and whose constituents are realized by *can you* sequence (e.g. *Can you write Morse code?*), and another that functions as a request, where *can you* is idiomatic (e.g. *Can you bring my glasses?*) (see Mairal and Ruiz de Mendoza 2009).

[11] Taylor explains the origin of the past tense as a mitigator as a cognitive process involving a double metaphorization. There is a first metaphor that structures the time domain in terms of space, as illustrated by expressions like *near future* and *distant past*, and a second metaphor that structures distance in terms of social involvement. Pérez further argues that the distance that triggers the mitigating effect has to be established both between the speaker and the speech act and between the intended speech act and the actual speech act.

REFERENCES

Brdar-Szabó, R. 2009. "Metonymy in indirect directives: Stand-alone conditional in English, German, Hungarian, and Croatian". In Panther, K.-U., Thornburg, L. and Barcelona, A. (Eds.), *Metonymy and Metaphor in Grammar*. Amsterdam/Philadelphia: John Benjamins, 323-338.

Butler, C.S. 1996. "On the concept of an interpersonal metafunction in English". In Berry, M., Butler, C.S., Fawcett and R., Huang, G. (Eds.), *Meaning and Form: Systemic Functional Interpretations*. Norwood: Ablex Publishing, 151-181.

Butler, C.S. 2009. "The Lexical Constructional Model: Genesis, strengths and challenges". In Butler, C.S. and Arista, J.M. (Eds.), *Deconstructing Constructions*. Amsterdam/Philadelphia: John Benjamins, 117-152.

Dancygier, B. and Sweetser, E. 2005. *Mental Spaces in Grammar: Conditional Constructions*. Cambridge: Cambridge University Press.

Dik, S.C. 1989. *The Theory of Functional Grammar. The Structure of the Clause*. Berlin: Mouton de Gruyter

Dik, S.C. 1997. *The Theory of Functional Grammar. Complex and Derived Constructions*. Berlin: Mouton de Gruyter.

Fauconnier, G. 1985. *Mental Spaces*. Cambridge: Cambridge University Press.

Goldberg, A.E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.

Goldberg, A.E. 2006. *Constructions at Work: The Nature of Generalization in Language.* Oxford: Oxford University Press.

Halliday, M.A.K. 1994. *An Introduction to Functional Grammar*. London: Edward Arnold.

Halliday, M.A.K. and Matthiessen, C. 2004. *An Introduction to Functional Grammar*. 3rd edition. London: Hodder Arnold.

Kövecses, Z. and Radden, G. 1998. "Metonymy: Developing a cognitive linguistic view". *Cognitive Linguistics* 9, 37-77.

Lakoff, G. 1987. *Women, Fire and Dangerous Things*. Chicago: Chicago University Press.

Langacker, R.W. 1993. "Reference point constructions". *Cognitive Linguistics* 4, 1-38.

Langacker, R.W. 1999. *Grammar and Conceptualization*. Mouton de Gruyter, Berlin.

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

Mairal, R., and Ruiz de Mendoza, F.J. 2009. "Levels of description and explanation in meaning construction". In Butler, C.S. and Arista, J.M. (Eds.), *Deconstructing Constructions*. Amsterdam/Philadelphia: John Benjamins, 153-198.

Panther, K.-U. 2005. "The role of conceptual metonymy in meaning construction". In Ruiz de Mendoza, F.J., Peña, S. (Eds.), *Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interaction*. Berlin/New York: Mouton de Gruyter, 353-386.

Panther, K.-U. and Thornburg, L. 1998. "A cognitive approach to inferencing in conversation". *Journal of Pragmatics* 30, 755-769.

Panther, K.-U. and Thornburg, L. 2003. "Metonymies as natural inference and activation schemas: The case of dependent clauses as independent speech acts". In Panther, K.-U. and Thornburg, L. (Eds.), *Metonymy and Pragmatic Inferencing*. Amsterdam/Philadelphia: John Benjamins, 127-147.

Panther, K.-U. and Thornburg, L. 2004. "The role of conceptual metonymy in meaning construction". *Metaphorik.de* 6, 91-111.

Pérez, L. 1996. "The cognition of requests". *Estudios Ingleses de la Universidad Complutense* 4, 189-208.

Pérez, L. 2001. *Illocution and Cognition: A Constructional Approach*. University of La Rioja Press, La Rioja.

Pérez, L. and Ruiz de Mendoza, F.J. 2002. "Grounding, semantic motivation, and conceptual interaction in indirective speech acts". *Journal of Pragmatics* 34, 259-284.

Risselada, R. 1993. *Imperatives and other Directive Expressions in Latin: A Study in the Pragmatics of a Dead Language*. Amsterdam: J.C. Gieben.

Ruiz de Mendoza, F.J. 1999. "La ilocución y la gramática". In Butler, C.S. Arista, J.M. and Ruiz de Mendoza, F.J. (Eds.), *Nuevas Perspectivas en Gramática Funcional*. Barcelona: Ariel, 99-171.

Ruiz de Mendoza, F.J. 2007. "High-level cognitive models: In search of a unified framework for inferential and grammatical behavior". In Kosecki, K. (Ed.), *Perspectives on Metonymy*. Frankfurt/Main: Peter Lang, 11-30.

Ruiz de Mendoza, F.J., and Baicchi, A. 2007. "Illocutionary constructions: Cognitive motivation and linguistic realization". In Kecskes, I. and Horn, L. (Eds.), *Explorations in Pragmatics: Linguistic, Cognitive, and Intercultural Aspects*. Berlin/New York: Mouton de Gruyter, 95-128.

Ruiz de Mendoza, F.J. and Gonzálvez, F. 2010. "Illocutionary meaning revisited: Subjective-transitive constructions in the Lexical-Constructional Model". In Stalmaszczyk, P. (Ed.), *Turning Points in the Philosophy of Language and Linguistics*. Frankfurt/Main: Peter Lang, 65-77.

Ruiz de Mendoza, F.J., and Mairal, R. 2008. "Levels of description and constraining factors in meaning construction: An introduction to the Lexical Constructional Model". *Folia Linguistica* 42, 355-400.

Ruiz de Mendoza, F.J. and Mairal, R. 2011. "Constraints on syntactic alternation: Lexical-constructional subsumption in the Lexical Constructional Model". In Guerrero, P. (Ed.), *Morphosyntactic Alternations in English. Functional and Cognitive Perspectives*. London/Oakville: Equinox, 62-82.

Ruiz de Mendoza, F.J. and Otal, J.L. 1997. "Communication strategies and realization procedures". *Atlantis: Revista de la Asociación Española de Estudios Anglonorteamericanos* 19, 297-314.

Searle, J. 1975. "Indirect speech acts". In Cole, P. and Morgan, J.L. (Eds.), *Syntax and Semantics 3*. New York: Academic, 59-82.

Sperber, D. and Wilson, D. 1995. *Relevance. Communication and Cognition*. 2nd edition. Malden: Blackwell.

Stefanowitsch, A. 2003. "A construction-based approach to indirect speech acts". In Panther, K.-U. and Thornburg, L. (Eds.), *Metonymy and Pragmatic Inferencing*. Amsterdam/Philadelphia: John Benjamins, 105-126.

Talmy, L. 1988. "Force dynamics in language and cognition". *Cognitive Science* 12, 49-100.

Taylor, J.R. 1995. *Linguistic Categorization: Prototypes in Linguistic Theory*. 2nd edition. Oxford: Clarendon Press.

Wittgenstein, L. 1978. *Remarks on the Fundaments of Mathematics*. Oxford: Blackwell.

BOOK REVIEW

English-Medium Instruction at Universities: Global Challenges
Aintzane Doiz, David Lasagabaster and Juan Manuel Sierra
Multilingual Matters: Bristol, 2013. 248 pages. ISBN: 9781847698148

Reviewed by **Annemieke Meijer**
A.A.Meijer@uu.nl
University College Utrecht, The Netherlands

This collection of articles is an interesting and timely addition to the growing literature on the use of English as the language of instruction at universities in non-native contexts. In the past two decades, EMI (English-Medium Instruction, as the educational practice tends to be called when referring to the tertiary level, as opposed to the more thought-through pedagogical approach of Content and Language Integrated Learning or CLIL at secondary level) has caught on in many different countries and settings. Motivated partly by the wish to attract international students, partly by the need to prepare home students for the international market, or, increasingly, with the aim of promoting the institution in an ever more competitive higher education market, universities have introduced English to replace the local language(s). Needless to say this is having immediate implications for teachers and students alike and research is badly needed on issues such as language proficiency demands, effective curriculum design, and quality assurance, not only because this may lead to important contributions to theory building but also, hopefully, because the results could feed into the decision-making processes of university administrators.

The aim of the present volume is, according to the editors, to "advance our awareness" of what is needed to improve EMI at tertiary level. It sets out to do so by providing a varied picture of current issues and practices, in contributions from eighteen authors from countries as diverse as China, Finland, Israel, the Netherlands, South Africa, Spain, and the USA. The selection may be somewhat arbitrary, but the emerging picture is highly interesting not only because of its diversity (in terms of, for instance, the societal position of English and the levels of language proficiency) but also because

unexpected connections appear and very different settings turn out to be faced with quite similar issues.

The volume has been organised into five parts which are not always very clearly delineated – some consist of just one chapter, while others contain chapters that might well have been more appropriate under the heading of one of the other sections. All contributions, however, add to the overall picture. The opening chapter is on Maastricht University in the Netherlands, a very early adopter, in which Robert Wilkinson emphasizes the crucial collaboration between 'content' departments and language specialists in the 25-year-old development of English-taught programmes there. This is followed by two case studies of the linguistic needs of students and teachers in the multilingual settings of South Africa and the Basque country respectively.

In the section on institutional policies, Taina Saarinen and Tarja Nikula use discourse analysis to study policy documents concerning language and internationalisation strategies in Finland, which has the highest number of institutions providing English-taught programmes in Europe. The authors' starting point is the apparent invisibility of language in such policies. The findings from their document analysis point at the 'self-evidence' of English, with "foreign language" often really meaning "English". They also ask the common-sensical question of "what kind of English" is referred to in lists of entry requirements, and find that apart from many departments mentioning vague criteria such as "a good command" or "a sufficient knowledge", Finnish polytechnics since 2011 have specified the setting in which prospective students should have learnt their English in such narrow terms that the qualifications of applicants from about 50 countries in which English is, in fact, an official language (India, Pakistan, South Africa) would not be accepted, thus creating "a hierarchy of different 'Englishes'". In the same section of the volume, Ofelia García, Mercè Pujol-Ferran and Pooja Reddy also make clear how language can be caught up, and become a factor in power relations. Studying a community college with immigrant students and a global research university, they describe a painful dichotomy between international and immigrant students. Using the somewhat comical (and rather poignant) abbreviation LOTE (Languages Other Than English), they conclude that "whereas the LOTEs of international students are taught, celebrated and used in academic pursuits, the use of LOTEs in education is often restricted by colleges and

universities serving immigrant students. Whereas international students are welcomed in university content classes and their English proficiency is not seen as an obstacle to learning, immigrant students are often excluded from academic content until they develop appropriate English vocabulary" (193).

It is this focus on language as a symbol of ethnic, cultural or national identity, which makes the volume such an interesting addition to the current body of literature on EMI. It informs a number of other essays in the volume, and the editors might have emphasized this in their introduction. Indeed, the book does not shy away from naming difficulties and potential negative effects of EMI, both at the level of the individual learner and the community: concerns about the loss of L1 (and 2) at the cost of English, about EMI programmes producing an "elite" whose skills and outlook sets them apart from others in society, and other sensitivities and anxieties. The buzzword in this book is "language ecology", emphasizing the role of language as a social practice, in interaction with its environment. All this comes to the forefront especially in the chapters on trilingual education, which together form the third part of the book.

The contexts of this section of the book are China and Spain, and the topics described are at times strikingly similar. David C.S. Li writes about the Chinese University of Hong Kong, where the question of whether the predominance of English in higher education should be viewed as hegemony or linguistic capital became quite urgent after a proposed increase in the percentage of English-taught courses. Local Cantonese speaking students, faced with the necessity to learn both Putonghua (Mandarin), which is the national language, and English, staged vehement protests. They feared that English would push out Cantonese and written Chinese and voiced their emotions in terms that made it very clear the issue went beyond the pragmatic intentions of the vice-chancellor who had made the proposal. Ofra Onbar-Lourie and Smadar Donitsa-Schmidt, whose contribution is mysteriously included in the section on language policies, study the same theme as Li but in the intricate linguistic scene of Israel, where English is tentatively being introduced at a small number of colleges. Because the language of instruction at almost all higher education institutions is Hebrew, Arabic speakers learn through their second language anyway, and English would be their L3. Through self-report questionnaires the views of students at one teacher training college are studied, one interesting finding being that English was not

being perceived as threat, perhaps, as the authors suggest, as a result of the unrivalled position of Hebrew in this particular educational setting.

Aintzane Doiz, David Lasagabaster and Juan Manuel Sierra investigate students' views on the introduction of English as L3 at the Basque University, where Spanish and Basque are the official languages. Again, linguistic capital turned out to play a major role. They found that "[l]ocal students show a manifest unwillingness towards being required to be proficient in English or to be proficient in two languages at the end of their studies", with Basque mother tongue students manifestly more negative than those with Spanish or Basque and Spanish as their mother tongue. Josep Maria Cots concentrates on the Catalan context with the aim of revealing "possible ambiguities and tensions" in the language policy of the bilingual University of Lleida. In this sensitive environment, the introduction of English as one of the languages of instruction met with scepticism: it is only in the past 30 years or so that there has been what the author calls a "reverse language shift" with Catalan gaining more prominence in society; moreover, there are serious fears that the university may not be ready for this, with the great majority of incoming students having an English language proficiency level of B1 or lower. The chapter illustrates the existing tensions nicely by pointing out some of the discourse surrounding the debate in which English is presented as a 'killer language' or a 'language predator', while the Catalan students are characterised as having a 'bunker attitude'.

The final word in the book is by Elana Shohamy, who presents a "critical view of EMI at university", identifying a number of educational and societal issues that may have been overlooked as universities rushed to implement English-medium programmes. It is a fitting conclusion to a collection of articles that is highly valuable because it contextualizes and because it problematizes: it places EMI firmly at the centre of a complex interplay of all kinds of socio-linguistic factors, and although it does offer ideas for overcoming some of the difficulties faced by universities, it is not simplistic or overly idealistic. Perhaps we may see it as a sign that EMI, both as an educational phenomenon and as an emerging field of study, has grown up.