Assessing Oral Production of English as a Lingua Franca in European Universities: Can Corpora Inform the Test Construct?

David Newbold (Università Ca’ Foscari Venezia, Italia)

Abstract This article examines the rapidly growing phenomenon of the use of English as a lingua franca (ELF) in European universities and the need to develop tests which reflect this use. After reporting on a research project which introduced a lingua franca element into a receptive skills test set at Levels B1-B2 of the CEFR (reflecting current entrance requirements to first and second level degree programmes), it goes on to consider the challenges posed by assessment of the productive skills, and the possible role of corpora (learner corpora and ELF corpora) in contributing to the identification of a test construct for oral production.

Summary

1 English in Higher Education in Europe

In 2013 a European linguistic taboo which had lasted since the 1999 commencement of the Bologna process was finally dropped. In a report to the Commission made by a ‘high level group on modernization of higher education’, English was named as an ‘indispensable component’ of higher education in Europe, along with a second foreign language and intercultural competences. This makes a clear break with earlier recommendations. Since the Lisbon declaration of 2000 (at least), the declared language policy of the EU had been to promote the learning of any two European languages, in an effort to acknowledge and safeguard the diversity of cultures, languages and education systems within Europe. By deliberately not naming English as the first of these two languages, the commission had avoided confronting the spectre of an English-only higher education system which had long been predicted (among others) by Phillipson (2003,

The need to rethink, or rather, clarify language policy for higher education has been prompted by the speed of change, as more and more higher education institutions throughout Europe take the decision to offer English Taught Programmes (ETPs), to promote student and teacher mobility, and to seek new sources of income. In 2008 Wächter and Maiworm reported more than 2,400 first and second level degree courses taught entirely through the medium of English. By 2011 the figure had risen to 3,700 courses being offered at Master’s Level (cfr. Brenn-White, van Rest 2012). What had begun as the preserve of northern European countries (especially Holland and Scandinavia) had begun to spread south of the Alps, with Italy and Spain offering a range of degree programmes through English; in 2011 the Polytechnic of Milan famously and controversially announced that from the next year all of its courses would be offered in English. Even France, a country which in 1994 legislated against the use of English in the workplace and stipulated that the only medium of instruction in higher education should be French (Loi Toubon) backtracked in May 2013 when MPs voted to make it possible for universities to introduce ETPs.²

But ETPs, intended primarily for incoming foreign students, are only the tip of the iceberg of the ‘Englishisation’ of European universities. Stay-at-homers, or local students who do not go on mobility programmes, are nonetheless likely to encounter English in a range of formal and less formal situations, whatever their course of study, in reading lists and while using the Internet, on their university website, at lectures given by visiting academics, or when socializing with international students; and they may also be required to write in English. It is with these students in mind that most European universities now require a minimum level of English for incoming students, usually B1 or B2 of the Common European Framework of Reference (CEFR) (Council of Europe 2001), though this may vary from one country to another, and according to the programme of study.

The requirement has brought with it the need for valid and reliable tests. Many institutions in Europe recognize international certification (such as IELTS or TOEFL, or Cambridge ESOL exams), which incoming students may have already acquired. But for those who have no certification, universities need to administer their own tests, which may be similar to those produced by the international agencies. But there is a problem of construct – the underlying competence, or competences, the test is designed to elicit – in attempting to replicate this kind of test of English as

---

a foreign language. Tests such as IELTS and TOEFL are designed for students who intend to study in an English speaking country, where they will have to interact with native speakers. In higher education in Europe, English has become a lingua franca, as the recommendation to the Commission referred to above acknowledges, and most interaction that students will have using English is likely to be with non-native speakers (whether teachers or students). A test designed to predict students’ ability to cope with the language requirements which will be made of them during their course of study should reflect this fact. In short, to have construct validity a European university entry test needs to tap into contexts of use – or target language use (TLU) domains, to use the term made familiar by Bachman and Palmer (1996) – of English as a lingua franca.

2 TEEUS Part One: a Test of the Receptive Skills with an ELF Element

This emerging role of English in higher education led to a needs analysis being carried out at the University of Venice Ca’ Foscari in 2010, with a view to developing an entrance test at levels B1 and B2 of the CEFR (cfr. Newbold 2012). Based on a questionnaire completed and returned by 275 third-year undergraduate students across the four faculties of the university (Economics, Humanities, Languages and Sciences) it showed that they had needed English primarily for reading textbooks and articles (70%), and using the Internet for research (53%), but also for watching videos (23%), attending lectures and seminars (21%) and interacting with foreign students (18%) (for a full report see Newbold 2012). Given the continued increase in student and teacher mobility in the university over the last decade, these figures are now likely to be higher.

The prototype computer-based test of reading and listening which was subsequently developed reflected these TLU domains. In the first, task-based, part of the test, students had (among other things) to make a (simulated, for test security reasons) Internet search and identify the result which was likely to be most useful to a given research area; to match course descriptions with subjects; and to retrieve information from a university website. Some of the texts used may have been written by non native speakers (NNS) – indeed, probably had been – but the source of the texts was irrelevant, given their authenticity.

When it came to selecting texts for the listening part of the test, however, a choice had to be made about whether or not to include native speakers. In the end, all texts were recorded by non native speakers, on the basis that more than 80 per cent of seminars and lectures in English at Ca’ Foscari are given by non native speakers of English. This, however, does not mean that native speakers will necessarily be excluded from further editions of the test. The texts used ranged from short semi-scripted mono-
logues and dialogues involving exchanges of information and opinions reflecting everyday university practices (such as asking about services and exchanging opinions about a lecture), to a 3 minute video of a Dutch professor introducing his course. A number of dialogues were recorded by international students (i.e. German, Italian, Indian, Israeli, Japanese, Swedish, Thai and Turkish), and the clearest (in terms of audio) and most spontaneous ones used.

Feedback from the 36 test takers suggested that the range of accents, far from being problematic, were «neither more nor less difficult to understand than native speaker accents» (23/36) or «easier to understand than native speaker accents» (9/36). The four students who claimed that the accents were more difficult than native speaker accents were the ones who scored lowest on the test overall, so this judgment may have been an attempt to justify their low scores. All test takers, however, considered the test to be either «fairly realistic» (17/36) or «very realistic» (19/36).

The test was not conceived as a ‘test of ELF’, whatever that might imply, but as a test of those receptive skills which incoming European students were likely to need. At best, we feel, it is a test with an ELF element – i.e. which includes oral and written texts produced by non native users. But this ELF element makes a fundamental contribution to the construct validity of the test.

The test is the first part of a project provisionally called TEEUS (Test of English for European University Students). The second part will involve the productive skills (speaking and writing), and could be set at level B2, which was introduced in 2013 by the University Senate as an exit requirement from the first degree cycle, as well as the level needed to enter Masters and other postgraduate programmes. The development of this part of the test poses challenges quite different from those of the first part (which ranged from text choice to software development to implementing digital literacies), as the focus shifts from test structure to assessing performance.

3 TEEUS Part Two: Spoken Production and Spoken Interaction

This paper is primarily concerned with the challenges posed by a test of speaking in a European university environment. Whereas academic writing is well served by a body of research and learner corpora which have been used to inform courses in EAP and learner dictionaries, and for which referring to a default, native speaker, standard variety of the language is, for the moment at least, still largely uncontroversial (cfr. Jenkins 2014, p. 208), many researchers into EAP, such as Swales and Feak (2004, p. 1), point out that the distinction between NS and NNS would be better replaced by the notion of a community of ‘expert users’ of the language, especially at the level of postgraduate research. Monologic production is likely to be more formal and prepared, but it is no longer the preserve of the teacher/
lecturer. Student presentations given in English in seminars and lessons, and not only on ETPs, have become a feature of academic life, especially at postgraduate level. Part of the rationale behind the practice is to prepare students for real-life tasks, the transmission of information in a knowledge-based society (as envisaged in the Lisbon Strategy of 2000), through the medium of English lingua franca.

Tasks such as these sit well within the B2 descriptors of the CEFR. For example the ‘low B2’ level for ‘Addressing Audiences’ within the category of ‘Spoken Production’ is:

Can give a clear, prepared presentation, giving reasons in support of or against a particular point of view and giving the advantages and disadvantages of various options. Can take a series of follow up questions with a degree of fluency and spontaneity which poses no strain for either him/herself or the audience. (p. 60)

‘Spoken interaction’ is divided into eight domains, most of which would be relevant for the proposed test (‘conversation’, ‘informal discussion’, ‘goal-oriented co-operation’, etc.). The overall descriptor for ‘spoken interaction’ at ‘low B2’ level reads:

Can interact with a degree of fluency and spontaneity that makes regular interaction, and sustained relationships with native speakers quite possible without imposing strain on either party. Can highlight the personal significance of events and experiences, account for and sustain views clearly by providing relevant explanations and arguments. (p. 74)

In terms of underlying skills typical of B2 level this can do statement, especially in its reference to supporting opinions through relevant arguments, seems appropriate for an academic setting. What is problematic is the reference to native speakers, and the notion of strain in native speaker-non native speaker (henceforth NS-NNS) interaction. References to native speakers occur frequently in the CEFR descriptors, and they get in the way of developing a test whose TLU domains do not include, or include only marginally, native speakers. What can cause strain in NS (native speaker) – NNS (non-native speaker) interaction (such as non standard formal features in the NNS, and low tolerance of those features by the NS) may on the contrary in an ELF (English as a lingua franca) interaction generate strategies which lead to successful communication. But at this point we need to look more closely at the nature of ELF communication, to be able to make decisions about what we want to assess.
4 Describing English as a Lingua Franca

Much of the debate associated with the development of ELF as a research field over the last decade has concerned the place of formal features in the description of ELF and the extent to which form «follows function» (Cogo 2008). Although much work has been done to identify recurring features of ELF, whether phonological, such as the core phonology proposed by Jenkins (2000), or lexicogrammatical (cfr. Seidlhofer 2004; Bjorkman 2008; Ranta 2009), no claim has been made that ELF is an emerging variety of English. For Seidlhofer (2011, p. 88) such a claim would be «meaningless and irrelevant» given that (a) there are no native speakers of ELF and (b) ELF communication transcends physical and linguistic boundaries, much of it taking place over the Internet. As such, it is best seen (at least in Europe\(^3\)) as «a liberating additional means of communication» whose «fluidity and flexibility […] strengthens the communicative robustness of cultural interactions» (Seidlhofer 2011, p. 80).

This is a useful caveat for language testers, who traditionally occupy the role of gatekeeper for a standard variety of the language. It offers a way of conceptualizing the recurring morpho-syntactic and phonological features which ELF researchers have isolated, not as constituting a simplified code or variety, or as providing evidence of deviations from a native speaker norm (and therefore errors, sanctionable in an assessment), but rather, as a strategic use of linguistic resources available to the speaker to co-construct meaning with the listener.

Most of the time ELF discourse seems very much like native speaker English (cfr. Meierkord 2004, p. 128; Mauranen 2010, p. 6). When they do occur, the non-standard forms which are typical of ELF communication are not usually an impediment to understanding – a fact that ELF researchers are quick to point out. They range from removing or adding redundancy (such as omitting the third person s or adding the preposition about after to discuss), to simplification (a single tag isn’t it? or no?), to transforming mass nouns into count nouns, such as evidences and accommodations,\(^4\) and in so doing staking out an extra, useful, level of meaning. Lexical creativity is also frequent, in the form of morphological overgeneralization (yielding patterns such as increasement, examinate), category shift (to precise), or semantic shift (actor to refer to anyone who does something, especially in an organization). The last two examples are taken from A brief list of misused English terms in EU publications, a list of 88 words compiled by

---

3 The case of English as a lingua franca in postcolonial contexts, where it might function as a simplified code alongside nativized varieties which sit at the top of a dialect continuum, has been problematized by Canagarajah (2007) who has opted for the term Lingua Franca English.

4 Accommodations is standard in American English.
EU translator Jeremy Gardner (2013), published by the EU Translation Directorate. Gardner shows how these words are commonly used in European circles in a manner deviant from native speaker norms (and therefore anathema to the native speaker), but at the same time he provides evidence (or evidences) of processes guiding the evolution of ELF in the non native speaker context of the EU.

Lexical inventiveness, of which the above appear to be examples, plugs gaps to meet immediate communicative needs, and is a necessary part of successful ELF interaction. Probably more important, both in frequency and impact, are the accommodation strategies used by speakers to co-construct meaning with their interlocutors. These include repetition, paraphrase, code-switching, and the use of non-verbal language, all of which crucially imply being sensitive to the listener’s needs, and adapting to them. In recent years, much descriptive research has focused on these strategies (e.g. Dewey 2007; Cogo 2009; Pitzl 2010; Basso 2012).

It seems reasonable, therefore, that a construct for speaking in an ELF context should allow both for the production of non standard forms, and accommodation strategies, within a framework of meaning co-construction.

5 Learner Corpora, ELF Corpora

The findings referred to above come mainly from the ELF corpora which have been built up over the past decade, in particular the VOICE project, housed at the University of Vienna, and the ELFA corpus of academic English at the University of Helsinki, as well as smaller scale corpora which individual researchers (cfr. Cogo, Dewey 2006; Prodromou 2008) have assembled. Unlike native speaker and learner corpora, they focus almost exclusively on spoken data, and as such pose quite different problems of collection and annotation from written corpora. They are organized differently, too. Learner corpora highlight deviance from a native speaker norm, to inform remedial teaching materials, academic writing modules, and the like, such as the 50 page section on improving writing skills which is a strong feature of the Macmillan English Dictionary for Advanced Learners (2007). Native speaker corpora, depending on the use made of them, might be divided into macro text types or genres, such as ‘conversation’ ‘academic prose’, ‘fiction’ or ‘news’ (which are the four register categories referred to in Longman (1999) Grammar of Spoken and Written English.

Of course, not all learner corpora are confined to written data. The LINDSEI corpus (Louvain International Database of Spoken English Interlanguage), for example, currently contains around 1 million words from speakers of eleven mother tongues. Data is elicited through a series of activities - a warm-up phase, in which informants choose a topic to talk about, a free discussion phase, and the description of a picture. This pro-
procedure offers corpus users a degree of comparability between learners, and (probably) a good sample of communicative functions. However, the activities all mirror familiar classroom, and, especially, testing, tasks. This is appropriate because learner corpora, by definition, provide photographs of learners at a specific stage in the learning process. In contrast, ELF is concerned with how English is used outside the language classroom; it has users, not learners, who co-construct meaning with interlocutors, by exploiting whatever resources are available to them. This means that information about interlanguage is likely to be of little use in the compilation of a test of proficiency in ELF. As Seidlhofer (2011) puts it, ELF is:

not a kind of fossilized interlanguage used by learners failing to conform to the conventions of Inner Circle native norms, but a legitimate use of English in its own right, an inevitable development of the globalization of the language. (2011, p. 24)

ELF corpora are thus ‘context aware’. They are organized by complete speech event, such as ‘lecture’ ‘meeting’, or ‘service encounter’ (cfr. Seidlhofer 2004, p. 219; Mauranen 2010, p. 12). VOICE and ELFA are of comparable size (each approximately one million words); both are organized according to speech event, and both use mainly speakers of European languages, in European settings, making them relevant to our test development purposes.

The immediate context of the speech event is crucial, to understanding the dynamics of ELF communication, and to this end both corpora, especially VOICE, provide paralinguistic information about what the speaker is doing (‘points at student’, ‘pours coffee’, ‘yawns’, ‘snorts’), thereby offering a sort of photograph of the event – although, inevitably, it is tantalizingly incomplete.

To tell us more about the communicative success of ELF users, and about the strategies they use, small scale studies in which the researcher has a personal knowledge of speakers offer qualitative insights. These include Podromou’s (2008) corpus of SUEs (successful users of English), the competent users analysed in Cogo, Dewey (2006), and the international students who were the subject of Basso’s (2012) study. Basso, for example, on the basis of 100 hours of recordings and interviews with students at Venice International University, identifies paraphrasing as the most useful strategy, followed by slowing down, repetition, and use of non-verbal language from a perceptive viewpoint, while from a productive viewpoint resorting to non-verbal language moves into second place behind paraphrasing.

Presumably most non-verbal language will elude annotation in the bigger data bases, but other strategies, such as paraphrasing, emerge, particularly in academic contexts. In the ELFA corpus, of which 33% is monologic, notably in the form of lectures and presentations, Mauranen (2012)
isolates a number of focusing devices including left dislocation (e.g. ‘This argument, we shall return to it later’), rephrasing, and meta-discourse, i.e. when the speaker reminds listeners of what he or she is trying to do. This last strategy resonates with accepted wisdom about structuring presentations (regardless of the native language of the speaker), such as telling an audience what you are going to speak about, then telling them, and then telling them what you have told them.\(^5\)

Of course, it may not always be clear whether these features are used primarily to promote comprehension, or as self-help strategies (such as buying time), or both, but this is irrelevant if in both cases the result is to prevent a breakdown of communication, and to keep it fluid and flexible. In terms of the presentation we have in mind to test oral production (see Section 3 above), focusing features could be incorporated into an assessment framework, as examples of strategies, rather than as evidence of dysfunctional non-standard language forms as the speaker gropes for the right words. Other strategies emerge in dialogic interaction, such as code-switching and other-repetition – in which it is the interlocutor who repeats a word or phrase which has just been said – (cfr. Cogo 2009, p. 270). These provide a useful reminder that the assessment framework for oral interaction has to include all participants in the communicative event, not just the one for whom an assessment is required, thus posing specific challenges for the assessor.

### 6 Towards a Test Construct for Assessing Oral Skills in an ELF Context

The first, and possibly greatest challenge in the development of a test of ELF, will be the abandonment of a NS standard. Tests need to be anchored to descriptions, which are in their turn based on standards, to the extent that standards «become statements of test constructs» (McNamara 2012, p. 199). McNamara goes on to suggest that the articulation of a construct for a test of ELF will pose challenges similar to those facing the theorizers of communicative language testing at the start of the 1980s.

The analogy is useful. Communicative language testing is (or should be) performance-based, and focused on successful communication, rather than on the language forms by which communication is (or is not) achieved; it embraces pragmatics and thus socio-linguistic skills, and in the well known framework proposed by Canale and Swain (1980) it includes *strategic competence*, which allows the speaker to plug the gaps when communication

\(^5\) See for example the guide to making academic presentations produced for students by the University of Southampton *Getting your point across*. Available at [http://www.learnwithus.southampton.ac.uk/academicSkills/pdfs/getting_your_point_across.pdf](http://www.learnwithus.southampton.ac.uk/academicSkills/pdfs/getting_your_point_across.pdf) (retrieved: 2013-09-29).
breakdown looms, using verbal and non-verbal repair strategies. All this is reminiscent of the strategies which emerge in ELF interaction, some of which we have referred to above.

Does this then mean that a test of ELF could look like an existing ‘communicative’ test of speaking? Elder and Harding (2008) suggest that the paired interaction format, which has become a familiar feature of institutionalized tests such as the Cambridge ESOL suite, offers a useful alternative to traditional one-on-one interviews in which there is an a priori imbalance of power (cfr. Elder, Harding 2008, par. 34.8), and in which the interviewer (in institutionalized high stakes tests) is often a native speaker.

The problem lies thus perhaps not so much in the test format (for oral interaction), but in the assessment criteria. In the Cambridge ESOL suite speaking is assessed for:
- Grammar and Vocabulary;
- Discourse Management;
- Pronunciation;
- Interactive communication.

The first three criteria appeal to traditional notions of accuracy (grammar and pronunciation), fluency (discourse management) and lexical range (vocabulary), which are assessed as if they were stable properties of speaker competence revealed in individual performance. Only the last criterion is (partially) concerned with the co-construction of meaning. In a test of oral interaction in an ELF context, by contrast, both parties have a shared responsibility in the communicative act which, as McNamara and Roever (2006, p. 46) memorably put it, is «a dance in which it makes no sense to isolate the contributions of the individual dance partners». We believe that a set of more interlocutor-oriented criteria, informed by ELF research and the empirical data provided by corpora, could be developed to replace more traditional notions of proficiency, so that:
- accuracy → appropriateness;
- fluency → flexibility;
- lexical range → lexical transparency;

where appropriateness might refer to strategies (such as accommodation) by which speakers align themselves to interlocutors, flexibility could include making use of a range of available resources, such as code-switching and extra-linguistic resources, which might be penalized in traditional tests, and lexical transparency which could involve the avoidance of unilateral idiomaticity as well as lexical creation.

The second great challenge in developing an ELF test, after defining the construct(s), will then be to develop a workable rating scale which maps onto the levels of the CEFR – since the CEFR (itself a product of the ‘communicative revolution’) has become the new gatekeeper to European...
universities. In essence, this will mean rewriting the CEFR level descriptors for spoken interaction and removing explicit and implicit references to a default native speaker model.

Spoken production appears to be less problematic than spoken interaction since the test task (‘making a presentation’) in our proposed TEEUS test closely reflects the TLU. It is also a markedly academic genre, familiar to teachers and to students, and within which presentations need to be structured, arguments and counter-arguments presented, and conclusions reached. Without the immediate feedback or support of an interlocutor, which are fundamental to interaction, the assessor’s focus is here likely to turn to self help strategies (e.g. self repair, paraphrasing, buying time, exploiting visual materials) as being fundamental to success – a fact which seems to be recognized by students who themselves have to go through the (for some) nerve-racking ordeal of making presentations before a class. In a survey of Masters’ level students asked to identify skills and strategies required to make a successful presentation, all 19 respondents recognized the ability to cope in a crisis as «important» (12/19) or «crucial» (7/19); a reminder, perhaps, that engaging with ELF means being constantly alert to communication breakdown, and veering a course to avoid it.

The rapid growth of English medium instruction, and the ubiquitous student presentations which it has spawned, have made the development of this part of the test a priority. An initial phase of the project will consist of drawing up a framework of features, rhetorical and paralinguistic, which contribute to a successful presentation, and which may range across such diverse elements as:
- Voice control;
- Transparent language;
- Effective repair strategies;
- Relating to audience;
- Use of visuals;
- Body language;
- Clear signposting.

Most, if not all, of these elements are part of the stock-in-trade of an accomplished native speaker, of course; but none of them are directly related to native ‘speakerism’. Voice control, for example, could refer to volume and speed of delivery rather than to phonology. Signposting could include both standard and non-standard features, such as the focusing features revealed in the ELFA corpus and referred to above (standard meta-discourse, non-standard left dislocation).

However, an empirical basis is needed for such a framework. Initial research will thus focus on the reception of student presentations by their

---

6 Survey carried out by the author at the University Ca’ Foscari Venice in December 2012.
intended audiences (fellow students), and the identification and isolation of assessable elements within an ELF context. A second stage would then involve drawing up an assessment grid, training raters to use it, and carrying out a pilot assessment.

This is, in fact, a daunting task. For experienced examiners it will mean a change of mindset, and involve substituting tick-boxes on score-sheets to reward formal features which approximate to native speaker production (e.g. stress timing, ‘difficult’ syntactical features, the use of idiommatic expressions) with the recognition of strategies which promote comprehensibility for a NNS audience. For some this may seem a fool’s errand.

However, if, as we argued at the beginning of this article, the role of English as the working language of academia is now established in the European Higher Education Area, in which most students are called to function as ELF users rather than as learners of English as a foreign language, and if universities need to seek evidence, through entrance tests and external certifications, that incoming students will be able to operate in this context, then it is now urgent to engage with assessing ELF not just for receptive skills, but also for production. Davies and Elder (2005) advised caution in trying to develop such a test, at least until ELF was better described. A decade down the line, with the considerable amount of research findings, and the insights provided by ELF corpora available freely on line, that time seems to have come.

Bibliography


