VOICE Recording –
Methodological Challenges in the Compilation of a
Corpus of Spoken ELF

Angelika Breiteneder, Marie-Luise Pitzl, Stefan Majewski, and Theresa Klimpfinger, University of Vienna

Abstract. VOICE, the Vienna-Oxford International Corpus of English, aims to provide a general basis for analyses of English as a lingua franca (ELF) talk on all linguistic levels. This paper discusses criteria that have to be fulfilled in order to achieve this aim. It illustrates that compiling a corpus of spoken ELF involves the reconciliation of different – and often conflicting – requirements: this means finding and keeping a reasonable balance between theoretical specifications, methodological considerations and practical limitations. The focus of this paper is on dilemmas encountered in the processes of data collection and selection and finding suitable ways of representing spoken ELF in written form.[23]

1. Introduction

At the beginning of the 21st century English is used by far more people who do not have it as their mother tongue than by its native speakers (e.g. Graddol 2006). It is used as a lingua franca worldwide and is a well-established means of communication, not only in public domains of use like politics, business, education, and science, but also in private interactions between individuals. However, although English as a lingua franca (henceforth ELF) is so widespread, relatively little work has been done on its description.

The Vienna-Oxford International Corpus of English (VOICE), a structured collection of language data capturing spoken ELF interactions, provides a basis for such a description. It is currently being compiled at the Department of English at the University of Vienna and initially aims at a size of one million words of spoken ELF. The VOICE project therefore paves the way for in-depth linguistic studies of ELF by creating the first general database of transcribed spoken ELF interactions in diverse settings all over the world[24]. The present paper reports on some methodological challenges we (i.e. the project director Barbara Seidlhofer and the research assistants Angelika Breiteneder, Marie-Luise Pitzl, Stefan Majewski and Theresa Klimpfinger) have encountered in the initial phases of our corpus compilation, and on how we have responded to them.

2. Finding spoken ELF

2.1. The general need for external criteria

One of the first methodological challenges in the compilation of any corpus is to identify and select data for inclusion. VOICE is a corpus of spoken English used as a lingua franca. As we have already indicated, ELF is very wide-spread. But how can one define and delimit this use of English? And how can it be satisfactorily sampled?

It soon becomes obvious that focusing on linguistic criteria when trying to identify ELF data is not a viable option. On the one hand, in the absence of any description of ELF we have no record of its distinctive features to refer to. On the other hand, even if we had such a description, focusing on linguistic criteria would lead to a dangerous circularity: using assumed linguistic features of ELF for the selection of texts would

---

[24] As a general corpus, VOICE is nicely complemented by the specialized corpus of English as a Lingua Franca in Academic settings (ELFA), compiled at the English department of Tampere University, Finland. For more information see http://www.uta.fi/laitokset/kielet/engf/research/elfa/. These two corpora are the only major ELF corpora to date that we are aware of.

lead to the occurrence of exactly those features in the corpus. Applying linguistic criteria to identify ELF data would thus inevitably invalidate the findings of any researcher working with the corpus.

Since a reliance on internal linguistic criteria is ruled out, we need to consider external criteria for the identification of ELF. As defined by Sinclair (2005), external selection criteria are not based on features of the language as such, but on the "communicative function of a text", while those "reflect[ing] details of the language of the text are called *internal criteria*" (para. 6). A focus on external criteria can thus be recognized as a way of circumventing the danger of circularity (e.g. Clear 1992, Sinclair 2005). The working definition of ELF as it is used for the VOICE project is based on such external criteria, namely the communication between fairly fluent interlocutors (cf. section 2.2.1) from different first language backgrounds, for whom English is the most convenient shared language (cf. http://www.univie.ac.at/voice). This definition is external since it is based on the characteristics and purposes of the speakers rather than on their linguistic output.

The distinction between internal and external criteria is, of course, not absolute since they are inherently interrelated (e.g. Clear 1992: 29). The point is, however, that the external criteria have to be given primary consideration in corpus compilation as a basis for the identification of internal features. How then can such criteria be defined?

2.2. Data collection and data selection

A limited set of clearly defined external criteria is needed in order to select those pieces of data which are to become part of the corpus. The collection of data, a central part of corpus building, is thus governed by these external criteria. When one is involved in building a corpus of spoken, rather than written, language, the process of **data collection**, however, is followed by another process of **data selection**. While the stage of data collection may more or less converge with that of data selection in building a written corpus, data collection and data selection are two separate stages in the compilation of a spoken corpus. Although they are interrelated to some degree, the process of data collection is a separate operation and, though preliminary, is crucial and needs to be systematically conducted.

2.2.1. External criteria for VOICE

Aiming at a general corpus of spoken ELF, we rely essentially on four main external criteria, which inform all decisions made and actions taken with regard to data collection. First and foremost, it is English as a **lingua franca** that VOICE seeks to capture. That is to say, the use of the language predominantly between speakers of different first languages, the majority of whom use it as an additionally acquired language system, and which is, therefore, in its purest form, a language use with no native speakers (cf. Seidlhofer 2001b: 146). VOICE researchers seek to gain access to interactions where people of different first language backgrounds meet and use English as their preferred language, i.e. as their lingua franca of choice. In this regard, the project operates with the definition 'fairly fluent non-native speaker of English,' which characterizes a speaker who is able to use the language effectively for his/her particular communicative needs. Fluency is therefore a function of effective communication and not based on external assessments of proficiency. Hence, it is the language produced by fairly fluent ELF speakers that VOICE seeks to record.

As a second criterion, the corpus aims to capture the **spoken mode** of real-life interactions rather than simply the spoken medium of speech. We therefore aim to collect non-scripted ELF speech, i.e. not written-to-be-spoken-aloud, and generally try to avoid occasions where scripted preparation is to be expected (e.g. presentations, speeches, etc.).

Third, it is a central objective of VOICE to record ELF speech not simply as it is produced by individual speakers but as it happens among speakers in the natural course of interaction. The corpus would otherwise be a misrepresentation of talk in that it would fail to record inherent features of spoken interaction such as online negotiation of meaning, back-channelling, providing feedback or signalling problems of understanding. **Interactivity** is therefore another main criterion for VOICE.

As a fourth guideline, it is the aim of the VOICE project to collect **naturally occurring** ELF speech, i.e. interactions which are not elicited or set up for research purposes, but "talk that would have happened anyway, whether or not a researcher was around to record it" (Cameron

2001: 20). Although it may prove worthwhile to also incorporate some interactions which could be categorized as semi- or quasi-natural conversations – especially if they are highly interactive and bring together speakers of a very wide range of L1 backgrounds – the predominant portion of the corpus (at least about 90 per cent) will consist of naturally occurring data.

These four criteria guide all VOICE data collection efforts. Yet, they can only be guidelines or points of reference, because as anyone involved in social science or linguistic field work knows, reality does not fit into neat categories and rarely, if ever, meets all the envisaged criteria at once. Consequently, the actual recordings, i.e. the stage of data collection, is followed by a phase of **data selection** in which the criteria are applied more stringently to establish in how far they are met by the actual recordings.

2.2.2. Matching data and criteria

It might indeed be tempting to assume that having recorded so many hours of carefully chosen naturally occurring speech one has already compiled a corpus. Yet, when it comes to corpora of spoken language this goal is not necessarily – in fact is rather rarely – reached at this point. For our own purposes, we therefore impose quality control by filtering the data through the criteria a second time before we eventually select them for inclusion in VOICE. This process of data selection, which could also be called 'reality check', functions as an analytic filter and feeds back again into the process of data collection at a later stage.

To provide an example, we aim to include only the most interactive and most widely ranging L1-constellations in the corpus, which means that not all the data which were collected in the course of the project end up as part of the corpus. This is the case because things may turn out differently than one could possibly foresee during the planning stages of the data collection phase. For instance, a panel discussion involving seven panellists three of whom (instead of two as stated in the conference programme) are native speakers of English who end up taking up more speaking time than the other four panellists taken together may have looked like a promising source for ELF initially. But it then turns out to be something that is not ideally suited for our ELF corpus, as it actually

constitutes intercultural communication between non-native and native speakers of English in equal proportions. While this is certainly very interesting linguistic data which allows for the analysis of a broad range of empirical questions, it is not suitable for inclusion in VOICE since it fails to sufficiently fulfil our criteria. Complications like these are unavoidable, because ELF interactions involve a considerable range of variables which are subject to variation in ways which cannot be predicted, particularly so in the case of naturally occurring conversations.

As a general rule, the VOICE team aims to record and transcribe individual speech events in their entirety in order to also allow for qualitative analyses[25]. Yet, an interaction can comprise different parts, some of which may be relevant for VOICE while others do not fulfil one of the four main criteria discussed above. In fact, we have to deal with two kinds of problem in the compilation of our corpus. One has to do with the content of the recorded interactions, while the other is related to their actual nature. Concerning the first problem, recorded data sometimes include parts which are highly sensitive content-wise (e.g. speakers discuss technical advances which may earn them a lot of money if they are the first to release them). When it comes to the second problem, recorded data sometimes turn out to be unsatisfactory in that they include parts which do not sufficiently fulfil our selection criteria (e.g. a highly interactive group discussion involves a presentation by one of the participants, which turns out not only to be monologic but also scripted). Given the labour-, time- and consequently also cost-intensive nature of collecting naturally occurring ELF data, however, it would seem imprudent to dismiss such interactions as irrelevant or unusable for the corpus altogether. What was pointed out by Atkins, Clear and Ostler almost fifteen years ago still holds true today:

> The difficulty and high cost of recording and transcribing natural speech events leads the corpus linguist to adopt a more open strategy in collecting spoken language. (Atkins, Clear & Ostler 1992: 2)

---

[25] For a discussion of the importance of clearly situated qualitative studies at this stage of ELF research see Seidlhofer, Breiteneder & Pitzl (2006).

Such a strategy means doing the least damage possible to the integrity of the interactions recorded, while at the same time adopting a realistic and manageable attitude towards data gathering, whereby we 'sacrifice' parts of otherwise available data to our criteria. In practical terms this means, e.g., that we may decide not to transcribe long monologic speaker turns in their entire length (but for example only ten minutes out of thirty) in order to secure a higher degree of interactivity, or leave portions of data untranscribed in order to secure informants' confidentiality. Procedures like these fall into the category of data selection rather than data collection and present a vital regulation phase in the compilation of the corpus.

2.2.3. Questions of corpus balance

When selecting data, it is of course not only important to consider how far the collected data satisfies the main criteria specified for the corpus, but subsequently also to think about how the selected data actually relate to each other. This second consideration opens up the issue of corpus balance.

> There is much talk of a 'balanced corpus' as a sine qua non of corpus analysis work: by 'balanced corpus' is meant (apparently) a corpus so finely tuned that it offers a manageable small scale model of the linguistic material which the corpus builders wish to study. (Atkins, Clear & Ostler 1992: 6)

Given that balancing a corpus "is not something that can be done in advance of the corpus-building process" (Atkins, Clear & Ostler 1992: 6), it is a long-term task which continually accompanies corpus compilation. Consequently, the issue of corpus balance has at the current stage of our project not finally been decided on. But it has been the subject of much discussion and, already at this stage of the project, we have identified two major areas within which some sort of balance and representativeness should be observed in VOICE, namely that of setting

and that of the speakers' first languages, both of which are purely externally defined.

When it comes to settings, the architect of a general (i.e. non-specialized) language corpus[26] needs to take an overall decision about whether the corpus should focus more on the reception or the production of the language in question. The distinction between reception and production, however, is by no means absolute, as it largely depends on the ratio of 'producers' and 'audience' whether something would be characterized as a 'reception' or a 'production' setting (e.g. Atkins, Clear and Ostler 1992: 5). With regard to VOICE, interactivity is one of the main criteria for our data, which means that (ideally) all 'receivers' should also be 'producers' and vice versa. Thus, we aim for a balanced numeric ratio between 'producers' and 'audience' in the settings we choose. While VOICE is thus a 'production' corpus in that it captures ELF as it is used (i.e. produced) by a great variety of speakers, it also covers the 'reception' aspect of ELF, as the interactivity criterion allows us to investigate how people react to and experience ELF (i.e. how they receive it). Thus typical 'reception' settings such as large parts of the mass media like webpages and radio or television news (with the possible exception of e.g. interviews on the radio or TV) are not relevant for our corpus, but neither are typical 'production' settings such as long monologic conference presentations with no opportunity for interactive participation.

Considering all the potential 'production' and 'reception' settings for ELF, we are, of course, restricted in our attempt to achieve a balance between them since not all settings are equally accessible for a researcher (e.g. There is no way of getting access and permission to record informal

---

[26] It is indeed the general, rather than specialized, nature of VOICE that leads to several additional concerns in the compilation of the corpus. More specialized corpora such as, e.g., the Michigan Corpus of Spoken Academic English (MICASE, http://www.lsa.umich.edu/eli/micase/index.htm) or the corpus of English as a Lingua Franca in Academic settings (ELFA, URL: http://www.uta.fi/laitokset/kielet/engf/research/elfa/) face fewer issues with regard to the balance and representativeness of settings (cf. e.g. the discussion of general language corpora in Sinclair 2005).

talk between high-level EU politicians). But even if it were possible to get permission and admission to all relevant settings, capturing comparable proportions of data from the various settings would not make our corpus more valid, since true representativeness would require us to measure the entirety of ELF conversations all around the world. And this is clearly not feasible. Even a large and well-designed corpus like the British National Corpus (BNC) only has 10% spoken language as opposed to 90% written language, thereby reflecting the very opposite of actual language use. So all one can do is to give what Simpson, Lucka and Ovens (2000: 45) call "an impressionistic measure" of which settings are more important than others and to what degree. As Sinclair puts it,

> [r]oughly, for a corpus to be pronounced balanced, the proportions of different kinds of text it contains should correspond with informed and intuitive judgements. (Sinclair 2005: para. 37)

As far as the balance of first language speakers is concerned, we are, once again, confronted with problems of practical feasibility and financial limitations. Given that our project is located in Vienna, Austrian speakers of German are inevitably over-represented in our corpus. This is, however, not necessarily seen as a disadvantage. On the one hand, given the computer readability of our transcripts, search facilities will make it possible to regulate how many Austrians one wants to include in one's query. And on the other hand, the fact that an Austrian-funded project offers the option of homing in on how Austrians communicate through ELF may indeed be an advantage. As Atkins, Clear and Ostler put it so pointedly, "knowing that your corpus is unbalanced is what counts" (1992: 6). Measures to counteract over-representation are taken in the continuous process of evaluating our data at the collection and selection stages while always retaining the integrity of the original interaction as far as this is practically possible.

2.3. Data collection: practical issues

So far we have been concerned with theoretical considerations involved in the compilation of VOICE. The present section will turn to more practical issues and consider the process of data gathering, i.e. actually recording ELF conversations. First of all, it should be noted that VOICE

is based on audio-recordings, wherefore non-verbal features are automatically not recorded systematically, but only via field notes. This decision is mainly based on practical considerations. Firstly, video-recordings do not only present a greater intrusion in and corruption of a naturally occurring situation, but they are also costly and rather complicated to deal with. Secondly, questions like how to install a video-camera, whose viewpoint to take, how to portray all participants in the same way on screen and how to actually make use of the amount of material in transcribing may sound trivial but have rather fundamental and far-reaching implications. Furthermore, there are ethical considerations of protecting the speakers' identities, which place additional obstacles in the process of getting permissions to do recordings.

In addition to the audio-recording of the event, a certain amount of background and contextual information is required by the corpus builders, the transcribers and also future corpus users for reasons of the kind Scott and Tribble point out with reference to educational purposes:

> [r]ather than seeing corpus data as entirely abstracted from its linguistic and social context, the studies [presented in this volume] stress the obligation on the researcher to re-connect with the text (and, where possible, with its context of production) in order to build accounts of language in use which will have value for teachers and students of language alike. (Scott & Tribble 2006: p. X)

Thus our field notes include information about the nature of the event and the interaction taking place as well as about the participants engaging in these ELF interactions. Information about the event refers to the setting in which the interaction occurs as well as its degree of interactivity. Relevant information concerning the participants includes their number, their respective first languages, approximate age and gender but also specifications concerning their functions and roles during the interaction, their (power) relationships, as well as the purpose of the interaction.

On the one hand, such information facilitates the process of data selection (cf. section 2.2). As has been indicated in the previous section, sometimes a recorded event does not turn out to be ideal for the inclusion in VOICE. Therefore, detailed information about the actually recorded interaction is of utmost importance to be able to select an event as either suitable or unsuitable for the corpus at a later stage.

On the other hand, contextual information including, e.g., knowledge of the purpose of the interaction or names of organizations, greatly supports the process of transcription. As Cameron points out,

> there is more to transcription than careful listening or a 'good ear'. There is also the issue of contextual knowledge. […] [I]f the transcriber lacks crucial background knowledge that is available to informants, s/he will find it far more difficult to understand certain parts of their talk. (Cameron 2001: 35)

For a better reconstruction of the conversation regarding who says what to whom at what time, researchers make a sketch of the room where the recording takes place and number the speakers according to their first contribution to the interaction.

Moreover, the availability of contextual information is also essential for the researchers doing both qualitative and quantitative linguistic analyses on the basis of VOICE. Of course, the analytic possibilities and the encoding of the collected contextual information also largely depend on the transcripts produced from the recordings. The transcription and representation of ELF data will thus be the topic of the next section.

3. Representing spoken ELF

3.1. The spoken-written dichotomy

After having recorded data, the next step in the compilation of a spoken corpus is to consider how to best represent them on paper and screen. Even though this may sound rather straightforward, actually transforming recorded talk into a form that can be analyzed presents a considerable challenge for the researcher. As Thompson (2005: para. 2) points out, collecting material for a corpus based on written language is relatively easy, as the written words already exist in advance. It is a different matter when dealing with spoken data, and especially when it comes to ELF (cf. section 3.3.). With spoken language, the transcript, i.e. the written representation of spoken language data, forms the basis for the systematic analysis of the data, which would otherwise be impossible. As such, the question of what a transcript should actually look like and which decisions have been made prior to and during the process of transcribing is crucial for all subsequent analyzes.

In the domain of corpus linguistics, one occasionally encounters transcripts in which naturally occurring spoken language has been made to fit the general structures of written text through "a high degree of normalization of false starts, hesitation, non-verbal signals, and other speech phenomena" (Atkins, Clear & Ostler 1992: 10). For example, the transcripts in the spoken part of the British English component of the International Corpus of English (ICE-GB) show some of these normalization procedures. The result of such a procedure is an idealized transcript, a 'script' transcript so to speak, which is clearly at a remove from the reality of natural speech, even though it might well be "sufficient for a wide variety of linguistic studies" (ibid.). At the same time, researchers in domains like sociolinguistics or conversation analysis seem to have reached agreement that the aim of a transcript is to represent the spoken data as precisely as possible. In their understanding, a transcript has to "make explicit what normally gets taken for granted" (Cameron 2001: 7), meaning that it has to include all features of speech that normally tend to be left out when transferred into writing, such as repetition, self-correction, false starts or pauses. These two approaches are obviously in opposition to each other and therefore create a dilemma for the compilers of a spoken corpus.

Although the cost and time required for transcription need to be taken into consideration, they should not override the theoretical and analytical motivations for building a corpus. Generally, it is desirable that the researcher should try to avoid imposing features of written language on the text when engaged in the process of transcribing spoken language. Writing is linear and organizes text into words, phrases and sentences, neatly separated and arranged by spaces and punctuation whereas speakers repeat, misspeak and omit parts of words and whole words. They make pauses and reorganize their intended message, they leave utterances unfinished, or cut somebody off and/or speak simultaneously. Transcribing all this is, of course, a challenging task. Since transcriptions are written, transcribers will always tend to convert the naturally untidy data of spoken language to a more orderly written form. It thus takes "a real effort to *not* hear spoken language in terms of the written model" (Cameron 2001: 33, original emphasis). When it comes to ELF, it takes an additional effort – and consequently thorough training of our transcribers – not to transcribe what they think they ought to hear, but to represent what they really do hear.

3.2. Human and computer readability

With regard to the practical decision of how much detail to include in a transcript, issues of readability also need to be taken into account. Readability in this context actually is a twofold concept, as it relates to human readability as well as computer readability. Considering the first, it is obvious that there is a limit to the amount of information the human eye can process effectively. Many researchers therefore stress the importance of preserving the information in a form which "enables the researcher to extract the main information as quickly as possible without overburdening short-term memory" (Edwards 1995: 23) and even suggest visual means of attracting reader attention (ibid.). Thus, as Cameron (2001:39) points out, there is clearly the "issue of knowing when to stop":

> It is worth remembering that too much detail can be as unsatisfactory as too little. There is a trade-off between accuracy and detail on the one hand, and clarity and readability on the other. In some systems, particularly those designed for searchable computerized corpora, transcripts can be very inaccessible because of both the amount of detail and the fact that so much information is represented by arbitrary symbols like $ and #. (Cameron 2001: 39)

Cameron's statement underlines the importance of preserving human readability and the need to make a conscious and informed selection of what to include in a transcript (see also e.g. Ochs 1999: 168). She also hints at the fact that there is a close link between human and computer readability, as computer readability is seen as one of the reasons why transcripts become increasingly difficult to read for humans. Yet, computer readability may in fact also offer some solutions for problems of human readability.

For a corpus that aims to be computer readable, unambiguous mark-up is the absolute requirement since computers cannot handle ambiguities as flexibly as a human reader – if they can at all. Flaws in the design of the transcription conventions, which could ultimately cause such ambiguities, would therefore seriously multiply the work required to transform a transcript into a well-defined computer-manageable format. Yet, transcripts with an unambiguous mark-up do not necessarily

have to be difficult to read for humans, because such unambiguous mark-up may eventually allow for different viewing options in the corpus layout. Computer-based processing makes it possible to display (or not to display for that matter) different levels of detail or mark-up of the transcript, thus easing the burden placed upon the human reader. A transcript can therefore be human readable and unambiguous at the same time.

To the compilers of VOICE, the reconciliation of these two needs – computer readability and human readability – seems vital. It was therefore deemed important to base the visual layout of the transcripts on a format that is easily accessible to users and transcribers. This begins with the choice of representing each speaker turn as a paragraph: While some transcription systems may require each word to appear on a single line, the VOICE Transcription Conventions place one word after the other, annotated with appropriate mark-up. Furthermore, reading is facilitated by the choice of symbols for our mark-up. Frequently occurring features, such as brief pauses (indicated in our transcripts as '(.)') require little mark-up, are easy to memorize and consequently do not disrupt the flow of reading. Less frequent tokens, like the indication of an interruption in the recording (indicated in our transcripts as e.g. '(nrec 00:00:45) {change of battery}'), are more verbose but relatively self-explanatory. The keyword with regard to computer readability in turn is consistency, which requires a clearly defined set of transcription guidelines consistently applied to all transcripts produced for a corpus. Some features of the transcription conventions specifically designed for VOICE will be discussed in the next section.

3.3. Transcribing spoken ELF

The previous sections discussed difficulties concerning the transcription of spoken language into written language and issues of human and computer readability of transcripts in general. But there are problems and challenges that have to do more specifically with the transcription of ELF. Considering the tricky question of how to best represent spoken ELF in writing therefore turned out to be one of the first major tasks that had to be faced in the early months of the VOICE project. As with general transcription, we were concerned with two kinds of conventions:

mark-up and spelling. The first of these refers to all features (tags, pauses, contextual information etc.) that are added to the actual words in the transcript, while the second refers to the way the actual words are represented, i.e. spelt. These conventions eventually emerged not only from our own experience in transcribing, where particular problems became apparent, but also in the light of discussions among the members of the VOICE team and others experienced in transcribing ELF data. Though limits of space do not allow for a complete account of the resulting VOICE Transcription Conventions [2.0], the present section offers the reader an insight into the factors accounted for in some of the decisions taken.

3.3.1. General principles guiding VOICE transcripts

The overall purpose of VOICE is to provide a relatively large-scale resource for descriptions of ELF on all linguistic levels which is accessible to linguistic researchers all over the world. Given this general goal, transcription conventions for a project like VOICE need to be broad enough to be useful to a large number of potential research foci, which often require different degrees of detail on different levels of language.

Acknowledging that selectivity is essential but "should not be random and implicit" (Ochs 1999: 168), VOICE transcripts are orthography-based and generally do not represent reduced, dialectal or accented pronunciations. With the exception of four wide-spread lexicalized phonological reductions (*cos*, *gonna*, *gotta*, *wanna*) and all standard contractions, words are represented in full standard orthographic form. Yet, specific mark-up features e.g. for lengthening, emphasis, speaking modes, rising and falling intonation allow for additional prosodic features to be included in the transcript. With regard to in-depth analyses on the phonological/phonetic level, it is envisaged to make selected audio files of the recordings available at a later stage of

the project, thereby opening up the possibility for the subsequent supplementation of our transcripts[27].

As we previously pointed out, our transcripts need to be human readable and the transcription conventions relatively easy to operate. This means that some aspects of face-to-face interaction, such as many non-vocal paralinguistic features, necessarily fall outside the scope of our conventions. As Cook puts it so pointedly, "[i]n transcription, there is an inevitable conflict between those elements which can be transcribed most exactly and those which must remain impressionistic, elusive and subjective" (1995: 36). Our compromise here has been that our recordings are always supplemented by detailed field notes which allow us to include some information about the non-verbal context in the transcripts. Additionally, our transcripts are complemented by elaborate transcriber's notes in which additional contextual information and observations about other features of the interaction not accounted for in the transcript are recorded. This information, however, is of course bound to be limited by the method of recording data (i.e. audio-taping) and is included only in so far as it is necessary in order to understand what is going on in the interaction. Consequently, verbal behaviour is clearly foregrounded over non-verbal behaviour in the transcripts, although our experience with ELF talk supports observations that non-verbal behaviour is an essential part of ELF interactions. The VOICE project is, after all, concerned with ELF in its verbal manifestation.

3.3.2. Orthographic representation of spoken ELF

VOICE transcripts thus focus primarily on the ELF speakers' verbal behaviour and represent their utterances in orthographic form. But according to which norms? As VOICE is concerned with ELF, a use of English that has hitherto not been described, let alone codified, the transcripts have to be modelled on some existing orthographic system.

---

[27] The Michigan Corpus of Spoken Academic English (MICASE) serves as a model in this respect (cf. e.g. Simpson, Lucka & Ovens 2000: 46).

As we have emphasized earlier, consistency in spelling is of prime importance for any computer readable corpus. Consequently, one needs to have a standard for spelling that transcribers – and also the computer – can refer to. It would seem then that the obvious choice for a spelling system to use would be one of the well-documented and extensively codified orthographic systems, i.e. Standard British or American English. But there are problems with adopting only one of these systems for ELF transcripts. As Cameron states,

> [t]hough standard English spelling is not a very exact representation of any kind of English pronunciation, its status as the default way of writing means it tends to bring to mind a 'standard' pronunciation […]. (Cameron 2001: 41)

And clearly, the majority of speakers recorded in VOICE by definition do not fully conform to either Standard British or Standard American English. The issue of according to which standard to actually spell words in our transcripts therefore sparked off one of the major discussions in the early stages of designing the new conventions.

For the reasons outlined above, we found that neither of the two standards was entirely appropriate for representing ELF. Consequently, the decision we came to was to rely on neither of these two standards in their entirety, but to introduce an element of fusion of both. The compromise we opted for in this matter is outlined below and has to be primarily seen as symbolic. In this symbolic sense, it was our aim to simultaneously dissociate ELF from both the British and the American standard. For our corpus, we wanted to be able to rely on an orthographic system which implies a certain independence of the existing national norms, but nevertheless offers the consistency and unambiguity required for a computer readable corpus.

The result of this practical as well as symbolic need for compromise was to rely on Standard British English as a main point of reference, while also including some specified items spelt in Standard American English. The decision to take British English (henceforth BrE) as a general reference point was mainly made on the basis of practical considerations such as the fact that VOICE is being built in Vienna, i.e. in the middle of Europe, and, at the current stage, has a certain focus on European speakers. Thus, from this location perspective, BrE seemed to be a more logical choice. The location factor was further enhanced by the fact that BrE is more widely taught than American English (henceforth

AmE) in Austrian schools as well as at the English Department of the University of Vienna, where VOICE is being compiled. Consequently, the majority of our transcribers have learnt BrE and would have difficulty to simply shift to American spelling. From a very practical perspective, adopting AmE as the main point of reference would thus have been likely to result in more erroneous – and therefore inconsistent – transcripts produced by our transcribers. Even though all transcripts are, of course, thoroughly checked by the members of the VOICE team, getting worse quality 'raw' transcripts would not only have implied a greater workload for those doing the checking but was also considered unfavourable for the general quality of our corpus.

However, in order to acknowledge the wide-spread presence of AmE, e.g. in the media or on the internet, and for the symbolic reason outlined above, we decided to introduce some elements into the VOICE spelling conventions which would be spelt – or rather spelled? – in AmE. Yet, we did not want to identify these items on the basis of some random intuitive selection but on the basis of a set of comprehensible and reproducible criteria. As a starting point, we studied word frequency lists of the British National Corpus (BNC) (cf. Leech, Rayson & Wilson 2001) to identify the most frequently occurring spoken BrE words which have different spellings in BrE and AmE. Subsequently, we tried to establish a cut-off frequency which would result in a manageable group of these words which should then be spelt in AmE in VOICE transcripts. We decided that all those words whose roots occur with a minimum frequency of 50 per one million words in the BNC were to form such a closed 'American spelling' list. The words identified in this way are *center*, *theater*, *behavior*, *color*, *favor*, *labor*, *neighbor*, *defense*, *offense*, *disk*, *program* and *travel (traveled, traveler, traveling)*. Not only the root form but also all the derivatives of the these words are spelt according to AmE standards in our transcripts.

Additionally, it became apparent when scanning the BNC word frequency lists that BrE makes use of both *-ise* and *-ize* morphemes in words such as *characterise–characterize* or *apologise–apologize* quite

extensively[28], which represents a general area of spelling ambiguity in BrE and thus an area of potential spelling inconsistency for VOICE. Consequently, we adopted a supplementary spelling rule that all words (verbs, nouns, etc.) which can be spelt with both *-ise* and *-ize* morphemes are spelt with the *-ize* variant in our transcripts. It is the *-ize* variant which is used in both BrE and AmE spelling, while the *-ise* variant is generally confined to BrE. It could be said that the *-ise* spelling variant is therefore rather British but is in fact being used less and less frequently even in the British context (e.g. the *Oxford Advanced Learner's Dictionary* (OALD) solely provides *criticize* as a main entry which is then supplemented by "BrE also *criticise*").

The spelling conventions as they are presently used for VOICE transcripts are therefore unique in that they were specifically designed to render the diversity of ELF speech in a standardized way. For the symbolic reason outlined above, BrE is employed for VOICE transcripts dissociated from any issues of correctness, identity or culture that might otherwise come with it. We clearly do not want to suggest that the ELF speakers in VOICE speak or sound more British than American – although this is in itself a question a researcher may well want to look into in the future.

3.3.3. Choosing a reference manual

Following these principles and decisions, we were further confronted with the fact that we needed an actual work of reference, a manual so to speak, which would be available likewise to ourselves and to our transcribers to look up all words other than the 'American spelling exceptions' specified above. The reference tool we chose for this purpose is the *Oxford Advanced Learner's Dictionary*, 7th edition (OALD 7). There are several prerequisites that such a reference work had to fulfil.

---

[28] The frequency of the tokens per million words spelt with the *-ize* morpheme in the BNC is usually not much lower than the frequency of the tokens of the same word spelt with the *-ise* morpheme (e.g. organi**s**ation 134 - organi**z**ation 93, characteri**s**e 14 - characteri**z**e 13) (cf. Leech, Rayson & Wilson 2001).

First of all, for purely practical reasons it needed to be accessible to ourselves and to our transcribers. A comprehensive dictionary such as the OED, which almost no one has in its printed form and which requires high registration fees in its online version, would thus have been impractical for our purposes. Secondly, it seemed appropriate that the work of reference should be comprehensive and yet not too detailed and should cover the general vocabulary which one would expect fluent ELF speakers to be in command of. The use of an 'advanced learner's dictionary' like the OALD 7 therefore seemed suitable. This is, however, by no means to imply that ELF users are in fact advanced learners of English. Quite the contrary, it is an essential part of the conceptualization of ELF – and thus also of VOICE – that ELF users are not learners of the language but language users in their own right, irrespective of whether or not they perceive themselves as learners or not (e.g. Seidlhofer 2001a, 2001b).

For the purposes of our project, the OALD 7 is thus not being used as a dictionary, i.e. an authority on matters of correctness, but rather as a manual, a stable and shared point of reference for practical reasons. In addition, from a theoretical and symbolic perspective, the OALD 7 was the first, and to our knowledge as yet only, dictionary to publicly acknowledge the existence of ELF. Its reference section includes an introduction to the phenomenon of ELF as well as a short description of VOICE and a list of initial observations regarding some lexicogrammatical features of ELF written by Barbara Seidlhofer (Hornby 2005: R 92). Practicalities like the fact that spelling variations are always marked as either BrE or AmE in the OALD 7 allow for giving proper instructions such as "If there are two separate entries for British and American spelling, the British entry is selected" in the VOICE spelling conventions (VOICE 2005b: 1). Since all VOICE transcribers are equipped with the *OALD7 Compass CD-Rom*, the primary source of reference for VOICE transcripts remains stable insofar as dictionary entries cannot be adapted at a later point (as they could in the online version of the OALD, for instance). With this in mind, instructions like "If an entry gives more than one spelling variant of a word, the first

variant is chosen" (VOICE 2005b: 1) guarantee consistency in spelling. Furthermore, the OALD 7 was only published recently (in 2005) and so represents an up-to-date point of reference which is clearly a favourable quality[29].

Bearing in mind that a learner's dictionary like the OALD 7, naturally, only captures a limited range of vocabulary, but leaves out most of the technical terms often recorded in expert conversations carried out in ELF, a specific tag is used in our transcripts to mark all those words in ELF talk that do not form part of the OALD 7. This tag, which is currently found under the heading "Pronunciation variations & coinages" (VOICE 2005a: 3), also highlights a certain number of technical terms and may at a later stage of the project be taken out again after comparing them with more specialized dictionaries. For example, terms which are now marked but which are likely to be recognized as part of general English vocabulary later on through reference to more specialized dictionaries include, for instance, *power-point presentation*, *commodification* or *papyrology*. Yet, this tag necessarily also includes words which are not part of any English dictionary at this point, but which represent newly coined words like *to examinate*, *to enfoster*, *importancy* or *forbideness*. This tag therefore presents what one could call a 'security device' for our transcribers in so far as its use guarantees that one of the members of the VOICE team checking the transcript will take a close look at the specific word. Additionally, the tag offers a valuable source for investigating productivity and creativity in ELF talk.

3.3.3. Some ELF-specific tags

Generally, the VOICE mark-up conventions are specifically designed to reflect what from our transcriptions so far has emerged as widespread features of ELF interactions. Apart from the tags already discussed, the nature of our data prompted us to devise a fairly detailed set of descriptors for, for instance, code-switching, onomatopoeic sounds and

---

[29] Yet, the choice of the OALD 7 is not necessarily final, as changes are always possible if another manual appeared to be more apt in the course of the work on corpus compilation.

laughter. As Klimpfinger (2005: 112) points out, code-switching is "an intrinsic element of ELF talk" used "to fulfil certain discourse functions, to compensate for perceived linguistic 'deficiencies,' and to signal group membership". In the light of these findings, VOICE transcripts include specific tags to mark non-English speech which also allow for a distinction between the speaker's first or other languages and to differentiate between forms expressing social distance or closeness which cannot easily be transferred into English (e.g. the German distinction between *du* and *Sie*). Additionally, translations into English are provided whenever this is possible in order to make the non-English speech accessible for those future users of the corpus who do not know the various languages used in the ELF conversations. The following extract provides an example taken from VOICE:

> S4: you understand what i mean (.) no? er i think yeah. er (.) well you er <L1fr> vous etes {you/dis are} </L1fr> er how do you say it

Dealing with a great variety of ELF data, it also appears that ELF speakers repeatedly produce noises in order to imitate something instead of using words. A specific tag was therefore designed to mark these onomatopoeic noises:

> S1: it may be quite HARMLESS and at the end of the day you (.) **<ono> d   d   d   </ono>** (.) somebody

Laughter is also emerging as quite central to ELF conversations – not only as such but as a prosodic feature of speech. Unlike other spoken corpora, VOICE transcripts therefore mark laughter in a fairly detailed manner by approximating syllable number and distinguishing between utterances spoken laughingly and laughter and laughter-like sounds.

At the end of this section dealing with the representation of spoken ELF, it becomes obvious once again that formulating transcription conventions and transcribing data is "effectively the first stage of analysis and interpretation" (Cameron 2001: 43). One realizes that in a large-scale project like the compilation of an ELF corpus every single, often minor, detail involves a decision that is – consciously or not – informed by particular assumptions and hypotheses. The VOICE transcription conventions [2.0] are thus not only based on our extensive experience with a wide range of ELF data as well as the knowledge we

have gained from initial analyses of some of these data but also, and in particular, on the general conceptualization of ELF provided by Seidlhofer (e.g. 2001b, 2004, 2005b, 2005c, 2006). Yet, even though the VOICE transcription conventions are tailor-made for ELF data, they are nevertheless open and broad enough to allow for transcripts to be analyzed with regard to a great variety of research interests on all linguistic levels. Furthermore, the conventions in their current form are not necessarily fixed, as allowance for modifications on the basis of our experience with more transcripts and further corpus work will always be made as far as this seems worthwhile and practical.

4. Outlook

The prime objective of the VOICE project so far has been the collection of ELF data in order to make it possible to explore the very nature of ELF talk. In this regard the ultimate purpose of VOICE is to render possible broadly-based, in-depth linguistic analyses of ELF on all linguistic levels. This versatility is reflected in our transcripts, which provide a solid foundation for accurate and comprehensive descriptions of ELF not only on a quantitative, but also a qualitative basis.

At the current stage of the project, qualitative analyses of the data available and the formulation of hypotheses to be checked against the corpus later on is most conducive to good progress. As pointed out by Seidlhofer, Breiteneder and Pitzl, at the present stage of ELF research,

> it is advisable to be tentative and circumspect and to proceed by way of clearly situated qualitative studies with a strong ethnographic element. As more qualitative, hypothesis-forming findings begin to emerge, it will become possible to introduce more controlled, quantitative procedures. (Seidlhofer, Breiteneder & Pitzl 2006: 21)

A number of VOICE-related case studies of ELF talk have already been conducted on several linguistic levels of description. So far, these have concentrated on various lexicogrammatical and pragmatic aspects of ELF, such as the redundancy of the 'third person -*s*' (Breiteneder 2005a, 2005b), types of miscommunication in business contexts (Pitzl 2004, 2005), the role of code-switching (Klimpfinger 2005), phatic communion (Kordon 2003), humour (Brkinjač 2005) and metalanguage (Wagner 2005), to mention but a few. One of the overall questions which VOICE

hopes to help find answers to is which features and strategies lead to successful communication in ELF. This comprises the speakers' preference for certain linguistic choices over others just as much as the question as to which features are characteristics of ELF while not impeding successful communication (e.g. Seidlhofer 2005a: 37-39).

As described in this article, important steps on the way to a comprehensive and reasonably representative ELF corpus have already been taken. While the anticipation and the motivation grows when new options for research come into sight, many tasks still remain to be accomplished and decisions need to be taken carefully in order to live up to the high expectations. One important step will be the finalization of the technical specifications of the corpus format. The choice of corpus format is mainly determined by aspects of interoperability and the tools one is going to use. As described before, the VOICE transcription conventions are designed to be unambiguous and thus entirely computer readable. Given the required specialist knowledge it is relatively easy to implement a converter between the human written transcripts and an arbitrary corpus format. Regarding the development of open standards, specifically the recommendations of the Text Encoding Initiative (TEI, http://www.tei-c.org/), which are based on the highly flexible eXtensible Markup Language (XML), the corpus format will most likely be based on or derived from these standards. As XML is an extensible and easily convertible format, it will be possible to adapt the corpus to the specific needs of various linguistic analysis tools around. When these issues are settled, it is planned to eventually make the corpus accessible through the VOICE website, where it will be possible for researchers all around the world to work online on the corpus. At a later stage, it is also envisaged to offer additional possibilities, apart from the transcripts, for investigating further aspects of ELF. Therefore selected audio files will be made available to enable researchers to supplement their work with phonetic transcription and analyses of the actual recordings. Given the growing interest in ELF, it appears likely that VOICE and other more specialized ELF corpora like ELFA will have a great impact not only on ELF research, but also on the conception of English in general, and on the nature and scope of English studies.

References

Atkins, Sue, Jeremy Clear and Nicholas Ostler. 1992. "Corpus Design Criteria." *Literary and linguistic computing* 7 (1): 1-16.

Breiteneder, Angelika. 2005a. Exploiting Redundancy in English as a European Lingua Franca: The Case of the 'Third Person -*s'*. MA thesis, University of Vienna.

Breiteneder, Angelika. 2005b. "The Naturalness of English as a European Lingua Franca: The Case of the 'Third Person -*s'*." *Vienna English Working PaperS* 14: 3-26. Available: http://www.univie.ac.at/Anglistik/Views0502ALL.pdf. 5 May, 2006.

Brkinjač, Tina. 2005. Humour in English as a Lingua Franca. MA thesis, University of Vienna.

Cameron, Deborah. 2001. *Working with Spoken Discourse*. London: Sage Publications.

Clear, Jeremy. 1992. "Corpus Sampling." *New Directions in English Language Corpora*. Ed. Gerhard Leitner. Berlin: Mouton de Gruyter. 21-31.

Cook, Guy. 1995. "Theoretical Issues: Transcribing the Untranscribable." *Spoken English on Computer. Transcription, Mark-up and Application*. Eds Geoffrey Leech, Greg Myers and Jenny Thomas. New York: Longman. 35-53.

Edwards, Jane A. 1995. "Principles and Alternative Systems in the Transcription, Coding and Mark-up of Spoken Discourse." *Spoken English on Computer. Transcription, Mark-up and Application*. Eds Geoffrey Leech, Greg Myers and Jenny Thomas. New York: Longman. 19-34.

Graddol, David. 2006. *English Next. Why Global English May Mean the End of 'English as a Foreign Language'*. The British Council. Available: http://www.britishcouncil.org/files/documents/learning-research-english-next.pdf. 4 May, 2006.

Hornby, A.S. 2005. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press.

Klimpfinger, Theresa. 2005. The Role of the Speakers' First and other Languages in English as a Lingua Franca Talk. MA thesis, University of Vienna.

Kordon, Kathrin. 2003. Phatic Communion in English as a Lingua Franca. MA thesis, University of Vienna.

Leech, Geoffrey, Paul Rayson and Andrew Wilson. 2001. *Word Frequencies in Written and Spoken English Based on the British National Corpus*. London: Longman.

Ochs, Elinor. 1999. "Transcription as Theory." *The Discourse Reader*. Eds Adam Jaworski and Nikolas Coupland. London: Routledge. 167-82.

Pitzl, Marie-Luise. 2004. "I Know What You Mean" – 'Miscommunication' in English as a Lingua Franca: The Case of Business Meetings." MA thesis, University of Vienna.

Pitzl, Marie-Luise. 2005. "Non-understanding in English as a Lingua Franca: Examples from a Business Context." *Vienna English Working PaperS* 14 (2): 50-71. Available: http://www.univie.ac.at /Anglistik/Views0502mlp.pdf. 5 May, 2006.

Scott, Mike and Christopher Tribble. 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: Benjamins.

Seidlhofer, Barbara. 2001a. "Brave New English?." *The European English Messenger* X (1): 42-48.

Seidlhofer, Barbara. 2001b. "Closing a Conceptual Gap: The Case for a Description of English as a Lingua Franca." *International Journal of Applied Linguistics* 11 (2): 133-58.

Seidlhofer, Barbara. 2004. "Research Perspectives on Teaching English as a Lingua Franca." *Annual Review of Applied Linguistics* 24: 209-39.

Seidlhofer, Barbara. 2005a. "Englisch als Lingua Franca und seine Rolle in der internationalen Wissensvermittlung. Ein Aufruf zur Selbstbehauptung." *Sprache(n) in der Wissensgesellschaft. Proceedings der 34. Jahrestagung der Gesellschaft für Angewandte Linguistik*. Eds Sabine Braun and Kurt Kohn. Frankfurt am Main: Peter Lang. 27-45.

Seidlhofer, Barbara. 2005b. "Language Variation and Change: The Case of English as Lingua Franca." *English Pronunciation Models: A Changing Scene*. Eds Katarzyna Dziubalska-Kolaczyk and Joanna Przedlacka. Bern: Peter Lang. 59-75.

Seidlhofer, Barbara. 2005c. "Standard Future or Half-baked Quackery? Descriptive and Pedagogic Bearings on the Globalisation of English." *Globalisation and the English Language Classroom*. Eds Claus Gnutzmann and Frauke Intemann. Tübingen: Narr. 155-69.

Seidlhofer, Barbara. 2006. "English as a Lingua Franca in the Expanding Circle: What It Isn't." *English in the World: Global Rules, Global Roles*. Eds Rani Rubdy and Mario Saraceni. London: Continuum. 40-50.

Seidlhofer, Barbara, Angelika Breiteneder and Marie-Luise Pitzl. (2006). "English as a lingua franca in Europe." *Annual Review of Applied Linguistics* 26. 3-34.

Simpson, Rita, Bret Lucka and Janine Ovens. 2000. "Methodological Challenges of Planning a Spoken Corpus with Pedagogical Outcomes." *Rethinking Language Pedagogy from a Corpus Perspective. Papers from the Third International Conference on Teaching and Language Corpora*. Eds Lou Burnard and Tony McEnery. Frankfurt am Main: Peter Lang. 43-49.

Sinclair, John. 2005. "Corpus and Text - Basic Principles." *Developing Linguistic Corpora: Guide to Good Practice*. Ed. Martin Wynne. Oxford: Oxbow Books. 1-16. Available: http://ahds.ac.uk/linguistic-corpora/ 16 Nov. 2005.

Thompson, Paul. 2005. "Spoken Language Corpora." *Developing Linguistic Corpora: A Guide to Good Practice*. Ed. Martin Wynne. Oxford: Oxbow Books. 1-16. Available: http://ahds.ac.uk/linguistic-corpora/ 16 Nov. 2005.

VOICE. 2005a. " Transcription Conventions [2.0]. Mark-up conventions." Available: http://www.univie.ac.at/voice/documents /VOICE_mark-up_conventions_v2-0.pdf 24 July, 2006.

VOICE. 2005b. "Transcription Conventions [2.0]. Spelling conventions." Available: http://www.univie.ac.at/voice/documents /VOICE_spelling_conventions_v2-0.pdf 24 July, 2006.

Wagner, Ingrid. 2005. Metalinguistic Features in the Use of English as a Lingua Franca for Professional Education Versus Casual Conversation. MA thesis, University of Vienna.