

The compilation of ICE Malta: State of the art and challenges along the way

Michaela Hilbert and Manfred Krug, University of Bamberg

1 Introduction

The idea to compile a Maltese English ICE component at Bamberg University along the principles laid out in Greenbaum (1991) and Nelson (1996) is owed to the fact that until recently no quantitative empirical studies existed beyond the domains of phonology (on which see Vella 1994ff.) and sociolinguistic statistics (on which see the census of 2005 and Sciriha and Vassallo 2003, 2006). As yet, no corpus-based studies exist on Maltese English morphosyntax (though compare the intuition-based monograph by Mazzon 1992). Hence, little is known about the status of written and spoken Maltese English, and about such fundamental questions as whether there is any evidence of an emerging standard with specific identity markers, or of stylistic variation with discernible quantitative correlates.

The corpus project became feasible thanks to a co-operation between Bamberg (Germany) and the departments of English and Linguistics at the University of Malta and thanks also to financial support granted by the Bavarian Ministry for Science and the Arts as well as the University of Bamberg Research Fund. Small speech communities have drawbacks for corpus compilation and representativeness because there may be a limited number of authors for some text categories. At the same time, they have the advantage that social networks are closely knit among the educated (Maltese) English-speaking classes, so that personal contacts of local academics facilitate data compilation.

2 Some facts about language use in Malta

Italian was the official language in Malta until 1934, when it was replaced by two co-official languages: English alongside Maltese, a Semitic language written in the Latin alphabet with a high proportion of Italian and Sicilian loanwords and an increasing number of English borrowings. In 2002, two years before

Malta's accession to the European Union, Maltese became an official language of the EU.

Maltese is the native language for over 95 per cent of the population, which currently totals an estimated 410,000 (roughly 380,000 of whom live on the main island, Malta, and some 30,000 on Gozo). Overall, some 90 per cent of the population claim competence in English, although degrees of proficiency vary considerably. The vast majority of the population have acquired English as a second language, but some five per cent use both English and Maltese as their main languages at home.¹ This will often come in the form of the one-parent-one-language strategy or in the form of varying language use with parents, siblings and grandparents. L1-speakers of English are found primarily in the higher socio-economic strata; in addition, higher usage rates of English vary regionally along the typical urban-rural cline. Particularly high rates are reported for such districts as Sliema and neighbouring St. Julian's, where both tourism and affluent households are concentrated.

Table 1: Native languages and multilingualism in Malta

Languages spoken in Malta			Native languages in Malta	
(2005 Census data; all degrees of proficiency considered)			(representative sample of 500 speakers; data collected by Sciriha and Vassallo in 2001, adapted from Sciriha and Vassallo 2006: 26)	
Language	speakers	%	Language	%
Maltese	354,664	97.9	Maltese	98.6
English	318,354	87.9	English	1.2
Italian	205,375	56.7	Italian	0.2
French	75,914	20.9		
German	20,110	5.5		
Arabic	14,046	3.9		
Other	15,159	4.2		

As can be seen from Table 1, bi-, tri- and multilingualism are widespread in Malta, an aspect that is well-documented in the relevant literature (e.g. Camilleri 1991ff.; Sciriha 2001). This is primarily due to historical language contact and to the importance of tourism in the country. Obviously, language and education policies play a role as well: schooling in Malta is mandatory until the age of 16. There is no official policy on the classroom use of languages, but the *National minimum curriculum* from 1999 issued by the Ministry of Education emphasizes

the importance of English and Maltese as official languages and states further that pupils in secondary schools are expected to learn a third or fourth language. Both Maltese and English are used from school entry, but Maltese is naturally more prominent in primary schools, while English is more prominent in secondary schools and dominant at tertiary level. There is a tendency for state schools to use less English than Catholic Church and private (so-called Independent) schools. The language of instruction also depends to a great extent on the language of the textbooks, most of which are in English. In particular, most reading and writing is done in English. For the same reasons, the subjects Maltese and History are almost exclusively taught in Maltese. In spoken interaction, code switching is widespread among both pupils and teachers, although a change in teacher education has shifted the balance somewhat towards Maltese: until the 1970s, teachers were trained by British religious orders, but more recently teachers have been trained by Maltese native speaker scholars at the University of Malta, which has generally led to an increasing proportion of Maltese in spoken classroom interaction.

In the remainder of the paper, we will discuss in more detail our approach to overcoming the challenges of compiling a Maltese English component for the International Corpus of English (ICE).

3 Status of the corpus and strategies of compilation

ICE-Malta is currently in the compilation and digitalisation stage for both the written and the spoken sections. That is, the written texts collected from the internet are selected (on the basis of the criteria given below), texts obtained in paper form are being digitalised, i.e. scanned, OCR-processed, proofread and manually edited. There is a range of written text types for which the collection of raw texts has not been completed yet. A considerable amount of collected spoken data is currently being transcribed. The collection of texts for the spoken section is less advanced than the collection of the written texts, for reasons that are given below. Annotation, however, is done together with the transcription process for the spoken data.

In order to facilitate the cooperation of members of the research project, we use Pax (Platform for Annotated Corpora), which was developed at the University of Augsburg, and which helps to document the individual steps and changes made in the compilation, digitalisation and annotation process.

In order to facilitate transcriptions, we use the *Transcriber* software. In an attempt to forestall labour-intensive post-editing, a number of additional features have been added that allow us to incorporate as many of the ICE transcription and annotation principles as possible.

4 *The written section of ICE-Malta*

The most advanced subsections of the written part are press reports and editorials. They were collected from the internet websites of four major daily Maltese newspapers published in English: *Malta Today*, *The Independent*, *Business Today*, and *The Times of Malta*. The individual texts were selected on the basis of the following criteria:

- publication after 2005
- exclusion of (parts of) news agency publications
- preference for the *Local News* section (or the *Opinion* section for the editorials)
- Maltese author (preference given to typical Maltese surnames like *Vella*, *Borg* etc.)

The first criterion is in accordance with the accepted time gap between the first-generation ICE components (UK, NZ etc.) and those more recent ICE components currently under compilation (Fiji, Trinidad, Nigeria etc.). This entails the disadvantage that comparisons of first and second generation ICE components have to be interpreted with caution from a synchronic perspective. The ideal world for corpus-oriented variationists should therefore be the recompilation of the first generation corpora. This would add a diachronic dimension to comparative studies of varieties of English around the world and it would make synchronic comparisons more reliable and up to date. The latter three criteria mentioned above (exclusion of news agency publications, preference for the *Local News* section, Maltese authors) are related and help to ensure that the texts selected are genuine examples of the English produced in Malta, since the articles in the *Local News* sections are almost exclusively covered by Maltese authors, whereas texts in other sections of the newspapers tend to consist of or to be based on British or international news agency releases.

The press report and editorial material has been expanded to 100,000 words for an extra-ICE Maltese newspaper corpus, for which parallel corpora for the UK, the Channel Islands and Gibraltar have been (or, in the latter case, are being) compiled. (These varieties are included in the Bamberg project, even though no complete separate ICE component is planned for the varieties other than Maltese English.)

Further relatively unproblematic text types for ICE-Malta are, firstly, the instructional administrative texts, since in Malta such texts are categorically published in both official languages, Maltese and English. Secondly, academic writing is easily available, since many researchers in Malta publish part of their

work with *Malta University Press*. This publisher, however, has a rather strong quantitative bias towards the humanities; technology and the natural sciences are less widely covered, which reflects the University's organisational and academic structure.

Thirdly, thanks to the cooperation with the University of Malta departments of Linguistics and English, the students' essays and exam scripts include writers from both linguistic and non-linguistic fields. This is possible because both departments are involved in the teaching of academic English for students from all subjects and fields. Fourthly, and in line with the other more recent ICE corpora discussed in the present volume, social letters will be replaced by emails (cf. e.g. Mukherjee *et al.* this volume), which will be compiled during a field trip to Malta in November and December of 2009 and the student tandems that are expected to result from this.² As a consequence, the texts in this subsection will be largely restricted to those produced by university students.

Several sections have proven to be more problematic, mainly on account of the small range of available texts: the criteria for the compilation of novels had to be extended, so that texts produced before 2005 (but no earlier than 1995) can be included. This is largely due to the fact that the vast majority of contemporary authors in Malta prefer to write in Maltese rather than English, while English fiction is imported from the UK.

Generally, a major problem arising in the compilation process is that a lot of specifically Maltese writing is done in Maltese, since texts on non-Maltese topics are available from the UK. This is particularly salient for the subsections of popular writing and instructional writing (skills, hobbies). It seems, though, that there will be enough material for these sections if concessions are made with regard to the variety of topics and authors.

The final remaining (and so far unresolved) problem is presented by the text category of business letters, for which the specifically Maltese challenges (i.e. the preference for Maltese rather than English as the medium of communication in this field) combines with general legal and privacy issues.

As for the authenticity and usefulness of the written data, there is one convenient peculiarity for future data collection: editing in Malta is not done extensively and in many cases not normatively, which is largely due to limited economic resources and a small community as target readership. Thus, a range of non-standard features, either general second-language or specifically Maltese features can be found rather frequently in the published written texts, for instance individual words such as *foreward* (for *foreword*), various morphological and syntactic features such as the pluralisation of mass nouns, non-standard use of articles or affixation resulting in words like *competitivity* (for *competitiveness*).

The current status (as of November 2009) of the written section of ICE-Malta is summarised in Table 2:

Table 2: The written component of ICE-Malta (as of Nov. 2009)

Written component	Collection	Digitalisation	Annotation
Student Essays (10) Exam Scripts (10)	almost completed expected	almost completed	
Social Letters (15) Business Letters (15)	scheduled for Nov.		
Academic (40)	completed	in progress	
Popular (40)	completed	in progress	
Press reports (20)	completed	completed	
Instructional (20)	in progress	in progress	
Editorials (10)	completed	completed	
Novels (20)	completed	in progress	

5 *The spoken section*

The compilation of the spoken section is subject to the challenges arising from the sociolinguistic situation in Malta (see Section 2 above). Most everyday conversations among the Maltese speech community take place in Maltese, which is also the language of the government and the court. Most TV stations broadcast exclusively in Maltese, since English programmes are available from British channels such as BBC, ITV and others (the same is true for Italian programmes). Thus, apart from very few individual shows, broadcasting in English is almost completely limited to interviews with foreigners, resulting in material that is inappropriate for our corpus. And finding the rare individual shows broadcast in English has so far been the major challenge for the spoken section.

This situation affects the subsections broadcast discussions, news and interviews. To some extent, the latter section might be covered by interviews broadcast in the form of podcasts. This could be regarded as a parallel strategy to finding other extensions of text types in the newer ICE components (like emails replacing social letters, as mentioned above).

Broadcast talks in English, on the other hand, were made available by the Campus FM radio station, located at and directed to listeners at the University of Malta. These talks are mostly but not exclusively rather academic. They cover a broad range of topics, including contributions from the Maltese Archaeological Society and the Maltese Historical Society, discussions of classic literary texts and theatre performances, and they include two more general series: *It's Debatable* and *Research Matters*. These sound files are currently being transcribed.

Even more problematic, if not impossible, is the collection of data for the categories parliamentary debates, legal presentations and legal cross-examinations, since these are exclusively done in Maltese. For the first of these categories we have contacted Maltese Members of the European Parliament (MEPs), but since Maltese was granted the status of an official EU language as early as 2002, the vast majority, if not all speeches and debates, are held in Maltese. Translations by Maltese interpreters could be made available, but this would probably stretch the category beyond the boundaries originally envisaged by the corpus designers. Possible solutions to this pending problem would be to collect data from text types as similar as possible to the original category, or to increase the number of texts of one or more of the other categories in the spoken section.

As to the other categories, private conversations will be recorded during the field trip mentioned above, when German students will record half-hour stretches of English conversation between Maltese university students and members of their families as well as their friends. The drawback of this method is that in both these contexts the language of choice would be Maltese for the overwhelming majority of the population (see Section 2), so that a certain unnaturalness attaches to conversations in English. This, however, is a stock problem for ESL corpora. Furthermore, in the presence of a speaker with no competence in Maltese, the language of choice in Malta is English, so that the situation is not entirely artificial.

Similarly to the student essays in the written category, class lessons at the University of Malta will be recorded by or with the help of our Maltese colleagues. Since the Institute of Linguistics has courses for students from all academic fields, a variety of topics and language backgrounds will be available.

Table 3: The spoken component of ICE-Malta (as of Nov. 2009)

Spoken component	Collection	Transcription	Annotation
Conversations (90) Phonecalls (10)	scheduled for Dec.		
Class Lessons (20)	in progress		
Broadcast Discussions (20)	in progress	in progress	
Broadcast Interviews (10)			
Parliamentary Debates (10)	EP meetings?		
Cross-examinations (10)			
Business Transactions (10)			
Commentaries (20)			
Unscripted Speeches (30)	in progress	in progress	
Demonstrations (10)			
Legal Presentations (10)			
Broadcast News (20)	in progress	in progress	
Broadcast Talks (20)	completed	in progress	
Non-broadcast Talks (10)	in progress	in progress	

6 Conclusion and outlook

This report has shown both the progress and the challenges in the compilation of the Maltese component of the International Corpus of English – challenges that probably many compilers have faced or will face who have chosen an ESL-variety as their contribution to the ICE project. In certain domains of usage, English as a second language is by its very nature excluded or employed only by a minority of the speech community. What turns a challenging project into an attractive prospect, however, is that this corpus – like any other second-language ICE component for its respective variety – is the first attempt to provide a systematic database of a not yet codified variety of English which develops in an exceptional contact situation. With its geographical setting and its history of language contact with Semitic and Romance languages and dialects, Maltese English is unique among the second language varieties of English and can thus be expected to expand the spectrum of the specific and general variation found in the so-called ‘New Englishes’ significantly.

Notes

1. The proportion of people claiming English to be their only native language is roughly 1 per cent.
2. Students from the German field trip group and from a Maltese university class have formed mixed German-Maltese mini groups, typically consisting of two people (maximum: four). The purposes are twofold: (i) helping the German students to record conversations for the ICE-Malta spoken section, and (ii) helping the Maltese students to write essays on Germany-related topics for an academic writing course. Email exchange between the two groups has started before the study trip to Malta and is to continue afterwards.

References

- Camilleri, Antoinette. 1991. Crosslinguistic influence in a bilingual classroom – the example of Maltese and English. *Edinburgh Working Papers in Linguistics* 2: 101–111.
- Camilleri, Antoinette. 1992. The sociolinguistic status of English in Malta. *Edinburgh Working Papers in Linguistics* 3: 14–24.
- Camilleri, Antoinette. 1995. *Bilingualism in education: The Maltese experience*. Heidelberg: Julius Groos.
- Camilleri, Antoinette. 1996. Language values and identities: Code switching in secondary classrooms in Malta. *Linguistics and Education* 8: 85–103.
- Census of population and housing 2005: Preliminary report*. 2006. Valletta: National Statistics Office. Source: http://www.nso.gov.mt/statdoc/document_file.aspx?id=1653
- Greenbaum, Sidney. 1991. *International Corpus of English. The compilation of the International Corpus of English and its components*. London: Survey of English Usage.
- Greenbaum, Sidney (ed.). 1996. *Comparing English worldwide. The International Corpus of English*. Oxford: Clarendon Press.
- Mazzon, Gabriella. 1992. *L'inglese di Malta*. Naples: Liguori Editore.
- Ministry of Education. 1999. *National minimum curriculum*. Malta: Ministry of Education.
- Nelson, Gerald. 1996. The design of the corpus. In S. Greenbaum (ed.). *Comparing English worldwide. The International Corpus of English*, 27–35.

- Schneider, Edgar. 2007. *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.
- Sciriha, Lydia. 2001. Trilingualism in Malta: Social and educational perspectives. *International Journal of Bilingual Education and Bilingualism* 4 (1): 23–37.
- Sciriha, Lydia and Mario Vassallo. 2001, repr. 2003. *Malta: A linguistic landscape*. Malta: Socrates.
- Sciriha, Lydia and Mario Vassallo. 2006. *Living languages in Malta*. Malta: Print It Printing Services.
- Thusat, Joshua, Emily Anderson, Shante Davis, Mike Ferris, Amber Javed, Angela Laughlin, Christopher McFarland, Raknakwan Sangsiri, Judith Sinclair, Victoria Vastalo, Win Whelan and Jessica Wrubel. 2009. Maltese English and the nativization phase of the Dynamic Model. *English Today* 25: 25–32.
- Vella, Alexandra. 1994. Prosodic structure and intonation in Maltese and its influence on Maltese English. Unpublished PhD Thesis. University of Edinburgh.