

MalToBI – Building an Annotated Corpus of Spoken Maltese

Alexandra Vella¹ & Paulseph-John Farrugia²

¹Institute of Linguistics, ²Department of Computer Science
University of Malta, Malta

sandra.vella@pt.lu, paulseph@paulseph.com

Abstract

Research on the phonetics and phonology of Maltese, and in particular on different aspects of its prosody, is, thus far, rather limited. This is in part due to the lack of structured resources for use in research. One resource which, to date, has been unavailable, is a corpus of spoken Maltese. Such a corpus, could, amongst other things, be used as a ready resource for the analysis of various aspects of the phonetics and phonology of Maltese. Moreover, given the limited research in the area of Maltese prosody, a speech corpus including an element of prosodic annotation could be particularly important for the continuing development of resources in the domain of Text-to-Speech (TTS) in the local context.

Recognition of the requirement of such a corpus gave rise to MalToBI, a project involving the collection of a relatively small body of spoken Maltese, together with the development of a *Tone and Break Indices* (ToBI) framework adapted for use with Maltese.

This paper gives a brief outline of some aspects of the prosody of Maltese and describes the development of this corpus. It discusses design considerations in some detail, and it outlines the progress made so far as well as the intentions for future work.

1. Introduction

Maltese prosody has received relatively limited academic attention. The list of works providing formal analyses thereof is quite short ([5], [19], [20], [21]). There is also a lack of resources to support such research, as well as little in the way of agreed-upon formalisms for annotation ([8]).

The MalToBI project outlined in this paper aims to address these issues through the collection of a corpus of spoken Maltese intended to make available a structured speech corpus, annotated through a version of ToBI ([17]) adapted for use with Maltese. This follows in the steps of similar efforts undertaken in the first instance for English ([3]), and later for other languages such as German ([2]) and Greek ([1]). The choice of ToBI also seems a natural one in the context of previous research on Maltese prosody ([19], [20]) which was undertaken using a tone sequence approach in the tradition of [14].

The first phase of the project has seen the compilation of the design of the corpus content, as well as of the recording of a select set of speakers of the language. The corpus is currently in the process of being analyzed and annotated for the purpose of formalizing the necessary adaptation to the ToBI framework and providing guidelines for use with Maltese.

This paper first describes some aspects of Maltese prosody in Section 2. It then proceeds to outline the steps involved in the corpus design in Section 3, while Section 4 documents the process of selecting the participants involved

in the subsequent recording of data. Finally, a summary of the status of the project is given in Section 5, including a sample analysis of two utterances from the corpus and a preliminary set of standards for a Maltese version of ToBI, MalToBI.

2. Maltese Prosody

Maltese is an intonation language in which tone is not lexically distinctive. Stress is also not generally used to distinguish between words, although a small set of bisyllabic minimal pairs exist in which stress on the first syllable, as opposed to the second, serves to distinguish a verb from a nominal form ([5]). A brief account of stress in Maltese is given in 2.1. This is followed by a brief description of Maltese intonation in 2.2 and by an indication of requirements for MalToBI in 2.3.

2.1. Stress in Maltese

All multisyllabic words in Maltese have one primary stress. Word stress generally occurs on the penultimate syllable, unless another syllable has a vowel realized as long, a morpheme suffix causing stress shift or is offset with either a geminate consonant or a consonant cluster ([5]).

It has been suggested in [21] that Maltese can be said to make a distinction similar to that suggested, for example, by [11] with respect to Palestinian Arabic, between light (C)V, heavy (C)VV/(C)V: and (C)VC, and superheavy (C)VVC/(C)V:C and (C)VCC syllables (where VV and V: represent a diphthong and a long vowel respectively, and CC represents either a consonant cluster or a geminate). Lexical stress in Maltese can then be said to be assigned to any syllable which is [+heavy] within the word.

In the case of final stress, the requirement of a final [+heavy] syllable is all that is necessary for stress to be assigned, as in the case of words such as “libertà” (freedom), “Mulej” (Lord), “verità” (truth) and “virtù” (virtue) ending in a final open syllable having the structure (C)VV/(C)V:. By contrast words ending in a closed syllable can only be stressed if this syllable happens to be [+superheavy] as in “hassejt” (I felt), CVVC, “Maltin” (the Maltese), CV:C and “berikt” (C)VCC. In words containing no [+heavy] or [+superheavy] syllable, stress is assigned penultimately by default.

Phrase stress in Maltese tends to gravitate as close to the end of the phrase as possible.

2.2. Intonation in Maltese

The body of work on the intonation of Maltese is not very extensive. [5] gives a characterization of different tunes that can be realized within the language in terms of a generic description of pitch rises and falls. However, the description in [5] is not set within a formal framework.

A more formal characterization of the intonation of Maltese can in fact be found in [19], which outlines a preliminary description of prosodic structure and intonation in Maltese using a tone sequence framework similar to that developed in [14] for English. Vella ([19]) originally identified H^*+L and L^* pitch accents which occur in nuclear or [+focus] position in Maltese, as well as a further L^* and $L+H^*$ which are associated with post-nuclear or [-focus] stretches of speech. Boundary tones L_P , H_P , L_I and H_I were also identified. As indicated by their subscripts, the latter are differentiated according to their attachment to two different phrase types, namely “phonological phrases” (P-phrases) and “intonational phrases” (I-phrases). Consequently, the identifiable tunes in Maltese were classified as H^*+L , L^* , $L+H^*$, L_P , H_P , L_I and H_I .

A prenuclear H^* was also found to occur preceding the nuclear tunes L^* , H_P or H^*+L within the P-phrase. [20] reinterpreted the tune $L+H^*$ as being better represented as $L+H-$ and L^*H_I as being better represented as $L-$, the dash representing a phrase accent. This redefinition arises from work on other similar “edge tones with a secondary association to an ordinary tone-bearing unit”, a phenomenon which has been shown to be common in a variety of European languages in [10] (pp 180). In both cases, a following H boundary tone is also possible.

2.3. Maltese Intonation and MalToBI

The description of Maltese intonation in [19] and [20] needs to be tested and refined by the exercise of building a set of ToBI framework annotation conventions for the language. The requirements for a Maltese version of ToBI, or MalToBI, are thus the following:

- The design and recording of a structured speech corpus.
- The formalization of the intonational descriptions available in previous research for use within the ToBI framework.
- The establishment of a break index scheme.
- The annotation of the recorded speech database.
- The compilation of annotation guidelines.
- The testing for inter-transcriber reliability.

The remaining sections in this paper report on the first step listed above, in other words, on the design and recording of a structured speech corpus of Maltese.

3. Corpus Design

The corpus is designed in such a way as to include data from each speaker in four different speech styles ranging from more formal text reading, through to more informal, quasi-conversational, speech.

The first two sections of the material intended as the basis for data collection involved individual speakers. In the first instance, speakers were required to read out a sequence of mixed utterances consisting of statements, imperatives, yes-no questions and tag questions. This section of the speech material constituted the most formal speech style in that it required speakers to read utterances out of context. The advantage of including the utterances in this set is that they yield data which are directly comparable.

Each individual speaker was then required to read out a short story. The story chosen was a Maltese version of the classic fable “The North Wind and the Sun” (“Ix-Xemx u r-

Rih”, [22]). This section of the data allowed the capture of a narrative style of speaking.

Working in pairs, speakers were then required to complete a map task inspired by the work in [12]. (Other approaches to the elicitation of natural sounding data, such as that undertaken in [18] which involved the design and use of a board game for a similar purpose, could also have been used.) The map task involved the use of two maps with selected, partly fictitious, Maltese place names.

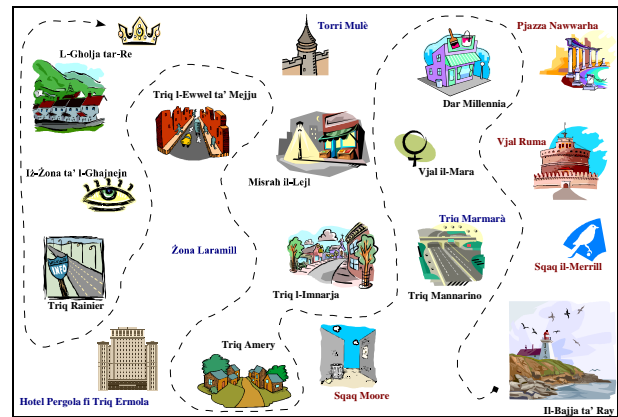


Figure 1: Leader's Map

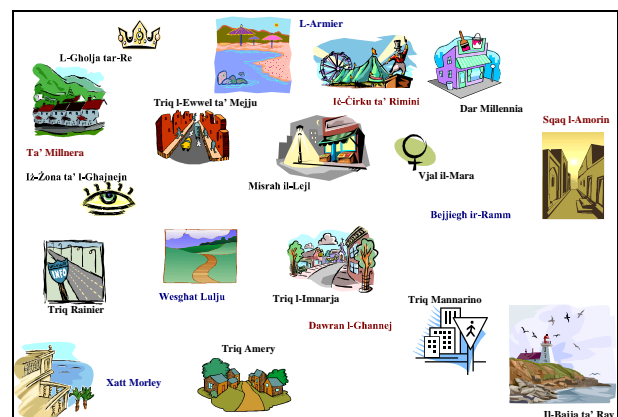


Figure 2: Follower's Map

The target items included in the maps were contained within phrases clearly identifiable to speakers as having a place name function in Maltese such as “Pjazza X” (“A Square named X”) “Sqaq Y” (“Y Alley”), “Triq Z” (“Z Street”), or slightly less commonly used, but also possible, phrases such as “Iż-Żona ta' A”, (“Zone A”) or “Bejjiegh il-B” (“The B Seller”). The target items themselves consisted of mostly real, but, in a small number of cases, made-up words, consisting solely of sonorants. Since it is to these target items within each of the phrases that speakers would be expected to assign stress, it was envisaged that F0 prominence realizations on these items would be easier to examine than had the target items also included obstruent segments.

Included amongst the final list of target phrases are target items involving words of native origin such as “l-ghajnejn” (the eyes) and “l-Innarja” (the traditional feast of St. Peter and St. Paul celebrated on the 29th of June each year), as well

as words of non-native origin such as English surnames which have long been used in actual street names, such as “Amery,” and nativized surnames of Italian or other Romance origins such as “Marmarà”. Resorting to such items was necessary in order to be able to come up with words conforming to the restriction of obstruent-free target items, particularly in the case of items having antepenultimate stress, since the latter is a relatively recent stress possibility in Maltese ([21]).

The final list of target items included in the maps consists of a balanced set of monosyllabic, disyllabic and polysyllabic (3 or 4 syllable) words, including as wide a range as possible of different syllable types for items having final, penultimate and antepenultimate stress. Some examples of the target phrases are listed below with the target items underlined and the stressed syllable within each target item indicated in bold:

- Il-Bajja ta' **Rav** (Ray's Bay)
- Iż-Żona ta' l-**Ghajnejn** (The Eye Zone)
- Triq **Marmarà** (Marmarà Street)
- Vjal il-**Mara** (Women's Avenue)
- Triq **Rainier** (Rainier Street)
- Triq l-**Imnarja** (L-Imnarja Street)
- Triq **Mannarino** (Mannarino Street)
- Triq **Amery** (Amery Street)
- Iċ-Ċirku ta' **Rimini** (Rimini Circus)
- Ta' **Millnera** (Millnera's Place)

One of the maps, Fig. 1, contained a path and was designed to be used by a task 'leader'. The other map did not have any path markings and was purposely designed to be slightly different from the leader's map, thus allowing for various possible information gaps requiring negotiation between the participants.

The latter map, illustrated in Fig. 2, was meant to be used by a 'follower' without access to the original path. Participants are required to co-operate in such a way as for the leader to verbally lead the follower through the various 'obstacles', thus enabling the latter to draw a corresponding path on her/his map. The immersion of both leader and follower in the endeavour is meant to provide a condition that elicits quasi-natural speech whilst ensuring the occurrence of the select list of target items embedded within the place names.

The fourth and final section of the corpus content consisted again of a paired speaker task, this time a task comprising an open, but topicalized, conversation. In this case, the speakers were required to pretend they knew someone who was prepared to give their partner a (fictitious) job and therefore needed to equip themselves with information about their partner to pass on to the person offering the job. Speakers took turns at simulating the different roles of the prospective employer's friend and job seeker.

4. Selecting and Recording the Speakers

Speakers of Maltese generally operate within an extremely complex linguistic environment in which English not only shares official status with Maltese, but, in certain domains such as the educational one, can possibly be considered as the predominant of the two languages, especially in its written form. For certain speakers, the situation is further complicated by the possible co-existence of a standard variety of Maltese with a dialectal form of the language. Given the complex linguistic environment in Malta, aspects of which

have been described in, amongst others, [6], [16] and [19], monolingual and/or monodialectal speakers of Maltese are therefore a relatively rare phenomenon.

Since the aim of this project was to collect a corpus of spoken Maltese, and given the relatively small size of the corpus, it was considered desirable to eliminate variation arising from intralinguistic dominance or intradialectal use in speakers. These were therefore carefully selected in such a way as to ensure as homogenous a corpus of standard spoken Maltese as possible, reflecting the speech of monodialectal speakers having Maltese as their dominant, and most frequent choice of language in everyday use.

In order to select suitable speakers of Standard Maltese, an email invitation was sent to a number of potential participants. These were requested to complete a questionnaire aimed at identifying their language bias. In addition to generic background details (such as sex, age, level of education, area of residence and use of other languages or dialects), the questionnaire also included a 5-point scale self-assessment section covering the four language skills of listening, speaking, reading and writing in both Maltese and English. The self-assessment section in the questionnaire is based on a self-rating test outlined in [15] and adapted successfully in the selection of participants having different language backgrounds by [7]. The absolute difference between the two languages for the corresponding averaged sum of points across the four skills in the self-rating assessment was used to determine speakers' dominant language.

Of the respondents, a total of 16 speakers were selected, half of these male and half female. These were in turn selected to contain a balanced number of speakers from within the age groups 18-29 and 30+, and having primary, secondary, post-secondary and tertiary levels of education. Recording sessions were carried out in male-female pairs within a professional recording studio.

Recordings were anonymized and stored in a structured set of 16 bit, 44.1kHz PCM WAV files. These constitute 1.1 GB of data, providing around 2 and a half hours of recorded speech data.

5. Project Status

With reference to the requirements for the development of a ToBI framework for Maltese discussed above, one can conclude that the speech corpus content design and recording have been completed. The next step sees the descriptions of prosodic structure and intonation in [19] and [20] being used as the basis for a preliminary annotation of the recorded data, thus providing groundwork for the formalization of MalToBI.

While the latter process is still in the very early phases of development, some early indications of the formalisation are presented through samples annotated in ToBI style using Praat ([4]) illustrated in Fig. 3 and Fig. 4 on the next page.

The tiers included in the analysis are a *Text* tier, a *Transcription* tier, a *Tone* tier and a *Break Index* tier. The *Text* tier contains an orthographic rendering of the utterance. Capitalisation is used to indicate that a syllable is accented within the phrase, while an asterisk preceding a capitalised stretch is used to show that the accented syllable is the one assigned [+focus] in the given phrase.

The *Transcription* tier provides an IPA phonetic transcription of the utterance utilising the Maltese phonemic inventory and transcriptional conventions suggested in [5].

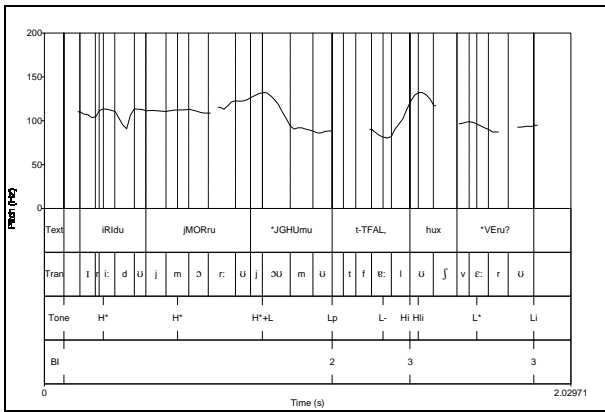


Figure 3

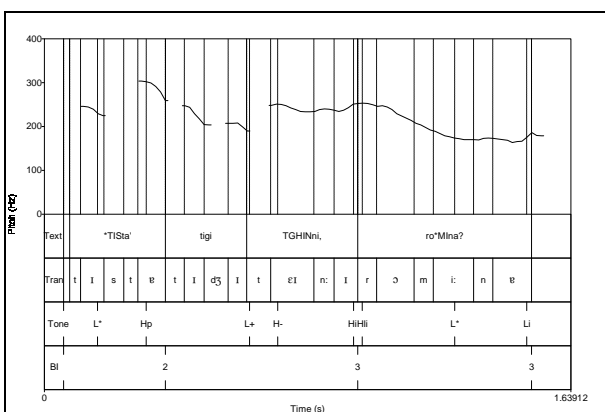


Figure 4

The *Tone* tier contains the intonational analysis. At this stage, a preliminary inventory of tones is being postulated on the basis of the analysis of Maltese intonation to date in [19] and [20]. Thus, nuclear H^*+L or L^* pitch accents occur in the context of [+focus] stretches of speech. These can be followed by $L-$ or $L+H-$ phrase accent tones and/or by L_P or L_I , H_P or H_I boundary tones, in line with the distinction between boundary tones which attach to P- or I-phrases suggested in [19]. An L^* pitch accent is also possible within a phrase having an H boundary tone at its left edge, designated as H_I and an L_I at its right edge. A prenuclear H^* is also possible.

As regards *Break Index* features for Maltese, further research is required before a comprehensive scheme can be formulated. Indices as follows have been used in the preliminary analysis presented here:

- 1 to mark the boundary between prosodic words.
- 2 to mark the edge of intermediate phrases, referred to as phonological phrases in [19] and [20].
- 3 to mark the edge of intonational phrases.

In the example in Fig. 3, the [+focus] “jghumu” ([they want to go] to swim) is assigned an H^*+L , with an L_P associated with the boundary of the phrase. This is followed by an $L-$ H_I sequence on the [-focus] “it-tfal” (the children). A separate intonational phrase follows containing

the positive tag “hux veru?” (isn’t that right?). The latter is assigned the sequence $H_I L^* L_I$.

The example in Fig. 4 involves use of the $L^* H_P$ sequence on the [+focus] “tista” (can you) followed by an $L+H-$ phrase accent on the [-focus] “tigi tghinni” (come and help me) and a final H_I . This is followed by a separate intonational phrase containing the vocative “Romina” and a sequence $H_I L^* L_I$.

The experience and insight gained from this endeavour can be expected to provide the basis for the compilation of annotation guidelines. The resulting framework will eventually require validation for inter-transcriber consistency, in the manner discussed in [13].

6. Conclusion

This paper presented the motivation for and the current state of MalToBI, an ongoing project intended to put together a corpus of spoken Maltese to be used as a structured resource for research on Maltese prosody, as well as an important element in the continuing development of TTS resources in the local context. This paper highlights, in particular, the progress so far in terms of the design of the corpus content and recording of a speech corpus of Maltese. It is the authors’ intention to eventually make the results of the project freely available on line.

7. References

- [1] Arvaniti, A.; Baltazani, M. 2000. Greek ToBI: a system for the annotation of Greek speech corpora. *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC)*. Athens, volume 2, 555-562.
- [2] Baumann, S.; Grice, M.; Benzmüller, R. 2000. GToBI – a phonological system for the transcription of German intonation. In *Prosody 2000: Speech Recognition and Synthesis*, Puppel, Stanislaw & Grazyna Demenko (eds.). Poznan: Adam Mickiewicz University, Faculty of Modern Languages and Literature, 21-28.
- [3] Beckman, M.; Elam, G.A. 1997. *Guidelines for ToBI Labeling*. 3.0 edition. The Ohio State University Research Foundation.
- [4] Boersma, P.; Weenink, D. 2006. *Praat: doing phonetics by computer* (Version 4.4.11) [Computer program]. Retrieved February 23, 2006, from <http://www.praat.org/>
- [5] Borg, A.; Azzopardi-Alexander, M. 1997. *Maltese. Descriptive Grammars*. London and New York: Routledge.
- [6] Camilleri, A. 1995. *Bilingualism in Education: the Maltese experience*. Heidelberg: Julius Groos Verlag.
- [7] Dandria, R. 2002. *Lexical Access and Prototypicality in the English of Maltese Bilingual Speakers*. Undergraduate dissertation, University of Malta.
- [8] Farrugia, P-J. 2005. *Text-to-Speech Technologies for Mobile Telephony Services*. Master’s dissertation, University of Malta.
- [9] Gatt, A. 2003. *Annotating Textual and Speech Data in Maltese*. Technical report, University of Malta. (Compiled in response to an International Standard Organisation call for contributions towards the construction of a Morphosyntactic Annotation Framework.)

- [10] Grice, M.; Ladd, D. R.; Arvaniti, A. 2000. On the place of phrase accents in intonational phonology. *Phonology*, volume 17, 143-185.
- [11] Hayes, B. 1995. *Metrical Stress Theory*. Chicago: Chicago University Press.
- [12] HCRC map task corpus. <http://www.hcrc.ed.ac.uk/maptask>.
- [13] Pitrelli, J.; Beckman, M.; Hirschberg, J. 1994. Evaluation of prosodic transcription labeling reliability in the ToBI framework. *Proceedings of the 1992 International Conference on Spoken Language Processing (ICSLP-92)*. Banff, Canada, volume 1, 123-126.
- [14] Pierrehumbert, J. 1980. *The Phonetics and Phonology of English Intonation*. Ph.D. thesis, Massachusetts Institute of Technology.
- [15] Romaine, S. 1995. *Bilingualism*. 2nd edition. Oxford: Blackwell.
- [16] Sciriha, L.; Vassallo, M. 2001. *Malta – a Linguistic Landscape*. Malta: Caxton Printshop.
- [17] Silverman, K.; Beckman, M.; Pitrelli, J.; Ostendorf, M.; Wightman, C.; Price, P.; Pierrehumbert, J.; Hirschberg, J. 1992. ToBI: a standard for labeling English prosody. *Proceedings of the 1992 International Conference on Spoken Language Processing (ICSLP-92)*. Banff, Canada, volume 2, 867-870.
- [18] Speer, S.; Warren, P.; Schafer, A. 2003. Intonation and sentence processing. *International Congress of Phonetic Sciences*. Symposium paper, Barcelona.
- [19] Vella, A. 1995. *Prosodic Structure and Intonation in Maltese and its Influence on Maltese English*. Ph.D. thesis, University of Edinburgh.
- [20] Vella, A. 2003. Phrase accents in Maltese: distribution and realization. *International Congress of Phonetic Sciences (ICPhS)*. Barcelona, 1775-1778.
- [21] Vella, A. 2003. Language contact and Maltese intonation. In *Aspects of Multilingualism in European Language History*, Braunmüller and Ferraresi (eds.). Amsterdam/Philadelphia: John Benjamins Publishing Company, 261-283.
- [22] Vella, E. B. 1934. *Ġabra ta' Ward – l-Ewwel Ktieb*. Oxford: Oxford University Press.