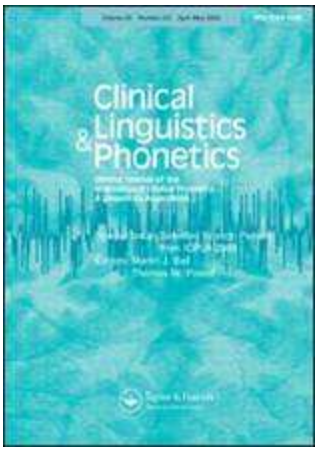
Article:
Balzan, P., Vella, A. and Tattersall, C. orcid.org/0000-0001-9796-7733 (2019) Assessment of intelligibility in dysarthria: development of a Maltese word and phrase list. Clinical Linguistics & Phonetics. ISSN 0269-9206

https://doi.org/10.1080/02699206.2019.1594383



Assessment of Intelligibility in Dysarthria: Development of a Maltese Word and Phrase List [ICPLA]

Journal:	*Clinical Linguistics & Phonetics*
Manuscript ID	TCLP-2019-0035.R1
Manuscript Type:	Original paper
Keywords:	dysarthria < speech disorders, intelligibility < speech, assessment, linguistics < language

Abstract

This paper describes the development of the Maltese Intelligibility Lists (MIL) for the assessment of word and phrase intelligibility in dysarthria. Two main tools were employed: the Frenchay Dysarthria Assessment-2 (FDA; Enderby & Palmer, 2008), and the Maltese Language Resource Server (MLRS). Three main criteria served as the basis for the construction of the word and phrase lists: frequency of occurrence of Maltese phonemes, word frequency and an analysis of syllable types and structures. The most common 500 words in the MLRS corpus (Korpus Malti v. 3, 2016) were broadly transcribed and an analysis of different types of syllables and their frequency of occurrence was carried out. Based on this analysis, the relevant proportion of different syllable types required for the word and phrase lists for Maltese was calculated in line with the number of items present in the FDA-2. With regards to phoneme frequency, the words chosen demonstrate a similar short-vowel and consonant distribution as reported in a previous large scale study. The MIL consists of 116 words and 50 phrases which are representative of Standard Maltese and can be used in the clinic to assess speech intelligibility in Maltese individuals with dysarthria.

Introduction

Dysarthria is an umbrella term for a group of disorders reflecting impairments in the strength, speed and precision of movements required for adequate control of the respiratory, phonatory, resonatory, articulatory and prosodic aspects of the speech mechanism (Duffy, 2013). The speech disorder can be classified into seven distinct subtypes, each characterised by diverse atypical features of speech production that underlie different neurological localisations and pathophysiology (Duffy, 2007). Irrespective of the underlying cause of the motor speech disorder, one of the principal consequences of all subtypes of dysarthria is diminished speech intelligibility (Dykstra, Hakel, & Adams, 2007).

Kent (1992) views speech intelligibility as forming part of communicative competence and 'is the sine qua non of spoken language' (p. 9). It is described as a measure of the degree of clarity of the speaker and comprehensibility of the spoken message by the listener (Yorkston, Stand & Kennedy, 1996). Recent views of the disorder extend the traditional focus on atypical physiological deviations in speech production to include issues of intelligibility and their implications on an individual's everyday functioning in society (Dykstra et al., 2007).

The clinical assessment of intelligibility in dysarthria

Given the relevance of intelligibility in both assessment and treatment of dysarthria, the collection of baseline intelligibility data commonly carried out in the speech-language clinic serves to index the degree and severity of impairment, to recommend direct intervention goals and to monitor disease progression or improvement (Gurevich & Scamihorn, 2017). Intelligibility testing often involves a variety of subtests all offering valuable insights into the influence of dysarthria on speech production. Phoneme and word intelligibility using sounds in isolation or minimal pairs

may equip the clinician with specific data on articulatory breakdown. By contrast, testing of intelligibility using single phrases, sentences embedded in a narrative passage or spontaneous speech may yield a better understanding of the influence of dysarthria on everyday communicative exchanges and interactions (Lillvik, Allemark, Karlstrom & Hartelius, 1999). Tasks included in the assessment of intelligibility often include reading or repeating of speech stimuli and spontaneous production of discourse (Kempler & Van Lancker, 2002; Weismer & Laures, 2002; Yorkston & Beukelman, 1978).

Intelligibility assessments should: (1) offer quantitative and qualitative data regarding atypical speech features; (2) consist of both words and phrases; (3) include a relatively large pool of test items to avoid familiarisation bias; (4) and be language-specific (Lillvik et al., 1999). Available perceptual standardised tests of dysarthria, such as the Frenchay Dysarthria Assessment-2 (FDA-2; Enderby & Palmer, 2008) include measures of intelligibility as one of the assessment sections. This assessment uses interval rating scales to estimate or rate the degree of impairment.

For the word intelligibility section of the FDA-2, Enderby and Palmer (2008) utilise a corpus of 116 phonetically balanced words. This relatively large pool of words reduces the possibility of examiners and individuals with dysarthria learning the words through frequent exposure. For sentence intelligibility, the authors make use of 50 carrier phrases and phonetically balanced key words. In addition, to control for the potential influence of word frequency effects on intelligibility, only words with a frequency of more than ten words per million were included.

Contrariwise, the Assessment of Intelligibility of Dysarthric Speech (ASSIDS; Yorkston & Beukelman, 1981) is specifically designed to evaluate intelligibility in single words and sentences by calculating percentage intelligibility, total speaking rate, intelligible words per minute

and communication efficiency ratio. During task administration, individuals are requested to read or imitate 50 select words and 22 sentences selected randomly.

The Maltese Language: some background

A detailed description of the linguistics of Maltese is beyond the scope of this paper. Nevertheless, some background information on elements of the language, as well as the context in which it is spoken, is provided with a view to aiding the reader's understanding.

Maltese and English are co-official languages of Malta. However, Maltese is the national language and remains the language spoken predominantly by most of the population. According to recent statistics, while 88% of the Maltese population can speak English, it is estimated that 98% of Maltese individuals speak Maltese (European Commission, 2012).

The Maltese language includes features from languages belonging to three different language families: Semitic, Romance and Germanic. Brincat (2004) and Spagnol (2011) suggest that Maltese consists of three strata. The Semitic stratum is the most salient as it underpins the phonology, morphology and basic lexicon of the Maltese language. The second stratum, the Romance, essentially encompasses lexical and syntactic structures. The last stratum stems from the period of British rule in Malta from 1800 to 1964 and involves predominantly lexical borrowing from English.

The vowel and consonant phoneme inventory of Standard Maltese as described by Borg and Azzopardi-Alexander (1997) is illustrated in Table 1. The Maltese language makes use of twenty-two consonants and eleven vowels, five short and six long, as well as seven diphthongs. Some phones, such as ʒ, but also dz, are sometimes argued to have peripheral status in the language. Nevertheless, since their status in the language is unclear (Borg & Azzopardi-Alexander

1997), they were not included in this study. Vowel length is phonemic and hence, all vowels are qualitatively and quantitatively distinct from each other (Azzopardi-Alexander, 2002).

Insert table 1 about here

Regarding syllable structure, Borg and Azzopardi-Alexander (1997) argue that the minimal syllable requirement in Maltese is a vowel. The maximum number of consonants within the onset is three and the maximum number of coda consonants is two. Thus, the maximal syllable in Maltese is: (C)(C)(C)V(C)(C) (Borg & Azzopardi-Alexander, 1997). Although research on the syllable in Maltese is continuing (see, e.g. Galea, 2016), estimates of the frequency of occurrence of different syllable types in Maltese are to date not available.

Rationale for current study

The limited linguistic research on Maltese poses significant challenges to clinical assessment generally, not least the assessment of intelligibility in Maltese individuals with dysarthria. Current local practice raises questions regarding the linguistic validity of utilised intelligibility measures in terms of their representativeness with respect to Maltese. Moreover, the lack of standard intelligibility lists limits comparability of within and between client evaluations.

This study constituted part of a larger study that aimed at adapting and normalising the FDA-2 for Maltese. When compared to the ASSIDS, the former assessment was deemed as encompassing a more comprehensive approach to the assessment of dysarthria. Hence, a Maltese

version of the intelligibility section of the FDA-2 was deemed necessary given the linguistic differences between Maltese and English.

The aims of the study are therefore twofold. Firstly, it attempts to provide an analysis of syllable structure and frequency of phonemes and words in Maltese. Secondly, it details the development of the Maltese intelligibility lists (MIL). The MIL will consist of useable Maltese word and phrase lists for the intelligibility section of the Maltese version of the FDA-2.

Materials and Methods

In line with the criteria proposed by Enderby and Palmer (2008) for the intelligibility section of the FDA-2, it was established that the developed Maltese lists should include 116 words and 50 phrases. Words included in the lists should account for word familiarity, phonetic distribution and the syllable types and structures of the Maltese language. To develop word and phrase lists which approximate the linguistic structures of Maltese, information on phoneme and syllable frequency and distribution is indispensable. To the authors' knowledge, there is no available data on the frequency of syllable types. In contrast, a study by Borg, Bugeja, Mangion and Gafà (2013) sheds light on the phoneme frequency of Maltese.

Frequency of occurrence of Maltese phonemes

To ascertain that the phonetic distribution of selected words for the word and phrase lists approximate that of Standard Maltese, it was established that the distribution of phonemes in both the word and phrase list be compared to the large scale study completed by Borg et al. (2013). A Chi-square test was used to determine whether there were statistically significant differences in the distribution of phonemes in the words included in the developed lists as compared to the frequencies reported by Borg et al. (2013).

Frequency of words

The criterion of frequency of words, as suggested in the FDA-2, was based on analysis of frequency of occurrence in the MLRS. This resource permits searches within a particular frequency range by using absolute frequencies. Gries (2010) argues that although absolute frequencies are the most basic corpus-based statistic, to be able to compare different corpora with each other, researchers need to normalise the frequencies to observed relative frequencies. Absolute frequencies are generally normalised and reported as frequencies per 1,000 or 1,000,000 words (Gries, 2010).

In order to identify words with 10 frequencies per million or more as presented in the FDA-2, absolute frequencies were normalised to a common basis of frequency of use per one million words using the mathematical formula:

$$\text{Frequency per million words} = \frac{\text{frequency}}{\text{number of tokens in corpus}} \text{X } 1{,}000{,}000$$

This equated to calculating the absolute frequency of a word, given its frequency per million, as follows:

$$\text{Frequency} = \frac{\text{frequency per million words}}{1{,}000{,}000} \text{ X number of tokens in corpus}$$

Details on the results of this analysis can be found in the section on Results and Discussion.

Analysis of syllable structure

A corpus of spoken Maltese is to date not available although some spoken details are included in the Maltese Language Resource Server (MLRS) corpus. Version 3 of the general corpus of Maltese in the MLRS, comprising mainly of written texts, was utilised to identify the most common syllable types in Maltese. This version of the MLRS (Korpus Malti version 3, 2016) contains 250 million tokens in a variety of sub-categories. All text categories available in the corpus, such as, academic, cultural, law, press and parliamentary debates were included in the search. However, to ascertain semantic neutrality, proper nouns and political, war or sexual words were excluded from the list.

A search was made to identify the most frequent 500 words in the general MLRS corpus. Syllable-by-syllable broad transcriptions were then made following Borg and Azzopardi-Alexander (1997). An analysis of the syllables in terms of their frequency and underlying CV structure was then carried out for the selected corpus, that is, the most frequent 500 words. A sample of the resulting analysis is presented in Table 2. In order to facilitate analysis, stressed syllables of transcribed words were always placed in the same database column.

Insert table 2 about here

The frequency of words containing different numbers of syllables was calculated. In other words, a count was made of the number of monosyllabic, disyllabic, trisyllabic etc. words, to be found amongst the most frequent 500 words. The location of lexical stress in these words was also noted. From this count, the number of words containing different numbers of syllables with

stress in different positions, mainly penultimate and final, required for each syllable type, was calculated using the formula:

$$\frac{\text{Frequency of words with the same number of syllables } \mathbf{X} \text{ Number of words in word/ phrase list}}{\text{Total number of words in selected corpus}}$$

The number of syllables of different types was then counted. The aim was to work out the proportion of the different types of syllables that should occur in the lists in order for them to be representative. This proportion was calculated using the formula:

$$\frac{\text{Frequency of words with the same syllable structure } \mathbf{X} \text{ Number of words in word/phrase list}}{\text{Total number of words in selected corpus}}$$

Prior to the selection of a set of word candidates for the final word and phrase lists, all words that did not match the criterion of familiarity were filtered out and excluded from the general MLRS corpus. Phoneme frequency was initially estimated through relative frequency of occurrence and then refined through the use of statistical methods. The proportional calculation of syllable structures to be included in the word list was followed rigidly in the wordlist.

Similar criteria were employed for the construction of words to be included in the phrase list. Although the number of words with different syllables was also calculated proportionally, syllable structures were not accounted for proportionally. The focus was primarily on the following supplementary criteria: phrases should contain between two and four words; the words in phrases should not exceed five syllables; articles should not be included in the syllable analysis; questions and exclamation phrases should be included to reflect naturalistic language production; and all phrases need to satisfy morphosyntactic rules. To permit future cross-linguistic

intelligibility assessment comparisons, the criteria for the phrase list were similar to those proposed by Enderby & Palmer (2008).

Results and Discussion

Frequency of occurrence of Maltese phonemes in the word and phrase lists

Percentage scores of the frequency distribution of Maltese phonemes in the developed word and phrase lists as compared to the Borg et al. (2013) study are presented in Table 3. The most common phonemes in MIL, ranked in order of frequency are /ɪ/, /ɐ/, /t/, /n/, /r/, /ɛ/ and /s/. These three vowels and four consonants make up one-fifth of the 33 Maltese phonemes and cover approximately 50% of all phonemes in the MIL. Although not in the same rank order, six out of the seven most common phonemes in the constructed lists match the most frequent phonemes in Borg et al. (2013). In contrast to the MIL, the Borg et al. (2013) results suggest a higher frequency for the phoneme /l/. One reason for this difference may be that the proclitic *il-* (equivalent to the definite article *the* in English) was included in the analysis carried out by Borg et al. but not in ours.

The least common phonemes in the words selected for inclusion in the MIL are /v/, /g/, /ts/, /ʧ/, /ʤ/, /ɛː/, /ɔː/ and /uː/. Each of the latter phonemes accounted for less than 1% of the total phonemes in the MIL. In contrast to Borg and Azzopardi-Alexander (1997), Borg et al. (2013) give full phonemic status to /dz/ and /ʒ/, including them in their frequency analysis. If we exclude these two sounds, which were not analysed in our study, the remaining six of the eight least common phonemes in the words selected for the MIL match with the least frequent phonemes in Borg et al. (2013).

Insert table 3 about here

Initial results evidenced significant differences between the phoneme distribution in the word list and the Borg et al. (2013) study, X^2 (34, N=151355575) = 406.92, p <.001. Similarly, the difference between the phoneme distribution in the phrase list and Borg et al. (2013) was found to be significant, X^2 (34, N=151355852) = 363.47, p <.001. This result was not what we had expected given the care that was taken to ensure representativity. We therefore looked again at the data.

Relative percentage differences in long vowel phoneme distributions were noted when comparing the results of the current study to those of Borg et al. (2013). These differences appear to be attributable to methodological differences between the two studies. While the current study employed manual transcriptions, Borg et al. (2013) made use of automated grapheme-to-phoneme conversion strategies which are likely to make mistakes in the case of vowel sounds for which a one-to-one sound-symbol correspondence is less consistent, notably the long vowels /i:/, /ɛ:/, /ɐ:/, /ɔ:/ and /u:/.

A Chi-square test was therefore performed again, this time excluding long vowels. When compared to Borg et al. (2013), the difference in phoneme distribution of short vowels and consonants in both the word and phrase list was found to be non-significant (wordlist - X^2 (26, N=144757089) = 30.87, p= .32; phrase list - X^2 (28, N=144757350) = 37.36, p = .11). Notwithstanding the fact that frequency of phonemes may be influenced by different syllables and words, it can be argued that when phonemes are counted in isolation, the short vowel and

consonant distribution of the lists developed adequately approximate the distribution of Standard Maltese.

Frequency of selected words

Given the size of the MLRS corpus - this contains 249256855 tokens - a word with 10 frequencies per million was found to have an absolute frequency of approximately 2492. In line with the FDA-2 criterion related to commonness of words which aims to control the effect of word frequency on intelligibility (Enderby & Palmer, 2008), only words with a frequency of more than 2492 were chosen for the MIL. Table 4 provides examples of words selected from the MLRS corpus that met the frequency criterion.

Insert table 4 about here

The application of this criterion and the subsequent filtering of words minimised the risk of selecting unfamiliar words to be included in the MIL and ascertain that the primary focus of the test is largely on intelligibility rather than on vocabulary or language competency. Nevertheless, given that the MLRS corpus makes use of words from academia, law and parliamentary debates, the possible inclusion of words not representative of everyday discourse cannot be fully excluded.

Syllable types and structures

Table 5 shows the frequency and respective percentages (corrected to 1 decimal place) of words having different numbers of syllables in the selected corpus and the developed word and

phrase lists. Disyllabic words are those which occur most frequently in these data (48.6%) followed by monosyllabic words (24.4%), trisyllabic ones (17.2%) and four-syllable words (7%). Not unexpectedly, five and six syllable words appeared less frequently. Six-syllable words were not considered for further analysis as their expected occurrence was below one word per list (word list–0.46; sentence list–0.60). This finding is reflected in the MIL as all included words do not exceed five syllables.

Insert table 5 about her

Unlike other speech tests that generally select the one or two most common syllable types to comprise a word or phrase list, the MIL includes words of all syllable types distributed proportionally according to frequency of occurrence. The decision not to match difficulty of words in the two lists ensures that factors crucial to the evaluation of speech execution and intelligibility are not excluded during testing. For instance, individuals with dysarthria who display impaired respiration may have more pronounced intelligibility deficits when producing long phrases as compared to shorter ones.

The most common syllable structures for words containing one to four syllables extracted from the selected corpus and their relative percentage of occurrence are presented in Table 6. Noticeable syllable occurrence patterns were noted for monosyllabic, bisyllabic and trisyllabic words. Contrariwise, patterns of syllable occurrence for four and five syllable words were not clear. While minor frequency variations were noted for four syllable words, all extracted five syllable structures had a minimal appearance of one in the selected corpus. For these syllable types

frequency of occurrence could not be calculated due to: (1) the small sample size of the selected corpus; (2) the low occurrence of four to five syllable words when compared to shorter syllable types; and (3) greater variability of syllable structures in these syllable types.

Insert table 6 about here

Tables 7 and 8, included in the Appendix, summarise the syllable structure types selected for inclusion in the final list of 116 words and the 50 phrases respectively. Principally, the variety of syllable structure included in the MIL guarantees coverage considered to be adequated, of the several syllable and word combination types that can occur in the language. Nevertheless, one cannot ignore the fact that the inclusion of significantly different structures in the lists may create an imbalance in the difficulty level involved in articulating the words and phrases.

General Discussion

In line with the FDA-2 criteria for the intelligibility section, the MIL consists of 116 words and 50 phrases. Three major factors were taken into consideration when developing the two lists. These are: (1) the frequency of Maltese phonemes; (2) word frequency; and (3) the number of syllables and syllable types in words. The analysis of phoneme distribution was carried out in line with a previous large scale study of the frequency of Maltese phonemes. Broad phonetic transcription and syllabic analysis of the most common 500 words taken from an online Maltese corpus were completed. Absolute frequencies were normalised to per-million frequencies so that only words with a frequency of at least 10 per million are included in the MIL. The frequency of occurrence of words in the MLRS was used as the main criterion for establishing familiarity.

This is the first known attempt at developing two formal Maltese intelligibility lists which include words and phrases whose phonological structure is adequately representative of Standard Maltese. The two lists are intended primarily to be utilised to assess speech intelligibility in dysarthria in a clinical setting. We suggest that it may also be possible to adapt or extend the use of the selected words and phrases to other clinical contexts, such as, speech recognition tests in audiological examination or repetition tasks in aphasia batteries. The lists could reduce reliance on informal methods of intelligibility assessment, enable clinicians to obtain objective, realistic and representative measures of intelligibility and provide key information for developing appropriate therapeutic input.

This study also allows for some avenues for future opportunities in this field of research to be brought to light. To start with, the lists should be tested on individuals with dysarthria and healthy individuals to obtain normative data, and verify the feasibility and applicability of using the MIL as a measure of speech intelligibility. Test-retest and intra-rater reliability of the MIL should be evaluated to determine whether the lists can replicate the same results over time and across different test administrators. It is also hoped that future studies will be based on a larger language sample, ideally taken from a spoken corpus which will help determine frequency of word and syllable structures better. This is particularly important in the case of words of more than four syllables.

Considerations to increase the current pool of words and phrases in the MIL should be taken into account so that risks of bias, particularly the influence of practice effects on test administrators and persons undergoing assessment, are minimised. Administration of English and Maltese intelligibility measures on bilingual Maltese speakers is also needed to shed light on similarities and language specific variations in word and phrase testing.

References

Azzopardi-Alexander, M. (2002). The vowel system of Maltese: From production to perception. In *Proceedings of the 5th Conference of the Association Internationale de Dialectologie Arabe (AIDA),* 321–333.

Borg, A., and Azzopardi-Alexander, M. (1997). *Maltese*. London: Routledge.

Borg, M., Bugeja, K., Vella, C., & Gafa, C. (2013). Preparation of a free-running text corpus for Maltese concatenative speech synthesis. In A. Borg, S. Caruana & A. Vella (Eds.), *Perspectives on Maltese Linguistics* (pp. 297-318). Berlin: Akademie Verlag.

Brincat, G. (2004). *Malta. Una storia linguistica*, Genova: Le Mani.

Duffy J. R. (2007). Motor speech disorders: History, current practice, future trends and goals. In G. Weismer (Ed.), *Motor speech disorders* (pp. 7-56). San Diego, CA: Plural Publishing.

Duffy, J. R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management*. St. Louis, MO: Elsevier

Dykstra, A.D., Hakel, M.E., & Adams, S.G. (2007). Application of the ICF in reduced speech intelligibility in dysarthria. *Seminars in Speech and Language, 28*, 308-311.

Enderby, P., & Palmer, R. (2008). *Frenchay dysarthria assessment*. (2nd ed.) Austin, TX: Pro-Ed.

European Commission (June 2012). *Special Eurobarometer 386: Europeans and Their Languages*. Eurobarometer special surveys. Retrieved from http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_386_en.pdf

Galea L. & Ussishkin. A. (2018). Onset clusters, syllable structure and syllabification in Maltese. In P.,Paggio & A.,Gatt (eds.), *The languages of Malta*, (pp.55–79). Berlin: Language Science Press.

Gries, S. T. (2010) Dispersions and adjusted frequencies in corpora: further explorations. In S. T., Gries, S. T., Wulff, S., & Davies, M. (eds), *Corpus linguistic applications: current studies, new directions* (pp. 197-212). Amsterdam: Rodopi.

Gurevich, N., & Scamihorn, S. L. (2017). Speech-Language pathologists' use of intelligibility measures in adults with dysarthria. *American Journal of Speech-Language Pathology, 26* (3), 873-892.

Kempler, D. and Van Lancker, D. (2002). Effect of speech task on intelligibility in dysarthria: a case study of parkinson's disease. *Brain and language, 80* (3), 449– 464.

Kent, R. (1992). Introduction. In R.D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement and management* (pp. 119-155). Amsterdam: John Benjamins Publishing.

Lillvik, M., Allemark, E., Karlstrom, P., & Hartelius, L. (1999). Intelligibility of dysarthric speech in words and sentences: development of a computerised assessment procedure in Swedish. *Logopedics Phoniatrics Vocology, 24*, 107-119.

Maltese Language Resource Server. (2016). Retrieved from: http://mlrs.research.um.edu.mt

Spagnol, M. (2011). *A tale of two morphologies. Verb structure and argument alternations in* Maltese (Ph.D. dissertation). Universität Konstanz.

Weismer, G., & Laures, J. S. (2002). Direct magnitude estimates of speech intelligibility in dysarthria: Effects of a chosen standard. *Journal of Speech, Language, and Hearing*

Research, 45, 421–433.

Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders, 11*, 499–512.

Yorkston, K. M., & Beukelman, D. R. (1981). *Assessment of intelligibility of dysarthric speech*. Tagard, OR: C.C. Publications

Yorkston, K. M., Beukelman, D. R., Strand, E. A., & Hakel, M. (2010). *Management of Motor Speech Disorders in Children and Adults (3rd ed.)* Austin, TX: PRO-ED.

Yorkston, K. M., Strand, E. A., & Kennedy, M. R. T. (1996). Comprehensibility of dysarthric speech: implications for assessment and treatment planning. *American Journal of Speech-Language Pathology, 5*, 55-66.

Table Titles

Table 1. Phonemes of Standard Maltese

Table 2. A sample of the phonetic citation form syllable-by-syllable transcriptions for the 500 most frequent words in the online Maltese corpus (Korpus Malti, v. 3)

Table 3. Phoneme counts of the constructed Maltese Intelligibility Lists (MIL - word and phrase list) as compared to Borg et al. (2013)

Table 4. Selection of words from the MLRS corpus meeting the frequency criterion

Table 5. Frequency and proportion of words with different number of syllables in the selected corpus, word and phrase list

Table 6. Most common syllable combinations found in words in the selected corpus and their relative percentage

Table 7. Syllable structure of words included in the word list of the Maltese Intelligibility Lists (MIL)

Table 8. Syllable structure and combination of words included in the phrase list of the Maltese Intelligibility Lists (MIL)

Class of Speech Sound	Number	Details
Vowels	11	/ɪ, ɪː, iː, ɛ, ɛː, ɐ, ɐː, ɔ, ɔː, ʊ, uː/
Consonants	22	/p, b, t, d, k, g, ʔ, m, n, f, v, s, z, ʃ, h, ʧ, ʤ, j, w, r, l, ts/
Diphthongs	7	/ɐʊ, ɐɪ, ɛʊ, ɛɪ, ɪʊ, ɔɪ, ɔʊ/

Word no.	IPA Syllable Type	IPA Syllable Type	IPA Syllable Type	Stressed Syllable (IPA) Syllable Type	IPA Syllable Type	IPA Syllable Type	Maltese Orthographic Transcription English Orthographic Transcription
74				ˈɐːn VVC	dɐ CV		għandha *she has*
75				ˈɪz VC	dɐ dɐ		iżda *but*
76				ˈʔɐ CV	bɛl CVC		qabel *before*
77				ˈɐn VC	kɛ CV		anke *also*
78				ˈzmɪːn CCVVC			żmien *age*
79			ɐr VC	ˈtiː CVV	kɔ CV	lʊ CV	artikolu *article*
80				ˈɐk VC	tɐr CVC		aktar *more*
81				ˈsniːn CCVVC			snin *years*
82				ˈmɛm CVC	brɪ CCV		membri *members*
83	ɐf VC	fɐ CV	rɪ CV	ˈjɪːt CVVC			affarijiet *things*
84				ˈɐh VC	nɐ CV		aħna *we*
85				ˈtɐht CVCC			taħt *under*

Phoneme	% in MIL - Wordlist	% in MIL- Phrase List	% in Borg et al.	Phoneme	% in MIL- Wordlist	% in MIL- Phrase List	% in Borg et al.
ɪ	9.50	11.90	12.09	t	8.26	8.08	8.10
ɐ	10.12	11.57	10.21	k	3.12	2.07	2.95
ɛ	5.14	4.37	5.14	d	3.74	3.93	2.74
ʊ	3.43	3.60	5.03	p	2.80	2.40	2.14
ɔ	3.43	3.28	3.37	b	1.87	1.64	1.66
ɐː	2.65	2.84	2.11	ʔ	1.09	1.09	1.12
ɪː	1.56	1.42	1.47	g	0.31	0.33	0.54
ɛː	0.00	0.11	0.50	ts	0.93	1.42	1.12
ɔː	1.09	0.76	0.11	ʤ	1.25	0.76	0.75
iː	1.40	1.53	0.11	ʧ	0.62	0.44	0.63
uː	1.56	0.87	0.05	n	6.07	7.42	6.44
s	4.67	4.15	4.19	m	3.58	3.82	4.39
f	1.40	1.75	2.01	l	5.61	6.22	8.30
h	1.40	2.73	1.43	r	6.85	4.48	5.06
ʃ	1.09	1.20	1.10	j	3.58	2.40	3.06
z	1.09	0.76	0.69	w	0.00	0.33	0.71
v	0.78	0.33	0.65	dz	0.00	0.00	0.00
ʒ	0.00	0.00	0.00				

Table 4. Selection of words from the MLRS corpus meeting the frequency criterion

Written Word	Phonetic Transcription	Written Word	Phonetic Transcription
nies *people*	nɪːs	bħal *like*	bhɐl
kliem *words*	klɪːm	ħaġa *thing*	hɐːʧɐ
bejn *among*	bɛjn	ħaddiema *labourers*	hɐdːɪːmɐ
fejn *where*	ʃɛjn	saħħa *health*	sɐhːɐ
tajjeb *good*	tɐjːɛb	tfal *chilren*	tfɐl
kulħadd *everyone*	kʊlhɐtː	*proġett* project	prɔʤɛtː
qatt *never*	ʔɐtː	*tweġiba* response	twɛʤiːbɐ

No. of Syllables	No. of words	% of 500 corpus	Distribution in word list	% of words in wordlist	Distribution in phrase list	% of words in phrase list
1 syllable	122	24.4	28	24.1	37	24.7
2 syllables	243	48.6	57	49.2	75	50
3 syllables	86	17.2	20	17.2	25	17.1
4 syllables	35	7	8	6.9	10	6.1
5 syllables	12	2.4	3	2.6	3	2.1
6 syllables	2	0.4	0	0	0	0

Monosyllabic Word Structures	Example	Relative % of occurrence in corpus
CVVC	bɪːʃ *so that*	29.5
CVCC	mɪnː *from*	21.3
CCVVC	dwɐːr *about*	14.8
CVC	sɐn *saint*	9
CVV	sɐː *until*	6.6

Disyllabic Word Structures	Example	Relative % of occurrence in corpus
CVC-CV	hɐfnɐ *a lot*	16.0
CVV-CV	kɪːnʊ *were*	10.7
CV-CV	ɛbdɐ *none*	7.8
CVC-CVVC	ʔʊddɪːm *front*	4.1
CVC-CVCC	mɐltiːn *maltese*	3.7

Trisyllabic Word Structures	Example	Relative % of occurrence in corpus
CVC-CVV-CV	pɛrsuːnɐ *person*	12.8
CV-CVV-CV	ɐlːuːrɐ *then*	8.1
CVC-CVC-CV	mɪnkɛjːɐ *in spite of*	7.0
CV-CVC-CV	nɐrɐw *we see*	5.8
CCV-CVC-CV	stʊdɛntɪ *students*	5.8

4-syllable Word structures	Example	Relative % of occurrence in corpus
CV-CVC-CVV-CV	dɛʧɪzjɔːnɪ *decision*	8.6
CV-CV-CVC-CV	rɛsɪdɛntɪ *residents*	5.7

Item No	Syllable Structure	Item No	Syllable Structure	Item No	Syllable Structure
1	CVVC	40	CVV-CV	79	V-CVCC
2	CVVC	41	CVV-CV	80	CVV-CVVC
3	CVVC	42	CVV-CV	81	CVC-CVV
4	CVVC	43	CVV-CV	82	CVCC-CV
5	CVVC	44	CV-CV	83	CCVC-CVCC
6	CVVC	45	CV-CV	84	CV-CVV
7	CVVC	46	CV-CV	85	CVC-CV
8	CVVC	47	CV-CV	86	CVC-CVV-CV
9	CVVC	48	CV-CV	87	CVC-CVV-CV
10	CVCC	49	CVC-CVVC	88	CVC-CVV-CV
11	CVCC	50	CVC-CVVC	89	CV-CVV-CV
12	CVCC	51	CVC-CVVC	90	CV-CVV-CV
13	CVCC	52	CVC-CVCC	91	CVC-CVC-CV
14	CVCC	53	CVC-CVCC	92	CVC-CVC-CV
15	CVCC	54	CVC-CVCC	93	CCV-CVC-CV
16	CCVVC	55	VC-CV	94	CCV-CVC-CV
17	CCVVC	56	VC-CV	95	CV-CVC-CCV
18	CCVVC	57	CVC-CVC	96	CV-CVC-CV
19	CCVVC	58	CVC-CVC	97	V-CVV-CV
20	CVC	59	CV-CVC	98	CVC-CV-CVVC
21	CVC	60	CV-CVC	99	V-CVC-CV
22	CVC	61	CV-CVVC	100	CVC-CVC-CCV
23	CVV	62	CV-CVVC	101	CV-CV-CV
24	CVV	63	CVC-CCV	102	CCVC-CV-CVVC
25	CCVCC	64	CVC-CCV	103	CV-CVC-CVVC
26	CCVCC	65	CVV-CVC	104	CVC-VC-CV
27	CCVC	66	VC-CVVC	105	CV-CV-CVVC
28	VCC	67	CCVV-CV	106	CV-CVC-CVV-CV
29	CVC-CV	68	VV-CV	107	CV-CV-CVC-CV
30	CVC-CV	69	CVVC-CV	108	VC-CV-CV-CVVC
31	CVC-CV	70	VVC-CV	109	VV-CV-CVV-CV
32	CVC-CV	71	VC-CVC	110	CVC-CV-CVC-CV
33	CVC-CV	72	V-CVVC	111	CCV-CV-CVV-CV
34	CVC-CV	73	VC-CVCC	112	CVC-CVC-CCVV-CV
35	CVC-CV	74	CV-CVCC	113	CVC-CV-CVV-CV
36	CVC-CV	75	CCV-CVCC	114	VC-CV-CV-CV-CVVC
37	CVC-CV	76	VV-CVC	115	V-CV-CVCC-CVV-CV
38	CVV-CV	77	CCVC-CV	116	V-CV-CV-CVV-CV
39	CVV-CV	78	VVC-CVC		

Item No	Word 1	Word 2	Word 3	Word 4
	Syllable Structure	Syllable Structure	Syllable Structure	Syllable Structure
1	CVVC	CVV-CVC	CCVC-CV	
2	VC-CCVV-CV	CCVC-CVVC	CV-CVC-CV	
3	VC-CV	CVV-CCV	CCVV-CV	
4	CV-CV-CVC-CV	CVC-CV	VC-VC-CVC-CV-CVV	
5	CCVVC	CVC-CV	CVC-CVC	
6	CV-CVVC	CVC-CVV-CV	CVV-CVC	
7	VVC-CV	CCVCC	CVCC	
8	CVVC	CCVV-CV	CVC-CV	
9	VC-CV	CVC-CV	CVC-CV-CVVC	
10	VC-CVV-CV	CCCV-CV	CVCC-CVV-CV	
11	CVV-CV	CVC-CCV-CV	CVC-CVC-CV-CV	
12	VVC-CV	CCVCC	CVV-CVC	
13	CVVC	CVC-CVC	CVC-CV	
14	CVC-CV	CVV-CVC	CVC	
15	CVVC	CVV-CVVC	CCVC-CVVC	
16	CVC-CV	CVC-CVC	VC-CVVC	
17	CVC-CVV-CV	CCVC-CVV	CV-CVC-CV	
18	VC-CVV-CV	CVC-CV	CVC-CV-CVCC-CVV-CV	
19	CV-CV	CVC-CVV	CV-CV-CVVC	
20	CVC-CV	CCVV-CVC	CVC-CVCC-CV	
21	CVV	CVC-VCC	CVC	
22	CVV	CVC-VCC	CVV-CV	CVCC
23	CV-CV	CVVC	CV-CVC-CVV-CV	
24	CVCC	CVVC	CVCC-VVC	
25	CVVC	CVC-CVV-CV	CVV-CVC	VCC-CVV-CV
26	CVC-CV	CVV-CV	CV-CVCC-CVV-CV	
27	CVVC	CVC-CV	CVV-CV-CV	
28	CVC-CV	CCVVC	CV-CVCC-CVV-CV	
29	VVC-CV	CVC-CVCC		
30	CV-CVC	CVVC	CV-CVVC	
31	VCC	CVC-CV	CV-CV	
32	CCVV-CV	CVCC	CCV-CV	
33	CV-CV	CVVC	CCVV-CCV	
34	CCV-CV-CVCC	CV-CV-CVC		
35	CVC-CV	CVC-CV	CVC-CV	
36	CVVC-VC	CVC	CVC-CVV-CV	
37	CVV-CVC	CVC-CV	CCVV-CV	
38	CVC-CV	CV-CV-CV-CV	CVC-CVC-CVC-CCV	
39	CV-CVC-CVVC	CVC-CVV-CV	CVC-VC-CV	
40	CVVC	CVV-CVC	CVCC	
41	CVVC	CVC-CVVC	CVV-CVVC	
42	CVCC	CCVVC	CVV-CVC	
43	CVC-CV	CVVC	CVVC	
44	CVCC	CVV	CVC-CV	
45	CVVC	CV-CV-CVCC-CVV-CV	CCV-CV	
46	CV-CV	CV-CVC-CVV-CV	CV-CVC-CVV-CV	
47	CVV-CV	CVCC	VV-CV	
48	CVVC	CVC-CV-CV	CV-CVC-CV	
49	CVC-VC	CCVC	VV-CV	
50	CVV-CV	CCVVC	VCC-CVCC	