COMPUTATIONAL TEXTSETTING ANALYSIS OF DUTCH FOLK SONGS

Varuṇ deCastro-Arrazola[1,2]**, Peter van Kranenburg**[1]**, Berit Janssen**[1,3]

[1]Meertens Instituut, Amsterdam [2]Universiteit Leiden [3]Universiteit van Amsterdam

`{varun.decastro-arrazola,peter.van.kranenburg,berit.janssen}@`
`meertens.knaw.nl`

ABSTRACT

The study of textsetting describes how words and tunes are aligned in songs. In languages with word stress such as English, it has been shown that the metrical prominence of melodies is aligned in a non-random way with the lyrics' word stress (Dell & Halle 2009). The present study addresses the textsetting rules of Dutch using a large dataset of folk songs and a novel methodology. The main findings are the following: (1) the combination of linguistic stress and metrical prominence moving in opposite directions is avoided; (2) textsetting rules do not apply across phrases, but they do across words; (3) the avoidance of stress and prominence combinations depends on factors such as phrase-finality and the presence of melisma. In contrast with previous textsetting studies, our approach does not use pre-defined mismatches, but induces the avoided stress / prominence combinations in a data-driven way. This allows for a more systematic understanding of how words and tunes are to be aligned in a given tradition.

1. INTRODUCTION

Songs can be perceived globally as homogeneous objects, but we can also consider them composite objects with two main components: a text and a tune. The analysis of textsetting describes how these two components are combined. Textsetting rules state whether particular combinations of linguistic and musical features are preferred or avoided in a given musical tradition.

Figure 1: Songs can be analysed as composite objects combining a text and a tune. This tune (with slight variations) is used several times in the same song, while different words are set to it.

The autonomous status of the tune can be observed, for instance, in cases where the same melody is used several times with varying lyrics. Figure 1 shows such an example taken from the MTC-FS corpus (van Kranenburg et al. 2014). Tunes can therefore be considered abstract templates to which words are set.

Avoiding stressed syllables in weak musical positions, for instance, is a common tendency in languages such as English or French (Dell & Halle 2009). In tone languages, where pitch contours are used to distinguish lexical or grammatical meaning (e.g. Chinese), textsetting rules can specify how to match linguistic and musical pitch contours (Schellenberg 2013).

Dutch is a non-tonal language, but it does have word stress; e.g. the first syllable in a word like *boter* 'butter' is more prominent than the second syllable. Similarly, in metered music some positions are more prominent than others. We will use *stress* to talk about linguistic prominence, and *prominence* to talk about musical prominence.

The goal of this paper is twofold. From a methodological point of view, it presents a systematic way of addressing the textsetting problem computationally (virtually all existing studies on the topic rely on manual analyses; cf. Temperley & Temperley 2013). Secondly, it provides a first description of textsetting in Dutch folk songs.

2. METHOD

2.1 Material

In order to study the textsetting rules of Dutch folk songs we analysed 3724 songs from the MTC-FS corpus (van Kranenburg et al. 2014). Most of the songs were collected through fieldwork between the 1950s and the 1990s as part of the radio programme *Onder de groene linde* led by Ate Doornbosch. The corpus also contains similar songs taken from 19th and 20th century songbooks. [1]

The original corpus contains 3861 songs. However, the features we focus on (stress and prominence) were not always obtainable. Songs encoded as having free rhythm (n = 125) were excluded because they lack the feature of prominence. Linguistic stress for the lyrics was obtained through a nearest-neighbour lookup in the e-Lex [2] lexical database (as specified in van Kranenburg & Karsdorp 2014). Thus, the database lookup is robust against minor variations in spelling. Cases in which the nearest neighbour in the e-Lex database has a different number of syllables than the word in the lyrics were discarded. Any phrase containing one such word has also been excluded from the analysis (n = 2451 phrases).

Every song in the dataset is divided into stanzas; stanzas are divided into phrases; phrases contain notes, which can then be associated to syllables. For the purposes of this pa-

---

[1] The melody, text and metadata for each song is openly available in several formats at `www.liederenbank.nl/mtc`.

[2] `http://tst-centrale.org/en/producten/lexica/e-lex/7-25`

per, stanzas are roughly equivalent to strophes, and phrases are also referred to as lines. The filtered dataset contains 3724 songs, 3973 stanzas, 20662 phrases, 185263 notes, and 176708 syllables. Syllables and notes are often in a one-to-one correspondence. Some syllables, though, span over more than one note; such syllables are referred to as *melismas*. In the filtered dataset, 4.46% of the syllables are melismas.

2.2 Corpus annotation

Stress is not a feature present in the original dataset, it was looked up at the e-Lex database. Stress is encoded in a binary way in the database; each syllable gets a value of either `0` (unstressed) or `1` (stressed). Secondary stress is not encoded. Figure 2 illustrates how this and the following features related to stress and prominence have been automatically annotated.
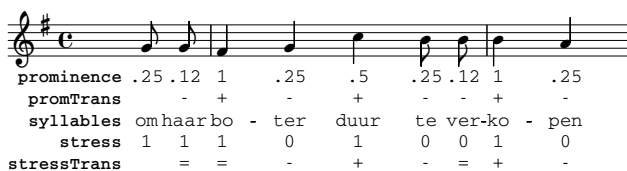

prominence	.25	.12	1	.25	.5	.25	.12	1	.25
promTrans		-	+	-	+	-	-	+	-
syllables	om	haar	bo	- ter	duur	te	ver-ko	- pen	
stress	1	1	1	0	1	0	0	1	0
stressTrans		=	=	-	+	-	=	+	-

Figure 2: Sample annotation of stress, prominence and their respective contours.

Musical prominence is also not explicitly encoded as a feature in the MTC-FS dataset. However, this feature can be inferred from the symbolic representation of the tunes. For each note, we know its position within the musical bar, and the time signature this bar belongs to (e.g. 6/8). Given that information, relative prominence can be derived. This was done using the `music21` software (Cuthbert & Ariza 2010). Prominence values range from `0` to `1`, the first position of the bar being assigned a `1`.

Both stress and prominence are relative notions, that is, given a syllable in isolation, its raw stress/prominence value is trivial. Hence, to capture how stress and prominence are aligned, it becomes necessary to compare a syllable with its neighbours. We have achieved this by computing the transition for the stress and prominence values of each syllable compared to the values of the preceding syllable. This produces three possible stress/prominence contours: decreasing (–), same (=) and increasing (+), as illustrated in Figure 2. Note that the first syllable of a song does not have any preceding syllable to compare its stress or prominence to; that is, its transitions values are left empty (set to `NA`).

2.3 Statistical analyses

In order to find out which combinations of stress and prominence are being avoided in the corpus, we first computed the frequency of each of the 9 possible transition combinations (decrease, same, increase for each feature). We then took the marginal probabilities of the contingency table to calculate the expected frequencies for each cell. Last, we calculated the ratio between observed and expected frequencies in order to detect under-represented combinations.

For instance, the proportion of syllables having a decreasing stress and an increasing prominence is 0.03. If we look at the marginal frequencies, we observe that 0.26 of all syllables show a decreasing stress, and 0.46 show an increasing prominence. The expected frequency for the combination of the two is the product of the marginal frequencies: $0.26 \times 0.46 = 0.12$. We can now compare the observed frequency (0.03) with the expected frequency (0.12), and note that that particular combination of features is under-represented in the dataset; i.e. if the alignment of text and tune was done randomly, that combination would be more frequent.

As a way of assessing the degree to which a combination is over- or under-represented, we calculated an association factor by dividing the observed frequency by the expected frequency. In the above example, this yields 0.26. Association factors can range from 0 to infinity, where 1 indicates that the combination is not controlled, as it occurs at chance level. We take the conventional threshold of 0.5 as a cut-off point to select combinations which are significantly avoided (Agresti 2013).

The same kind of analysis was performed on conditioned subsets of the data in order to refine the textsetting rules in two ways. First, the domain on which textsetting rules operate was addressed. Given that our analyses are computed on two-syllable windows (by comparing every syllable with its preceding syllable), some of these windows will go across boundaries. Figure 3 illustrates the two cross-boundaries contexts which were analysed: beginning of phrases and beginning of words. Second, we investigated whether the alignment of stress and prominence is influenced by other features, namely, being sung to a melisma, and being placed at the end of phrases.
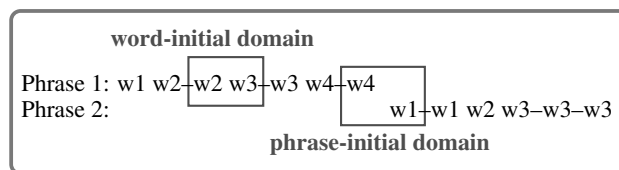

Figure 3: We define the textsetting domain with a two-syllable window. In the above example, the word-initial domain includes the last syllable of word 2 (w2) and the first syllable of word 3 (w3). The phrase-initial domain includes the last syllable of phrase 1 and the first syllable of phrase 2.

3. RESULTS

Table 1 shows the association factors for all possible combinations of stress and prominence transitions. Two of the nine possible alignments are being avoided: [stress+, prom–], and [stress–, prom+]. That is, when aligning words to a tune, stress and prominence should not go in opposite

directions. When stress or prominence show a = contour, association values are close to 1, suggesting these alignments are not controlled.

	stress-	stress=	stress+
prom-	1.77	0.93	**0.36**
prom=	0.85	0.95	1.24
prom+	**0.26**	1.08	1.60

Table 1: Association factors for stress and prominence contours. Combinations below the 0.5 threshold, marked in bold red, are considered avoided alignments.

3.1 Relevant domain for textsetting

Next, we use this measure of association to further investigate the domain at which textsetting rules operate. Is the alignment of stress and prominence controlled across phrases? In order to address this question, we divided the data into phrase-initial contexts (illustrated in Figure 3) and non-initial contexts, and then calculated the association factors between stress and prominence.

The right panel in Figure 4(a) shows that the transition of prominence from the last syllable of a phrase to the first syllable of the following phrase is not associated to the stress transition of those two syllables (values do not go below the 0.5 threshold). Alignment of stress and prominence is controlled within phrases (left panel), but not across them (right panel).

In what follows, transition values for phrase-initial contexts are discarded, because they do not appear to be actively controlled. In the left panel in Figure 4(a), we also observe that both stress– and stress+ mismatches are now equally avoided within phrases, which was not the case in Table 1, where phrase-initial contexts were included in the analysis.

The second question on the textsetting-relevant domain asks whether stress and prominence transitions are controlled across words. Figure 4(b) shows that alignment is controlled both within and across words. Besides, there is one additional constraint when aligning stress and prominence word-internally (left panel): [stress+, prom=] is avoided. That is, in polysyllabic words, an increase of stress must be aligned with an increase of prominence.

3.2 Additional textsetting factors

How do other factors make stress / prominence mismatches more or less acceptable? Figure 6(a) shows the data divided into two subsets: phrase-final and non-final contexts. Mismatches involving stress+ are much more avoided phrase-finally (right panel), while the ones with stress– are more accepted in this context. Figure 5 displays one of the rare examples of a [stress+, prom–] at the end of the phrase.

In Figure 6(b) the data has been divided according to whether the current syllable has or has not a melisma (m1, m0), and whether the preceding syllable has or has not a melisma (mp1, mp0). The condition where a syllable has

Figure 5: The alignment [stress+, prom–] is much more avoided phrase-finally than elsewhere.

a melisma *and* is preceded by a melisma (m1mp1) is excluded from the plot because it turns out to be statistically uninformative due to its infrequency.

When a syllable is sung to a melisma (rightmost panel), [stress+, prom–] mismatches are even more infrequent than without a melisma. In contrast, [stress–, prom+] mismatches become more acceptable when sung to a melisma (while still being avoided).

Regarding the effect of a preceding melisma, the most salient one is shown in the left column of the middle panel of Figure 6(b). A [stress–, prom+] alignment is not avoided if the first of the two syllables is a melisma. This is illustrated in the penultimate word of Figure 5 (*woorden*).

4. DISCUSSION

The avoided combinations highlighted in the previous section can be reformulated in terms of constraints which are active when a Dutch speaker creates a new song, and inversely when a song is judged as well- or ill-formed. Minimally, the window where stress and prominence are compared must comprise two adjacent syllables. This window is reset afresh at the beginning of each phrase, but not at the beginning of words. Further analysis should determine whether the relevant textsetting domain is smaller than the musical phrase, maybe corresponding to a phonological phrase (Proto & Dell 2013).

Within this two-syllable domain, the general rule is that stress and prominence should not move in opposite directions. The [stress–, prom+] misalignment constraint can be relaxed phrase-finally and when the second syllable is sung to a melisma. If the first syllable is sung to a melisma, the constraint does not apply at all. The [stress+, prom–] misalignment constraint is *stricter* phrase-finally, and when the second syllable is sung to a melisma.

The alignment of stress+ with prom= is also avoided within polysyllabic words, a constraint not explicitly mentioned in the previous textsetting literature. This might suggest that two adjacent positions with the same degree of prominence according to a traditional analysis (e.g. positions 2 and 3 of a 6/8) are actually not treated as equal, but rather the second one being weaker, hence asking for a less stressed syllable.

Future work will determine whether Dutch listeners are sensitive to the rules stated above. Besides, the reason why constraints are relaxed in certain contexts remains open. For instance, the expectation that phrases should finish in a more prominent note (because of closure) may account for the observed phrase-finality effects.
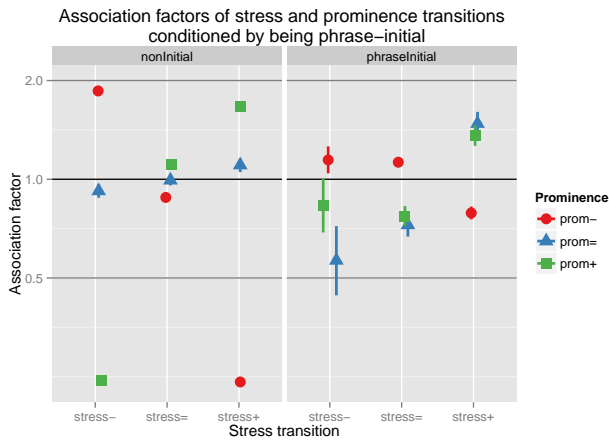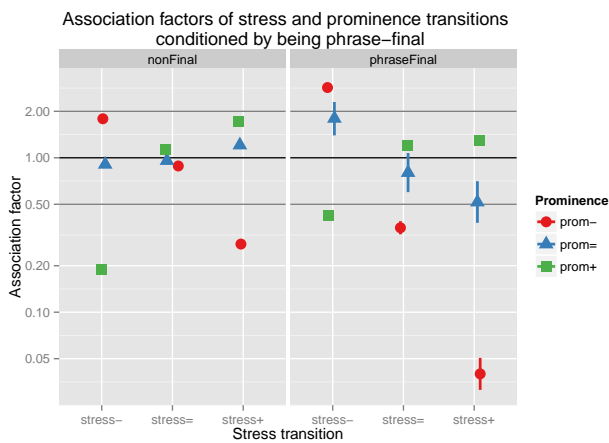
(a) Association between stress and prominence contours for syllables in phrase-initial and non-initial contexts.
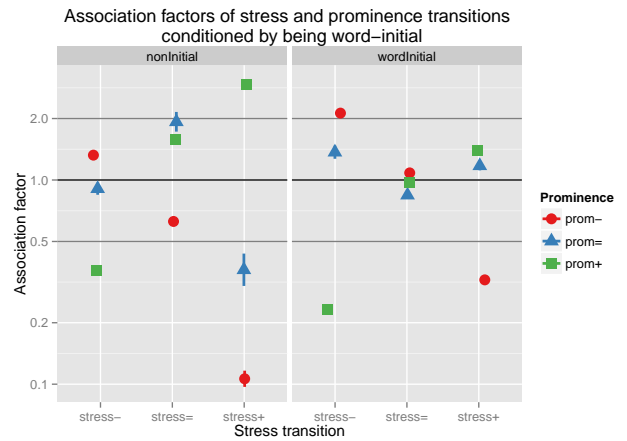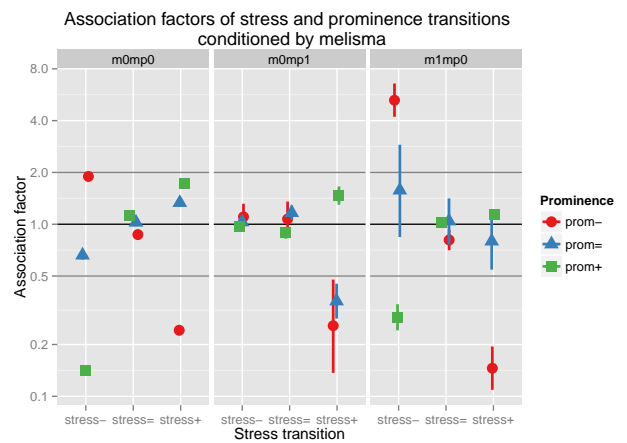
(b) Association between stress and prominence contours for in word-initial and non-initial contexts.

Figure 4: Association factors for boundary-crossing contexts compared to non-crossing contexts: (a) phrase-initial context, (b) word-initial context. A value of 1 indicates random alignment. Upper and lower significance thresholds are set at 2 and 0.5.

(a) Association between stress and prominence contours for syllables in phrase-final and non-final contexts.

(b) Association between stress and prominence contours for syllables with(out) melisma (m1, m0), preceded by syllables with(out) melisma (mp1, mp0).

Figure 6: Effect of (a) phrase-finality and (b) melisma on the association between stress and prominence contours. A value of 1 indicates random alignment. Upper and lower significance thresholds are set at 2 and 0.5.

In order to follow up the present study, further linguistic and musical features can be considered. Dutch is said to have secondary stress, which means that some words contain more than one stressed syllable (Booij 1995). This could account for some of the examples currently analysed as mismatches. On the musical side, note duration and pitch can be of particular interest, since they can contribute to the perception of prominence (Müllensiefen et al. 2009). Also, the effect of factors such as melisma or phrase-finality have been studied here in isolation; however, more complex interactions between them and with other features can be expected.

A detailed description of this kind of interactions can eventually shed some light on the cognitive processes involved when simultaneously processing music and language. The fact that we unconsciously control how linguistic prosody and musical features are to be aligned may suggest that these two domains are processed by the same neural resources (Zatorre & Baum 2012; Hausen et al. 2013).

Finally, an understanding of how words and tunes are aligned in a given language provides the basis for a number of applications. For instance, historical lyrics for which the tune is unknown can be matched to well-aligned melodies of the same period (and vice versa for text-less tunes). In order to address this kind of automatic alignment, the observed avoidance values can be straightforwardly converted to weighted penalties.

5. CONCLUSION

This paper constitutes an initial study on the textsetting of Dutch folk song. We show that, as in other languages, linguistic stress and musical prominence moving in opposite directions is avoided. The domain where this constraint is active is smaller than the musical phrase, and bigger than the word. A number of factors (phrase-finality, melisma) make these mismatches more or less acceptable. The proposed methodology has the advantage of being able to infer avoided combinations of linguistic and musical features, and to make explicit predictions about which contexts would be perceived as more or less ill-formed. Careful inspection of the existing songs where avoided combinations occur remains crucial in order to systematically test further features which may interact with stress and prominence.

6. REFERENCES

Agresti, A. (2013). *Categorical Data Analysis*. Hoboken (New Jersey): John Wiley & Sons.

Booij, G. (1995). *The Phonology of Dutch*. Oxford: Clarendon Press.

Cuthbert, M. S. & Ariza, C. (2010). music21: A toolkit for computer-aided musicology and symbolic music data. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)* (pp. 637–642). ISMIR.

Dell, F. & Halle, J. (2009). Comparing musical textsetting in french and english songs. In J.-L. Aroui & A. Arleo (Eds.), *Towards a Typology of Poetic Forms* (pp. 63–78). Benjamins.

Hausen, M., Torppa, R., Salmela, V. R., Vainio, M., & Särkämö, T. (2013). Music and speech prosody: a common rhythm. *Frontiers in psychology*, *4*, 1–16.

Müllensiefen, D., Pfleiderer, M., & Frieler, K. (2009). The perception of accents in pop music melodies. *Journal of New Music Research*, *38*(1), 19–44.

Proto, T. & Dell, F. (2013). The structure of metrical patterns in tunes and in literary verse. evidence from discrepancies between musical and linguistic rhythm in italian songs. *Probus*, *25*(1), 105–138.

Schellenberg, M.-H. (2013). *The Realization of Tone in Singing in Cantonese and Mandarin*. PhD thesis, The University of British Columbia.

Temperley, N. & Temperley, D. (2013). Stress-meter alignment in french vocal music. *Journal of the Acoustical Society of America*, *134*(1), 520–527.

van Kranenburg, P., de Bruin, M., de Grijp, L., & Wiering, F. (2014). The meertens tune collections. *Meertens Online Reports*, *1*.

van Kranenburg, P. & Karsdorp, F. (2014). Cadence detection in western traditional stanzaic songs using melodic and textual features. In *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)* (pp. 391–396). ISMIR.

Zatorre, R. J. & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biology*, *10*(7), e1001372.