Typological tendencies in verse and their cognitive grounding

De Castro Arrazola, V.

Author: De Castro Arrazola, V.
Title: Typological tendencies in verse and their cognitive grounding
Issue Date: 2018-05-03

Part III

How words are set to templates

6 A computational analysis of textsetting

6.1 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 autonomous status of the tune can be observed, for instance, in cases where the same melody is used several times with varying lyrics, be it in the same or in different songs. Example 6.1 shows such a case (song `NLB073196`) taken from the MTC-FS corpus of Dutch songs (van Kranenburg et al. 2014) on which we have based this study. Tunes can therefore be considered abstract templates to which words are set. This makes songs similar to poems in that a template (e.g. iambic pentameter) is used productively to create new instances of verse. Compared to the instantiation of templates discussed in Chapter 4, though, song templates are less abstract (i.e. they can be whistled), but they are also more complex in that they contain additional variables (e.g. pitch).

The analysis of textsetting describes how the two components of songs (i.e. text and tune) are combined (see Proto 2015 for an overview). This is often done by formulating textsetting rules, which state whether particular combinations of linguistic and musical features are preferred or avoided in a given musical tradition. Textsetting studies typically focus on three dimensions common to language and music which can be either aligned or misaligned when setting words to a tune: (1) prominence, (2) pitch, and (3) constituency.

In their seminal paper on the analysis of English textsetting, Halle & Lerdahl (1993) develop an algorithm which describes how a line of text is to be aligned with a given tune, addressing exclusively the dimension of prominence (further elaborated in Halle 1999). As a case study, they discuss an example where a single tune is set to a wide range of words: the well-known sea-chanty *The Drunken Sailor*. The first step in their textsetting algorithm, slightly simplified, specifies: "assign all accented syllables to available [strong musical] beats from left to right"

Example 6.1: 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 (`NLB073196`, van Kranenburg et al. 2014), while different words are set to it.

(Halle & Lerdahl 1993:15). This can be observed under Example 6.2, where both lines are aligned so that their stressed syllables (marked with an accent) match the four beats (marked with asterisks) of the *drunken sailor* tune. Note, for instance, that the weak musical position where the word *his* is placed (first line) gets 'skipped' by the stressed syllable *there* in the second line so that it falls on the beat.

The kind of procedure described by Halle & Lerdahl (1993) seeks to produce text-to-tune alignments which are accepted by subjects familiar with the tradition by avoiding, for instance, placing stressed syllables in musically weak positions. This is based on the intuition that music and language share some notion of prominence (Jackendoff 2009), and that textsetting grammars avoid 'incongruent' alignments (Hayes 2009).

Pitch is another dimension shared by speech and music, and it also plays a role in the textsetting grammar of at least some languages. Constraints on pitch alignment have been proposed particularly for tonal languages (for a review, see Schellenberg 2012), where pitch contours are used to create lexical contrasts (Yip 2002). A standard method of evaluating the degree of textsetting control is to calculate the proportion of cases in a song where the tonal and melodic contours move in a similar direction. In a sample of 20 Vietnamese popular songs, Kirby & Ladd (2016) found that cases of opposing pitch contours (i.e. linguistic and musical pitch moving in opposite directions) are exceedingly rare, occurring less than 5% of the times. Similar degrees of correspondence have been reported for other Asian (e.g. Cantonese) and African (e.g. Ewe) tonal languages (Schellenberg 2012).

Finally, grouping and constituent structure is yet another dimension shared by language (Nespor & Vogel 2007; Matthews 2007) and music (Deutsch 2013; Lerdahl & Jackendoff 1983). Compared to stress and tone, however, the alignment

Example 6.2: Two sample lines of the song *The Drunken Sailor*. Stressed syllables (marked with accents) are aligned with the musical beats (marked with asterisks). See Halle & Lerdahl (1993) for details.

*			*		*		*	
stíck	*on*	*his*	*báck*	*a*	*mús*	*tard*	*plás*	*ter*
kéep	*him*		*thére*	*and*	*máke*	*him*	*báil*	*her*

of constituent boundaries has received less attention in textsetting studies (but see e.g. Dell & Halle 2009; Halle 2004). Nevertheless, the logic to evaluate the correspondence between text and template remains comparable to the domains of prominence and pitch; the (mis)alignment of constituent boundaries is taken as evidence for textsetting control or lack thereof.

This chapter deals exclusively with the first dimension, i.e. the alignment of linguistic stress and musical prominence, in Dutch folk songs. The Dutch language uses word stress (Booij 1995); e.g. the first syllable in a word like *bóter* 'butter' is more prominent than the second syllable, exactly as in its English translation. Similarly, in metered music some positions are more prominent than others (London 2004). In the remaining of the chapter, we will use *stress* or its abbreviation *s* to talk about linguistic prominence, and *prominence* or just *p* to talk about musical prominence. As a notational shorthand, we indicate linguistic stress with an acute accent ($\sigma\acute{\sigma}$), and underline the syllable which receives the highest metrical prominence within a given context ($\sigma\underline{\sigma}$).

The goal of this chapter is twofold. From a methodological point of view, it presents a systematic way of addressing the textsetting problem computationally (most studies on the topic rely on manual analyses; but see, e.g. Temperley & Temperley 2013). Standardised methods of this kind are critical to test and compare the extent to which textsetting rules apply in the languages of the world. Secondly, as a case study, it provides a first description of textsetting in Dutch folk songs. We demonstrate that the alignment of stress and prominence is not random, and that it interacts with other musical and linguistic factors (presence of melisma and part of speech).

6.2 Method

6.2.1 Material

In order to study the textsetting rules of Dutch folk songs we analysed 3,724 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 3,861 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 lexical database[2] (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 = 2,451 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 chapter, stanzas are equivalent to musical strophes, and phrases are also referred to as lines. The filtered dataset contains 3,724 songs, 3,973 stanzas, 20,662 phrases, 185,263 notes, and 176,708 syllables. Syllables and notes are often in a one-to-one correspondence. Some syllables, though, span over more than one note; such a syllable is referred to as a *melisma*. In the filtered dataset, 4.46% of the syllables are melismas.

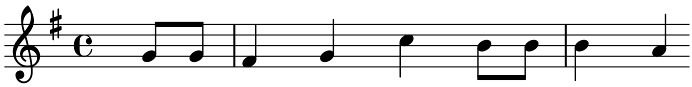
6.2.2 Corpus annotation

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). Dutch is described as containing secondary stress too (Booij 1995:105), but this is not explicitly encoded in the e-Lex database. Example 6.3 illustrates how stress and the following features related to stress and prominence have been automatically annotated. We also added part-of-speech information using the CELEX database (Baayen, Piepenbrock & van Rijn

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

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

Example 6.3: Sample annotation of stress, prominence and their respective contours.



prominence	.25	.12	1		.25	.5	.25	.12		1		.25
p. contour		–	+		–	+	–	–		+		–
syllables	*om*	*haar*	*bo*	-	*ter*	*duur*	*te*	*ver*	-	*ko*	-	*pen*
stress	1	1	1		0	1	0	0		1		0
s. contour		=	=		–	+	–	=		+		–

1995). Words were categorised into two classes: content words (nouns, verbs), and function words (articles, prepositions, conjunctions).

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 (Lerdahl & Jackendoff 1983). 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 cannot be determined. 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 (–), level (=) and increasing (+), as illustrated in Example 6.3. Note that the first syllable of a song does not have any preceding syllable to compare its stress or prominence to; hence, its transitions values are not computed (set to `NA`).

6.2.3 Statistical analyses

The goal of the analyses is to find alignment patterns of linguistic features and musical features which are arguably avoided by the individuals who created the songs in the corpus (and other speakers of Dutch as an extension). A simplified illustration of our analysis goes us follows. We count the number of syllables following a combination of a linguistic stress contour and a musical prominence contour (e.g. `s–p+`), we then calculate the expected count of syllables with the given combination assuming a random alignment of text and tune, and finally compare the observed frequency to the expected one. If a given alignment shows a much lower frequency in the corpus compared to the expected baseline, we conclude that it is an avoided alignment in the textsetting grammar under study.

In order to calculate the baseline probability of a combination of any 2 features f_1 and f_2, we compute the product of their individual frequencies: $p(f_1, f_2) = p(f_1) \times p(f_2)$. For instance, in a subset of our data, the p+ prominence contour shows up in 35% of the syllables, and the s– stress contour in 15%. The expected frequency of s–p+ syllables would be $.35 \times .15 = .05$, i.e. roughly 5% of the data. The *observed* frequency, however, is substantially lower, at 0.3%, indicative of it being an ill-formed setting. In principle, the same general procedure applies for any additional number of features, allowing us to introduce further linguistic and musical variables and fine-tune the analysis:

$$p(f_1 \ldots f_n) = \prod_{i=1}^{n} f_i \tag{6.1}$$

Following this approach, however, does not take into account that tunes and texts can behave as independent objects, as evidenced e.g. by everyday conversation and instrumental music. In order to evaluate the well-formedness of a text-to-tune *alignment*, it is irrelevant whether a particular musical or linguistic context is exceptional on its own. To avoid that, we treat the frequency of the set of musical features ($m_{1...j}$), and the frequency of the set of linguistic features ($l_{1...k}$) in an autonomous fashion. Their product provides the expected frequency of a textsetting alignment assuming the lyrics and tunes are, individually, well-formed objects:

$$p(m_{1...j}, l_{1...k}) = p(m_{1...j}) \times p(l_{1...k}) \tag{6.2}$$

Once we have obtained the expected frequencies for all the combinations of musical and linguistic features we are interested in, we compute the relation (in the form of a log ratio) between observed and expected frequencies. In cases where the log ratio is close to zero, we lack evidence to argue that that particular alignment is controlled by textsetting constraints, because the observed frequency of the alignment corresponds to that expected by a random textsetting. Here, instead, we focus on under-represented alignments, setting a conservative significance threshold at a log ratio of -2, i.e. cases where the observed frequency is less than four times the expected one (Agresti 2013:56).

Besides the main variables of interest (stress and prominence), we refine the analysis by including two additional linguistic features (word boundary and word

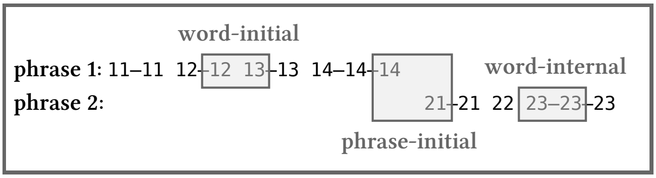


Figure 6.1: We define the textsetting domain with a two-syllable window: a target syllable on the right, and its preceding syllable. In the above figure, two abstract phrases are represented as a sequence of two-digit codes standing for syllables; the first digit indicates the phrase, and the second digit the word; syllables within the same word are connected by a hyphen. In the *word-initial* domain, the target syllable is preceded by a syllable from a different word; in the *phrase-initial* domain, by a syllable from a different phrase; and in the *word-internal* domain, by a syllable from the same word.

class), and two musical features (phrase boundary and melisma).

The boundary features serve to narrow down the relevant textsetting domain. Given that the stress and prominence variables are contour values over a two-syllable window, we can test whether textsetting constraints apply *across* musical/linguistic constituents. Does the prominence contour between the last note of a phrase and the first of the following line constrain the choice of words to be set? We compare three different domains, as illustrated in Figure 6.1: phrase-initial, word-initial, and word-internal.

Finally, we add two additional features which can affect the stringency of the textsetting constraints. We hypothesise that these are less stringent for so-called function or grammatical words, as it has been observed in Italian nursery rhymes (Proto & Dell 2013:112). Similarly, melismas can interact with the alignment of stress and prominence. In French, for instance, final unstressed syllables set to a relatively prominent note may be more acceptable if preceded by a melismatic syllable (Dell & Halle 2009:71).

6.3 Results

By combining the six features in our analysis, we obtain a six-dimensional textsetting model with all the 288 possible feature-value combinations and their associated log ratio, i.e. the extent to which this set of songs differs from a random textsetting alignment. Some of these combinations are completely absent from the corpus (e.g. there is no word-internal `s+p–` syllable preceded by a melisma).

Others are exceedingly rare, occurring only once in the whole corpus (e.g. a melismatic, word-internal s–p+ syllable belonging to a function word). A problem common to all corpus-based studies is that they are unable to provide negative evidence (Schütze 2011; 2016); hence, it is hard to determine whether absent alignments are accidental gaps, or whether they offer meaningful insights about forms perceived as ill-formed. Similarly, alignments with very low counts (such as single occurrences) present very wide confidence intervals, which make their interpretation unreliable.

6.3.1 Textsetting domain

In order to determine the domain in which textsetting is most constrained, we focus on a large subset of the data (content words with no melismas, n = 16,552 syllables) and compare the alignment of stress and prominence in the three domains depicted in Figure 6.1: line-initial, word-initial and word-internal. As shown in the Tables 6.1a–c (Supplementary Information), the alignment of stress and prominence is increasingly constrained in these three domains. None of the alignments exceeds the under-representation threshold (log ratio = −2) in the line-initial window. For the word-initial domain only one type of alignment is significantly avoided (s+p–, log ratio = −2.74). This alignment shows an even more extreme log ratio in the word-internal domain (−3.60), where an additional combination, s–p+, also exceeds the significance ratio (−4.14). This alignment, hence, represents the primary difference between a word-initial and a word-internal target syllable.

In absolute terms, the alignment s–p+ occurs approximately as often word-internally as word-initially (270 and 282 respectively). However, the word-internal examples are much more exceptional in comparison, because, assuming a random alignment, word-internal s–p+ are expected to occur more than four times as often as word-initial s–p+ (4,768 vs 1,092). This is explained by the trochaic bias in the Dutch lexicon, i.e. the prevalence of words with initial stress. Hence, the word *méisje* 'girl' with a s–p+ alignment on the second syllable (e.g. in song NLB072300) is more likely to be perceived as ill-formed, than the word *gemóed* 'mind' where the first, unstressed syllable is preceded by a stressed syllable, also exhibiting a s–p+ alignment (e.g. in song NLB161811; see Example 6.4; this and all subsequent examples are included in the Supplementary Information).

Among the two opposing alignments avoided word-internally, s–p+ is more strongly under-represented than s+p–, as established by the log ratios (−4.14 vs

−3.60, see Table 6.1c, Supplementary Information). This suggests native speakers would deem the alignment of the word *dróef<u>heid</u>* 'sorrow' in song `NLB074530` as ill-formed more readily than the alignment of the word *<u>ge</u>lóof* 'faith' in song `NLB111653` (Example 6.5). In the remaining analyses, we focus exclusively on word-internal syllables, as this context shows the strongest evidence for alignment constraints.

6.3.2 Linguistic context: content and function words

In a sample of 37 Italian nursery rhymes analysed by Proto & Dell (2013:112), function words exhibit textsetting misalignments more often than content words. Our results (Figure 6.2) support this difference in textsetting stringency between word types.

In the top panel we observe that only the `s–p+` alignment is significantly avoided in function words (e.g. *ónder* 'under'), while the bottom panel shows that both `s–p+` *and* `s+p–` are avoided in content words (e.g. *méisje* 'girl', *tambóer* 'drummer'). The total number of target syllables belonging to function words is much lower than those belonging to content words (814 vs 16,552), as reflected in the wider confidence intervals. Most observations of function words exhibit `s–` contours (top leftmost facet), which provides sufficient evidence for the avoidance of `s–p+`. Note, for instance, that the prepositions *óver* 'over' and *ónder* 'under' appear 227 and 188 times respectively, of which more than 90% follow a `s–p–` alignment. Function words with increasing stress contours are less frequent, and are often compounds, e.g. *totdát* 'until', *zoláng* 'as long as'. More than 20% of function words with an increasing stress contour exhibit a `s+p–` alignment, which does not allow for a robust avoidance rule. Consider, for instance, the word *totdát* 'until', which is the most frequent iambic function word in our sample: 9 cases follow a `s+p+` alignment, and 12 cases follow a `s+p–` alignment (Example 6.6). This particular example and the overall results indicate that we cannot posit a definite textsetting rule for iambic function words.

6.3.3 Musical context: melisma

The association between prominence and stress can also be mediated by features of the musical context. Here we address the problem of melismatic syllables, i.e. syllables sung to more than one musical note. Overall, syllables preceded by a melisma are (1) less frequent, and (2) less controlled from a textsetting point of view, as shown in Figure 6.3. This loosening of rules introduced by melisma replicates previous observations made for French textsetting (Dell & Halle 2009:71).

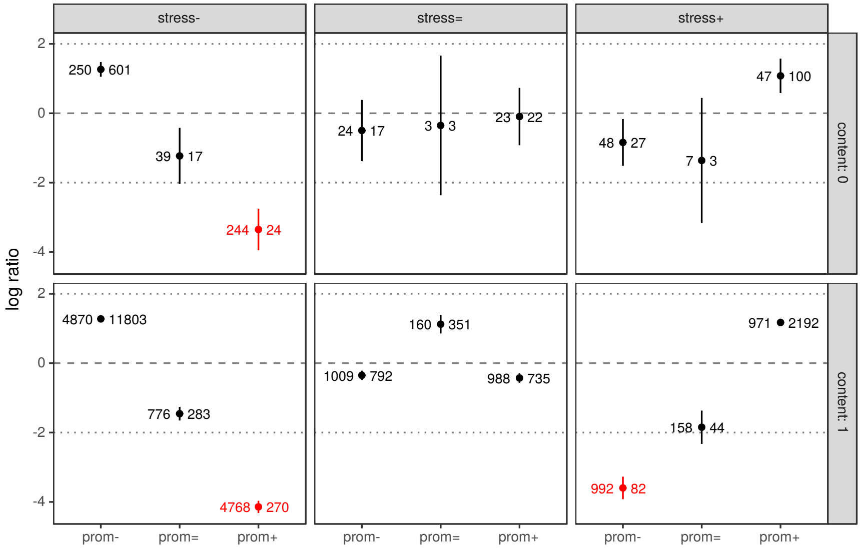


Figure 6.2: Association between stress and prominence in content words (bottom panel) vs function words (top panel). For each combination of stress and prominence values, the data points plot the log ratio between the observed frequency (number to the right of the point) and the expected frequency (left number). Error bars indicate the 95% confidence interval. Combinations below a log ratio of -2 are highlighted and interpreted as being significantly avoided.

As with function words, the relatively low frequency of contexts where a melismatic syllable precedes the target syllable produces less reliable results, as indicated by the larger error bars in the lower panel of Figure 6.3. The melismatic context is approximately 15 times less frequent than the non melismatic one, with the vast majority of the cases (92%) involving a decreasing stress contour. This lower leftmost facet provides the most reliable results: all three possible alignments (`s–p–`, `s–p=`, `s–p+`) are over-represented, even the `s–p+`, which is robustly avoided in the non-melismatic context (top leftmost facet). Consider the word *méisje* 'girl': in non-melismatic contexts, only 1% of the cases follow a `s–p+` alignment, while the proportion rises to 16% in the melismatic context.

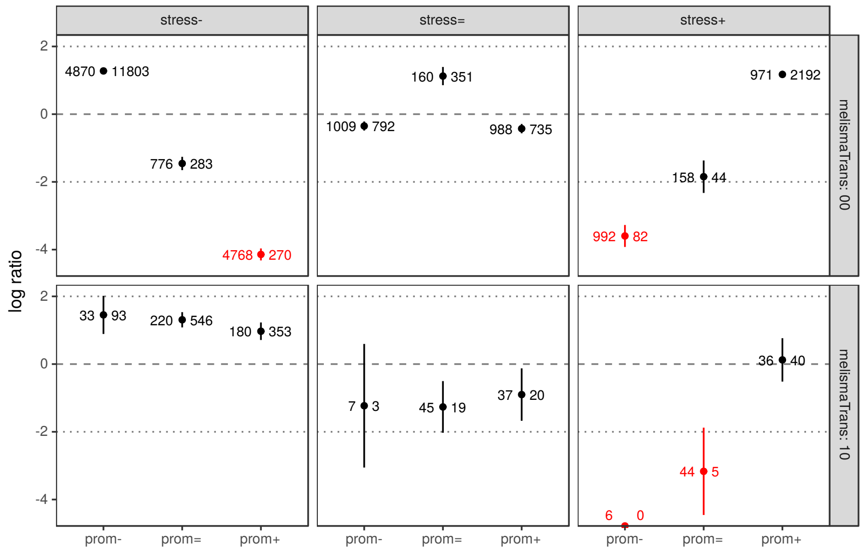


Figure 6.3: Association between stress and prominence in syllables preceded by a melisma (bottom panel) vs preceded by a non-melismatic syllable (top panel). For each combination of stress and prominence values, the data points plot the log ratio between the observed frequency (number to the right of the point) and the expected frequency (left number). Error bars indicate the 95% confidence interval. Combinations below a log ratio of −2 are highlighted and interpreted as being significantly avoided.

6.4 Discussion

The avoided combinations described in the previous section can be reformulated in terms of constraints which are active when a Dutch speaker creates a new song. Analogously, these constraints are arguably used in perception to judge a text-to-tune alignment as well- or ill-formed. Corpus studies like this one can put forward working hypotheses on such textsetting constraints, but their psychological reality is better tested through production (Hayes & Kaun 1996) and perception experiments (Gordon, Magne & Large 2011, and Chapter 7).

The fact that stress and prominence are relative entails that, minimally, the window where their alignment is compared must comprise two adjacent syllables. Our results suggest that this window is reset afresh at the beginning of each phrase, as evidenced by the lack of avoided alignments phrase-initially. Word-initially, the alignment is somehow controlled, although less so than within word-internal windows. Further analysis is needed in order to determine with more precision the relevant domain where textsetting rules apply (e.g. the phonological phrase, as argued for Italian by Proto & Dell 2013). In this respect, the current annotation of the corpus only offers information on word boundaries; in some cases, however, a word boundary can also mark the beginning/end of a higher-order constituent, such as a phonological or intonational phrase (Nespor & Vogel 2007). The decrease of stringency we observe word-initially may be the result of a gradual loosening of textsetting constraints across increasingly higher-order constituents.

In a comparable way to the textsetting rules proposed for tonal languages (Schellenberg 2009; Kirby & Ladd 2016), the general rule in our corpus is that stress and prominence should not move in opposite directions. Further, there is an asymmetry between s–p+ and s+p– alignments, the former being more heavily avoided. These tendencies, however, apply to a lesser extent in function words, and after the presence of a melismatic syllable.

By including these two finer-grained linguistic and musical contexts to the analysis, we have illustrated how the simple alignment of stress and prominence can be conditioned by further variables. The methodology straightforwardly allows for the inclusion of additional features, although interpreting the interactions between numerous variables can become complex, especially because it leads to an increasing number of absent (and thus uninformative) combinations in the corpus. All in all, there are at least two additional features which can prove fruitful for future studies.

On the linguistic side, we have followed a binary treatment of stress, as provided by most lexical datasets, but more detailed annotations can include further degrees, taking into account secondary stress and phrasal context (Hulst 1984; Booij 1995). Moving beyond stress binarity places the linguistic prominence on a gradual scale comparable to that of the musical metre (see Hayes, Wilson & Shisko 2012 for such an analysis in a corpus of English verse).

On the musical side, note duration and pitch are essential elements of the template, which we have ignored from the analysis so far. This omission is common in textsetting studies of stress languages, as it is generally regarded that pitch is less relevant for non-tonal languages (Halle & Lerdahl 1993, but see Särg & Ambrazevičius 2007). Nonetheless, there are reasons to further investigate the alignment of words to the melodic tier. First, pitch is one of the main cues for stress in languages such as Dutch or English (Sluijter & Heuven 1996), even if the associated contours are not as fixed as in tone languages (Hyman 2006). Second, pitch contours are also a source of prominence in musical tunes, creating so-called melodic accents (Thomassen 1982; Müllensiefen, Pfleiderer & Frieler 2009). These usually correlate and hence reinforce metrically prominent positions, but, crucially, they have the potential of producing prominence independently of metre (Särg & Ambrazevičius 2007 for a case study in Estonian and Lithuanian songs).

By discovering the musical and linguistic features relevant for text-to-tune alignment, we may shed some light on the cognitive processes involved in the simultaneous processing of music and language. Ultimately, the fact that speakers of a language are readily sensitive to the co-occurrence of features from these two domains needs to be accounted for by cognitive models aiming at explaining the shared or independent neural resources involved (Zatorre & Baum 2012; Hausen et al. 2013; Peretz 2012).

Finally, a thorough description of the textsetting grammar of any song style represents a valuable back-engineering tool. When presented with a historical set of lyrics for which the original tune is missing, the alignment constraints can be effectively used to search among a set of plausible melodies the one yielding the least violations (see e.g. the contributions within Proto, Canettieri & Valenti 2015). Even in the absence of potential melodies, a textsetting grammar can serve as a base model to generate, from scratch, well-formed templates (Conklin 2016), constrained by the specifications of the tune-less lyrics we are interested in. Naturally, the same procedure applies to the reverse process of generating a new text (or locating a missing one) given an available melody.

6.5 Conclusion

We have described a method which is general enough to describe in a systematic way textsetting constraints in the song traditions of different languages. The MTC-FS corpus of Dutch songs has been used as a case study, where we have described a number of avoided combinations and how they are affected by the linguistic and musical context. Further experimental validation and cross-linguistic studies will determine the most relevant features involved in textsetting grammars, and their typological prevalence.

References

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

Baayen, R. H., R Piepenbrock & H van Rijn. 1995. *The CELEX lexical database (release 2)*. Web version (last consulted: 2016.06.24). Philadelphia: Linguistic Data Consortium.

Booij, G. 1995. *The phonology of Dutch*. Oxford: Clarendon Press.

Conklin, D. 2016. Chord sequence generation with semiotic patterns. *Journal of Mathematics and Music* 10(2). 92–106.

Cuthbert, M. S. & C. Ariza. 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)*, 637–642. ISMIR.

Dell, F. & J. Halle. 2009. Comparing musical textsetting in French and English songs. In J.-L. Aroui & A Arleo (eds.), *Towards a typology of poetic forms*, 63–78. Benjamins.

Deutsch, D. 2013. Grouping mechanisms in music. In D. Deutsch (ed.), *The psychology of music*, 183–248. Elsevier.

Gordon, R. L., C. L. Magne & E. W. Large. 2011. EEG correlates of song prosody: A new look at the relationship between linguistic and musical rhythm. *Frontiers in Psychology* 2(352).

Halle, J & F Lerdahl. 1993. A generative textsetting model. *Current Musicology* 55. 3–23.

Halle, J. 1999. *A grammar of improvised textsetting*. Columbia University Doctoral dissertation.

Halle, J. 2004. Constituency matching in metrical texts. In *Proceedings of the Words and Music conference*. University of Missouri-Columbia.

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

Hayes, B. 2009. Textsetting as constraint conflict. In J.-L. Aroui & A Arleo (eds.), *Towards a typology of poetic forms*, 43–61. Benjamins.

Hayes, B. & A. Kaun. 1996. The role of phonological phrasing in sung and chanted verse. *The Linguistic Review* 13. 243–303.

Hayes, B., C. Wilson & A. Shisko. 2012. Maxent grammars for the metrics of Shakespeare and Milton. *Language* 88(4). 691–731.

Hulst, H. Van der. 1984. *Syllable structure and stress in Dutch*. Foris.

Hyman, L. M. 2006. Word-prosodic typology. *Phonology* 23(2). 225–257.

Jackendoff, R. 2009. Parallels and nonparallels between language and music. *Music Perception* 26(3). 195–204.

Kirby, J. & D. R. Ladd. 2016. Tone-melody correspondence in Vietnamese popular song. *Tonal Aspects of Languages 2016*. 48–51.

Lerdahl, F & R Jackendoff. 1983. *A generative theory of tonal music*. Massachusetts Institute of Technology.

London, J. 2004. *Hearing in time, psychological aspects of musical meter*. Oxford University Press.

Matthews, P. H. 2007. *Syntactic relations: a critical survey*. Vol. 114. Cambridge University Press.

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

Nespor, M. & I. Vogel. 2007. *Prosodic phonology: with a new foreword*. Walter de Gruyter.

Peretz, I. 2012. Music, language, and modularity in action. In P Rebuschat, M Rohrmeier, J.-A. Hawkins & I Cross (eds.), *Language and music as cognitive systems*, 254–268. Oxford University Press.

Proto, T & F Dell. 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.

Proto, T. 2015. Prosody, melody and rhythm in vocal music: the problem of textsetting in a linguistic perspective. *Linguistics in the Netherlands* 32(1). 116–129.

Proto, T., P. Canettieri & G. Valenti. 2015. *Text and tune: on the association of music and lyrics in sung verse*. Bern: Peter Lang.

Särg, T & R Ambrazevičius. 2007. Melodic accent in Estonian and Lithuanian folk songs. In K Maimets-Volt, R Parncutt, M Marin & J Ross (eds.), *Proceedings of the 3rd conference on interdisciplinary musicology*. Published at: http://www-gewi.uni-graz.at/cim07. Tallinn (Estonia): Estonian Academy of Music & Theatre.

Schellenberg, M. 2009. Singing in a tone language: Shona. In *Selected proceedings of the 39th annual conference on African linguistics*, 137–144. Cascadilla Proceedings Project Somerville, MA.

Schellenberg, M. 2012. Does language determine music in tone languages? *Ethnomusicology* 56(2). 266–278.

Schütze, C. T. 2011. Linguistic evidence and grammatical theory. *Wiley Interdisciplinary Reviews: Cognitive Science* 2(2). 206–221.

Schütze, C. T. 2016. *The empirical base of linguistics: grammaticality judgments and linguistic methodology*. Berlin: Language Science Press.

Sluijter, A. M. & V. J. van Heuven. 1996. Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Fourth international conference on spoken language, icslp 96*, vol. 2, 630–633. IEEE.

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

Thomassen, J.-M. 1982. Melodic accent: experiments and a tentative model. *Journal of the Acoustical Society of America* 71. 1596–1605.

van Kranenburg, P & F Karsdorp. 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)*, 391–396. ISMIR.

van Kranenburg, P, M de Bruin, L. de Grijp & F Wiering. 2014. The Meertens Tune Collections. *Meertens Online Reports* 1.

Yip, M. 2002. *Tone*. Cambridge University Press.

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

Supplementary Information

Table 6.1: Log ratios between observed and expected frequencies of stress and prominence alignment in three different domains: line-initial, word-initial, and word-internal.

(a) Line-initial domain (n = 2,032).

	prom-	prom=	prom+
stress-	-0.28	-1.65	-0.33
stress=	-1.43	-1.12	-0.40
stress+	-1.70	-0.13	0.72

(b) Word-initial domain (n = 31,616).

	prom-	prom=	prom+
stress-	0.90	0.15	-1.95
stress=	-1.10	-0.60	0.64
stress+	-2.74	-0.08	1.03

(c) Word-internal domain (n = 16,552).

	prom-	prom=	prom+
stress-	1.28	-1.46	-4.14
stress=	-0.35	1.13	-0.43
stress+	-3.60	-1.85	1.17

Table 6.2: Log ratios between the observed and expected frequencies of stress and prominence alignment in function and content words.

(a) Target syllable belongs to a function word (n = 814).

	prom-	prom=	prom+
stress-	1.27	-1.23	-3.35
stress=	-0.50	-0.35	-0.10
stress+	-0.84	-1.36	1.08

(b) Target syllable belongs to a content word (n = 16,552).

	prom-	prom=	prom+
stress-	1.28	-1.46	-4.14
stress=	-0.35	1.13	-0.43
stress+	-3.60	-1.85	1.17

Table 6.3: Log ratios between the observed and expected frequencies of stress and prominence alignment in two musical contexts: preceded by a melismatic or by a non-melismatic syllable.

(a) Target syllable preceded by a melismatic syllable (n = 1,079).

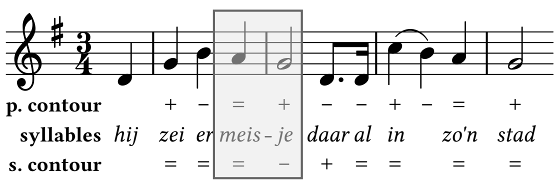
	prom-	prom=	prom+
stress-	1.45	1.31	0.97
stress=	-1.23	-1.27	-0.90
stress+	-Inf	-3.17	0.12

(b) Target syllable preceded by a non-melismatic syllable (n = 16,552).

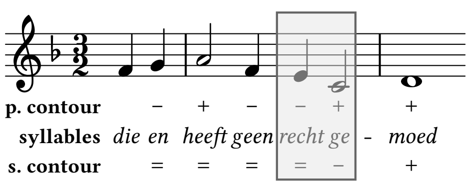
	prom-	prom=	prom+
stress-	1.28	-1.46	-4.14
stress=	-0.35	1.13	-0.43
stress+	-3.60	-1.85	1.17

Example 6.4: A decreasing stress contour aligned with an increasing prominence (`s–p+`) is heavily avoided word-internally, though not word-initially.

(a) Word-internal `s–p+` (song `NLB072300`).

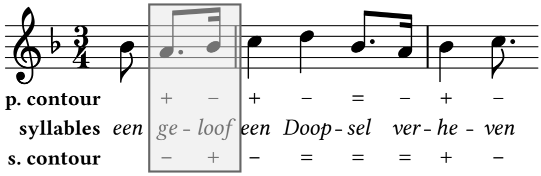


(b) Word-initial `s–p+` (song `NLB161811`).

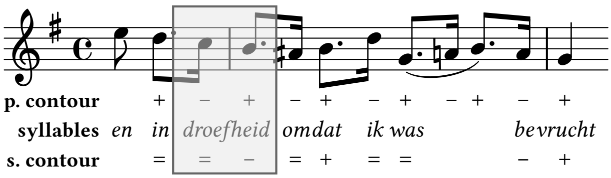


Example 6.5: In content words with no melisma, the alignment `s–p+` is more heavily avoided than `s+p–`.

(a) The word *geloof* with `s+p–` (song `NLB111653`).

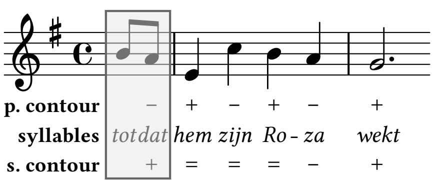


(b) The word *droefheid* with `s–p+` (song `NLB074530`).

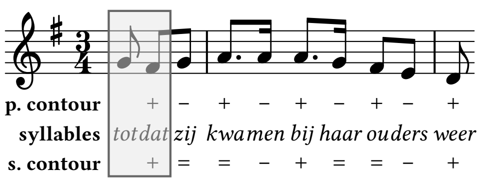


Example 6.6: In function words, the alignment `s+p–` is not robustly avoided. The word *totdat*, for example, shows 9 instances of `s+p+`, and 12 instances of `s+p–`.

(a) Function word *totdat* with `s+p–` (song `NLB070749`).

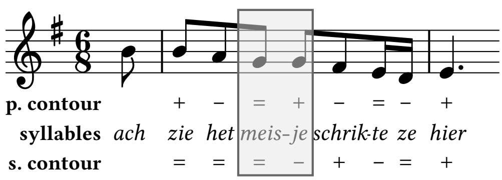


(b) Function word *totdat* with `s+p+` (song `NLB075303`).



Example 6.7: The alignment `s–p+` is not avoided when the target syllable is preceded by a melisma. This example illustrates the case of the word *meisje*: in non-melismatic contexts, only 1% of the cases follow a `s–p+` alignment, while the proportion rises to 16% in the melismatic context.

(a) The word *meisje* with `s–p+` in a non-melismatic context (song `NLB070805`).



(b) The word *meisje* with `s–p+` in a melismatic context (song `NLB072415`).