Comparing musical textsetting in French and in English songs

François Dell & John Halle*
CRLAO, EHESS-CNRS, Paris, France/Bard Conservatory of Music, Anondale-on-Hudson, NY, USA

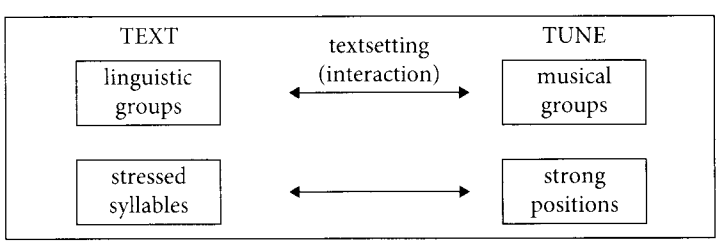
A strophic song is in general a sequence of repetitions of the same tune, one repetition for each stanza. Strophic songs in French have the following properties: when two stanzas have the same tune, they have the same number of lines, and for any *i*, the *i*-th line in one stanza has the same number of syllables as the *i*-th line in the other, and it also has the same distribution of melismas. These properties follow from a more general requirement on textsettings that we call Positional Parallelism. Whereas violations of Positional Parallelism are rather infrequent in traditional French songs, they are quite common in English songs.

We propose to relate this difference between French and English songs with another difference in textsetting practice. English matches stress and musical beat anywhere in a line. French enforces the stress/beat match in a rigid manner only at the end of lines, which is presumably a reflection of the fact that in French, stress is easily perceptible only before major breaks.

1. Preliminaries

A song is a composite which combines two objects each with its own structure, a linguistic object (text) and a musical object (tune). In addition to providing a description of each component, one must specify how the two objects interact, see (1):

(1) *Composite* (two independent objects)



*We thank Morris Halle, two anonymous reviewers and conference participants for various helpful criticisms and suggestions. Any remaining problems are the responsibility of the authors.

In this paper, the terms "alignment" and "textsetting" are used as synonyms to refer to the correspondence between text and tune. In our view of textsetting, alignments make use of two formal similarities between language and music. A first similarity is hierarchical constituent structure. Like syllables and words in language, the notes in a tune form groups and these groups in turn form larger groups, resulting in a hierarchy of nested groups. The second similarity between language and music resides in the fact that in both domains certain units are heard as relatively prominent compared to others in their immediate vicinity: "stressed" or "accented" syllables in language, metrically strong positions ("strong beats") in music. Textsetting must achieve the two independent matches represented by the arrows in (1). One match has to do with constituency while the other pertains to prominence: linguistic groups must match with musical groups on the one hand, and on the other hand certain stressed syllables must fall on strong beats.

To give a rough idea of what we mean by constituency matching, let us consider the text at the beginning of the nursery rhyme *Au clair de la lune*,[1] which is represented in (2):

(2)
```
[---------- P ----------]
[--------- J ----------][-------- K --------][---------- N ----------]
                                [--------- L ----------][--------- M -------]
au clair de la lune, mon ami Pierrot, prête-moi ta plume pour écrire un mot.
```

According to (2), the text is a group P which breaks down into three groups, viz J (*au clair de la lune*), K (*mon ami Pierrot*) and N (*prête-moi ta plume pour écrire un mot*). N is in turn comprised of groups L and M, and so on down to lower levels which are not represented in the diagram. (2) is a partial representation of what linguists call prosodic structure,[2] and in this paper the expression "linguistic group" means the same as "prosodic constituent".

Moving to musical groups, (3) below is a representation of that portion of the tune which is a carrier for the text in (2):

(3)
```
a. (_____VII_____)
b. (_____V_____)(_____VI_____)
c. (_____I_____)(_____II_____)(_____III_____)(_____IV_____)
d. x x x x x x x x  x x x x x x x x  x x x x x x x x  x x x x x x x x
e. G G G A B  A  G B A A G  G G G A B  A  G B A A G
```

The individual pitches and their sequencing in time are represented on lines (3d–e). The other lines in the diagram show how they are organized into hierarchical groupings.

1. Charpentreau (1976: 93).

2. See e.g. Hayes (1989).

Each successive note in the tune is defined by two parameters, its pitch and the time of its inception. The capital letters in the bottom line in (3) represent pitch. The "x" marks in the line above represent evenly-spaced points in time. A note is represented directly underneath an "x" mark in order to indicate that that "x" is the moment when the note in question *begins*.

According to (3a–c), VII, the portion of the tune under consideration, is a musical group and it breaks down into groups V and VI. V in turn breaks down into groups I and II, while VI breaks down into groups III and IV. The hierarchical structure displayed in (3) is what Lerdahl & Jackendoff (1983) call "grouping structure".

Text-to-tune alignment requires that linguistic groups and musical groups match to a certain extent. (4) below represents the correspondence between the linguistic groups in (2) and the musical groups in (3):

(4) *Constituency Matching*

a. (_____ VII _____)
b. (_____ V _____) (_____ VI _____)
c. (_____ I _____) (_____ II _____) (_____ III _____) (_____ IV _____)
d. x x x x x x x x x x x x x x x x x x x x x x x x x x x x x x x x
e. G G G A B A G B A A G G G G A B A G B A A G

au clair.....lune mon ami Pierrot prête.....plume pour écrire un mot

[-------- L --------][-------- M ---------]
[--------- J -------][------- K --------][------------------ N ----------------]
[------------------------------------ P ------------------------------------]

In (4) the edges of musical group I coincide with those of linguistic group J, those of musical group II coincide with those of linguistic group K, and so on. As a rule, mismatches between linguistic constituent structure and musical constituent structure yield ill-formed composites. Here is for instance the beginning of the composite which results from singing the text of *En passant par la Lorraine*[3] to the tune of *Au clair de la lune*:

(5) (_____ I _____)(_____ II _____)(_____ III _____)(___ IV _
G G G A B A G B A A G G G G A B A G B A

en passant par la **Lor--rain'** avec mes **sa--bots** rencontrai trois **ca--pi**taines
[-------------------------][---------------][-------------------------]

In (5) as in (4), the edges of musical groups are indicated by parentheses and those of linguistic groups are indicated by square brackets. Musical groups I, II, III and IV

3. Charpentreau (1976: 220).

are the same as those in (4c), and the notes which make up each group (cf. (4e)) are printed above the corresponding vowels in the text of *En passant par la Lorraine*. The words in boldface give the impression of being torn apart by the music, as their first two syllables are sung to notes which do not belong to the same musical group.

We will not say more about constituency matching, about which see Halle (2003). We now turn to the concordance between musical beat and linguistic stress.[4] Let us first introduce the metrical grid, which will be used to represent the alternation between strong and weak beats in music.

In singing, the beginning of each syllable in the text coincides with that of a note in the tune.[5] (6) below represents the beginning of the nursery rhyme *Ne pleure pas Jeannette*.[6]

(6)

G1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	C		A			G	A		Bb		A			G	F		
	|		|			|	|		|		|			|	|		
	ne		pleu-				re	pas		Jean-		net-			–	te	

(6) is an adequate representation of the objective structure produced by someone who sings the song, but not of the internalized structure which underlies the singer's performance and which listeners infer from it. Using the information contained in (6), their perceptual apparatus constructs the following representation. (Rows and columns are numbered for ease of reference):

(7)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
G4			x								x					
G3			x				x				x				x	
G2	x		x		x		x		x		x		x		x	
G1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	C		A			G	A		Bb		A			G	F	
	|		|			|	|		|		|			|	|	
	ne		pleu-				re	pas		Jean-		net-			–	te	

(6) is changed into (7) by constructing lines G2–G4 above line G1. The layering of lines G1–G4 is what is called a metrical grid, and the columns in that grid are called metrical positions. The metrical grid of a melody represents the fact that certain positions are felt to be stronger than others. The strength of a position is represented by

4. In this paper we use "stress" and "accent" as strict synonyms.

5. This is a simplification. Strictly speaking, the points in the speech chain which are associated with note attacks are the beginnings of vowels, rather than those of syllables. As each syllable contains one vowel and only one, this simplification is without consequence in this paper.

6. Charpentreau (1976: 233).

the height of the column of "x" tokens. For instance, the strongest positions in (7) are those numbered 3 and 11. The grid plays the same role as time signatures and bars in conventional musical notation.

Like its grouping structure, the metrical grid of a tune is independent of the words associated with it: a purely instrumental rendition of the tune on a violin or flute will manifest the same metrical grid as the tune with a text assigned to it. Our only reason for including the lyrics in diagrams (6) and (7) was to facilitate the comprehension of the song for those who are not fluent in musical notation. For an excellent discussion of the various factors at play in the construction of metrical grids, see Lerdahl & Jackendoff (1983).

2. Prominence matching in English

In languages in which certain syllables are more stressed than others, there is a preference for songs to associate stressed syllables with strong metrical positions. This is what we call stress-to-beat matching. In (7), for instance, the syllable *pleu-* has stronger stress than the adjacent syllables *ne* and *-re*, and metrical position 3, which is occupied by *pleu-*, is stronger than the positions occupied by *ne* and *-re*, viz positions 1 and 6. We shall see below that in French, except at the end of lines, the preference of stressed syllables for strong metrical positions is only a statistical tendency. Stress-to-beat matching is more strict in English, where certain stress/beat mismatches are prohibited in all environments. (8) below represents three English sentences sung to the tune of (7):

(8)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
G4			x								x					
G3(TACTUS)			x				x				x				x	
G2	x		x		x		x		x		x		x		x	
G1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	|		|			|	|		|		|		|	|		
	C	(C)	A		(G)	G	A		Bb		A			G	F	
	|	|	|		|	|			|		|			|	|		
a.	my		neigh-			bor's	dog		is		bar-			king	now	
*b.	my		old			**neigh-bor's**			dog		barks			a	lot	
c.	my	old	neigh-			bor's	dog		–		barks			a	lot	
d.	my		old		neigh-	bor's	dog		–		barks			a	lot	
e.	the		king			of	France		died		yes-			ter-	day	

As will be explained below, parenthesized capitals represent notes added to the original tune: C at position 2 for the tune of composite (8c), and G at position 5 for the tune of composite (8d). (8e) is not relevant for the time being and will be discussed later.

In (8a) the eight syllables in the sentence *my neighbor's dog is barking now* are sung to the eight notes in the tune of (7), and the resulting composite sounds natural to English speakers. In (8b) the sequence associated with the same eight notes is *my old neighbor's dog barks a lot*, which is also comprised of eight syllables, but in this case the result sounds awkward to English speakers. The difference in well-formedness between the two composites has to do with the alignment of *neighbor*, a word with stress on its first syllable. In the well-formed composite (8a) the metrical position of the stresssed syllable is stronger than that of the unstressed syllable, but it is the other way round in (8b). It is the mismatch between stress and beat which makes (8b) ill-formed.

If one wants to sing the eight syllables in *my old neighbor's dog barks a lot* to the tune in (7), one must make a slight change to that tune. (8c) and (8d) are two composites that English speakers find acceptable. The tune is comprised of nine notes in either composite. The tune in (8c) was obtained by inserting a C between the first two notes in the original tune and the additional note is compensated for by using the syllable *dog* as a carrier for two notes in sequence (A and Bb). The note inserted at position 2 is associated with *old*, which frees position 3 for the first syllable of *neighbor's*. The tune of (8d) is derived from that of (7) by inserting a G at position 5, and in the resulting composite the metrical position of *neigh-* is stronger than that of *-bor's*.

In the metrical grid in (8), line G3 represents what Lerdahl & Jackendoff (1983) call the tactus. The tactus of a tune is the regular beat which listeners naturally fall into when for example they clap their hands or choregraph dance steps to that tune. In the case of (8), the handclaps would land on positions 3, 7, 11 and 15, i.e. on all those positions at which there is a token on line G3. Whether one defines the tactus as a certain layer in the grid (line G3), or as the set of all those positions in the grid which have a token belonging to that layer (the set {3,7,11,15}), the result is the same. In our view the tactus plays a special role in the definition of stress/beat mismatches in English. Let us first state what counts as a stress/beat mismatch in English. We then indicate certain environments in which such mismatches are prohibited.

(9) *Stress mismatch in English (definition).*[7]

Let *S* and *s* be two syllables occurring in any order, one accented (*S*) and the other unaccented (*s*). A stress mismatch occurs if *S* and *s* are adjacent within the same line and the metrical position associated with *s* belongs to the tactus, but not that associated with *S*.

According to this definition, the alignment of *neighbor's* in (8b) is a stress mismatch, as the adjacent syllables *neigh-* and *-bor's* belong to the same line in the song, one is

7. This definition is relatively close to that implied by the constraint MATCHSTRESS proposed in Hayes (2008). One difference worth noticing is the fact that our definition accords a special status to the tactus.

accented and the other is unaccented, and furthermore the metrical position associated with the unaccented syllable (*-bor's*) belongs to the tactus, whereas that associated with the accented syllable (*neigh-*) does not.

Here is what we mean by "accented" and "unaccented" in definition (9). We say that a syllable is accented (*S*) when it is the only syllable of a content word (e.g. *boy*) or it bears the main stress in a polysyllabic word, and furthermore if those words are included in a compound word, the syllable bears the main stress of the whole compound. All other syllables will be said to be unaccented (*s*). Here are the accentual contours of a few noun phrases according to the definition just given: *the boy*, s-S; *my hat*, s-S; *a reptile*, s-S-s; *in paradise*, s-S-s-s; *a cowboy*, s-S-s; *green cowboy hats*, S-S-s-s.

We now indicate two classes of contexts in which stress mismatches are prohibited.

(10) *Mismatches illicit in English*: a composite is ill-formed if it contains a stress mismatch which meets one of the following conditions:

a. the two syllables are not separated by a word boundary ; or
b. the two syllables are separated by a word boundary and *s* precedes *S*.

Condition (a) prohibits mismatches involving words such as *dinner*, *balloon* and *rebate*, but it does not say anything about words such as *cowboy* and *software*, which have an internal word boundary.[8] Condition (b) prohibits those mismatches which involve sequences such as *the boy*, but not those which involve sequences such as *see me*.

Let us return to (8). (8b) is ill-formed because the alignment of *neighbor's* gives rise to a stress mismatch of type (10a). One can avoid the lethal mismatch by singing (8c) or (8d) instead. The tunes in (8c) and (8d) are not identical with the original tune (that of (8a)), but they bear such a close resemblance to the original tune that many listeners do not hear the difference. The tunes of (8a), (8c) and (8d) give the impression of being "the same tune", or at least variants of the same tune.[9]

What conditions must two tunes satisfy, for listeners to perceive them as variants of the same tune? The answer does not seem to depend on the linguistic material aligned with the tunes, i.e. the question is about tunes considered from a purely musical point of view. Preliminary investigations suggest in particular the following generalization:

(11) *Melodic Contour Conservation*: for two tunes to be perceived as variants of the same tune, they must have the same melodic contour.

---

8. Hayes and Kaun (1996) observe that mismatches of type (10a) sporadically occur line-finally. The attested mismatches all involve words with a falling stress contour, e.g. *morning*. On these mismatches, see also Hayes (2002).

9. For an attempt at a formal description of similarity in strophic songs see Halle (forthcoming).

What we are calling the melodic contour of a tune is the sequence of pitch changes in that tune, abstracting away from the number of note attacks and their timing.[10] In (3), for instance, musical group I is the sequence G-G-G-A-B-A, but the melodic contour in that group is simply G-A-B-A, for the initial sequence G-G-G does not contain any pitch change. Similarly, in musical group II, which is G-B-A-A-G, the melodic contour is G-B-A-G. Returning to (8), composites (a), (c) and (d) have different tunes, but all three tunes have the same melodic contour, namely C-A-G-A-Bb-A-G-F. The tune in (c) is derived from that in (a) by replacing the initial C in (a) by a sequence of two Cs, an operation which does not change the melodic contour.

Another manner of formulating Melodic Contour Conservation, is to say that when a variant tune is created by adding or deleting a note, that note must have the same pitch as a note adjacent to the addition or deletion site. Note that (11) is a necessary condition, not a sufficient one: not all tunes which abide by (11) are perceived as variants. As already stated, Melodic Contour Conservation is a requirement of a purely musical nature.[11] We shall see below that it interacts with Positional Parallelism, a requirement regulating text-to-tune alignment.

Returning to alignments, let us compare composites (8a) and (8c). The texts in these composites have the same number of syllables, namely 8, but accented and unaccented syllables are distributed differently, and the price to pay in order to sing the texts to "the same tune", is a departure from a property which we call positional parallelism:

(12) *Positional Parallelism (definition):*
Two alignments are positionally parallel if the distributions of their syllables along the grid are identical.

In (8), (e) and (a) are positionally parallel, whereas (c) and (a) are not. Every position in the grid which is associated with the beginning of a syllable in (e) is associated with the beginning of a syllable in (a), and conversely, whereas the same does not hold for (c) and (a). Position 2 coincides with a syllable onset in (c), but not in (a), and position 9 coincides with a syllable onset in (a), but not in (c). In (c), the syllable *dog* is sung to two notes the second of which begins at position 9. In (c), then, position 9 does not correspond to *the beginning* of a syllable, whereas it does in (a).

Definition (12) implies that two composites which do not have the same number of syllables cannot be positionally parallel.

---

10. For a discussion of formal aspects of melodic contour, see Quinn (1999).

11. (11) implies an underlying structure of tonal melodies reminiscent of the reductional scheme proposed by Heinrich Schenker. For an introduction to Schenkerian theory see, for example, Cadwallader & Gagne (1998). On the cognitive status of that theory, see Lerdahl & Jackendoff (1983), Chapters 5–7.

3. Prominence matching in French

In this paper we limit ourselves to a singing idiom which we call "traditional" for lack of a better term. This idiom includes nursery rhymes and most commercial songs until the advent of French imitiations of Anglo-American pop songs in the early 1960s. It does not include "serious" music, e.g. art-songs by Debussy and Fauré. In French the vowel schwa (ə) is subject to various deletion and epenthesis processes. As some of these processes are optional, the same word can be pronounced in more than one way in a given environment. In *tu retournes les cartes* "you turn up the cards", for instance, *retournes* has four acceptable pronunciations, namely [rəturnə], [rturnə], [rəturn] and [rturn]. How optional schwa deletion and optional schwa epenthesis operate depends on the style of delivery, and it is the style of pronunciation appropriate in singing which provides the richest range of alternate pronunciations.[12] In traditional French songs, alignments are rigidly constrained at the end of lines:

(13) At the end of lines, certain melodies call for a form which is phonetically feminine, while others call for a form which is phonetically masculine.

What we mean by a form is a particular pronunciation of a word. A form is phonetically feminine if it is polysyllabic and its last vowel is schwa (ə); otherwise it is phonetically masculine. Consider for instance the imperative *détourne* 'divert!'. This word can be pronounced [deturnə] or [deturn]. The former pronunciation is a phonetically feminine form, and the latter, a phonetically masculine form. *De* [də] 'of', which ends in a schwa but has only one syllable, is a masculine form. From now on all mentions of gender must be understood as refering to phonetic gender as just defined. Let us use the nursery rhyme *Compagnons de la Marjolaine* to illustrate (13).[13]

(14a) below is the alignment of the first line in a stanza of the song. The other lines in (14) are composites obtained by changing the end of the text but keeping the same number of syllables.[14] The melody of (14) requires a text ending in a masculine form.[15]

12. For a summary of the facts about the behavior of schwa in singing, see Dell (1989: 122–124).

13. Charpentreau (1976: 204).

14. Our French examples are given in the conventional spelling, with the following adaptations to indicate the relevant facts concerning the pronunciation: an apostrophe represents an orthographic "e" which does not have a corresponding vowel sound in the pronunciation of the example, and line-final occurrences of orthographic "e" are replaced by "ə" to indicate that the example is pronounced with a final schwa.

15. This is a simplification. The melody also accepts lyrics ending in a feminine form, provided that the next-but-last syllable be the carrier of more than one note, e.g. *que demand' son pè-è-re*. Our discussion of stress-to-beat matching in French is limited to line ends without melismas, i.e. without syllables which carry more than one note.

For that reason, the original composite (a) sounds natural to the ears of French speakers, and so does the construct (b), while construct (c) sounds awkward:

(14)

	G1		x				x				x				x				
	G0		x	x	x	x	x	x	x	x	x	x	x	x	x	x			
			A			A	E		E	D			E	C					
			|			|	|		|	|			|	|				
a.	MASC		que			de-mand'			le	che-			va-	**lier**					
b.	MASC								les	deux			com-	**tess'**					
*c.	FEM								la	com-			**tes-**	sə					

(15a) below is the alignment of the second line in the stanza. The other lines in (15) are constructs obtained by changing the words at the end of the original line. As the melody of (15) calls for a text ending in a feminine form, composites (a) and (b) sound natural to the native ear. On the other hand (c) and (d) sound awkward, due to the fact that they end with masculine forms.

(15)

	G1		x				x				x				x				
	G0		x	x	x	x	x	x	x	x	x	x	x	x	x	x			
			B			C	D	D	D	C			D	E			E		
			|			|	|	|	|	|			|	|			|	
a.	FEM		com-			pa-	gnons	de	la	Mar-			jo-	**lai-**			nə		
b.	FEM								ma	chèr'			Ni-	**co-**			lə		
*c.	MASC								ce	cher			Ni-	co-			**las**		
*d.	MASC								ma	chè-			re	Ni-			**col'**		

A common view is that feminine forms are accented on their penultimate syllable, and masculine forms on their last syllable. Line-final accented syllables are in boldface in (14)–(15). If one adopts this view of French accentuation, generalization (13) may be construed as following from the following prominence-matching requirement:

(16) *Local Maximum:*
The grid position associated with the last accented syllable in a line must be stronger than the positions associated with adjacent syllables belonging to the same line.

In (15b), for instance, condition (16) is met, as the feminine form *Ni-co-lə* is accented on its penultimate syllable (*-co-*) and the grid position occupied by that syllable is metrically stronger than those occupied by syllables *Ni-* and *-lə*. Condition (16) is not met in (15d), on the other hand, for in the masculine form *Nicol'* it is the last syllable which is accented, and its position is weaker than that of *Ni-*.

Like the restrictions formulated in (10) for English, *Local Maximum* excludes certain mismatches between linguistic prominence and musical prominence. The excluded mismatches are not the same in English and in French, but they bear a definite family resemblance. There does not exist at present an overwhelmingly-accepted analysis of accentuation in French. There are even researchers who deny that French has anything like stress.[16] To adopt that view would make it impossible to interpret the facts summarized in (13) in terms of stress, which would in turn make it impossible to capture the relatedness between the restrictions on line ends in French songs on the one hand, and the concordance between stress and musical beat in English, Italian or Russian songs on the other hand. Here is a case where the facts of text-to-tune alignment in songs provide important evidence as to the phonological structure of the language.

Local Maximum only concerns line ends. A tendency to match stress and beat is also found elsewhere in traditional French songs, but it is no more than a tendency. Whereas conflicts between stress and beat at the end of lines are extremely rare, and they grate on the ear when they occur, such conflicts are commonplace in other positions, where most of the time they escape the listener's notice, see Dell (1989) for examples and some discussion.

4. Positional parallelism in strophic songs

In strophic songs each stanza is sung to the same tune or a slightly different note sequence which listeners perceive as a close variant of the basic tune. Resorting to such variants is much more frequent in English than in French, as we shall now see.

Consider again *Ne pleure pas Jeannette*. The song is comprised of seven stanzas, and the tune is exactly the same in every stanza. That tune can be broken down into four musical groups, and these groups divide the composite into four successive lines. (17) below displays the alignments for the stanza-initial lines in the first four stanzas of the song.

(17)

			x								x					
			x				x				x				x	
	x		x		x		x		x		x		x		x	
	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	C		A			G	A		Bb		A			G	F	
1.	ne		pleu-			re	pas		Jean-		net			–	tə	
2.	a-		vec			le	fils		d'un		prin-			–	cə	
3.	je		ne			veux	pas		d'un		prin-			–	cə	
4.	je		veux			mon	a-		mi		Pier-			–	rə	

16. See Di Cristo (1999) for a survey of the literature on the accentuation of Modern French.

All four composites in (17) contain seven syllables, and the syllables are distributed along the grid in the same manner. Note in particular that the syllable with a melisma is always the next-but-last. (17) illustrates the following generalization:

(18) In a French strophic song, as a rule, all stanzas are positionally parallel (see definition (12)).

A song with positionally parallel stanzas has the following properties, among others: (a) all stanzas have exactly the same tune; (b) if two lines occupy the same position in the stanza, they have the same number of syllables, e.g. if the third line in one stanza has nine syllables, the third line has nine syllables in every stanza in the song; (c) the distribution of melismas is the same in all stanzas. In essence, these three properties are mere consequences of combining positional parallelism with Melodic Contour Conservation (11), as a little thought should convince the reader. Note in particular that as far as property (b) is concerned, there is no need to assume that the text has a metrical structure of its own, which would be responsible for regularities in syllable count. In traditional French songs, the only device for "counting" syllables is their alignment with the notes in the tune, as explained in Dell (2004).

One occasionally comes across French songs with departures from positional parallelism, but such songs are a small minority, see the article just cited for some examples. In English, on the other hand, such departures are commonplace. We give two examples. The first is from *The Star-Spangled Banner*, the national anthem of the United States. Whereas the first line in the first stanza has five syllables, the first line in the second stanza has six syllables. The notes are represented by digits to save space:

(19)		1	2	3	4	5	6
	a.	**O-**	**-oh**	say	can	you	see
	b.	**on**	**the**	shore	dim-	ly	seen

The two lines have exactly the same melody. In (19b) there is a one-to-one correspondence between the notes and the syllables. In (19a), on the other hand, there are only five syllables, and the first syllable carries two notes. Our second example, in (20) below, is from the nursery rhyme *The Farmer in the Dell*.[17] Whereas the first line in the first stanza has 6 syllables, the first line in the second stanza has only 5, and its tune has one note less than the original tune (the deleted note is indicated by parentheses):

17. Linscott (1939/1993: 8).

(20)

			x						x					
			x			x			x			x		
	x	x	x	x	x	x	x	x	x	x	x	x	x	
	D	G		(G)	G		G	G						
	|	|		|	|		|	|						
a.	the	far-		mer	in		the	dell						
b.	the	wife			takes		the	child						

A final example of the departures from positional parallelism which are allowed in English is the sea chantey discussed in Halle & Lerdahl (1993). The tune in that song assumes variant shapes which allow it to accommodate texts whose lengths vary between 7 and 13 syllables.

(21)

				x				x				x				x								
	x	.	x		x		x		x		x		x		x		x			x			
	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			
	|	|	|		|	|	|		|	|	|		|	|	|	|	|				
	(A)	(A)	A		A	A	A		A	A	A		D	(D)	F	(F)	A							
a.			What		shall	we	do		with	a	drunk-		en		sail-		or							
b.			Keel-				haul		him		till		he's		sob-		er							
c.	Scrape	the	hair		off	his	chest		with	a	hoop		ir-	on	raz-	or								

5. Conclusions

We have presented two differences between textsetting in English and in French songs. The first difference concerns the concordance between linguistic accents and beat strength. Whereas stress-to-beat matching is in force in all contexts in English, in French it is strictly enforced only at the end of lines. The second difference concerns parallelism in settings with the same melody. Positional parallelism across stanzas is the norm in traditional French songs, whereas it is routinely violated in English.

At first sight, these two differences appear unrelated. The first pertains to a property of settings which is of a very local nature. In order to determine whether a prominence mismatch occurs at a given location in a setting, one needs only to examine three adjacent syllables, i.e. a stressed syllable and the syllables adjacent to it (see (9) and (16)). All the examples in this paper are from strophic songs, but nonstrophic songs are regulated by the same restrictions on prominence mismatches.[18]

18. For French, see for instance the poems by Jacques Prévert set to music by Joseph Kosma.

The second difference concerns properties of settings which are of a more global nature. The definition of positional parallelism formulated in (12) does not say anything about the size of the settings involved, but generalization (18) is about whole songs.

We believe that these two differences are not unconnected facts, and that the second difference follows from the first. Here is how. Let us assume that all singing idioms are under a universal pressure for positional parallelism:

(22) *P-Parallelism (universal):*
To the extent possible, identical melodies give rise to positionally parallel settings.

The prevalence of parallelism across stanzas in French is a direct reflection of (22). What about English, then, which routinely violates (22)? We would like to suggest that in English, (22) comes into conflict with the demands of prominence matching. These demands are far more stringent in English than in French, and it is in order to meet them that English allows itself greater freedom with *P-Parallelism* (22). If English strictly enforced *P-Parallelism*, in addition to prominence matching, the range of well-formed settings would be very narrow, which would presumably make songs too difficult to compose.

Why is it that prominence matching is enforced in a much more rigorous manner in English than in French? Rather than merely a matter of poetic convention, we believe that the difference has a phonological basis: stress has more perceptual salience in English than in French.

Table (23) sums up the commonalities and differences in the machinery which governs textsetting in the two languages. As in diagram (1), the components in our account pertain to three domains which must be carefully distinguished from each other: tune and text are independent objects each governed by specific principles, and their alignment yields a third object, what we call a composite.

(23)

	ENGLISH	FRENCH
TUNE	musical syntax	
TEXT	grammar of English	grammar of French
COMPOSITE	constituency matching (4)	
	prominence matching	
	all positions (10)	line ends only (16)
	P-PARALLELISM (22)	

In (23) the shaded boxes indicate components of our analysis which are different for English and for French. These are the two linguistic systems on the one hand, and the constraints on matching stress and beat on the other.

One aspect of our approach to textsettting is worth emphasizing. It must be noted that the only constraints on the texts are those deriving from the linguistic grammars. Considered by itself, independently from the tune, the text of a song has no intrinsic organization in feet, lines, stanzas, or any other of the units of poetic form as it is conventionally understood. Songs can indeed be broken down into chunks analogous to the lines and stanzas of literary poetry, but these, in our view, are byproducts of text-to-tune alignment. Rather than being features of the text which exist independently of its association with a tune, these divisions are features of the composite.[19]

In both languages there are obvious similarities between textsetting and literary poetry. In French, versification in classical literary verse requires that lines which occupy homologous positions in a poem have exactly the same number of syllables, and we have seen that the same is true of strophic songs (v. (18)), where it is a consequence of the rigorous enforcement of *P-Parallelism*. In English, on the other hand, the distribution of stressed syllables is strictly regulated in literary verse as well as in songs; strict syllable count is only required for certain meters. We have argued that certain properties of the texts of songs are but consequences of their alignment with music. As similar properties are found in literary poetry, this suggests that literary metrics has a closer likeness to musical textsetting than is generally thought. "The music of poetry" may after all be more than a metaphor.

References

Cadwallader, Allen & David Gagne. 1998. *Analysis of Tonal Music: A Schenkerian Approach*. New York: Oxford University Press.

Charpentreau, Simone. 1976. *Le Livre d'Or de la chanson enfantine*. Paris: Les Éditions Ouvrières.

Dell, François. 1989. Concordances rythmiques entre la musique et les paroles dans le chant. L'accent et l'*e* muet dans la chanson française. In Marc Dominicy (ed.), *Le souci des apparences*, 121–136. Brussels: Editions de l'Université de Bruxelles.

Dell, François. 2004. Singing as Counting Syllables: Text-to-Tune Alignment in Traditional French Songs. Submitted for publication in the proceedings of the conference *Words and Music*, March 14, 2003, University of Missouri-Columbia.

19. According to Table (23), *P-Parallelism* (22) legislates over composites. Here is why it cannot be considered a purely musical matter: the adoption of strophic form for a song is indeed a purely musical matter, but (22) takes this choice for granted and indicates one of its consequences for the correspondence between text and tune (see definition (12)).

Di Cristo, Albert. 1999. Vers une modélisation de l'accentuation du français : première partie. *Journal of French Language Studies* 9. 143–179.

Halle, John. 2003. Constituency Matching in Metrical Texts. Submitted for publication in the proceedings of the conference *Words and Music*, March 14, 2003, University of Missouri-Columbia.

Halle, John. Forthcoming. Music, Language and Metrical Form. Ms.

Halle, John & Fred Lerdahl. 1993. A Generative Textsetting Model. *Current Musicology* 55. 3–23.

Hayes, Bruce. 1989. The Prosodic Hierarchy in Meter. In Paul Kiparsky & Gilbert Youmans (eds.), *Rhythm and Meter*, 201–260. Academic Press, San Diego.

Hayes, Bruce. 2008. Faithfulness and Componentiality in Metrics. In Kristin Hanson & Sharon Inkelas (eds.), *The Nature of the Word: Essays in Honor of Paul Kiparsky*, 113–148. Cambridge, MA: MIT Press.

Hayes, Bruce & Abigail Kaun. 1996. The Role of Phonological Phrasing in Sung and Chanted Verse. *The Linguistic Review* 13. 243–303.

Lerdahl, Fred & Ray Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Linscott, Eloise H. 1939/1993. *Folk Songs of Old New England*. New York: Dover.

Quinn, Ian. 1999. The Combinatorial Model of Pitch Contour. *Music Perception* 16, 439–456.