Textsetting as Constraint Conflict[*]

Bruce Hayes

Department of Linguistics, UCLA
July 2005

Submitted for the proceedings volume of the April 2005 Paris conference,
"Typology of Poetic Forms", edited by Jean-Louis Aroui

Abstract

Halle and Lerdahl (1993) lay out the problem of **textsetting**: when singers encounter a novel stanza for a song that they know, they have consistent intuitions about where the syllables of the stanza ought to be aligned in time when the new stanza is sung. In other words, people have productive textsetting ability. Halle and Lerdahl's work offers the first explicit proposal for modeling this ability. The present paper likewise proposes a formal model of textsetting, but using a different theoretical approach.

I argue that many well-formed textsettings represent the best possible resolution between conflicting metrical principles. These involve: (a) matching of stress to strong position; (b) avoidance of long lapses (sequences where no syllable is initiated); (c) avoidance of extreme syllable compression; and (d) alignment of phonological phrase boundaries with line boundaries. A good textsetting often must sacrifice perfect realization of one of these goals in order to satisfy another goal that takes higher priority. For instance, many lines place stressed syllables in weaker rhythmic positions, and stressless syllables in stronger positions, in order to avoid a long syllable lapse, thus sacrificing (a) to satisfy (b).

I formalize this approach under Optimality Theory (Prince and Smolensky 1993/2004). Using data from Hayes and Kaun (1996), in which native English speakers spontaneously set many lines of verse, I show that an approach based on constraint conflict offers considerable improvement in the accuracy with which native speakers' settings are predicted.

---

[*] I would like to thank John Halle, Patricia Keating, Colin Wilson, Kie Zuraw, and members of the conference audience at the Typology of Poetic Forms conference for helpful input. None of them should be held responsible for remaining errors.

1. Introduction

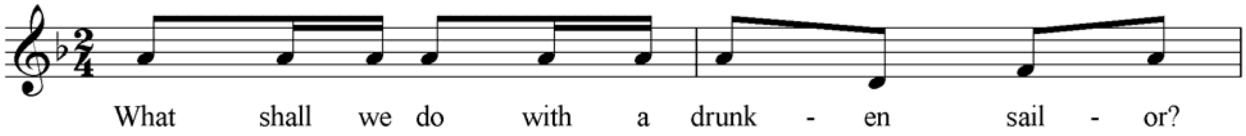
The **textsetting problem**, proposed by Halle and Lerdahl (1993), concerns how lines of linguistic text are arranged in time against a predetermined rhythmic pattern. It arises in the context of sung and chanted verse. We suppose when a person knows at least one verse of a particular song or chant, she has internalized its rhythmic pattern. Later verses are generally compelled to adhere to this pattern, even when they have different stress patterns or syllable counts than the original lines. Whenever speakers use their native intuition to arrange the syllables of novel lines into an existing pattern, they are engaged in textsetting.

The rhythmic patterns needed to understand textsetting have been formally explicated using **grid** notation by Liberman (1975), Lerdahl and Jackendoff (1983), and many later workers. In a grid, units arrayed in rows depict series of isochronous beats on a hierarchy of levels, and the columns indicate the strength of individual beats:

```
(1)   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
```

The textsetting of a particular line can be depicted by aligning its syllables against the grid. Thus, the familiar first line of "What Shall We Do with a Drunken Sailor?" is arranged to the rhythmic grid of (1) as in (2a); (2b) gives the equivalent in the more familiar—but less explicit—standard musical notation:

```
(2)a.  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
       |         |    |    |         |    |    |         |         |         |
      What     shall  we   do       with  a  drunk-      en       sail-      or?[1]
```

b.

The ability to set text is a productive one. Native speakers familiar with a particular tradition of sung or chanted verse can readily text-set novel lines to existing rhythms, and their choice of textsetting will agree to a fair extent with that preferred by other participants in the tradition. Thus, while not every English speaker who knows the song "What Shall We Do with a Drunken Sailor?" knows that *Stick on his back a mustard plaster* is a line of this song, even speakers who don't know this line can readily come up with the following setting:

---

[1] If the rhythm intended is not intuitively clear from the notation, please consult http://www.linguistics.ucla.edu/people/hayes/textsetting/ for audible versions.

```
(3)     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
        |           |     |     |           |           |           |           |           |
        Stick       on    his   back        a           mus-        tard        plas-       ter
```

Moreover, textsetting involve well-formedness intuitions, as seen throughout natural language. Thus, a native speaker confronted with a hypothetical setting like (4):

```
(4)     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
        |     |     |     |     |                 |                 |     |     |
        Stick on    his   back  a                 mus-              tard  plas- ter
```

will immediately reject it as ill-formed.

Now, it is true that within certain limits, native speakers do differ in their preferred textsettings. For example, Lerdahl and Halle (1993) give the following textsetting as the best one for its line:

```
(5)     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
        |     |     |     |     |           |     |     |           |           |           |
        Put   him   in    the   guard-      room  till  he          gets        so-         ber
```

I personally feel (as a native speaker) that while (5) is acceptable, I would prefer (6):

```
(6)     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
        |     |     |     |     |           |           |     |     |           |           |
        Put   him   in    the   guard-      room        till  he    gets        so-         ber
```

But such differences tend to be minor. Hayes and Kaun (1996), who elicited textsettings of 640 lines of traditional English folk song from native speaker consultants, found that the responses given were generally limited to a small number of choices, an average of 2.4 per line from a group of nine consultants.

In fact, it seems fair to say that, if we are willing to abstract away from a modest amount of free variation, *textsetting is predictable*. This means that if we have the right theory, we should be able to set up a system of rules which inputs a metrical grid and a given line, and outputs the textsetting (or small group of settings) that are produced by native speakers of the language who are familiar with the metrical tradition in question.

Halle and Lerdahl (1993) were, to my knowledge, the first to recognize this, and their pioneering rule system is discussed below. In this paper, I will offer a different system and argue that it improves on Halle and Lerdahl's initial attempt. The crucial aspect of my analysis is it that uses **conflicting**, **ranked constraints**, under the general approach of Optimality Theory (Prince and Smolensky 1993/2004).

2. Halle and Lerdahl's Analysis

Although the textsetting system of Halle and Lerdahl (1993) was the first explicit metrical grammar for textsetting, there has been much other work directed toward an understanding of how textsetting works, including Dell (1975, 2004), Stein and Gill (1980), Oehrle (1989), Hayes and Kaun (1996), Temperley (1999), and Dell and Halle (this volume). The Halle-Lerdahl (1993) analysis has since been amplified by Halle (1999, 2004). In the discussion that follows, I will summarize the version given in Halle (1999), because it is the most explicit—to the point of being implemented in software code.

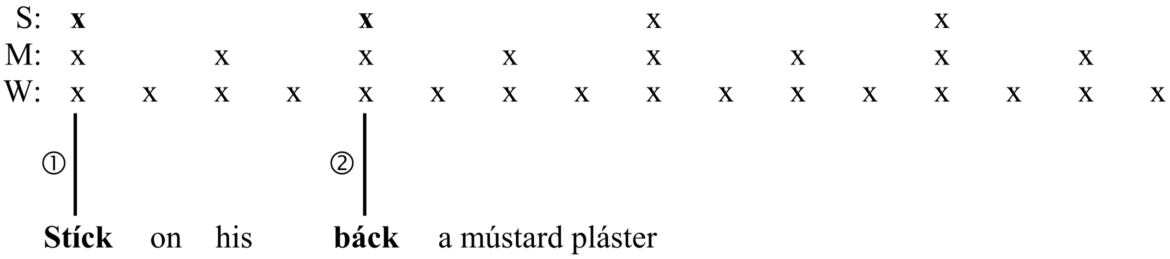
Before we start, some preliminaries. The grids given so far have had four rows of *x*'s. In what follows, the highest row will play no role, and so I will omit it henceforth. The remaining rows, labeled in (7), will be called the "S", "M", and "W" rows, standing for Strong, Medium, and Weak. Moreover, column of *x*'s in the grid will be termed S, M, or W, according the row occupied by its highest *x*.

(7)	S	W	M	W	S	W	M	W	S	W	M	W	S	W	M	W
S row:	x				x				x				x			
M row:	x		x		x		x		x		x		x		x	
W row:	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

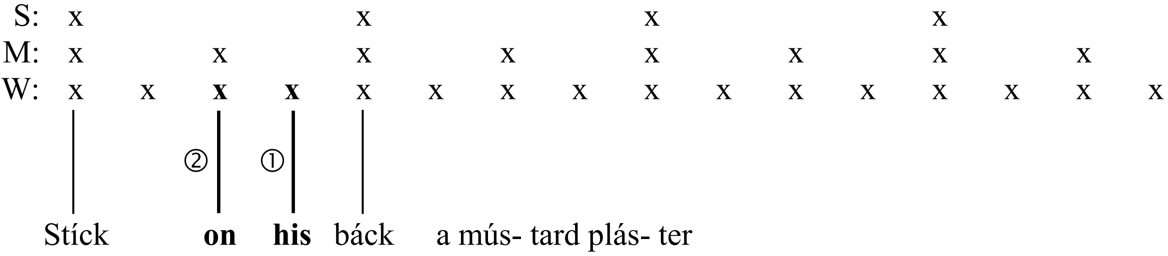
In linguistic representations, I will use ordinary orthography, depicting stress with acute accents.

The versions of the Halle-Lerdahl system given in Halle (1999) is an explicit algorithm, called the **Syllabic Distribution Algorithm**. This algorithm consists of a set of **mapping rules**, together with principles that govern when and how the rules apply. The rules map the syllables of the phonological representation onto columns of the metrical grid. They alternate between placing stressed, then stressless syllables.

At the start of the Syllabic Distribution Algorithm, the line is scanned, going from left to right, until a stressed syllable is found. This syllable is then placed in the first unoccupied S position of the grid. Then the algorithm iterates, seeking the next stressed syllable and placing it likewise in S. To give an example, here are the first two iterations as applied to the line *Stick on his back a mustard plaster*.

(8)


Whenever the assignment of a stressed syllable to S skips over a sequence of stressless syllables (as in *on his* in (8)), an additional rule is invoked: the algorithm counts the number of stressless syllables before the stressed syllable that was just mapped, then maps these stressless syllables one-to-one, right-to-left onto the *highest grid level having enough marks available to accommodate them*.[2] For instance, if there are two free syllables, the M level of the grid will not suffice, since there is only one free M slot available in the relevant domain. Therefore, mapping must take place at the W level, where there are three slots available. In the present case, right-to-left mapping works as in (9):

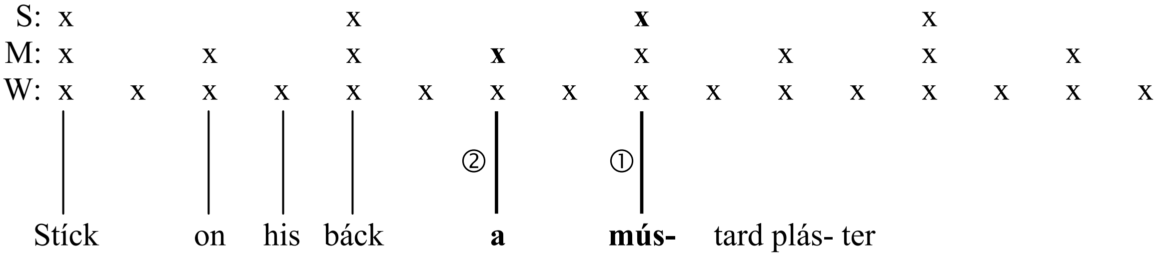
(9)


To see a three-syllable case, consult (5), where the syllables mapped would be "…③*him* ②*in* ①*the…*".

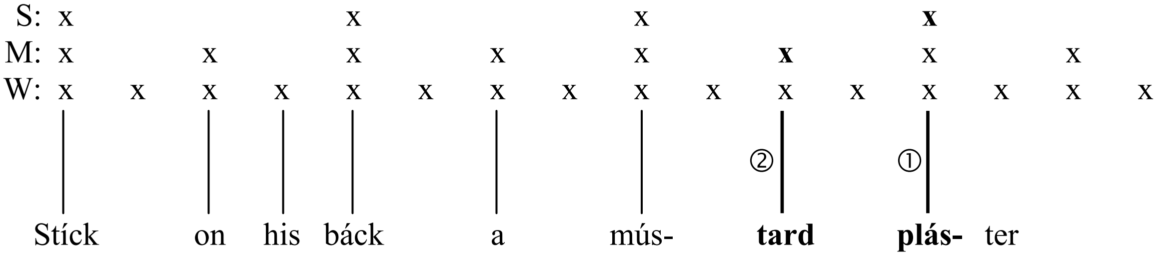
When there is just one syllable to be mapped, it goes on the M row, which is the highest row that can accommodate it; thus for the next iteration of the algorithm:

---

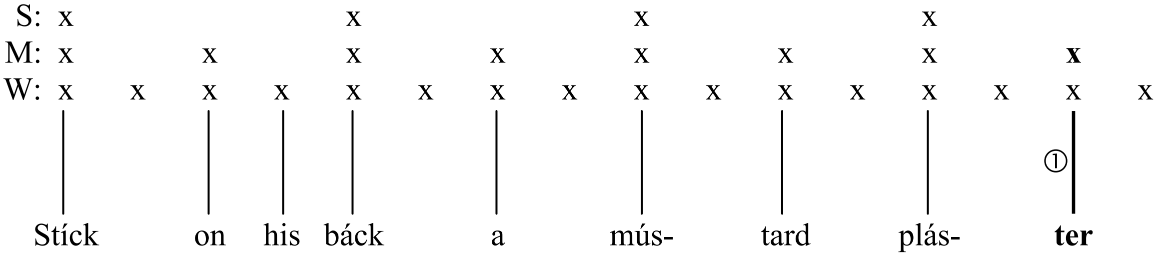
[2] Formally, the algorithm does this using the formula: $y = -(int(\log_2(C) + 1)$. *y* is a negative number, specifying how many rows down from the Strong row we must search to find the row onto which the *C* stressless syllables will be mapped.

(10)


Then there is one further iteration just like the previous one, shown in (11) below:

(11)


This brings us to the last stage of the algorithm: once all of the stressed syllables have been mapped, any trailing stressless syllables are mapped to remaining positions, again preferring the highest grid row that can accommodate them,[3] but this time from left to right. In the present case, there is just one stressless syllable left over, so that it can be accommodated at the M level:

(12)


This completes the derivation.

The Syllabic Distribution Algorithm is summarized below (cf. Halle (1999, 43)):

---

[3] This is not quite accurate; the row is actually picked by the same formula as that in fn. 2, which gives different results in certain cases. Since the version given above in the text is very close, and actually performs better in some cases where it matters (see (15w-z) below), I have retained it for the discussion here.

(13) *Syllabic Distribution Algorithm*

a. Map stressed syllables one-to-one, left-to-right onto S positions;

After each iteration:

b. Map stressless syllables one-to-one, right-to-left onto the highest grid level able to accommodate them;

Once (a) and (b) have applied as many times as they can:

c. Map any remaining stressless syllables one-to-one, left-to-right onto the highest grid level able to accommodate them.

3. Evaluating the Syllabic Distribution Algorithm

In the original publications, the Syllabic Distribution Algorithm was applied only to a fairly small set of cases. To put it to a sterner test, I have evaluated it against a much larger set of data gathered by Hayes and Kaun (1996). Hayes and Kaun asked nine native speakers of English[4] each to chant 640 lines from English folk song, largely unknown to them. Their renditions were taken down in a shorthand notation, reflecting their alignment to the grid. From the Hayes/Kaun corpus I selected 364 lines, by the following criteria. First, they were all "four-beat" lines (in the sense of Hayes and MacEachern 1998), so that we can exclude the extraneous factor of final line truncation, which according to Hayes and MacEachern, is used to define stanza structure. Second, I used only lines that received a consensus setting, involving at least four of the nine native speaker consultants. Both criteria were intended to provide a set of clear cases, suitable for initial testing of a theory.

The lines had been transcribed by Hayes and Kaun for stress, largely following the system of Chomsky and Halle (1968), and achieving reasonably good intersubjective agreement. The corpus of lines, with their stress assignments, is posted on my Web site, currently http://www.linguistics.ucla.edu/people/hayes/. It should be noted that these lines employed a slightly different grid from "Drunken Sailor" lines just noted. This grid begins with a Medium rather than a Strong position. Here is an example:

(14) S:			x				x				x				x	
M:	x		x		x		x		x		x		x		x	
W:	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	|	|	|		|	|	|		|	|	|		|		|	
	It	was	láte		in	the	níght		when	the	squíre		came		hóme	

---

[4] In truth, there were ten, but one consultant consistently offered highly creative textsettings, contrary to our instructions. It seems legitimate to confine attention here to the nine consultants who cooperatively provided the ordinary, workaday textsettings that form our present subject matter; cf. Halle and Lerdahl (1999, 9).

This grid is better represented in the English folk song corpus than the "Drunken Sailor" grid, so that ample material is available for study. The two grids are so similar that it would seem reasonable to use the same rules for both.

I programmed the Syllabic Distribution Algorithm in the version described above ((13)), and used the program to test the 364 lines just mentioned.[5] I found that it set 84 of them, or 23.1%, correctly, where correctness is defined as agreement with the plurality of the Hayes/Kaun consultants.

In (15) below I enumerate the stress and syllable patterns of these 84 lines, giving one example for each pattern. For brevity the textsettings are given in brief formulae which may be read as follows: /Σ/ = a grid slot filled with a stressed syllable, /σ/ = a grid slot filled with a stressless syllable, /./ = an empty grid slot.

(15)	a.	Our orders came on board, my boys,	σ.Σ.σ.Σ.σ.Σ.σ.Σ.	plus 34 more
	b.	Mother, mother, make my bed	..Σ.σ.Σ.σ.Σ.σ.Σ.	4 more
	c.	O Barleycorn is the choicest grain	σ.Σ.σ.Σ.σσΣ.σ.Σ.	4 more
	d.	There was Sydney Smith and Duncan too	σσΣ.σ.Σ.σ.Σ.σ.Σ.	4 more
	e.	A squire, a squire, he lived in the woods	σ.Σ.σ.Σ.σ.Σ.σσΣ.	3 more
	f.	He's up to the rigs, he's down to the rigs,	σ.Σ.σσΣ.σ.Σ.σσΣ.	2 more
	g.	O sir, I will accept of you the keys of your heart,	σ.ΣσσσΣσσσΣ.σσΣ.	2 more
	h.	Go and fetch me my pony, O!	..Σ.σ.Σ.σσΣ.σ.Σ.	1 more
	i.	Ho! Is my dinner ready now?	..Σ.σσΣ.σ.Σ.σ.Σ.	1 more
	j.	I fight for my king and country too	σ.Σ.σσΣ.σ.Σ.σ.Σ.	1 more
	k.	Yet here I will stay, nor ever from thee part,	σ.Σ.σσΣ.σ.ΣσσσΣ.	1 more
	l.	Full of care, yet I swear	..Σ.σ.Σ...Σ.σ.Σ.	
	m.	Seven are the seven stars in the sky,	..ΣσσσΣ.σ.Σ.σσΣ.	
	n.	The king's permission granted me	σ.Σ.σ.Σ.σ.Σ.σ.σ.	
	o.	We hoist our colors to the top of the mast	σ.Σ.σ.ΣσσσΣ.σσΣ.	
	p.	And there he espied his a-lady, O!	σ.Σ.σσΣ.σσΣ.σ.Σ.	
	q.	He rode and he rode till he came to the town,	σ.Σ.σσΣ.σσΣ.σσΣ.	
	r.	O where are you going to, my pretty maid,	σ.Σ.σσΣσσσΣ.σ.Σ.	
	s.	Again I'm a bachelor, I live with my son	σ.Σ.σσΣσσσΣ.σσΣ.	
	t.	O where have you been roving, Henery my son?	σ.ΣσσσΣ.σ.ΣσσσΣ.	
	u.	Eleven and eleven are the keys of heaven,	σ.ΣσσσΣσσσΣ.σ.Σσ	
	v.	In the month of May, in the month of May	σσΣ.σ.Σ.σσΣ.σ.Σ.	
	w.	And today I will sup and dine with you	σσΣ.σσΣ.σ.Σ.σ.σ.	
	x.	And the best of them all I will sell to thee	σσΣ.σσΣ.σσΣ.σ.σ.	
	y.	The king's permission granted me	σ.Σ.σ.Σ.σ.Σ.σ.σ.	
	z.	This day you shall sup and dine with me	σ.Σ.σσΣ.σ.Σ.σ.σ.	
	aa.	And he loved her as dear as he loved his life	σσΣ.σσΣ.σσΣ.σ.Σ.	
	bb.	I will sing you a ditty that will cause you to smile	σσΣ.σσΣσσσΣ.σσΣ.	

---

[5] Source code is available on request.

However, 260 lines, or 76.9%, were not set correctly by the algorithm. Among these lines, a number of instructive cases can be found which illustrate why the algorithm is not getting a higher score.

3.1 Leftward greed

Lines do not always have as many as four stresses to fill the four S positions; indeed, sometimes there are just two. The Syllabic Distribution Algorithm, based on left-to-right mapping, places the stresses as far to the left as possible, even if this leaves large gaps later on in the line:

```
(16) S:         x               x               x               x
     M: x       x       x       x       x       x       x       x
     W: x   x   x   x   x   x   x   x   x   x   x   x   x   x   x   x
        |       |   |   |   |   |               |               |
       *I      próm-ised her I'd már-           ry              her
```

The resulting scansion seems ill-formed to me;[6] what the Hayes/Kaun consultants preferred was this:

```
(17) S:         x               x               x               x
     M: x       x       x       x       x       x       x       x
     W: x   x   x   x   x   x   x   x   x   x   x   x   x   x   x   x
        |       |       |       |       |       |       |       |
        I       próm-   ised    her     I'd     már-    ry      her
```

The significance of this kind of error will be discussed below.

3.2 Squeezing the stressless syllables

Sometimes a line includes a sequence of more than three stressless syllables between two stressed ones. The Syllabic Distribution Algorithm, seeking to install the next stressed syllable, will squeeze all of these syllables into the inter-strong interval, by adding the next row down on the grid (which we can call the "Extra Weak" row); this leaves the end of the line empty:

```
(18) S:         x               x               x               x
     M: x       x       x       x       x       x       x       x
     W: x   x   x   x   x   x   x   x   x   x   x   x   x   x   x   x
    EW: 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
        |       |       | | | | |       |       |
       *En-     quír-   ing for his a- lá-      dy,     Ó
```

Such a setting is plainly impossible; the Hayes/Kaun consultants preferred (19) instead:

---

[6] This scansion is what is derived by the rules in (13) above; the slightly different rules given by Halle (1999) derive /σ.ΣσσσΣσσ......./. The point at hand remains essentially the same, i.e. that leftward greed is problematic.

```
(19) S:                  x                               x                               x                               x
     M:  x               x               x               x               x               x               x               x
     W:  x       x       x       x       x       x       x       x       x       x       x       x       x       x       x       x
         |               |               |               |               |       |       |               |               |
         En-             quír-           ing             for             his     a-      lá-             dy,             Ó
```

3.3 *More than four stresses*

When the Syllabic Distribution Algorithm is applied to a line that has more than four stressed syllables, then the provision requiring mapping of stressed syllables into S position ((13a)) causes the additional stresses to be spilled over into the next line, as for example in the following case:

```
(20)                     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
         |               |               |               |               |               |               |               |
         *I              pláced          my              báck            a-              gáinst          the             óld

                         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
                         |               |               |
                         gár-            den             gáte
```

The setting preferred by the Hayes/Kaun consultants was (21):

```
(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
         |               |               |               |       |       |       |       |               |       |       |
         I               pláced          my              báck    a-      gáinst  the     óld             gár-    den     gáte
```

3.4 *Consecutive stressed syllables*

It is not at all uncommon for lines from the corpus to include consecutive stressed syllables. When the Syllabic Distribution Algorithm encounters such sequences, it maps both stressed syllables onto S positions, thus creating long “lapses,” that is to say, long sequences containing no syllable. Here is an example.

```
(22)              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
        |         |                   |                   |         |    |    |
       *To      cóurt               yóung               mái- dens   I   was  bént
```

The Hayes/Kaun speakers preferred the setting in (23) for this line:

```
(23)              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
        |         |         |         |         |         |         |
       To       cóurt     yóung      mái-     dens          I        was       bént
```

Lapses are quite rare among the settings given by the Hayes/Kaun consultants; in fact, they are limited to lines that are so short that they cannot be set without a lapse. An example is the line *Thése last wórds thús he spáke*, set as /..Σ.σ.Σ...Σ.σ.Σ./.

3.5 Altering stress?

It is true that some of the problems just mentioned could be addressed by giving the Syllabic Distribution Algorithm the power to alter stresses (see Halle and Lerdahl 1993, 19). Thus in (22), for instance, if the algorithm could somehow ignore the stress on *young*, and promote to stress the (normally stressless) pronoun *I*, then it would set the line correctly. Yet I suspect that such phonological changes would have *no motivation other than that of rescuing the metrics*; and, as I hope to show, they are also unnecessary. A more articulated textsetting algorithm can deal with the stresses as they stand.

4. Toward an alternative

In determining why the Syllabic Distribution Algorithm ends up mis-setting so many lines, it is useful to proceed first at a purely intuitive level: what are the *goals* that the system is trying to accomplish? The algorithm is in a sense trying to get the stressed syllables into the strong positions; that is what its foremost rule ((13a)) infallibly accomplishes, though often with ill result.

But in fact, there are many other "goals" in metrics. For instance, as already noted, English sung/chanted verse strongly tends to avoid lapses, the gaps in the grid seen in examples like (22). It also strongly tends to align the beginnings and endings of intonational phrases (or similar units) with the beginnings and ending of lines, thus avoiding "run-on" settings like (20) (Hayes and MacEachern 1996). And there is also a strong tendency to avoid extreme syllable compression, of the kind seen in (18). The errors pointed out in the preceding section arguably arise because the Syllabic Distribution Algorithm is pursuing just *one* goal, namely stress-to-strong alignment, when it should be pursuing several at once.

Moreover, the "goals" of textsetting system evidently have different priorities. For instance, the well-formed setting in (23) makes a kind of strategic sacrifice, placing the stressed syllable *young* in M position and the stressless syllable *I* in S. The payoff is the avoidance of two lapses. From this, it appears that lapse avoidance is more important to the system than the occasional misplacement of stressed vs. unstressed syllables with respect to the grid. Indeed, I would claim that "strategic sacrifices" are ubiquitous in textsetting.

If we accept this conclusion, we must ask what the algorithm is that can find the textsetting that maximally satisfies multiple constraints. This is not a new problem; Lerdahl and Jackendoff (1983), in their pioneering work on musical rhythm, set up a great variety of conflicting constraints but deliberately refrained (p. 53-55) from proposing an explicit algorithm to deploy them. Halle and Lerdahl (1993), in contrast, take a different tack, claiming that a constraint-satisfying model is "inefficient" (p. 11), so that we should instead deploy a sequence of rules that directly finds the correct answer.

However, the recent history of linguistics suggests that severe difficulties arise when we attempt to write rule-based, sequential algorithms whose basic purpose is to adjudicate between conflicting goals. Typically one finds oneself trapped in a nest of redundancies ("Do X unless Y", where "Do Y" is also needed in the system) or encounters "look ahead" provisions: "Do X unless, later on in the derivation, it will turn out that that would keep you from doing Y".[7]

The response to this problem that was proposed by Prince and Smolensky (1993/2004) was **Optimality Theory**, which has proved very influential. In Optimality Theory (henceforth OT), the core of the analysis is made up by the "goals" of the system, as expressed formally in constraints. The constraints are prioritized; that is, placed in a rank order. The output is defined as the candidate that best satisfies the ranked set of constraints, in a specific sense defined under the theory.[8]

OT draws a firm distinction between defining the solution—as just given—and the problem of searching for it. The issue of search will be glossed over here, but is the subject of a growing literature in computational linguistics (e.g. Ellison 1994, Tesar 1995, Eisner 1996, Riggle 2004, Albro 2005), which suggests it is not insoluble.[9]

In fact, I believe it likely that real people also execute a kind of search. Often, when a person sets a novel line, he gets stuck, having followed a garden path; or adopts a poor solution

---

[7] For "do unless" and "look-ahead", see Prince and Smolensky (1993/2004, ch. 4). An example of look-ahead in the Syllabic Distribution Algorithm is right-to-left mapping of stressless syllables: these are mapped onto M positions, *unless* it would later turn out that this would ultimately cause the system to run out of grid slots, in which case they are mapped onto W positions, and so on down the levels.

[8] The procedure is this: the full set of logically-possible candidates is culled to the subset that violates the top-ranked constraint the fewest times; this subset is further culled to the sub-subset that violates the second-ranked constraint the fewest times, and the process is iterated until only a single candidate (or set of tied candidates) remains.

[9] For textsetting the search space is usually small enough that it can be searched fairly easily by machine: if the grid has 16 positions, no line will have more than about 13,000 possible settings. I found it useful first to locate "contender" candidates (Samek-Lodovici and Prince 1999, Riggle 2004), then search possible rankings using just the contenders.

and fixes it later. Such behaviors are what we would expect if people set text by conducting a somewhat hard-to-perform search, but would not be expected if they are guided by a simple deterministic algorithm.

Below, I will give a tentative account of textsetting under Optimality Theory. My goal is not to present a complete analysis of textsetting (a work in progress), but simply to make the case that textsetting is appropriately considered as involving the satisfaction of conflicting, prioritized constraints.

5. Analysis: Constraints and Ranking

Let us first consider what constraints could "translate" the basic analytic intuitions that underlie the Syllabic Distribution Algorithm.

First, there is an affinity between stress and strong position: as (13a) expresses, stressed syllables "want" to be in S position. To this I wish to add an elaboration: if a stressed syllable is *not* to be placed in S, then M is far preferred to W as a second choice.

Second, the *absence* of a syllable (meaning: a grid position filled with pause, or the phonetic continuation of a syllable initiated earlier) has the opposite affinity. It is preferred for empty positions to be W positions; failing this, they should be M positions. The core idea here—that the onset of a rhythmic event serves as a cue to metrical strength—was stated earlier in the pre-OT, constraint-based framework of Lerdahl and Jackendoff (1983, 76); with a constraint they called "MPR 3 (Event)".

These two tendencies can be formalized using six constraints, which can perhaps best be understood by placing them at the corners of a 3 x 3 array:

(24)	**S Position**	**M position**	**W position**
Stressed syllable		*STRESSED IN M ←	*STRESSED IN W ↓
Stressless syllable	*STRESSLESS IN S ↑		*STRESSLESS IN W
Absence of syllable	*NULL IN S →	*NULL IN M	

The constraint names begin with asterisks, which can be read "is disfavored" or "assess a violation if this is present." In the chart, the shaded cells represent configurations that are penalized by no constraint; nor should they be: S is the canonical location for stressed syllables, M for stressless, and W for unfilled; and indeed the most common kind of line for the grid we are examining is /σ.Σ.σ.Σ.σ.Σ.σ.Σ./, as in (15a) above. The arrows indicate "natural" rankings: it is worse for a stressed syllable to occur in W than in M, so we expect that *STRESSED IN W will outrank (be stricter than) *STRESSED IN M; and similarly for the other arrows.

Nothing in the system so far corresponds to Halle and Lerdahl's principle of mapping stressless syllables from right to left ((13b)). I believe that in fact, directional mappings is not really what is happening here, but rather that we see in the data the effects of a very general principle propounded earlier by Lerdahl and Jackendoff (1983, 80-84): rhythmically *strong* units tend to be *long* as well. For example, the grid sequence (25a), which in music and chant is considerably more common than (25b), is favored because its strongest element (the first syllable) is its longest (it is uttered over two grid positions):

```
(25)a.  S:   x                      b.  S:   x
        M:   x     x                    M:   x     x
        W:   x  x  x  x                 W:   x  x  x  x
             |     |  |                      |  |  |
             σ     σ  σ                      σ  σ  σ
```

If this is right, the task at hand is to formalize a Strong is Long constraint in an appropriate way. Here, I adopt a rough-and-ready approach: we simply count the number of positions on the W level that follow an S position going up to the next S, and (since constraints in OT must penalize, not reward), deduct this value from 3. In this view, (25a) violates Strong is Long twice (one empty position, subtracted from three), and (25b) violates it thrice (3 – 0). Violations are summed across the four S positions in a line.

I assume three further constraints, all of which cover phenomena already mentioned. For the dispreference against lapses (long grid continuities) noted in section 3.4 above, I assume a constraint *Lapse, which is violated whenever there are three empty grid positions in a row. I also assume a constraint *Fill Extra Weak, violated whenever a syllable is placed in an extra weak position, as in (18), and a constraint *Run-On, violated whenever the syllables of a line exceeded the allotted grid for that line, as in (20).

Thus, the constraints in the system number ten, as follows:

(26) *Fill Extra Weak
*Run-on
*Lapse
*Null in S
*Stressless In S
*Stressed in W
*Stressless in W
Strong Is Long
*Stress in M
*Null in M

The order in which the constraints are listed is in fact the order that I believe[10] gives the most accurate textsettings of the Hayes/Kaun data.

---

[10] I found this ranking with a greedy hill-climbing search, and cannot guarantee that my search found the best ranking. Quite a few other rankings give the same outputs that (26) does.

6. Analysis: Assessment

I implemented the grammar just given as a computer program and ran it on the 364 lines of the corpus. The fit to the data was improved: 70.6% of this grammar's outputs matched the favorite setting of the native speaker consultants,[11] vs. 23.1% for the Syllabic Distribution Algorithm. While further improvement is possible (see below), the substantial increase by itself suggests that the approach taken here, with constraint conflict and OT, has promise.

It is worth examining the proposed grammar's treatment of a line mentioned above (see (22)-(23)) as problematic, namely *To court young maidens I was bent*. As is conventional in OT, comparisons are made using a table ("tableau"), whose columns list the constraints in rank order and whose rows indicate the (relevant) candidates. The cells of the table include asterisks, which indicate the number of times the candidate for that cell's row violates the constraint for that cell's column.

(27) *Candidate table for "To court young maidens I was bent"*

Input: *To court young maidens I was bent*, stress pattern [σΣΣΣσσσΣ]	*Fill Extra Weak	*Run-On	*Null in Strong	Avoid Lapse	*Stressless In S	*Stress In W	*Stressless in W	*Stress In M	*Null in M	Strong Is Long
(23) `[σ.Σ.Σ.Σ.σ.σ.σ.Σ.]`					*			*		**** ****
(22) `[σ.Σ...Σ...ΣσσσΣ.]`				**						*** **

The table shows candidate (23) defeating (22). This is because the highest-ranking constraint that favors (23), Avoid Lapse, outranks the highest-ranking constraint that favors (22), Stressless in S. In other words, just as proposed above, the preference for matching stress with metrical strength is given up here to some degree, in order to satisfy the more important requirement of lapse avoidance. Although this is not shown in the table (but was confirmed by my program), candidate (23) also defeats every other candidate; that is, every other possible alignment of syllables to grid positions.

Here is a table showing how the correct output for (16)-(17), *I promised her I'd marry her*, is favored over the erroneous output generated by the Syllabic Distribution Algorithm:

---

[11] An additional 6.9% of the settings derived also matched the consultants' settings, but they were tied with another candidate as the winning output of the grammar.

(28) *Candidate table for "I promised her I'd marry her"*

Input: *I promised her I'd marry her*, stress pattern [σΣσσσΣσ]	*Fill Extra Weak	*Run-On	*Null in Strong	Avoid Lapse	*Stressless In S	*Stress In W	*Stressless in W	*Stress In M	*Null in M	Strong Is Long
(17) `[σ.Σ.σ.σ.σ.Σ.σ.σ.]`					**					**** ****
(16) `[σ.ΣσσσΣ...σ...σ.]`				**	**					*** **

Here, since at least two S positions will fail in any event to be filled with stress, the lapses of (16) are almost gratuitous, yielding only minor benefits with regard to the low-ranked Strong is Long.

The remaining two cases of section 3 work similarly: (18), with extreme syllable crowding, is bound to lose due to its violation of high-ranking *Fill Extra weak; the winner (19) imposes only modest costs in its violations of *Stressless in S and *Stressless in W. Likewise, (20) is bound to lose due to its violation of high-ranking *Run-On; the winner (21) avoids a run-on at a relatively modest cost in violations of *Stressless in W and *Stress in M.

In conclusion, I've argued that the Syllabic Distribution Algorithm includes only a subset of the factors that determine textsetting. When we examine a larger set, and deploy them in a ranked Optimality-theoretic grammar that permits prioritization of the constraints, we obtain a more articulated system that makes better predictions. It is precisely the ability to override otherwise-active constraints, when a higher-ranked constraint violation needs to be avoided, that makes the algorithm work.

7. Postscript: More Metrics Needed

Still, an accuracy percentage of 70.6% is not good enough. I believe to obtain a better account of textsetting, we will need to draw on the resources of the theory of **generative metrics** (Halle and Keyser 1969 and much subsequent work). This research tradition, though seldom focused specifically on the textsetting problem, has yielded many results that (as the findings of Hayes and Kaun 1996 indicate) are directly applicable to the formulation of an adequate constraint set for textsetting. Here are some findings that I believe ought to be considered.

- Stress is not just a binary distinction, but rather involves **multiple levels** (Chomsky and Halle 1968). Moreover, these multiple levels are metrically relevant (Jespersen 1900, Halle and Keyser 1969, Kiparsky 1977, Hayes 1983). Thus, a medium-stressed syllable must count as stronger than a weakly stressed syllable, but weaker than a fully stressed syllable. The works just cited therefore use a *relative* system of stress assessment, examining the "ups" and "downs" between syllables, rather than the absolute stressed-stressless distinction seen here.

- Kiparsky (1975), (1977) demonstrated powerful metrical effects of both **word boundaries** and **phrase boundaries**; moreover, these effects were also shown to be present in the textsetting data gathered by Hayes and Kaun (1996).
- Hayes and Kaun (1996) argue that textsetting involves **duration matching**; i.e. the number of grid positions assigned to a syllable (all else being equal) tends to match that syllable's natural phonetic duration, as determined by pre-boundary lengthening (Klatt 1973, Wightman et al. 1992) and syllable weight (Kenstowicz 1994, §6.1). These effects motivate constraints that can override the default preference specified by STRONG IS LONG. Halle (2004), addressing similar data, proposes a different approach involving boundary alignment.

In my current research, I have constructed larger OT grammars for textsetting, making use of constraints taken from the above three areas, and have been able to raise the accuracy score of my grammars on the same corpus as described above to a fair extent. This work is still in progress, and I hope to report on a best-tuned grammar in a future paper.

Beyond the goal of improving the raw accuracy of the model, there is an additional phenomenon that must be accounted for (Halle 1999, 4): native speakers often are **ambivalent** between possible textsettings. A fully accurate analysis should reflect this fact and therefore generate multiple outputs, ideally with weighted preferences that correspond to human preferences. The theory of stochastic Optimality Theory (Boersma 1997, Boersma and Hayes 2001) appears to have some potential toward fulfilling this goal, and I hope in future work also to make progress on this issue.

References

Albro, Daniel (2005) *Studies in Computational Optimality Theory, with Special Reference to the Phonological System of Malagasy*, Ph.D. dissertation, UCLA.

Boersma, Paul (1997) How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21:43-58. http://www.fon.hum.uva.nl/paul/papers/learningVariation.pdf

Boersma, Paul and Bruce Hayes (2001) Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45-86.

Chomsky, Noam and Morris Halle (1968) *The Sound Pattern of English*. New York: Harper and Row.

Dell, François (1975) Concordances rythmiques entre la musique et les paroles dans le chant: l'accent de l'e muet dans la chanson française. In *Le soucie des apparences*, Marc Dominicy (ed.), 121-136. Brussels: Editions de l'Université de Bruxelles.

Dell, François (2004) Singing as counting syllables: text-to-tune alignment in traditional French songs. Ms.

Dell, François, and John Halle (this volume) Comparing musical textsetting in French and in English songs.

Eisner, Jason (1997) Efficient generation in primitive Optimality Theory. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pp. 313-320

Ellison, T. Mark (1994) Phonological derivation in Optimality Theory. Kyoto: *COLING 1994*, vol. 2, pp. 1007-1013.

Halle, John and Fred Lerdahl (1993) A generative textsetting model. *Current Musicology* 55:3-23.

Halle, John (1999) *A Grammar of Improvised Textsetting*. Ph.D. dissertation, Columbia University.

Halle, John (2004) Constituency matching in metrical texts. xxx

Halle, Morris and S. Jay Keyser (1966) Chaucer and the theory of prosody. *College English* 28:187-219.

Hayes, Bruce (1983) A grid-based theory of English meter. *Linguistic Inquiry* 14:357-393.

Hayes, Bruce and Abigail Kaun (1996) The role of phonological phrasing in sung and chanted verse. *The Linguistic Review* 13:243-303.

Hayes, Bruce and Margaret MacEachern (1996) Are there lines in folk poetry?, *UCLA Working Papers in Phonology* 1:125-142. [http://www.linguistics.ucla.edu/people/hayes/metrics.htm]

Hayes, Bruce and Margaret MacEachern (1998) Quatrain form in English folk verse. *Language* 74:473-507.

Jespersen, Otto (1900, English translation 1933). Notes on meter. In *Linguistica*, 249-274. Copenhagen: Levin and Munksgaard.

Kenstowicz, Michael (1994) *Phonology in Generative Grammar*. Oxford: Blackwell.

Kiparsky, Paul (1975) Stress, syntax, and meter. *Language* 51:576-616

Kiparsky, Paul (1977) The rhythmic structure of English verse. *Linguistic Inquiry* 8.189-247.

Kiparsky, Paul and Gilbert Youmans, eds. (1989) *Rhythm and Meter.* Orlando, FL: Academic Press.

Klatt, Dennis H. (1973) Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America* 54:1102–1104.

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

Liberman, Mark (1975) *The Intonational System of English*, Ph.D. dissertation, MIT. Published 1979: New York, Garland Publishing, Inc.

Oehrle, Richard (1989). Temporal structures in verse design. In Kiparsky and Youmans (1989), pp. 87-119.

Prince, Alan, and Paul Smolensky (1993/2004) *Optimality Theory: Constraint interaction in generative grammar*. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Revised version published 2004: Oxford: Blackwell.

Riggle, Jason (2004) *Generation, Recognition, and Learning in Finite State Optimality Theory*, Ph.D. dissertation, UCLA.

Samek-Lodovici, Vieri, and Alan Prince (1999). Optima. Rutgers Optimality Archive ROA-363, http://roa.rutgers.edu.

Stein, David, and David Gil. 1980. Prosodic structures and prosodic markers. *Theoretical Linguistics* 7. 173-240

Temperley, David. 1999. Syncopation in rock: a perceptual perspective. *Popular Music* 18:19-40.

Tesar, Bruce (1995) Computing optimal forms in Optimality Theory: basic syllabification. Technical Report CU-CS-763-95, Department of Computer Science, University of Colorado, Boulder.

Wightman, Colin, Stefanie Shattuck-Hufnagel, Mari Ostendorf and Patti J. Price (1992) Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91:1707-1717.