

Explanation of Features and Formalism

For Hayes, Wilson and Shisko, “Maxent grammars for the metrics of Shakespeare and Milton”

1. Preamble

Our paper is written with intelligibility in mind, but lurking behind it is a somewhat abstruse system used to express constraints and evaluate violations. The system involves a feature system and symbols that express the feature content of scanned syllables.

2. Features

The features used are [wb], [strong], [segment], [accent], [rise], [fall], [J2], [J3], [J4], [J5], and [RealSyl]. They are defined as follows.

[+wb]	Holds only of the line boundary symbol #, which is automatically placed by the program at the end of a line. # is underspecified for all other features (0 value).
[+strong]	Syllable is scanned in S metrical position; – is W.
[+segment]	This is borne by either real syllables or the dummy symbols LB2-LB5, explained below. # is [–segment].
[+accent]	Syllable has some degree of stress; in the data files created by human transcribers this is either 2, 3, or 4; 1 is [–accent].
[+rise]	The next syllable is stronger in stress. This feature value is computed from numerical stress values by the program <i>ConvertStressesAndJuncturesToSingleSymbols.exe</i> .
[+fall]	The next syllable is weaker in stress, analogous to [+rise]. [–rise,–fall] denotes a level stress sequence.
[+J2]	The syllable bearing [+J2] ends a simplex word. A syllable internal to a simplex words would be [–J2].
[+J3]	Syllable ends a Clitic Group.
[+J4]	Syllable ends a Phonological Phrase.
[+J5]	Syllable ends an Intonational Phrase.
[+RealSyl]	This feature value is borne by real syllables (there are 10 or 11, depending on extrametricality)
[–RealSyl]	<p>This value is borne by the four dummy symbols LB2, LB3, LB4, LB5. These dummies are placed at the beginning of each line, and are used to indicate that the last syllable of the <i>preceding</i> line has juncture level 2, 3, 4, or 5. This juncture level is needed for some constraints, like inversion constraints.</p> <p>When a line begins a poem, of course there is no preceding syllable. The software we use insert LB5 for such cases, meaning (correctly) that the line-initial syllable begins an Intonational Phrase.</p>

3. Symbols in input files

The symbols of the data files can be understood in terms of these features. Thus Shakespeare's line *Mark how one string, sweet husband to another*, as transcribed prosodically by Hayes, is given as:

LB5 MwSdJ3 MsU1J2 MwUuJ2 MsSdJ5 MwSuJ3 MsSdJ1 MwU1J4 MsU1J2 MwUuJ1 MsSdJ1 MwU1J5

The initial LB5 is a dummy that means, "previous line ended at the end of an Intonational Phrase". ("LB" is mnemonic for "line boundary".)

The next symbol in the line:

MwSdJ3

means:

"in metrical Weak position" (**Mw**)

"stressed" (**S**)

"first syllable of a falling stress sequence" (*mark* is more stressed than *how*; **d** for "down")

"word final, but non-final in its Clitic Group" (**J3**, the third of the five levels of phrasing)

In other symbols we find:

Metrical Strong is **Ms**.

Stress is **S**.

Rising stress is **u** (for up) .

Level stress is **l** (for level).

4. Constraints

Constraints are sequences of feature matrices preceded by an asterisk, for example:

*[-J5][-Strong,+Accent,-J2]

This means, "Don't have a stressed syllable in weak position that is nonfinal in its word and is noninitial in its Intonational phrase."

There is a small difficulty with making your own constraints. The true content of a feature matrix is the set of symbols (like **MwSdJ3** etc.) that it abbreviates, but there are usually a great number of ways you can use the features to form a "label" for any given set. For *Maxent.jar* it is essential that you use the *exact label that the program expects*. You can find the list of legal labels in the program folder in the file **NaturalClassMemberships.txt**.