Diphthongs versus monophthongs in English

Renate Raffelsiefen, Anja Geumann
Institut für Deutsche Sprache (IDS), Mannheim, Germany
raffelsiefen@ids-mannheim.de, geumann@ids-mannheim.de

Abstract

We present evidence for the analysis of the vowels in English <say> and <so> as biphonemic diphthongs /ɛi/ and /əu/, based on neutralization patterns, regular alternations, and foot structure. /ɛi/ and /əu/ are hence structurally on a par with the so-called “true diphthongs” /ɑi/, /ɐu/, /ɔi/, but also share prosodic organization with the monophthongs /i/ and /u/. The phonological evidence is supported by dynamic measurements based on the American English TIMIT database. Calculations of F2-slopes proved to be especially suited to distinguish the relevant groups in accordance with their phonologically motivated prosodic organizations.

Introduction

Rhymes in English (E.) sigh, soy, and sow (pig) are often classified as “true diphthongs”, in contrast to those in say and so, which have been claimed to pattern with monophthongs (Clements/Hertz 1996, Giegerich 1992, Harris 1994, Kenyon/Knott 1953, Ladefoged 1999, Lehiste/Peterson 1961). We argue here that phonological evidence supports the parallel biphonemic representations in (1a,b), where a vowel associated with the nucleus (N) is followed by /i/ or /u/ in the coda (C), thus building a complex rhyme (R) (Phonological and phonetic evidence for the representations of the diphthongs in (1a) is discussed in Raffelsiefen and Geumann 2016). The choice of /i/ and /u/ is motivated by markedness (affinity between high vowels and syllable margin). Syllabic /i/ and /u/ are assumed to generally co-associate with the coda in E. (cf. (1c)), unlike other monophthongs, which associate with the nucleus only (cf. 1d)).

\[ (1)a. \quad R \quad R \quad R \\
\begin{array}{ccc}
\wedge & \wedge & \wedge \\
N & C & N \\
| & | & | \\
/α \ i/ & /υ \ u/ & /ɔ \ i/ \\
(sigh) & (sow) & (soy) \\
\end{array} \]

\[ (1)b. \quad R \quad R \\
\begin{array}{c}
\wedge \\
N & C \\
| \\
/ɛ \ i/ & /ɔ \ u/ \\
(say) & (so) \\
\end{array} \]

\[ (1)c. \quad R \quad R \\
\begin{array}{c}
\wedge \\
N & C \\
\wedge \\
/\ i/ & /\ u/ \\
(see) & (sue) \\
\end{array} \]

\[ (1)d. \quad R \quad R \\
\begin{array}{c}
| \\
N & N \\
\wedge \\
/\ɔ/ & /\ɑ/ \\
(saw) & (shah) \\
\end{array} \]

The prosodic structures in (1) are supported by neutralization patterns in prevocalic position. Systematic replacements of simple nuclei by the type of complex rhymes shown in (1a,b,c) in loan word adaptation serve to satisfy the constraint *NN (No adjacent nuclei). Examples are given in (2).


"/␣/-insertion" in non-rhotic dialects also satisfies *NN by supplying a final coda as is shown in (3). No /␣/ is inserted when the morpheme ends with a coda already according to the representations in (1a,b,c), as is illustrated in (4).
(3) law[x] is, shah[x] is (cf. (1d));
(4) tie is, cow is, boy is (cf. (1a)), day is, toe is (cf. (1b)), zoo is, bee is (cf. (1c))

The segmental representation of /ɛɪ/ in (1b) is supported by American E. alternations between word-final /ɛɪ/ and /ɛ/ before /u/ due to historic schwa-loss illustrated in (5). The /ɛɪ/-/ɛ/ alternation results then from simple deletion of /ɪ/ ("=/ r" rhymes with):

(5) /ˈprei/ <pray> vs. /ˈprɛɪ/ <pray> 'words of praying' = /ˈbɛɪ/ <bear> (cf. /ˈpɛiə/ <prayer> 'one who prays')

Schwa is often posited as the initial member of the diphthong in so in British English, as opposed to initial [o] in American English (cf. Wells 2000). The choice of schwa as initial member of this diphthong at the phonemic level also in American E. is based on evidence from /u/-allophony. Assuming that vowels in the weak syllable of the foot are restricted to {i/±i/±o/±l/}, the distribution of flapped versus aspirated /t/ in (6) is explained in terms of distinct foot structure, giving rise to foot-initial flapping in (6a) versus foot-initial aspiration in (6b). The prosodic structures for motto versus latte are presented in (6c,d).

(6b) <kow[t]ow>, (t).[ou] <tutu>, (sæ[t]ou) <sætəre>

(6c) <motto>

(6d) <latte>

Data

We use the TIMIT corpus of American English (Garofolo et al. 1993) which represents 8 separate dialect areas and includes verified segmental annotations. The corpus is based on recordings of 630 speakers, who read 10 sentences each.

For our calculations we used the Burg algorithm, searching for 5 formants in the range from 0-5500 Hz for females. Outliers were not excluded. The calculations refer to 6 equidistant points for each vowel, limited to 25% - 75% of vowel duration, to minimize coarticulatory influences. Formant values were calculated in Hz and Bark.

Results

The studies presented below are limited to the 31 female speakers of the northern dialect region (DR 2) and the 36 female speakers of the southern dialect region (DR 5). Figure 1 and 2 show the smoothed trajectories. The values are averaged over means.

Figure 1. Formant trajectories for Southern American English are smoothed over mean values at 6 points between 25%-75%.

Figure 2. Formant trajectories for Northern American English are smoothed over mean values at 6 points between 25%-75%.
For the individual trajectories we calculated Euclidean distance, F2-slope and vowel duration, to establish potential ways in which these data support or contradict the representations based on strictly phonological evidence in (1). The Euclidean distance refers to distance in the F1xF2 plane and has been calculated with the formula in (7):

(7) Euclidean distance in Bark and Hertz
\[ \sqrt{((F1_{25\%} - F1_{75\%})^2 + (F2_{25\%} - F2_{75\%})^2)} \]

The F2-slope refers to the difference in the F2 values between 25% and 75% divided by duration and can be positive (movement to the left) or negative (movement to the right) and has been calculated with the formula in (8).

(8) F2-slope:
\[ (F2_{75\%} - F2_{25\%}) / 0.5 \times \text{vowel duration} \]

The data in Figure 3-6 are boxplots made with the graphics-package in R (R Core Team 2013), showing the first and third quartiles and the median. The notches extend to +/-1.58 IQR/sqrt(n) (Chambers et al.1983:62). This is considered roughly an equivalent to the 95% confidence interval.

Figure 3. Boxplot for the duration of all vowels. Northern American English (DR2). Median and first and third quartile are shown.

Figure 4. Boxplot for the Euclidean distance in the F1xF2 vowel space between 25% and 75%. Northern American English (DR2). Median, first and third quartile shown.

Figure 5. Boxplot for the F2-slope between 25% and 75% of vowel duration. Northern American English (DR2). Median, first and third quartile shown.
Discussion

Based on various types of phonological evidence we posit two types of vocalic rhyme constituents for English: complex rhymes consisting of a nucleus followed by a coda as in (1a,b,c) versus simple rhymes consisting of only a nucleus in (1d).

The trajectories in the F1xF2 plane (see Figures 1, 2) support the parallel representations in (1a, b, c) versus (1d): complex rhymes are reflected in outward-pointing movements, directed towards the left top corner for the rhymes with /i/ codas, and the right top corner for the rhymes with /u/ codas. By contrast, monophthongs not followed by a coda vowel as in (1d) point towards the center (except for schwa, which exhibits a short outward movement). (These effects are missed in the studies of steady states based on F2 trajectories alone (Lehiste/Peterson 1961, Clements/Hertz 1996)).

Our calculations of the F2-slopes in Figure 5 show high positive values for rhymes ending in /i/ in (1) (/i/, /ei/, /ai/, /ai/) as opposed to negative values around zero for /i/, /e/ and /a/. Our calculations of the F2-slopes in Figure 6 show high negative values for rhymes ending in /u/ in (1) (/u/, /au/, /au/), as opposed to positive values closer to zero for /a/, /o/ and /u/. These groups correlate directly with the two types of rhymes posited on phonological grounds, thereby supporting these abstract representations.

As for duration, a corresponding grouping effect would not be expected. Figure 3 shows that /i/ and /u/, which form complex rhymes, are somewhat longer than the corresponding lax vowels, which form simple rhymes. But the data also confirm well-known correlations between vowel openness and length.

Also for calculations of Euclidian distances shown in Figure 3 there is no predicted group effect. The fact that the so-called “true diphthongs” in (1a) appear to stand out does not necessarily reflect on some fundamental difference compared to the complex rhymes in (1b,c), but can be explained as a function of the longer distances between the relevant articulatory vowel targets specified in (1).

References


