SOME REFLECTIONS ON THE NOTION OF DIPHTHONG

FERNANDO SÁNCHEZ MIRET
University of Salamanca, Spain

“If long vowels produce methodological headaches, diphthongs are a positive migraine.” (Lass 1984: 95).

1. Introduction

The notion of ‘diphthong’ is not well defined in the linguistic literature (Lass 1984: 95; Maddieson 1984: 161).1 If linguists try to make an absolute definition they are confronted with the following dichotomy: are diphthongs single sounds or sequences of two sounds? This becomes evident if we compare the following definition by Catford (1977: 215): “A diphthong may be defined as a sequence of two perceptually different vowel sounds within one and the same syllable” with the one by Ladefoged (1982: 171): “[diphthongs are] single vowels with continuously changing qualities.”

This conflict between unity (single vowel) and duality (sequence) is present in all discussions of diphthongs.2 In this sense, a well-known field of controversy is the question of mono- or biphonematicity which has been a favorite topic in the structuralistic phonology and in different works on experimental phonology.3

---

1 A preliminary version of this paper was presented at the 30th Poznań Linguistic Meeting (1-3 May 1997). I am very grateful to Hans Christian Luschützky and Jesús Fernández González, who made some valuable comments to a second version of the paper. As usual, the responsibility for any errors rests with me. My attendance at the 30th Poznań Linguistic Meeting was supported in part by the Spanish Ministry of Education.

2 Greek ὀὐδόνως ‘double sound’ points to the duality, so that Peeters (1991: 213) considers that it is inadequate for the diphthongs /ai, au/ in the different Germanic languages. Such diphthongs are in his opinion “inherently dynamic monophthongs”.

3 The problem is unavoidable in every practical analysis. Nevertheless, Rischel (1991: 241-242) expresses a pessimistic opinion about the possibility of solving it. On the one hand, he thinks that this question is not interesting, because the decision is nearly always arbitrary. On the other hand, the internal representation of one speaker must not be necessarily coincidental with the representation of the rest of the speakers (a fact...
(Slama-Cazacu 1958; Cowan et al. 1985; Berg 1986; Bertinetto 1988; Bertinetto et al. 1994). From this point of view, diphthongs, as well as long vowels, affricates, geminate consonants, and prenasalized stops are very problematic entities (see the epithets suspicious sounds, Pike 1947a: 251; or "suspect" complex phonetic events, Maddieson 1984: 161).\footnote{This indeterminacy arising in analyses of concrete languages (see, for example, Pike 1947b) points to a fundamental characteristic of diphthongs from the point of view of a theory of language universals, namely its Janus-like character: diphthongs share characteristics of both sequences and single segments, and this fact constitutes its essential nature. Besides, there is another great amount of debate about the different types of diphthongs and their legitimacy to be so called. There is for example little agreement whether entities like /je, wo/ (rising diphthongs) in Romance languages should be classified as authentic diphthongs. In this issue different factors are involved: a) the exact phonetic nature of what is transcribed with /j, w/; b) the traditional idea that authentic diphthongs are always falling (like [ai], see below section 2); c) the mostly unavoidable “English-point-of-view”, since for this language the similar sequence /ju/ is generally analyzed as consonant + vowel (CV); d) the historical factors play an important role also at this point, as already Sievers (1901:\footnote{For an extensive analysis of the affricates-problem see Luscha"utzky (1992).} § 422) pointed out that: in the Romance languages the sequences /je, wo, we/ are called “rising diphthongs”, since they historically derive from single vowels. Moreover, the process of diphthongization of Latin ò, è created in some Romance languages alternations diphthong ~ monophthong (e.g., Spanish vuelo ‘I fly, flight’ ~ volar ‘to fly’), that reinforce the diphthongal analysis of these sequences.}

For the reasons I have mentioned so far, an absolute definition seems to be impracticable. In response to this problem, many linguists tend to classify diphthongs or potential diphthongs into different dichotomies. I will refer only to two examples:

1) Andersen (1972: 18) differentiates between ‘segmental diphthong’ ("a single segment whose central phase is acoustically heterogeneous in its temporal development, rather than presenting a steady state") and ‘sequential diphthong’ ("a sequence of segments, usually forming part of the same syllable").

2) Lehiste and Peterson (1961: 277) include under the general term of “complex nuclei” two types of entities: ‘glides’ and ‘diphthongs’. A glide is a vocadic syllable nucleus consisting of one target position, with associated formant transitions to the target, and formant transitions from the target. [...] A diphthong is a vocadic syllable nucleus containing two target positions.\footnote{\textsuperscript{7} For Lehiste and Peterson (1961: 277) ‘target position’ means: “The time interval within the syllabic nucleus where the formant are parallel to the time axis”. Another more frequent term for this is ‘steady-state’. In other cases ‘target’ has the meaning of canonical form of a phoneme (cf. Keating 1988: 6). Lehiste and Peterson (1961) use the term ‘glide’ in a particular form for designating one kind of complex nuclei. Further, ‘glide’ has also in the literature two other more usual meanings. On one hand, ‘glide’ can refer to the transition between two steady-states or to the transition from a steady-state to or to a steady-state (cf. Peeters 1991: 302). For these entities we also have the term ‘transition’, which I shall prefer here. On the other hand, ‘glide’ designates one of the components of the diphthong as opposed to the ‘nucleus’ (see below section 3.4) and in this sense will be used here.}

The idea behind such classifications, as shown in Table 1, is that some diphthongs behave like units (or show more unitary features) and some others behave like sequences (or show more dualistic features).

<table>
<thead>
<tr>
<th>Units</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>segmental diphthong: a single segment</td>
<td>sequential diphthong: a sequence of segments</td>
</tr>
<tr>
<td>glide: one target position</td>
<td>diphthong: two target positions</td>
</tr>
</tbody>
</table>

2. “True” and “false” diphthongs

One finds from the beginning of the investigation about diphthongs some kind of idea that there should be good and bad examples of diphthongs. Two terms have often been used: ‘true’ and ‘false’ diphthong (e.g. vs. unecht Sievers 1901:\footnote{For a different analysis see Ladefoged (1982\textsuperscript{2}: 77-78).} § 418; eigentlich vs. uneigentlich Jespersen 1897-99 [1904]: § 211). From this perspective some diphthongs are considered to be good representatives of the ideal diphthong, and are said to be true members of that category, while other diphthongs are viewed either as another type of entity or as bad examples of the category. I shall show only a few examples.

Sievers (1901:\footnote{"Remember that there is nothing sacred about the phonetic value of a symbol" (Ladefoged 1982\textsuperscript{2}: 36).} § 418) considers true diphthongs instances like [ai, ej, ay, oy], which give an impression of unity (“[...] beide Teile für das Ohr mehr zu einer Art glatt verlaufender Einheit zusammenschmelzen”). On the other hand, Sievers considers that [ia, uy, 2y] are false diphthongs, since they are more like sequences (“die beiden Glieder mehr selbständig und unvermittelt neben einander zu stehen scheinen”).

Another use of this dichotomy affects the division between ‘falling’ (or ‘offgliding’) and ‘rising’ (or ‘ongliding’) diphthongs. The traditional idea is that only the falling ones are true diphthongs (cf. Sievers 1901:\footnote{For Lehiste and Peterson (1961: 277) ‘target position’ means: “The time interval within the syllabic nucleus where the formants are parallel to the time axis”. Another more frequent term for this is ‘steady-state’. In other cases ‘target’ has the meaning of canonical form of a phoneme (cf. Keating 1988: 6). Lehiste and Peterson (1961) use the term ‘glide’ in a particular form for designating one kind of complex nuclei. Further, ‘glide’ has also in the literature two other more usual meanings. On one hand, ‘glide’ can refer to the transition between two steady-states or to the transition from a steady-state to or to a steady-state (cf. Peeters 1991: 302). For these entities we also have the term ‘transition’, which I shall prefer here. On the other hand, ‘glide’ designates one of the components of the diphthong as opposed to the ‘nucleus’ (see below section 3.4) and in this sense will be used here.} § 422, who does not
share this common opinion; Jespersen 1897-99 [1904]: § 211). An explicit argumentation for this claim is made by Donegan (1978 [1985]: 190), who takes into account prosodic arguments like the equivalence between falling diphthongs and long vowels (both count as two moras vs. rising diphthongs, which usually count only as one mora), or the different role of falling and rising diphthongs in rhyme as in the examples in both parts of falling diphthongs count in rhyme, but the glide of rising diphthongs does not count.

\[ \text{paid [peid]} : \text{raid [reid]} \quad \text{rhyme} \]

\[ \text{paid [peid]} \neq \text{red [red]} \quad \text{do not rhyme} \]

\[ \text{feud [fiud]} : \text{mood [mud]} \quad \text{rhyme} \]

In her analysis of Italian diphthongs, Marotta (1988) distinguishes three kinds of tautosyllabic vocalic sequences, of which only one can be considered a "true" diphthong. Marotta divides these Italian vocalic sequences as follows: a) the rising diphthong /wo/ (e.g., uomo ‘man’), that behaves as a vocalic nucleus; b) the rising diphthong /je/ and the other diphthongs with /i/ (e.g., ieri ‘yesterday’, piano ‘slowly’, fiune ‘river’...), that behave like a CV sequence; c) the falling diphthongs (e.g., mai ‘unca’, poi ‘afterwards’, sei ‘six’, pausa ‘break’, Europa ‘Europe’), that are analyzed as divided between nucleus and coda. 8

\[ \begin{align*}
\text{a)} & \quad \text{R} & \text{b)} & \quad \text{O} & \text{c)} & \quad \text{R} \\
\text{N} & \quad \text{N} & \quad \text{N} & \quad \text{C} \\
\text{w} & \quad \text{x} & \quad \text{x} & \quad \text{a} & \quad \text{y} \\
\text{e} & \quad \text{x} & \quad \text{x} & \\
\end{align*} \]

For Marotta (1988: 407) only diphthongs that are entirely dominated by the node nucleus on the representation of the syllable can be considered as "true" diphthongs. Consequently, the only "true" diphthong in Italian would be the rising /wo/. 9

Another such case of dichotic classification can be found for Dutch diphthongs. This language is said to have the ‘genuine diphthongs’ /ei, aɪ, au/, and the ‘pseudo-diphthongs’ /aj, oj, uj, ew, iu/. The genuine ones are more like unitary segments, while the pseudo-diphthongs act like sequences of vowel + consonant (VC) (Mioni 1973: 378-379). Phonetically the genuine diphthongs seem to be the product of a single gesture, while the pseudo-diphthongs seem to be the result of two gestures (Collier et al. 1982).

Finally, as I have said, Lehnste and Peterson (1961) distinguish in American English two types of complex nuclei. 10 On the one hand, they refer to [æj, əj, ɔj] (e.g., tight, loud, voice), which have two target positions and can be called ‘diphthongs’. On the other hand, they find that [ej, əʊ, ɔ] (e.g., fate, lope, hurt) contain a single target position, “and therefore should not properly be classed as diphthongs” (Lehnste and Peterson 1961: 275).

I summarize the preceding sketched classifications in Table 2.

Table 2. Classifications of ‘true’ and ‘false’ diphthongs.

<table>
<thead>
<tr>
<th>criterion</th>
<th>true diphthongs</th>
<th>false diphthongs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siervers</td>
<td>auditory impression</td>
<td>unity [æj, əj, ɔj]</td>
</tr>
<tr>
<td>Jespersen</td>
<td>tradition</td>
<td>falling</td>
</tr>
<tr>
<td>Donegan</td>
<td>prosody</td>
<td>falling</td>
</tr>
<tr>
<td>Marotta</td>
<td>syllabic structure</td>
<td>branching nuclei [wo]</td>
</tr>
<tr>
<td>Dutch diphthongs</td>
<td>phonetics, articulation</td>
<td>VV /ei, aɪ, au/</td>
</tr>
<tr>
<td>Lehnste and Peterson</td>
<td>acoustics</td>
<td>diphthongs [æj, əj, ɔj]</td>
</tr>
</tbody>
</table>

As can be inferred from Table 2, in each case the distinction has been made from a different point of view. As a consequence, contradictions arise. For example, if for Jespersen and Donegan falling diphthongs are the best representatives of the category, in the analysis of Marotta a rising diphthong is the only one that deserves this label.

Instead of viewing the problem in terms of "true" and "false" diphthongs, I would like to propose a dynamic interpretation based on some principles of Natural and Cognitive Phonology. I use ‘dynamic’ in the sense that one can better understand an object if one knows the rank of its possible origins and the plausible changes it may subsequently suffer: the phonological processes that create, change and eliminate diphthongs. My claim is that we can consider the diphthong a prototypical category governed by such processes. From this point of view some diphthongs are more prototypical than others, or they match better than others some of the characteristics of the category. In addition to this, the processes of diphthongization and assimilation can explain the prototypical features of the category and the relationship between their members. I think that this approach to the problem allows one to see the variety of facts from one single point of view, as well as the synchronic and diachronic dynamic aspects of these sounds.

---

8 Some of these Italian falling diphthongs are not easily differentiated from hiatuses (cf. Marotta 1987).

9 Through its layered formalization Autosegmental Phonology offers a representation of the conflict between unity and duality in diphthongs and in other complex entities (the same applies for affricates, cf. Luschatzky 1992: 136). One finds a hint of this idea in Pike (1947b: 158).

10 See a similar analysis in Pike (1947b).
I will also argue, in the vein of the studies of typology and universals (Crothers 1978; Maddieson 1984; 1991), with data collected from different phonological inventories.

3. Diphthongs and other categories

We need to distinguish between diphthongs and other vocalic entities, which are placed at two opposite points in the scale of unity/duality, namely the monophthong and the hiatus. This distinction is not easy. Diphthongs are characterized by the presence of formant movement. However, it has been reported that monophthongs also exhibit some quantity of formant movement (Strange 1989; Nearney 1989; Andruski and Nearney 1992; but see the discussion by Harrington and Cassidy 1994; see also section 5.2). Likewise, there are difficulties when trying to discriminate between diphthong and hiatus in concrete analyses (Marotta 1987; see below also section 5.2).

A look at the historical evolution shows the different changes that can involve diphthongs: diphthongs can monophthongize (Latin *au > Proto-Romance *e), can evolve into a hiatus (Latin *au > Romanian au), or consonantize their pre- or post-vocalic glide (Italian uomo 'man' > dialectal vomo; Romance *ei > Raeto-Romance ek). We also find changes with the opposite direction: monophthong, hiatus, VC and CV > diphthong (Latin *e > Spanish ie, Spanish ma.iz > dialectal maiz; Latin multu > Portuguese muito, Latin plenu > Italian pieno).

This suggests that diphthongs cover an area of the vocalic possibilities of human beings situated in the middle position between a monophthong and a hiatus, and also between a VC sequence and a CV sequence. This is represented below.

```
  CV
   ↓
monophthong ←→ diphthong ←→ hiatus
   ↑
  VC
```

Diphthongs can be situated in the space delimited by these four categories and have the possibility to place themselves close to every one of these poles.

4. Data

I draw the data that will support my analysis from three different sources. The UPSID (UCLA Phonological Segment Inventory Database; Maddieson 1981, 1984 and UPSID 1992)²¹, the SPhA (= Stanford Phonology Archive; Crothers et al. 1979) and Weeda (1983). In order to attain some homogeneity among the data I have made a few adaptations of the transcriptions given in the sources.²² The numbers in round brackets that follow some diphthongs in the tables and in the text refer to the number of languages that attest the diphthong (if no number follows, the diphthong is attested only in one language of the inventory). I present in figures (D1-D5) the collected diphthongs from the mentioned sources. The first elements of diphthongs are shown on the vertical axis; the second elements are listed on the top of the figure; the oblique line of boxes separates falling (below the line) from rising (above the line) diphthongs;³³ the dashes inside the boxes symbolize that a sequence of two equal sounds is not a diphthong. The double line in figure (D1), (D2), (D3) and (D5) separates the diphthongs involving a central vowel from the other diphthongs.

UPSID (1992) includes 451 languages. The diphthongs listed in the graph (D1) belong to 48 languages.⁴⁴ The reason why so few languages are considered to have diphthongs is that UPSID takes into account only monophonemetic entities (cf. Maddieson 1984: 133).

The SPhA includes 197 languages; 55 of them have some kind of diphthong. This inventory offers, besides the phonological system, also information about allophonic realizations and biphonic diphthongs (however, sometimes we find only the reference to the existence of such diphthongs without concrete listing, which makes it impossible to take them into account in my computations).⁵⁵ In the figures D2-D4 I present the data of the SPhA about diphthongs under three different status: monophonemetic, allophonic and biphonic diphthongs.⁶⁶

---

²¹ Personally I prefer transcriptions like [ai, ja], where [j] signals the glide (cf. Sievers 1901: § 410). In the sources one finds (among others) [ai, aj, a]. I have respected the transcriptions of the glide and have altered only in some cases the symbols for the nucleus.

²² Diphthongs that are placed very close to the line (e.g., [iu, ui, oe, o]) fluctuate between falling and rising realizations.

²³ Bladon (1985: 147) says to have found 78 languages with diphthongs in Maddieson (1981). I find in Maddieson (1981) and (1984) only 23 languages with diphthongs (cf. Maddieson 1984: 133). Bladon gives a table with the data he has found: the number of languages which attest every kind of diphthong is bigger than in my figure (D1), but the general tendencies one can extract from the data are the same.

In (D1) I do not take into account the nasal, pharyngized and breathy voiced diphthongs. Also, for reasons of space, the following attested diphthongs involving secondary vowels are not shown in the figure: [ey, ay (3), ey (2), ea, ia, aw, wi, oy, u, y, ae, au (2), u, wa, w (2), ia, ia (2), ia].

²⁴ Very often the SPhA just refers to sequences vowel + glide (VG) or glide + vowel (GV), as for instance in Portuguese, Spanish or Romanian. All these cases are not included in the figures, since the SPhA does not provide the concrete form and number of such VG or GV sequences, which means that the number of languages with possible diphthongs in the SPhA is actually bigger than 55.

²⁵ In (D2-D4) I have fused SPhA's distinction between [e, o] (mid vowels) and [e, o] (mid high vowels) (this may have little consequences in the frequency of the diphthongs with mid high vowels); cf. the same policy by Maddieson (1984: 203) for the inventories of the UPSID. Somewhat in contradiction to this position, but in order to give an acceptable representation to SPhA [e, o] I transcribe these diphthongs in the figures as [e, o]. I have also excluded from the figures the nasal and retroflexed diphthongs and I do not differentiate between long and short diphthongs.

For reasons of space, I do not include in (D3) the following diphthongs [uw, ro] (mid vowels) and [u, o] (mid high vowels) (this may have little consequences in the frequency of the diphthongs with mid high vowels); cf. the same policy by Maddieson (1984: 203) for the inventories of the UPSID. Somewhat in contradiction to this position, but in order to give an acceptable representation to SPhA [e, o] I transcribe these diphthongs in the figures as [e, o]. I have also excluded from the figures the nasal and retroflexed diphthongs and I do not differentiate between long and short diphthongs.

²² Personally I prefer transcriptions like [ai, ja], where [j] signals the glide (cf. Sievers 1901: § 410). In the sources one finds (among others) [ai, aj, a]. I have respected the transcriptions of the glide and have altered only in some cases the symbols for the nucleus.

²³ Diphthongs that are placed very close to the line (e.g., [iu, ui, oe, o]) fluctuate between falling and rising realizations.

²⁴ Bladon (1985: 147) says to have found 78 languages with diphthongs in Maddieson (1981). I find in Maddieson (1981) and (1984) only 23 languages with diphthongs (cf. Maddieson 1984: 133). Bladon gives a table with the data he has found: the number of languages which attest every kind of diphthong is bigger than in my figure (D1), but the general tendencies one can extract from the data are the same.

In (D1) I do not take into account the nasal, pharyngized and breathy voiced diphthongs. Also, for reasons of space, the following attested diphthongs involving secondary vowels are not shown in the figure: [ey, ay (3), ey (2), ea, ia, aw, wi, oy, u, y, ae, au (2), u, wa, w (2), ia, ia (2), ia].

²⁴ Very often the SPhA just refers to sequences vowel + glide (VG) or glide + vowel (GV), as for instance in Portuguese, Spanish or Romanian. All these cases are not included in the figures, since the SPhA does not provide the concrete form and number of such VG or GV sequences, which means that the number of languages with possible diphthongs in the SPhA is actually bigger than 55.

²⁵ In (D2-D4) I have fused SPhA's distinction between [e, o] (mid vowels) and [e, o] (mid high vowels) (this may have little consequences in the frequency of the diphthongs with mid high vowels); cf. the same policy by Maddieson (1984: 203) for the inventories of the UPSID. Somewhat in contradiction to this position, but in order to give an acceptable representation to SPhA [e, o] I transcribe these diphthongs in the figures as [e, o]. I have also excluded from the figures the nasal and retroflexed diphthongs and I do not differentiate between long and short diphthongs.

For reasons of space, I do not include in (D3) the following diphthongs [uw, ro] (mid vowels) and [u, o] (mid high vowels) (this may have little consequences in the frequency of the diphthongs with mid high vowels); cf. the same policy by Maddieson (1984: 203) for the inventories of the UPSID. Somewhat in contradiction to this position, but in order to give an acceptable representation to SPhA [e, o] I transcribe these diphthongs in the figures as [e, o]. I have also excluded from the figures the nasal and retroflexed diphthongs and I do not differentiate between long and short diphthongs.
Some other diphthongs are not included in (D3): [iæ, eæ, oæ] and [i véhicule, z, ə, æ, oæ, oæ]. According to principles about diphthongs most widely defended (see section 5) the above sequences can hardly be accepted as diphthongs. However, the cases [iæ, eæ, oæ] can be considered well-formed diphthongs under specific conditions (see section 5). 
Not included in the figure (D4): [aæ, uæ, yæ, yæ, uæ, æ, æ, øæ, øæ, øæ, øæ, øæ, øæ].

---

**Figure D1.** Diphthongs in UPSID (1992).

**Figure D2.** Monophonemic diphthongs in SPhA.

**Figure D3.** Allophonic diphthongs in SPhA.

**Figure D4.** Biphonematic diphthongs in SPhA.
Figure D5. Diphthongs in Weeda (1983).

5. A set of preferences for the diphthong

I would like to propose that the diphthong can be seen as a prototype category (for the applicability of Cognitive Theory in phonology see Nathan 1986, 1989, 1996; Hurch and Nathan 1996). This means that the category 'diphthong' cannot be defined by the presence or absence of some necessary and sufficient conditions of membership. Instead, it is necessary to find out a series of features that contribute in different degrees to the building up of the category. According to those features, some diphthongs match better than others the characteristics of the prototypical diphthong. This conception has the following advantages:

a) it allows a simple taxonomy to be exceeded, and the different types of diphthongs that have been discussed in the linguistic literature can be organized within one category;

b) it explains the possible evolutionary routes of a diphthong (as well as its possible origins);

c) it situates the diphthong in an intermediate point between the categories of monophthong, hiatus, VC and CV.

I shall propose several prototypical features (universal preferences) of the diphthong as criteria for the organization within the category.18

Arguments for the determination of a preference are:

a) the number of languages showing the pattern and
b) the possible motivation (Maddieson 1991).19

Natural Phonology claims that processes act as “shapers of phoneme systems” (Hurch and Nathan 1996: 241). On one hand, a group of foregrounding processes (the so called prelexical processes) are responsible for the form of the phonological systems and the phonotactics of the languages and “constitute, at least in part, universal prototypicality judgments about sounds” (Hurch and Nathan 1996: 241). On the other hand, backgrounding processes build an explanation of the form these pho-

---

18 Some of these universal preferences have already been discussed (especially by Weeda 1983). Some aspects less developed by Weeda will be further investigated in this section.

19 The frequency of some concrete types of diphthongs points to the universal preferences of the category. However, frequency does not explain anything in itself, on the contrary, it has to be explained (Dressler 1989: 111, 1996; Luschnitzky 1992: 133). The explanatory arguments for the frequency of one concrete type of diphthong are based in general principles as the functions of the phonology (to make languages pronounceable and perceptible). Such functions are served by the phonological processes (Dressler 1984). In the case of diphthongs two processes have a fundamental importance: diphthongization and monophthongization (assimilation).
nemes show in surface representations. I shall argue that the processes of diphthongization and its converse monophthongization or assimilation explain the patterns shown by diphthongs. I shall discuss the following questions (some of them formulated, but not discussed by Lass 1984):

5.1. Are all vowel combinations possible in diphthongs? Which are better? And why?
5.2. Is there any preference for a particular height relationship between the two members?
5.3. What is the structure of diphthongs?
5.4. Is there any preference for the position of the nucleus?

5.1. Are all vowel combinations possible in diphthongs? Which are better? And why?

The first preference of diphthongs consists of their tautosyllabic nature. In this case, we are dealing with an indispensable condition. Consequently, the best diphthong is that which is clearly tautosyllabic. However, not every diphthong fulfills this condition to the same degree. In some cases the language cannot determine if one is dealing with a tautosyllabic or a heterosyllabic entity: this is the case, for example, with the sequence [ig] in the South German dialects (cf. Sievers 1893: § 392), or in Leonese (Northern Spain). The same problem affects the Italian falling diphthongs (cf. Marotta 1987).

The principal factor involved in this issue is the sonority of the components of the diphthong. Every syllable has only one nucleus and, for this reason, diphthongs formed by two sounds of great sonority are disfavored, since both sounds could be nuclei and that means that they tend to form a hiatus. The data support this statement, since the combinations of the most open vowels [a, e, o, ë] with themselves [ae, ao, ae... ] are scarcely attested, which is no wonder, since they are not good examples of diphthongs. If such a sequence remains as a diphthong, the tendency is to increase the difference of sonority between the two members via the loss of sonority in one of them (a.e > ae > 0, e.g., aere > Spanish aire 'air'). In this few cases when we come across such diphthongs in the inventories one expects that in the actual realization one of the elements (the glide) decreases its sonority in some way.

Besides the sonority of the elements, another decisive factor for the issue of tautosyllabic is the organization into nucleus and glide (see below section 5.3-5.4). Note that the area of conflict between tauto- and heterosyllabic mostly

20 As Menzerath (1941: 11) states, "in der Tat ist eine richtige Diphthongdefinition nur mit Bezug auf die Silbe möglich."

21 In Leonese [ia] is usually described as a hiatus with a diphthongal realization (in weak contexts, polysyllabic words and words outside the phrasal accent, Krüger 1923: 22).

22 In some other cases the indeterminacy between diphthong and hiatus is due to historical or morphological factors, as in the case of Spanish /ia, uu/ (cf. RAE 1973: § 1.4.11e, g): /ia/ is a hiatus in verb forms (costruimos 'we build'), in derived forms with a suffix -ita (jesuita 'Jesuit') or in cases with conservation of the Latin accentuation (gratia-to 'free of charge').

affects falling diphthongs. As I have said, this is the case in Italian. One reason can be the fact that falling diphthongs usually are longer than rising ones. Equally the glide of falling diphthongs shows more vocalic features than the glide of rising diphthongs (Avram 1975; the historical evolution confirms this point, since the consonantization of the glide occurs more often in rising than in falling diphthongs). The consequence is that from this angle, rising diphthongs could be seen as more prototypical than falling ones, since they fulfill more clearly the prerequisite of tautosyllabic and in some analyses a rising diphthong falls completely under the nucleus, while a falling diphthong is split between the nucleus and the coda. This is the view partially taken by Marotta (1988) regarding the Italian diphthongs (cf. section 2). Such an analysis finds some support from the results of a psycholinguistic experiment carried out by Bertinetto et al. (1994): they analyze the cohesion between the two members of several vowel clusters, and find out that rising diphthongs are less separable than falling ones.

Summing up, the tautosyllabic and the cohesion of the two members play a role in the unity of the diphthong. We shall see that for the other tendency, namely the need for duality, other preferences are in place.

5.2. Is there any preference for a particular height relationship between the two members?

The components of a diphthong must show a difference in quality, otherwise there would be no case to speak about diphthong (= two sounds; observe the formation of a diphthong in Spanish no pellé tanto 'I didn't fight so much' ≠ American Spanish no pelé tanto, Alcina and Blecua 1975: 413). These two aspects, tautosyllabic (section 5.1) and difference in quality, are obligatory and appear in every definition of the diphthong. They form the essential characteristic of this entry, which can be formulated as unity in duality or duality in unity (it depends on what one wants to stress).

Once we have stated the necessity of a qualitative difference between the two members of a diphthong, the following question is: how big must it be? Weeda (1983: 149) and Lindblom (1986: 36) propose that those diphthongs are preferred

<table>
<thead>
<tr>
<th></th>
<th>diphthong type</th>
<th>split</th>
<th>cohesion</th>
</tr>
</thead>
<tbody>
<tr>
<td>falling diphthong</td>
<td>laigo → lalaigo (50.8%)</td>
<td>lalaigo (45%)</td>
<td></td>
</tr>
<tr>
<td>rising diphthong</td>
<td>zaiko → ziak (12.5%)</td>
<td>ziak (85.4%)</td>
<td></td>
</tr>
</tbody>
</table>

The results for the task of substitution of the first syllable with the sequence /vai/ are:

<table>
<thead>
<tr>
<th></th>
<th>diphthong type</th>
<th>split</th>
<th>cohesion</th>
</tr>
</thead>
<tbody>
<tr>
<td>falling diphthong</td>
<td>laigo → vuigo (43.3%)</td>
<td>vuigo (52%)</td>
<td></td>
</tr>
<tr>
<td>rising diphthong</td>
<td>zaiko → vako (10.8%)</td>
<td>vako (75%)</td>
<td></td>
</tr>
</tbody>
</table>
which exhibit a greater trajectory in the vowel space or, in other words, those diphthongs with the maximum perceptual differentiation of endpoints. Lindblom (1986: 36) proposes the hierarchy: /aj/, aw/ » /ei/, ow/ » /uj/, iw/.24

A slightly different point of view is taken by Maddieson. Instead of focusing on the distance between the two members of the diphthong, Maddieson (1984: 134) states that "diphthongs that begin or end with a high vowel element are preferred over those which lack such an element", and he insists on the fact that it does not mean that the diphthongs with a maximum trajectory are preferred, since diphthongs as /ei, ie, ou/ are quite often attested in the languages of the world.

As I have already mentioned, if both members of a diphthong have a high degree of sonority, the tendency is to have a hiatus (section 5.1). This fact can be seen as the reason for the preference of a high vowel in diphthongs, which allows the other element to have a greater degree of sonority and also usually to be the nucleus.

Further, one can suppose that the trajectory in the vowel space must be, as a tendency, greater in the case of biphonomatic diphthongs, than in those which are still allophonic. The data from the SPhA seem to confirm this tendency: on one hand /ai, au/ are the most frequent biphonomatic diphthongs, cf. (D4). On the other hand, the most frequent allophonic diphthong is [εi], cf. (D3).25 The data are given in Table 3.26

Table 3. Diphthong frequency.

<table>
<thead>
<tr>
<th>diphthong type</th>
<th>diphthong frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D4) biphonomatic</td>
<td>ai (8) au (7) ui (5)</td>
</tr>
<tr>
<td>(D3) allophonic</td>
<td>ei (8) ou (5) ij (4)</td>
</tr>
</tbody>
</table>

The fundamental question at this point would be the notion of sufficient vs. maximal contrast between the two members of diphthongs. The studies on the perception of diphthongs have mainly concentrated on the parameters that listeners utilize for discriminating different diphthongs. For example, Bladon (1985) suggests that the spectral change which necessarily takes place in a diphthong plays two roles: on one hand it acts as a "weighting flag" which alerts the auditory system of the presence of a diphthong; on the other hand, it appears to be a pointer to regions (steady-states) of spectral significance for the recognition of what diphthong is being perceived. However, as far as I know, little attention has been paid in the phonetic studies to the amount of change necessary to pass from a monophthong to a diphthong ("threshold change", Nearey and Assmann 1986: 1305).27 More attention has been devoted to the matter of the duration of the transition. As Peeters (1991: 313) points out, it seems that a transition longer than 80-100 ms. is needed in order to perceive a diphthong.28 The same author argues that it is not clear to what extent non-phonetic factors (phoneme- or language-specific characteristics) impose some constraints on what is perceived as a monophthong, a diphthong or a VC sequence.

It is difficult to speak about universal minimal and maximal contrast between the members of the diphthong. However, it is feasible to obtain some prototype effects from the distribution of the different types of diphthongs in the data (Nathan 1989: 59). In this sense, the observation of Maddieson on the frequency of [j, u, i, w] in these entities is significant and is confirmed in the different data in Table 4, where the numbers correspond to tokens.

Table 4. The frequency of [j, i, w, u] in diphthongs. Token numbers.

<table>
<thead>
<tr>
<th>sources</th>
<th>diphthongs with [j, i, w, u]</th>
<th>diphthongs without</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D1)</td>
<td>104</td>
<td>18</td>
</tr>
<tr>
<td>(D2)</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>(D3)</td>
<td>45</td>
<td>16</td>
</tr>
<tr>
<td>(D4)</td>
<td>37</td>
<td>26</td>
</tr>
<tr>
<td>(D5)</td>
<td>82</td>
<td>28</td>
</tr>
</tbody>
</table>

27 Already by Sievers (1893: §391): "Ein allgemeineres Abstandsminimum oder -maximum der Komponenten lässt sich nicht angeben." The question seems now to be more complicated. The traditional opinion was that diphthongs and monophthongs can be distinguished by the fact that monophthongs are stable sounds along their duration (Vowel targets = canonical forms of vowels: represented articulatorily as static vocal tract shapes, acoustically as points in the acoustic space). Against this view, some phoneticians have pointed to several dynamic aspects (change over time in spectral structure), which should be relevant for the perception of monophthongs. On one hand, supposed monophthongs, such as the ones found in Canadian English /æ, æ/ show significant formant change (comparable to the change of the diphthongs [æi, ou] = /æ, ο/). On the other hand, dynamic cues, as intrinsic duration, formant change and consonantal transitions have been shown to play a role in the vowel classification (cf. Nearey and Assmann 1986; Strange 1989; Nearey 1989).

This hypothesis has been criticized by Harrington and Cassidy (1994). In their perceptual experiments the dynamic factors play a role only for the perception of diphthongs, but not monophthongs. As they point out, the proponents of the dynamic theory consider the notion of vowel target to be irrelevant, but for Harrington and Cassidy this concept is crucial in the distinction between monophthongs and diphthongs.

For the time being the only possible conclusion seems, as Harrington and Cassidy (1994: 369) say, "that the theory that vowels are inherently dynamic warrants further investigation". The different studies in experimental phonetics have revealed part of the huge complexity of parameters involved in the production and perception of sounds, but we have not yet a unitary model that could coherently articulate all these findings.

28 This threshold is the result of different studies on Germanic languages.
The question is what does this fact tell us and how can we explain it? I see an answer to this question in the theory of diphthongization of Stampe (1972) and Donegan (1978). The prelexical process of diphthongization defines the prototype effects for this particular class of sounds. According to these authors, the diphthongization of a vowel consists in the polarization of its two fundamental features, sonority and color, which dissimilate to the extremes of the vowel space (Stampe 1972: 584). Consequently, an optimal diphthong has a maximal sonorous nucleus and a maximal colored glide. We can observe the effect of such a process in its diachronic manifestation, for example in the evolution of Latin ë > French ej > oJeremy. In the first phase we find polarization of sonority: the second part of the diphthong rises. In the second phase the polarization affects the color of the first element; see the representation in Diagram 1.

Diagram 1. Diphthongization of vowels e > ej > oJeremy

At later phases of the evolution of the Ancient French diphthong [Jeremy] we find the other process involved in the history of diphthongs, namely assimilation: OJeremy > Oe > ge > Oa > ya; see the representation in Diagram 2.

Diagram 2. Assimilation of the diphthongs OJeremy > Oe > ge > Oa > ya.

This process of assimilation probably began in less stressed contexts and then extended to all positions. Note that in the last phase the glide rises from [Jeremy] to [ya]. Here we see a manifestation of the tendency towards high glides, which is active not only in the process of diphthongization of a monophthong but also during the posterior evolution of the diphthong.

In the sense of Stampe’s theory of diphthongization, the high frequency of [Jeremy, w, i, u] in diphthongs is an effect of the glide’s tendency to be colored. Further, Stampe’s idea of an optimal diphthong as an entity which polarizes its two members finds corroboration in Figures (D1), (D3), (D4), (D5), where we can ascertain that diphthongs tend to concentrate on the most distant points from the oblique line, that is, where the distance between the two members is greatest (in terms of sonority or in terms of color, or both).

We can try to go further with this investigation of the possible combinations of vowels into diphthongs asking what is the role of the different phonetic features in this issue. The most important ways of qualitative distinction in the vowel space are:

a) difference in height;

b) difference in frontness vs. backness.

One can suppose that for the aim of producing maximal differentiation, some collaboration of these features should take place. In fact, both differences involving height and frontness vs. backness are almost always present in biphonemic diphthongs, cf. (D4). This is expected, since such diphthongs tend to be maximally differentiated. A more balanced situation is expected for the allophonic diphthongs; in this case the difference in height seems sufficient to account for a great number of diphthongs, cf. (D3). The frequencies of each type are shown in Table 5.

Table 5. The frequencies of various differentiation features for two types of diphthongs: biphonemic, and allophonic.

<table>
<thead>
<tr>
<th>diphthong type</th>
<th>total number</th>
<th>height + front vs. back</th>
<th>height</th>
<th>front vs. back</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D4) biphonemic</td>
<td>38</td>
<td>25</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>(D3) allophonic</td>
<td>38</td>
<td>23</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

These data point to a predominance of height (vs. the feature frontness/backness), that is of differences in sonority, in the formation of diphthongs. There are a few cases where the two parts are only distinguishable by means of the opposition front vs. back, as in [iu, ui, oe, eo]. The diphthongs [æ, əe, æu, uə... ] are not attested; these hypothetical diphthongs are formed with elements of great sonority, a combination which is strongly disfavored (section 5.1). Further, it must be noted that the diphthongs formed only by means of the difference front vs. back pose the problem of differentiating between nucleus and glide (as in the case of /iu, ui/ in Spanish, RAE 1973: § 1.4.11a).

On the other hand, diphthongs containing the central vowels [a, ə, u] are attested in the SPhA only as monophonemic or allophonic, cf. (D2) and (D3). The most frequent central vowel in these diphthongs is [æ] and it is usually found in the glide position, cf. (D6). Furthermore diphthongs formed only with central vowels are uncommon: in the UPSID we find [æi, əi, ər] and in Weeda (1983) [əi, ər], but every case is exemplified only in one language. The reason for this avoidance of 'central diphthongs' can be seen in the fact that central vowels are acentric, so that they are not very adequate for cases of polarization.

In conclusion, some distance between the members is needed for the perception of a diphthong. The greater it is, the more perceivable the diphthong is. The prototypical biphonemic diphthong has the maximal distance between its two members (as one can conclude from the data of the SPhA). The fact that the great majority of diphthongs present a high element can be explained as the consequence of two factors:
a) the process of diphthongization, which consists in the creation and increase of the differences between the two members of the diphthong;
b) the fact that two elements with a high degree of sonority tend to convert into a hiatus.

5.3. What is the structure of diphthongs?

The point I want to discuss is the internal organization of diphthongs into nucleus and glide. The nucleus is the most perceptible part and tends to coincide with the element with the greatest degree of intrinsic sonority. This is called the sonority principle (Sonoritätsprinzip; Jespersen 1897-99 [1904]: § 198) and predicts that the diphthongs which are below the oblique line in (D1-D5) are falling, and those above this line are rising. This is a traditional idea (Luick 1891: 337; Sievers 18934: § 392; more recently Donegan 1978 [1985]: 191), and this is the reason why I have excluded diphthongs like [iə, eə, oə] and [iɪ, ɛɪ, ɔɪ, ɛɪ, ɔɪ] from the tables (D1-D5) (see fn. 16, 17), since they imply an implausible organization. Nevertheless, cases like [iə, eə, oə] can correctly be interpreted as diphthongs. In [iə, eə, oə] the postulated glide has more intrinsic sonority than the nucleus, a fact that runs against the sonority principle. The same applies to diphthongs like [iɔ, uə]. As Schubiger (1977 [1989]: 116) points out, the glide in these cases is produced with less expiratory intensity, i.e. [i, ɛ, ɔ, ə] can correctly be less sonorously than the nucleus [i, e, o, u]. Note that these diphthongs were called “enclitic” by Sievers (section 2) and they are the product of the so called “centering” or “gliding” diphthongization, which mostly affects lax vowels (Sánchez Miret 1996: § 2). On the other hand, diphthongs like [iɪ, ɛɪ, ɔɪ, ɛɪ, ɔɪ] are actually unpronounceable, no matter how much one can diminish the expiratory intensity on the first element. This points to an important restriction on the articulatory capabilities of humans.

The sonority principle finds its parallel in the historical evolution in cases of syllabic change, as in rege > Old French [roi] > [roie] > [roie] > rot [rət] ‘king’; Spanish maih ‘corn’ > dial. maih; fii.ii.ii.ii.ii > fii.ii.ii.ii.ii > ‘small son’ > (Italian figliuolo, French filleul, Spanish hijuelo); (Luick 1891: 338, Jespersen 1897-99 [1904]: § 198, Passy 1891: § 475, Sievers 18934: § 395).29

The other element of the diphthong is the glide. It has sometimes been stated that the glide of a diphthong must consist of a high vowel. From the data it is easy to prove that this is not always the case (see also Sievers 18934: § 389; Ruhlen 1975: 52). However, it can be safely stated that the majority of diphthongs and especially the most frequent ones tend to have a high glide, cf. (D6) and section 5.2., which agrees with Stampe’s idea that the best glide is maximally colored.

On the other hand, it is commonplace in the phonetic literature that the actual realization of these glides is not as extreme as the transcriptions would suggest (Sievers 18934: § 389; Jespersen 1897-99 [1904]: § 212; Solomon and Sara 1984; Jha 1985). In fact, the realization of the glide presents different degrees of accuracy in different styles: see, for example, the situation of the pronoun I in English below, (Stampe 1972: 585).

very careful speech ordinary speech casual speech
[aj] → [æ] → [ɑː] (I’m here [ɑːmir])

Equally the German diphthongs (e.g., Zeit ‘time’, Haus ‘house’, Schau ‘fear’) are usually transcribed as [ai], [au], [ɔi], [ai], [ɔi], or [æi], [æi], and one can also find the phonetic transcription [æi], [æi], [æe] (Mioni 1973: 149; Iivonen 1989: 24-27; Moosmüller 1997). We see that linguists have used four possible degrees of height for the transcription of the glide.

In conclusion, we see in these facts, first the tendency to the polarization in a sonorous nucleus and a colored glide (whose results are tentatively represented in the current phonological transcriptions of diphthongs), and then the action of postlexical assimilation processes, which are responsible for the observed less extreme realization of glides, and also of nuclei.

5.4. Is there any preference for the position of the nucleus?

Finally, I want to address the question of preferences in the position of the nucleus relative to the glide. Traditionally, diphthongs are considered to be sequences of nucleus + glide, that is, falling diphthongs (see section 2). In fact, this seems to be the most frequent form of a diphthong. The distribution of falling and rising diphthongs in the data is shown in Table 6.30

<table>
<thead>
<tr>
<th>Table 6. The distribution of falling and rising diphthongs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>source</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>UPSID</td>
</tr>
<tr>
<td>Weeda</td>
</tr>
<tr>
<td>SPHa monoph</td>
</tr>
<tr>
<td>SPHa alloph</td>
</tr>
<tr>
<td>SPHa bipho</td>
</tr>
<tr>
<td>SPHa total</td>
</tr>
</tbody>
</table>

The reason for this can be seen again in the process of spontaneous diphthongization. As Donegan (1978 [1985]: 196) states: “The original diphthongization of a simple vowel […] typically produces a falling diphthong: V(–) → VV.” This is so because a falling diphthong generally has a greater duration than the equivalent ris-


30 The sources do not always clearly state the syllability of the diphthongs. I count as rising diphthong every one whose second element has greater sonority than the first one, and as falling those diphthongs with the inverse pattern. I do not count diphthongs with elements of the same sonority: [ii, iu, ui, oe, eo, se, ii, ii].
ing one, thus making it easier to perceive (see measurements in Gottfried et al. 1993: 218-219, about English diphthongs; Salza et al. 1987, about Italian; Borzone de Manrique 1979, about Spanish).

Falling diphthongs are the normal result of a diphthongization process, since this is a foregrounding process and it is expected to preserve or even increase the duration of the strengthened vowels. Consequently, a rising diphthong is a bad solution for diphthongization, since it implies some amount of shortening (Donegan 1978 [1985]: 101). In fact, falling diphthongization seems to be the most common form of diphthongization in the history of Germanic and Romance languages (Sánchez Miret 1996: § 2.4). Furthermore, we arrive at the same conclusion if we look at the cases of allophonic diphthongization registered by the SPhA; the data are shown in Table 7.31

Table 7. Allophonic diphthongization in SPhA.

<table>
<thead>
<tr>
<th>falling diphthongs</th>
<th>rising diphthongs</th>
</tr>
</thead>
<tbody>
<tr>
<td>a &gt; aj</td>
<td>o &gt; oj</td>
</tr>
<tr>
<td>æ &gt; æ9</td>
<td>ç &gt; ç9</td>
</tr>
<tr>
<td>ë &gt; ej</td>
<td>ë &gt; ej (4)</td>
</tr>
<tr>
<td>â &gt; æ9 (2)</td>
<td>o &gt; og</td>
</tr>
<tr>
<td>ö &gt; o9</td>
<td>o &gt; ou</td>
</tr>
<tr>
<td>o &gt; og (2)</td>
<td>o &gt; ow</td>
</tr>
<tr>
<td>õ &gt; õ9 (2)</td>
<td>o &gt; ow</td>
</tr>
<tr>
<td>ð &gt; 9</td>
<td>i &gt; ej (2)</td>
</tr>
</tbody>
</table>

w = approximant back unrounded

Moreover, the most frequent evolution among diphthongs is falling > rising (and not rising > falling), as in French ë > æj > aj > og > ge > ye > ya; e.g., rege > roi ‘king’; or Spanish aj > og > ge > ye; e.g., coño > coito > cuero ‘leather’.32 The reason is that falling diphthongs are easily reducible33, and the evolution falling > rising implies a temporal reduction. On the other hand, rising diphthongs have difficul-

31 I do not include two cases where the SPhA does not make explicit the syllabicity, which otherwise is not deducible: Amuesha e > ye; Mandarin Chinese r > or. I have also excluded the cases alluded above (see fn. 16): Chukchi e > ge, a > ge, u > ou, ø > go, i > gi; Nez Percé i > gi, i > gi.

32 A counterexample would be the evolution je > 9, which is reconstructed for the history of several South Italian dialects, but see in Sánchez Miret (1996: in press) the arguments for the more plausible reconstruction: 9 > je.

33 Cf. the results by Gottfried et al. (1993: 219) for American English diphthongs, where the differences in duration between slow vs. fast tempo are of 27% for /ai, au/ but only of 19% for /ju/.

6. Conclusion

This contradiction focuses on the conflicting requirements of the category 'diph-
thong'. I think we can overcome this contradiction if we look at the dynamics of the facts. The diphthongization of a vowel tends to increase its latent duality, and the best way of doing this is for it to develop into a falling diphthong with maximal distant endpoints. Extreme evolutions lead to hiatus or to VC sequences. Furthermore, the realization of a diphthong varies, so that in most casual styles it can be produced as a monophthong or, if the circumstances are given, it can be abbreviated via syllabic change. The resulting rising diphthong can monophthongize (French ë > ye > a; e.g., cœur > coeur ‘heart’) or can evolve into a CV sequence. Some of these evolutions, as such diphthongization and conversion into hiatus are favored in foregrounding contexts (like stressed syllable and phrasal accent). In such contexts also an increase in duration is expected. On the opposite side, falling > rising and diphthong > monophthong are favored in backgrounding contexts, which usually are accompanied by decrease in duration. All these processes take place inside the supracategory of vocalic sounds. The evolution: diphthong > CV, VC implies the apperition of elements of another different supracategory, namely consonantal sounds. This is what I try to represent in Diagram 3.

Diagram 3. Processes inside then supracategory of vocalic sounds.
During these processes some diphthongs can be analyzed as allophonic manifestations of monophthongs, or as phonemic units or as biphonematic sequences. All through these evolutions into the interior or to the exterior of the category the preferences of the diphthong are fulfilled to a different degree. These preferences are represented below.

<table>
<thead>
<tr>
<th>unity</th>
<th>duality</th>
</tr>
</thead>
<tbody>
<tr>
<td>tautosyllabic</td>
<td>perceptual</td>
</tr>
<tr>
<td>cohesion</td>
<td>duration</td>
</tr>
</tbody>
</table>

a) Tautosyllabic has the aim of ensuring some degree of unity.
b) Syllabic cohesion serves unity.
c) Perceptual distance has the aim of providing some degree of duality. A maximal degree of distance serves perceptibility. A minimal degree serves pronounceability. The process of diphthongization polarizes the incompatible features of a monophthong (color and sonority). The process of assimilation has the opposite function.
d) An optimal diphthong is longer than a short vowel. Falling diphthongs fulfill this condition better than rising ones.
e) The category of diphthong shows a gradation between unity and duality. The prototypical diphthong should exhibit both features to a certain degree.

In conclusion, diphthongs are complex phenomena that show both unity and duality features, and which can arise in many forms in the human languages. I think that we come closer to their understanding if we look at their dynamics.

REFERENCES


Nathan, G. S. 1986. “Phonemes as mental categories”. In Nikiforidou, V. et al. (eds.). 212-223.


Slama-Cazacu, T. 1958 [1959]. “The experimental reversed speaking, with special view to diphthongs”.

In Rosetti, A. (ed.). 123-134.


