Syllable, accent, rhythm: typological and methodological considerations for teaching Spanish as a foreign language

José María Lahoz
Universidad Complutense de Madrid
josemarialahoz@yahoo.es

Resumen: El ritmo es una propiedad del habla relacionada con la organización temporal de los sonidos en términos de agrupamiento. Las unidades de segmentación son específicas de cada lengua y emergen de propiedades fonológicas tales como la estructura silábica, la fonotáctica, y los contrastes prosódicos en los niveles léxico y postléxico.

Las diferencias rítmicas entre las lenguas plantean problemas para la adquisición de segundas lenguas, debido a la compleja combinación de claves acústicas, las dificultades perceptivas causadas por la sordera fonológica, y las interferencias con la organización de los contrastes segmentales y con el acceso al léxico.

Este artículo ofrece una comparación tipológica que incluye descripciones de las distintas clases de ritmo existentes (temporización silábica, acentual, y moraica), así como de los distintos sistemas de contraste prosódico en el nivel léxico (lenguas tonales, de acento tonal, y de acento tipo «stress»). Este análisis permite predecir los errores típicos que afectan a los estudiantes de español de distintas procedencias lingüísticas.

Además, se describe la estructura silábica y el sistema acentual del español, y se sugieren algunas estrategias para practicarlos en clase.

Palabras clave: Ritmo; Estructura silábica; Lenguas tonales; Acento tonal; Acento; Tipología; Adquisición de segundas lenguas; Español para extranjeros.

Abstract: Rhythm is a speech property related to the temporal organization of sounds in terms of grouping. Segmentation units are language-specific and emerge from phonological properties such as syllable structure, phonotactics, and prosodic contrasts at the lexical and postlexical level.

Rhythmic differences across languages pose problems for second language acquisition, given the intricate combination of acoustic cues, the perceptual difficulties caused by phonological deafness, and the interferences with the organization of segmental contrasts and with lexical access.

This paper provides a typological comparison that includes descriptions of the attested rhythm classes (syllable-timed, stress-timed, and mora-timed), as well as of word prosody systems (tone, pitch-accent, and stress languages). This analysis yields predictions regarding typical errors for learners of Spanish from different linguistic backgrounds.

Additionally, Spanish syllable structure and stress system are described and some strategies are suggested to practice these in class.

Keywords: Rhythm; Syllable structure; Tone languages; Pitch-accent; Stress; Typology; Second language acquisition; Spanish for foreigners.

0. Introduction

0.1 Rhythm typology

Rhythm is a speech property with differing characteristics across languages. Although rhythm in speech is a concept difficult to grasp, it is generally assumed to relate to speech cadences, namely the temporal organization of sounds in terms of grouping (Jun 2005).

The first reference to a rhythm typology is attributed to Pike (1945), who coined the terms «syllable-timed» and «stress-timed». So-called syllable-timed languages devote a similar amount of time to pronouncing each syllable in an utterance. Of course, the exact duration of a specific syllable depends on factors such as the segmental nature of its sounds (fricatives are longer than taps, for example), or the number of segments that form part of it, which in turn depends on the syllable structures allowed in a particular language. In fact, syllable-timed languages are usually more restrictive about what syllable structures are legal and often exclude complex consonant clusters. This observation is certainly related to the trend toward syllable isochrony, although it is not clear which is the cause and which is the consequence. Some examples of this rhythm type can be encountered in Spanish, Italian, Portuguese, French, Korean, or Mandarin.

On the other hand, stress-timed languages tend to keep stressed syllables at fairly regular intervals. As a result, the more intervening unstressed syllables there are between two consecutive stressed syllables, the more they will become reduced to accommodate stress isochrony. This reduction, which increases the variability of syllable length, especially affects vowels[1]. Conversely, and contrary to syllable-timed languages, stress-timed languages generally tolerate consonant clusters. English, German, Dutch, Swedish, Arabic or Greek are exemplars of this rhythm class.

A third type, «mora-timed», was proposed by Bloch (1950). A mora is a unit of quantity. For example, a short vowel corresponds to one mora, whereas a long vowel amounts to two morae. The presence of postnuclear consonants also makes a syllable heavy or bimoraic. In mora-timed languages, each mora takes about as long as each other and hence the duration of a bimoraic syllable approximately doubles that of a monomoraic one. Prototypical mora-timed languages include Japanese, Ancient Greek, or Sanskrit.

The different rhythm types determine different strategies for the segmentation of the sound continuum (Jun 2005). Thus, temporal organization of sound

---

1. Although stress-timing does imply vowel reduction, the reverse is not necessarily true. For instance, Catalan is a syllable-timed language with vowel reduction in unstressed positions.

pivots on syllables, feet or morae, respectively, where a foot means a unit of contrast between a prominent (stressed) and a less prominent (unstressed) part. These units also condition different strategies for lexical access to the native speakers of these languages (Cutler & Norris 1988; Cutler *et alii* 1986; Mehler *et alii* 1981; Otake *et alii* 1993), and have been proved to interfere with speech segmentation processes when individuals listen to foreign languages belonging to other rhythmic groups (Cutler & Otake 1994).

0.2 Rhythm measures

The traditional account of rhythm links this concept to the temporal distribution of suprasegmental elements, such as the syllable or stress. However, recent approaches incorporate the idea that it may also be important to look at the duration patterns of segments. More specifically, rhythm is associated to the durational variability of vowel groups, on the one hand, and consonant groups, on the other. Recall that syllable-timed languages usually disprefer consonant clusters; that stress-timing triggers vowel reduction; and that mora-timed languages typically have contrastive vowel length.

Some authors (Dellwo 2006; Grabe & Low 2002; Ramus 2002; Ramus *et alii* 1999) have attempted to establish quantifiable parameters to capture rhythm differences by means of these related phonological and phonotactic properties of segments. For example, durational variability indices obtained for stress-timed languages show high values for both consonantal and vocalic intervals. This is as expected, since the possibility of consonant clusters in the language does not preclude simpler syllable onsets, and is therefore conducive to a greater variability. Vowel reduction, on the other hand, also correlates with greater contrasts in the duration of vowels. The exact opposite pattern is characteristic of syllable-timed languages, with more stable durations for both consonant and vowel groups. And for mora-timed languages, the variability index is low for consonants and high for vowels, as corresponds to the phonemic contrast between short and long vowels.

Whether or not these measures are truly measures of rhythm or something else (albeit correlated) is a matter of debate, but whatever they capture provides relevant cues for auditory discrimination in experimental tasks comparing languages of different rhythmic groups and, more surprisingly, also languages within the same group (Vicenik 2011). Not in vain has the rather rigid division between rhythm types of the first proposals given way to the less strict interpretation of isochrony and a classification of rhythm based on a gradual scale. According to Ramus (2002), for example, subtle but significant differences can be found in vocalic duration variability between Spanish, on the one hand,

and Catalan, French and Italian, on the other, despite the fact that they are all syllable-timed languages. The latter exhibit greater variability due to vowel reduction phenomena (Catalan and French) or a heavy reliance on vowel duration as a cue to convey stress (Italian).

For these reasons, a careful comparative analysis is necessary to predict potential sources of error in L2 acquisition processes, even among neighboring languages.

0.3 Relevance of rhythm in L2 acquisition

Rhythm is an organizational property of every particular language, taken as a whole. Therefore, it is not subject to within-language variation[2] and cannot be used to convey differences in meaning (contrary to intonational changes, for example). However, in the context of L2 learning, a deficient acquisition of the rhythm of the target language can actually affect meaning. As explained above, stress-timing causes hypoarticulation of unstressed syllables, and the resulting segmental reduction may have morphological and semantic effects. For example, an English student of Spanish may neutralize the gender distinction, based on the opposition of vowel timbres (-o / -a).

On the other hand, rhythm has implications for the way that a listener segments the speech continuum into discrete units and for lexical access (as mentioned before). Accordingly, proper acquisition of rhythm is crucial for effective oral comprehension.

An early acquisition of rhythm is also believed to set the basis for a correct organization of other aspects of the language, such as phonemic categories. In fact, infants do not organize the perceptual space of vowels until they are in their fifth month, and that of consonants until they are 8-10 months old. Nevertheless, they show signs of language discrimination based on rhythm starting from birth for languages that belong to different classes, and starting from the fifth month for languages from the same class, as long as their L1 is one of them (Vicenik 2011).

In contrast, rhythm is one of the traits most difficult to acquire when learning a foreign language as an adult. According to Vicenik (2011), adults weigh intonation over rhythm and paying excessive attention to the former may impair their ability to accurately discriminate the cues to the latter. On the other hand, rhythm is a complex property that emerges from syllable structure and stress. It

---

2. To be more precise, there may be some cases of rhythm divergence between dialects of a single language, as exemplified by English, whose Singapore dialect presents lesser vowel reduction than other dialects, such as British (Grabe & Low 2002).

sounds reasonable that, in order to learn the rhythm of a foreign language, the student must first be able to adequately produce the syllable structures and the accent patterns of that language.

Accordingly, the following sections treat these two aspects in turn (section 1 focuses on syllable and section 2 on types of accent and, particularly, stress), and then section 3 suggests some specific drills to exercise rhythm in class.

1. Syllable structure

1.1 Syllable structure in Spanish[3]

Spanish syllables conform to the following general structure:

$$(1) \quad [_{syll} \pm Onset + [_{rhyme} \ Nucleus \pm Coda]]$$

The onset is an optional component that may comprise one or at most two consonants. In the case of a simple onset, any consonant in the phonemic inventory of Spanish may occupy that position, /p, b, t, d, k, g, f, θ[4], s, x, tʃ, m, n, ɲ, l, r, ɾ, j/, except in word-initial onsets, where the tap /ɾ/ is not licensed. As for complex onsets, the first element must be one out of /p, b, f, t, d, k, g/ and the second element is always either /ɾ/ or /l/, with the caveat that *tl and *dl clusters are forbidden in most Spanish dialects. In these dialects, /tl/ and /dl/ sequences are heterosyllabic, that is to say, /t/ or /d/ belong to the previous syllable's coda, and /l/ fills the onset.

The nucleus is the only required constituent and consists of any one vowel of the phonemic inventory /a, e, i, o, u/. Unlike other languages, sonorant consonants cannot be part of the nucleus. But Spanish syllables do not always have a singleton vowel; rather, diphthongs and, to a lesser extent, triphthongs may appear. There are two kinds of diphthongs, namely rising (ia, ie, io, ua, ue, uo, iu, ui[5]) when the glide precedes the nuclear vowel, or falling (ai, ei, oi, au, eu, ou[6]) when the glide follows the nuclear vowel. Triphthongs (iai, iei, uai, uei) typically obtain in the conjugation of 2nd person plural[7] of the present tense –indicative or subjunctive– of those verbs that have a vowel before the infinitive ending –ar. There are also some cases of triphthongs outside the verb category, but they are

---

3. The following search engine (Alves, Rico & Roca 2011) can be used to look for examples of the various phonotactic patterns of Spanish: <http://www.estudiosfonicos.cchs.csic.es/fonetica/bufon>.

4. In standard peninsular Spanish.

5. The latter two are considered to be rising due to the general tendency of Spanish to open syllables.

6. Only one non-loanword has this diphthong, formed across a morpheme boundary: *estadounidense*.

7. Again, in standard peninsular Spanish, since this form is barely used in the Canary Islands or in America.

very rare (*guay, Paraguay, Uruguay, buey, vieira*). The analysis of diphthongs and triphthongs in terms of the structural position that each segment occupies varies depending on the authors (e.g., Colina 2012: 141ff.; Hualde 2005: 70ff.), but without major consequences for the practice of pronunciation training, as will be argued in section 1.2.

Thirdly, and like the onset, the coda position is noncompulsory, but if there is a coda it may be either simple or complex, the latter being relatively more frequent in word-medial position. At the end of a word, the most common consonants are /r, s, l, n, d, θ/ because historically they triggered apocope of final vowel /e/. Less typically, /b, m, f, k, x/ may also appear (mostly in loanwords). Word-medial coda may be filled with /r, s, l, n, d, θ/ as well, but the range of possibilities is enlarged to also cover /p, b, f, t, d, k, g/. Finally, a complex coda consists of one of the following consonants /p, b, d, n, l, r, k/ followed by /s/.

1.2 What to teach about syllable structure and how

When teaching a second language, the goal is surely not to form linguists, but to help students to enhance their comprehension and production of the oral and the written language. This means that all the explanations above are necessary for the teacher to anticipate likely errors, diagnose the causes, and apply effective solutions, but they need not be taught to students. With this in mind, teachers should prepare exercises according to the real difficulties of learners, which typically depend on their native languages.

For example, Japanese does not license onset clusters, so when Japanese speakers listen to a foreign language with clusters they perceive an epenthetic vowel (Dupoux *et alii* 1999), and hence it is natural that they also reproduce it in their pronunciation. In order to remove it, the student may try to articulate that part increasingly faster. The same is true for simple codas: only a nasal[8] is possible in Japanese, so the appearance of a different consonant triggers the insertion of a vowel. In this case, in addition to the mentioned strategy, the student should start trying with the most favorable contexts, such as word-final coda with the following word starting with a vowel, where resyllabification occurs naturally in Spanish, as will be explained below.

Another frequent problem arises with students who are native speakers of languages with vowel reduction. According to the verbotonal method of phonetic correction (Renard 1971), the origin of this difficulty lies in a lack of tension, and can be solved by means of some postural exercises. The student should seat upright, even stiff, for the pronunciation drills. Standing up may also help. As a

---

8. Or the first part of a geminate consonant, which is not a valid pattern in Spanish anyway.

last resort, pushing against a wall of the classroom while pronouncing may be a good idea.

On the other hand, speakers of these languages usually reduce vowels as an outcome of their rhythmic sequencing, but do not necessarily have problems pronouncing them in an isolated fashion. If this is the case, the teacher may write a series of vowels on the blackboard (e.g., a u e i o) and have the student pronounce them separately, hyperarticulating them, that is to say exaggerating the tongue and lip gestures. When the timbres are acceptable, then the teacher will introduce a real word made up with those vowels (e.g., *abuelito*), and he should underline the idea that if the student was able to do it properly before, he can also expect to achieve it now. English students may benefit from this drill in an additional manner, namely the avoidance of final vowel diphthongization (which would yield *abuelit*[əʊ] instead of *abuelit*[o]). Two more tips can be offered to prevent final diphthongs. Firstly, when the student is about to start gliding his tongue, the teacher may close his fist in the air so to signal the learner to fall silent. Secondly, the student may lean his elbow on the desk and, in turn, his chin against his thumb. This way, he will have a finer perception of his jaw movements and he can aim at locking his mandible during the production of vowels.

Yet another common problem is that of students who pronounce every single syllable separately. For example, this tendency is especially marked for Chinese learners of Spanish, owing to the fact that Chinese is a tone language[9], which gives every syllable a relative autonomy. Moreover, even though it must be acknowledged that Mandarin Chinese has a high number of compound words (formed to disambiguate otherwise homonym morphemes), it is relatively close to a prototypical isolating language, where morpheme-word ratio is 1:1. Given that Mandarin morphemes are monosyllabic, this strengthens syllable independence. A possible solution consists of having the students speak at a very slow rate, slurring the words together so as to break their staccato rhythm, and then steadily go back to a regular speech rate.

The need to break the boundaries across words also relates to a major difficulty for many learners of Spanish, namely resyllabification (Gil Fernández 2007: 272). When a word begins with a vowel (or the silent letter h-), it must be pronounced as linked to the previous word, forming a syllable across both words. For example *mis amigos* is syllabified as *mi.sa.**mi**.gos* (the inherently unstressed possessive merges with the noun into a single accent phrase, with the stress on the boldface syllable). Similarly, *el arte antiguo* becomes *e.**lar**.tean. **ti**.guo* (two accent phrases, distorting the word boundaries).

---

9. Or macrolanguage, following the *Ethnologue* (Lewis 2009).

Such a relaxation of boundaries poses problems to speakers of English and German, for example, who typically reinforce words beginning with a vowel by means of the insertion of a glottal stop [ʔ]. On the other hand, the phenomenon of resyllabification is responsible for the subjective impression that Spaniards speak very fast, yet in fact it is not a matter of speech rate but of rhythm. This temporal grouping does not make word division evident and, subsequently, it impairs lexical access and makes oral comprehension more difficult.

An effective approach to work on this phenomenon in class may be to present a first sentence without any such rearrangement, and then another one with the same cadences but with instances of resyllabification. The student must attempt to keep the rhythmic grouping, matching the number of phonetic syllables to that of the first sentence. The rationale is that if the student can pronounce, e.g., *Mi primo trabaja con Cristina y mi prima con su madre*, then he should also be able to get it right with *Mi prima trabaja_en_un colegio y mi primo_es_abogado*. This kind of comparison has the effect of boosting the learner's morale, which is also an important component to success.

2. Prosodic contrast, types of accent, and stress

2.1 Language typology regarding contrastive uses of prosody

According to Jun (2005), a language typology must account for two aspects of prosodic variation. One of them is the units of temporal grouping that determine rhythm, such as the syllable, the foot and the mora (if we take units below the word-level, which is the focus of this paper) or the accent phrase, the intermediate phrase and the intonational phrase (if we take units above the word-level)[10]. The other aspect of variation is the contrastive use of prosody –to mark *prominence*, in Jun's terms– at both the lexical and the postlexical level. This section will focus on the lexical prosodic contrasts.

The lexicon of any language may be characterized as using prosody for paradigmatic contrasts, syntagmatic contrasts, both or none, as will be explained next. In any case, this contrast is inherent to the words that bear it as a lexicon-based phonological property (though there may be atonic words too), and its type or position (where applicable) must be learnt by memory.

Tone languages exploit prosodic contrasts paradigmatically, namely each syllable in a word is specified for a tone (that may be level or a contour), and

---

10. These units are defined by the presence of a pitch contour that is particular to each language. The intonational phrase is the largest grouping unit delimited by a specific contour. The accent phrase and the intermediate phrase are smaller units, and they differ from each other on whether they comprise only one or more than one content word, respectively.

the chosen tone contrasts with all the range of possible tones that might have occurred in that syllable[11]. Thus, no syllable stands out over the rest because every single syllable has a tone. This is the case of Cantonese, for example. Mandarin, on the other hand, is primarily a tone language, yet there is sometimes a syllable whose tone is unspecified (or, as it is usually put, whose tone is neutral), thus creating a syntagmatic contrast as well. Strictly speaking, Mandarin is therefore a language that uses both tone and stress by way of prosodic contrast (Jun 2005). Some other exemplars of tone languages are Vietnamese, Thai, Punjabi (a language of India), most sub-Saharan African languages (with some exceptions, such as Swahili), and also some native American languages.

Instead of a paradigmatic contrast, languages may make use of prosody for syntagmatic contrasts in the lexicon, and this is what is called accent proper (Beckman 1986)[12]. Accent can in turn be divided in two types: pitch accent, and stress accent or simply stress.

Pitch-accent languages are characteristic for enhancing one syllable within the word (as is distinctive of accent languages by definition). Accordingly, a variation in the position of the accent may determine a semantic change. Yet not only is that position (syllable or mora) specified in the lexicon, but so too is the pitch contour associated to it. Some languages only have one possible contour in their inventory (e.g., Japanese)[13], whereas others have more than one (e.g., Swedish, with two), which introduces an additional paradigmatic contrast in the system. Moreover, despite being a pitch-accent language, Swedish also has stress-like contrasts[14] (recall that stress is compatible with the other systems, as in Mandarin). On the other hand, Japanese has a series of rules that determine the pitch manifestation of the unaccented morae (Rogers 2000: 277-278). Following these rules, it may well be the case that there are two or more high-pitched morae in one word, although only one is the accent bearer (the one preceding the change to low pitch), as in _a|tarashi|i_ 'new', with the accent on *shi*. The extreme case is that of inherently unaccented words, where no mora marks the transition from high to low (e.g., _to|modachi_ 'friend')[15].

In pitch-accent languages, the phonetic manifestation of pitch contours is barely affected by the general intonation pattern of an utterance, in contrast

---

11. As is widely known, provided that tone is a lexicon-based feature, a variation in the tone choice entails a change in meaning as well. E.g., Mandarin contrasts *mā* 'mother', *má* 'hemp', *mǎ* 'horse', *mà* 'insult'.

12. Accent can also refer to the particular way of pronunciation characteristic of a region, a social class, or an individual.

13. Beckman (1986) argues that it is not clear whether in these cases the pitch contour or only its position is specified in the lexicon, since the contour could be derived by rule.

14. See Jun (2005).

15. In Japanese there is a relatively high number of unaccented content words.

to stress-accent languages, as will be explained next. Stress refers to a type of accent, and hence implies a syntagmatic or positional contrast within the word. However, unlike pitch-accent, only the position is specified, but not the manifestation of accent as a particular pitch-contour[16]. According to Beckman (1986), that manifestation is resolved post-lexically, depending on the intonation contour that is applied to that word in a sentence. And if intonation sets a low pitch on the stressed syllable for independent reasons, then that syllable will be low-pitched, regardless of its stressed nature (it is indeed possible to have flat pitch over entire words).

As a result, the need to keep the syntagmatic contrast dictated by stress makes other cues come into play. Depending on which cue is used, a difference can be made between quantitative accent (duration), dynamic accent (amplitude or intensity) and qualitative accent (spectral tilt and timbre[17]), in parallel to the designation of pitch accent. Anyhow, stress usually surfaces as a combination of two or more of these cues, yielding what we could call a multidimensional accent. Such simultaneous cueing typically leads to a ranking between a primary cue and secondary cues, as will be explained below.

The presence of these additional cues does not necessarily preclude the appearance of a pitch peak but in any case it will be contingent on the general intonation contour. Even when a peak does occur, it is mounted on the melodic curve, so it represents a local maximum, but it is not necessarily the same height as other stress peaks in the sentence, in contrast to pitch-accent languages (Beckman 1986). Also, unlike pitch-accent languages, the pitch peak (if there is any) affects only one syllable, and it is not possible to find several consecutive syllables with a high pitch.

On the other hand, given the highly intonation-dependent nature of stress, the occurrence of a pitch peak does not necessarily mean that it will be centered on the stressed syllable, but it may be advanced or delayed. This alignment setting is language-specific and can be a source of errors for learners whose native language differs on this point. For example, Bissiri & Pfitzinger (2009) report the difficulties experienced by Italian learners of German due to the fact that Italian presents an early peak whereas German has a delayed peak. Italian listeners reanalyze the late German peak as being the early peak corresponding to the *following* syllable, thus giving rise to a misinterpretation of stress location.

---

16. Some examples of stress languages are: English, German, Dutch, Spanish, Portuguese, Italian, Greek, Arabic.

17. As explained in section 1, stress-timed languages (which are only a subset of languages with stress-accent) are characteristic for vowel reduction in unstressed position. This typically leads to a centralization in timbre and possibly to a loss of energy in the high frequencies of the spectrum, resulting in a steeper spectral tilt. Conversely, vowels in stressed syllables are more peripheral and present a softer tilt.

On the other hand, the primary cue in Italian is duration (while pitch peak is a secondary cue). In contrast, vowel length is contrastive in German, so duration is no longer an available cue for stress, and intensity (along with pitch) is enhanced instead. When Italians hear a compound word, they tend to perceive stress as though it were on the second component (and not the first), especially if there is a syllable with a long vowel. This illustrates the need for in-depth cross-linguistic studies in the field of L2 acquisition, in order to compare primary and notably secondary cues (given that their redundant nature may make them go unnoticed, thus making them more difficult to teach).

In some languages, stress always falls on the same position for every accented item (e.g., in Hungarian, it is the first syllable of the word). This is called fixed stress. In some other languages, the stress position is not fixed, but can be predicted by rule, depending on the syllable structure, the sequence of phonemes, or some other phonological property. This happened in classical Latin, for example.

It should be noted that, in spite of its postlexical completion, stress must not be confused with phrase accent, another prominence mechanism that also operates at the postlexical level but that affects groups of words, not single words (Beckman 1986; Jun 2005). In English, for example, the most important content words of a sentence tend to be emphasized, in a similar way to broad focus[18].

Finally, some languages only have phrase accent, but not lexical accent, as is the case of French or Korean (Jun 2005). In French, for example, a pitch rise delimits the right edge of intermediate phrases, a unit consisting of one content word possibly preceded by other content or non-content words. This usually gives non-native listeners the impression that there is a stress located on the last syllable of content words (thus wrongly generalizing to content words that are not phrase-final), but in fact this apparent oxytonicity is not determined by the lexicon (Beckman 1986).

2.2 Typology-based prediction of errors for learners of Spanish

Speakers of tone languages may have trouble changing from a system based on paradigmatic contrasts to another that makes use of syntagmatic contrasts. However, this difficulty may disappear if their language combines tones and stress. Secondly, they need to understand that only one syllable can be high-pitched within a word, or else they will sound singsong. And thirdly, they have to learn to mark the contrast not only by means of pitch, but also using the

---

18. Of course, narrow focus can also be applied to one element with contrastive purpose.

secondary cues, namely duration, peak delay (where applicable, see section 2.3), and intensity.

As for speakers of pitch-accent languages, they will likely perceive the stress position correctly (at least, as long as the pitch peak is allowed by intonation), but then they will apply their own rules to deduce the pitch of the rest of the syllables, which may possibly result in words with more than one high-pitched syllable. And like the former group, they also need to master the use of cues other than pitch that accompany stress in Spanish.

Even within stress languages, the students' mother tongues may produce various interferences. It goes without saying that speakers of fixed-stress languages must succeed in placing stress on different positions. Additionally, English speakers tend to overemphasize phrase accent. On the other hand, Italian relies on pitch and duration as cues to stress, but it is duration that prevails as the primary cue, and stressed syllables become exaggeratedly long as compared to the Spanish norm. Both Italian and Spanish use duration to cue stress. Therefore, since the cue itself is not incompatible, it may prevent students from realizing that the problem lies in a question of degree, and that phonetic detail does matter when they are trying to perceive the features of a foreign accent.

In general, regardless of the native language of the learner, it is hard to acquire a good command of secondary cues as well as of proper peak alignment, which in turn depends on an adequate mastery of the intonational contours of the language. And, needless to say, the stress position must be learnt by memory. In verbs it can be morphologically predicted based on the tense, but although certain tendencies may be identified in other words, these are not an infallible guide (Rico 2012).

Textbooks of Spanish as a foreign language at best only explain some of these rules, and typically illustrate the relationship between the use of graphic accent and the placement of stress. Yet knowing the stress position does not imply that an individual can perceive nor reproduce the adequate acoustic cues. The following sections offer a brief summary of these cues (2.3) and make suggestions on how to help students to improve the perception and production thereof (2.4).

2.3 Stress cues in Spanish

At least from a perceptual point of view, the primary cue for stress in Spanish is a pitch peak (Llisterri *et alii* 2003a, b), which does not necessarily mean that such a peak is always present in every stressed syllable in running speech, although it certainly occurs in isolated words. Duration, and to a lesser extent intensity, enhance the perception of stress as secondary cues (*ibid.*).

Longer duration and greater intensity, when present, always appear on the exact syllable that bears the lexical stress. In contrast, as is characteristic of stress systems, the pitch peak is not necessarily aligned with that syllable and, in fact, this is usually the case. Spanish shows a strong tendency toward late peaks[19]. It has been suggested that, in many languages, peak delay may be a secondary cue for high peak, via the inference that, given a constant slope, the longer it takes to reach the summit, the higher it is (Gussenhoven 2002: 52).

Pamies & Amorós (2005), on the other hand, observe that peak occurrence can be better predicted if related to morphosyntactic boundaries than to word stress; that is, it typically appears on the edge of words and syntactic phrases, such as the subject of the sentence[20]. Following these considerations, it might be proposed that pitch peak here is performing a demarcative function for words or phrases, rather than a culminative or a contrastive function.

Under this view, it makes sense that the peak is not delayed in isolated words, since they need not be separated from any other. And this, in turn, has a practical consequence for language teaching, which is that stress should not be practiced with the traditional minimal-pair drills, but in full sentences, so that students can get used to peak delay, which is a natural feature of running conversation.

Apart from in isolated words, the peak is not delayed when the word is under narrow focus (Face 2001), and strict alignment is the norm in the speech of Buenos Aires and its area of influence, which extends to Uruguay (Sosa 1999).

2.4 Correcting stress pronunciation

More often than not, pronunciation problems in a second language derive from perception problems. In other words, inasmuch as they are not trained to recognize certain acoustic cues that are not used in their mother tongue, students fail to discriminate acoustic differences that are relevant for the target language and consequently cannot reproduce them either. The concept of phonological deafness, first proposed by Petar Guberina and incorporated to the verbotonal method for phonetic correction (see Renard 1971), alludes to the difficult, or virtually impossible, task that listeners face when trying to recognize a phonetic difference based on an acoustic distinction that is not relevant to their own L1.

---

19. The peak usually appears on the syllable following the stress bearer. However, oxytone words do not show such a strong tendency, so the peak is often aligned with the stress position. On the other hand proparoxytone words sometimes show a greater delay (2-syllable step). As a consequence, the peak almost always occurs toward the end of the word.

20. At the end of declarative sentences, the expected downward tilt of the pitch curve situates the local peak at the beginning, not the end, of the syntactic phrase, even if it corresponds to a clitic (an inherently unaccented word), such as a preposition.

Or put differently, because listeners do not expect a certain acoustic trait to be present in the signal, it goes unnoticed.

Students unconsciously adapt what they hear to their own expectations, assimilating it to their closest native category. In the recent decades it has been proposed that, by way of a solution, it is very effective to start with the student's own preconceptions and dismantle them by recording his voice, manipulating the relevant acoustic cues with the aid of a computer, and having him listen to his own voice pronouncing correctly so that he can hear the difference and repeat accurately (Bissiri & Pfitzinger 2009; Felps, Bortfeld & Gutiérrez-Osuna 2009; Lahoz 2011, 2012; Nagano & Ozawa 1990; Sundström 1998). Also, according to Bissiri & Pfitzinger (2009), using the student's own voice and not a pre-recorded model, or possibly a recast version of the intended sentence made by the teacher, has the effect of eliminating speaker-dependent voice-quality traits that may distract the learner from the real target, namely acoustic cues of stress. Surprisingly, using this kind of synthetic speech is more effective as a training model than a natural, native voice.

Next follows an explanation of how to carry out this manipulation[21] with the free software *Praat* (Boersma & Weenink 2011), which can be downloaded from the internet[22]. When the program is launched, two windows appear: the Objects window, and the Picture window. The latter may be closed because it is not necessary for the current purposes. First of all, the student's natural voice must be recorded by going to the menu and selecting *New > Record mono sound*. A typical recorder interface will appear. After recording the learner's pronunciation of a target sentence, the recording should be saved into the objects list, and the recorder may be closed. Alternatively, if a recording of the student already exists on the local drive, it can be retrieved with *Open > Read from file*. In either case, a new Sound object will appear in the main window list, and some buttons will come into view on the right-hand column as well. The Sound object must now be turned into a Manipulation object. To do so, select the Sound and press *To Manipulation…*, whereby a dialog window will pop up. Pitch floor settings may need to be changed in that window depending on the speaker's gender (75 for male, 100 for female). Press OK and then a Manipulation object will be added at the end of the list. With this object selected, press *View & Edit*. In the emerging window there are three main horizontal parts, namely the waveform, the pitch contour and the duration tier. The latter two will be the focus of manipulation.

---

21. A similar proposal can be found in Lahoz (2011).

22. The program website is <http://www.praat.org>.

In the menu *Pitch* there are several helpful options. First of all, the curve should be stylized with *Stylize pitch (2 st)*. As a result, only the important turning points will remain. Furthermore, points can be inserted or removed with the appropriate functions from the menu. Additionally, a point can be moved from one location to another by clicking and dragging. With all this in mind, the aim is to alter the original contour as pronounced by the student and turn it into a Spanish-like contour (the contour that it should have been), following the indications about pitch cues in the previous section. A graphical comparison of the new melodic curve, in green, with the original, gray one in the background, may provide beneficial feedback to the student (Chun 1998; Martin 2005).

Additionally, durations may also be manipulated by adding points to the duration tier from the *Dur* menu. A value of 1, namely points situated on the reference line, keeps the duration unchanged. Higher values increase the original duration and lower values reduce it, by way of multiplicative factor. Since values are interpolated between one point and another, in order to change the duration of a segment without altering the surroundings, the modifiable portion should be delimited by 1-valued points, and then within these a different coefficient can be specified. Obviously, changes must be oriented toward shortening the incorrectly lengthened syllables and prolonging the stress-bearer syllables.

Once the manipulation has yielded the expected results (they can be listened to by clicking on the lower gray bars), the Manipulation object can be resynthesized as a Sound by means of the PSOLA algorithm by pressing *Get resynthesis (overlap-add)* in the objects window. The outcome may then be played to the student or saved to the computer or a pendrive with *Save as WAV file*.

Thus the student can keep a collection of sounds to continue practicing at home and consolidate any improvements. Moreover, by keeping a record of all these materials, the learner can gain awareness of his progress and thus be further motivated.

Sometimes the manipulation takes longer than others, and the teacher will have to adjust the cues by ear until it sounds native-like. In order to avoid boring the students at class, manipulations can be prepared at home and brought in the following day. By doing so, it is possible to combine a correction customized for each student and a dynamic class.

3. From syllable and stress back to rhythm

In the remainder of this paper, some exercises will be proposed to practice rhythm as a global phenomenon.

3.1 The accents of corporal movement

When we speak, we usually accompany our ideas and our attitudes with bodily movements. We also tend to parallel our rhythmic phrasing with such movements, especially of our hands and shoulders. There are likely to be differences between cultures in this respect, but it is certainly a marked tendency among speakers of Spanish and is therefore a cultural and non-verbal aspect of communication that students should know and try to imitate, mainly because it may help them to develop proper rhythm.

In class, the teacher may play a video (e.g., from a TV show) that features an expressive individual. The activity consists of listening and paying attention to the individual's movements and seeing how these correlate with groups of words that are packed together, somehow accompanying the prosodic segmentation of an utterance. These gestures may also reinforce the emphasized phrases. Then the teacher will turn down the volume and play the video again. The students will have the script in front of them and they will attempt to dub the video, being careful to imitate the gestures. Alternatively, they may change the script and improvise the dubbing, which can be great fun for a class.

3.2 Rhythm in songs

Songs highlight speech rhythm. In the Spanish case, they may be useful to help students understand that stresses are not equidistant and that there is no vowel reduction. Also, they may be useful for practicing the resyllabification rule. The teacher must think carefully about his choice of song given that some Spanish songs do change stress positions with respect to the norm so that they can fit the lyrics to the musical rhythm. Other than that, songs provide an excellent and amusing way of practicing rhythm, just by listening to them or singing along.

Songs are especially advisable for learners whose mother tongue has fixed stress. For example, Hungarian students will be tempted to always stress the first syllable of the word. The teacher can then play a song such as the traditional *Que llueva, que llueva, la Virgen de la cueva…* where the intrinsically atonic conjunction *que* starts the song, before the beginning of the first beat, thus constituting what in music is called *anacrusis*. Admittedly, the word *llueva* bears the stress on the first syllable but, to all effects and purposes, the phonological word (accent phrase)

includes the clitic and more importantly, if the student is able to pronounce *que llueva*, he should also get it right for *manzana*, or any other word with the stress on the second syllable.

3.3 Perception of rhythm in synthetic speech deprived of segmental information

As explained in section 0.2, adults can auditorily discriminate between languages on the basis of durational flexibility in both consonantal and vocalic groups (Vicenik 2011). In these experiments, participants were presented synthetic speech where the identity of segments had been removed. On the other hand, according to the verbotonal method, emphasis should be put on perception drills in L2 acquisition settings so that students can overcome phonological deafness and start using that knowledge for production as well. For these reasons, sensitization exercises with that kind of synthetic speech may be of help.

With the aid of a *Praat*[23] script[24] and the also freely available program MBROLA (Dutoit *et alii* 1996), a natural speech recording can be transformed so that all consonants are substituted for by /s/ and all vowels by /a/. For instance, *Mis padres se fueron de vacaciones* would become *sas sá.ssas sa saá.sas sa sa.sa. saá.sas*. The choice of these segments (/s/ and /a/) tries to maximize perceptual difference between a fairly closed (obstruent) articulation and another maximally opened (vowel) articulation, while the preference for a fricative instead of a stop lies in the fact that fricatives can prolong their duration more naturally. The so-called *sasasa* speech removes segmental information, and thereby also possible distractors, but preserves most of the prosodic features[25].

For example, all durational cues are respected. Syllable duration is key to perceiving syllable-timed rhythm. Vowel duration allows the student to perceive that there is no vowel reduction. Durational contrasts among vocalic groups, on the one hand, and consonant groups, on the other, are related to the quantitative measures of rhythm discussed in section 0.2. And a greater duration signals the location of the stressed syllables. Additionally, the pitch curve is also retained, thus giving the necessary cues for stress and its particular alignment properties in Spanish. Altogether, duration and pitch cues of stress allow the student to perceive that stress is not evenly spaced in Spanish; that is, it is not isochronic.

---

23. Boersma y Weenink (2011).

24. The text of the script and some more detailed instructions can be found in Lahoz (2012), along with extra exercises on rhythm and intonation.

25. Intensity is normalized to a mean of 70 dB.

This kind of input helps students to focus on the relevant, prosodic cues, instead of trying to identify word meanings above all, a very common attitude in learners of a foreign language. Repeated exposure to these synthetic stimuli may enhance the acquisition of the intricate factors that give rise to the particular rhythm of a language (Spanish, in this case).

However, Vicenik (2011) notes that, between intonationally similar languages, the presence of the pitch contour may actually act as a distractor, since adults weigh intonation over the subtler cues of rhythm. It is therefore necessary to conduct further research to compare the success attained by different L1 groups with normal *sasasa* speech versus flat-pitched *sasasa* speech, where sheer durational information can still partially cue rhythmic grouping.

References

Alves, H.; Rico, J. & Roca, I. (2011). *BuFon: Buscador de patrones fonológicos.* Madrid. CSIC. <http://www.estudiosfonicos.cchs.csic.es/fonetica/bufon>.

Beckman, M. E. (1986). *Stress and non-stress accent.* Dordrecht. Foris.

Bissiri, M. P. & Pfitzinger, H. R. (2009). «Italian speakers learn lexical stress of German morphologically complex words». *Speech Communication*, 51 (10): 933-947.

Bloch, B. (1950). «Studies in Colloquial Japanese IV: Phonemics». *Language*, 26 (1): 86-125.

Boersma, P. & Weenink, D. (2011). *Praat: doing phonetics by computer.* [Computer program.] Version 5.3.03, retrieved on 12 December 2011 from: <http://www.praat.org>.

Chun, D. M. (1998). «Signal analysis software for teaching discourse intonation». *Language Learning & Technology*, 2 (1): 61-77.

Colina, S. (2012). «Syllable structure». In: Hualde, J. I., Olarrea, A. & O'Rourke, E. (eds.). *The Handbook of Hispanic Linguistics.* Oxford. Wiley-Blackwell: 133-151.

Cutler, A.; Mehler, J.; Norris, D. & Segui, J. (1986). «The syllable's differing role in the segmentation of French and English». *Journal of Memory and Language*, 25: 385-400.

Cutler, A. & Norris, D. (1988). «The Role of Strong Syllables in Segmentation for Lexical Access». *Journal of Experimental Psychology: Human Perception and Performance*, 14 (1): 113-121.

Cutler, A. & Otake, T. (1994). «Mora or phoneme? Further evidence for language-specific listening». *Journal of memory and language*, 33: 824-824.

Dellwo, V. (2006). «Rhythm and speech rate: A variation coefficient for ΔC». In: Karnowski, P. & Szigeti, I. (eds.). *Language and language processing: Proceedings of the 38th Linguistics Colloquium, Piliscsaba 2003*. Frankfurt am Main. Germany. Peter Lang: 231-241.

Dupoux, E.; Kakehi, K.; Hirose, Y.; Pallier, C. & Mehler, J. (1999). «Epenthetic vowels in Japanese: A perceptual illusion?». *Journal of Experimental Psychology: Human Perception and Performance*, 25 (6): 1568-1578.

Dutoit, T.; Pagel, V.; Pierret, N.; Bataille, F. & Van der Vrecken, O. (1996). «The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes». *Proceedings of the Fourth International Conference on Spoken Language (ICSLP'96)*. Philadelphia. IEEE: 1393-1396.

Face, T. (2001). «Focus and early peak alignment in Spanish intonation». *Probus*, 13 (2): 223-246.

Felps, D.; Bortfeld, H. & Gutierrez-Osuna, R. (2009). «Foreign accent conversion in computer assisted pronunciation training». *Speech Communication*, 51 (10): 920-932.

Gil Fernández, J. (2007). *Fonética para profesores de español: de la teoría a la práctica*. Madrid. Arco/Libros.

Grabe, E. & Low, E. L. (2002). «Durational variability in speech and the rhythm class hypothesis». In: Gussenhoven, C. & Warner, N. (eds.). *Laboratory Phonology*,7. Berlin. Mouton de Gruyter: 515-546.

Gussenhoven, C. (2002). «Intonation and interpretation: phonetics and phonology». In: Bell, B. & Marlien, I. (eds.): *Proceedings of the 1st International Conference on Speech Prosody, Aix-en-Provence*. Aix-en-Provence. Laboratoire Parole et Langage: 47-57.

Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge. Cambridge University Press.

Jun, S. A. (2005). «Prosodic typology». In: Jun, S. A. (ed.). *Prosodic typology: The phonology of intonation and phrasing*. Oxford. Oxford University Press: 430-458.

Lahoz, J. M. (2011). «Manipulación de claves acústicas para la corrección del acento léxico en la enseñanza de lenguas extranjeras». In: Sevilla Muñoz,

J., Fernández-Pampillón, A. & Poves Luelmo, A. (eds.). *El laboratorio de idiomas y la enseñanza - aprendizaje de lenguas*. Madrid. Editorial Complutense: 161-169.

— (2012). «La enseñanza de la entonación, el ritmo y el tempo». In: Gil Fernández, J. (ed.). *Aproximación a la enseñanza de la pronunciación en el aula de español*. Madrid. Edinumen: 93-132.

Lewis, M. P. (2009). *Ethnologue: Languages of the world.* 16th edition. Dallas, TX. SIL International. Online version: <http://www.ethnologue.com>.

Llisterri, J.; Machuca, M. J.; De la Mota, C.; Riera, M. & Ríos, A. (2003a). «The perception of lexical stress in Spanish». In: Solé, M. J., Recasens, D. & Romero, J. (eds.). *Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona, 3-9 August 2003*: 2023-2026.

Llisterri, J., Machuca, M. J., Ríos, A., de la Mota, C. & Riera, M. (2003b). «Algunas cuestiones en torno al desplazamiento acentual en español». In: Herrera, Z. & Martín Butragueño, P. (eds.). *La tonía: dimensiones fonéticas y fonológicas*. México. Colegio de México: 163-185.

Martin, P. (2005). «WinPitch LTL, un logiciel multimédia d'enseignement de la prosodie». *Apprentissage des Langues et Systèmes d'Information et Communication*, 8 : 95-108.

Mehler, J.; Dommergues, J. Y.; Frauenfelder, U. & Segui, J. (1981). «The syllable's role in speech segmentation». *Journal of Verbal Learning and Verbal Behavior*, 20 (3): 298-305.

Nagano, K. & Ozawa, K. (1990). «English speech training using voice conversion». *First International Conference on Spoken Language Processing (ICSLP 90), Kobe, Japan*: 1169-1172.

Otake, T.; Hatano, G.; Cutler, A. & Mehler, J. (1993). «Mora or syllable? Speech segmentation in Japanese». *Journal of Memory and Language*, 32: 258-278.

Pamies, A. & Amorós, M. C. (2005). «Pico tonal, acento y fronteras morfo-semánticas: experimento con hablantes granadinos». *Estudios de Fonética Experimental*, 14: 202-223.

Pike, K. L. (1945). *The Intonation of American English*. Ann Arbor, MI. University of Michigan Press.

Ramus, F. (2002). «Acoustic correlates of linguistic rhythm: Perspectives». In: Bell, B. & Marlien, I. (eds.). *Proceedings of the 1st International Conference*

on Speech Prosody, Aix-en-Provence. Aix-en-Provence. France. Laboratoire Parole et Langage: 115-120.

RAMUS, F.; NESPOR, M. & MEHLER, J. (1999). «Correlates of linguistic rhythm in the speech signal». *Cognition*, 73: 265-292.

RENARD, R. (1971). *Introduction à la méthode verbo-tonale de correction phonétique*. Paris. Didier.

RICO, J. (2012). «El acento y la sílaba en la clase de ELE». In: GIL FERNÁNDEZ, J. (ed.). *Aproximación a la enseñanza de la pronunciación en el aula de español*. Madrid. Edinumen: 75-92.

ROGERS, H. (2000). *The sounds of language. An introduction to Phonetics*. Harlow. UK. Pearson Education.

SOSA, J. M. (1999). *La entonación del español. Su estructura fónica, variabilidad y dialectología*. Madrid. Cátedra.

SUNDSTRÖM, A. (1998). «Automatic prosody modification as a means for foreign language pronunciation training». *Proceedings of the ISCA Workshop on Speech Technology in Language Learning (STILL 98), Marholmen, Sweden*: 49-52.

VICENIK, C. J. (2011). *The role of intonation in language discrimination by infants and adults*. Ph. D. Dissertation. Los Angeles. UCLA.

Mentioned websites

BUFON (ALVES, RICO & ROCA 2011): <http://www.estudiosfonicos.cchs.csic.es/fonetica/bufon>

MBROLA (DUTOIT *et alii* 1996): <http://mambo.ucsc.edu/psl/mbrola/>

PRAAT (BOERSMA & WEENINK 2011): <http://www.praat.org>