**Modeling second language prosody acquisition:
A dynamic systems theory approach**

Emily Nava and Joseph Tepperman
Rosetta Stone Labs
enava@rosettastone.com; jtepperman@rosettastone.com

I. Introduction

1.1 Approaches to second language prosody acquisition

Among the plethora of second language (L2) acquisition models that have moved in and out of circulation (Larsen-Freeman & Long 1991 estimate around 40), a number concern the L2 acquisition of phonology-related processes, and a handful of these have been applied to research and findings on suprasegmental aspects. The Speech Learning Model (SLM) (Flege 1995) posits a directionality of difficulty for the L2 acquirer such that similarity poses more of a problem than difference. In order to acquire a sound (phoneme or allophone) that is present in the target language but not in the native language, the acquirer must first identify the sound as sufficiently different from existent sounds in the native system; failure to do so will result in substitution of a sound from the L1. Likewise, Major's Ontogeny Phylogeny Model (OPM) (2001) claims that acquirers are more successful at acquiring sounds that represent the greatest contrast to those found in the L1, while those approximating existing units more closely are hypothesized to present the biggest challenge. See Mennen (1999) and Gut (2003) for examples of prosody-related studies that provide support for and against, respectively, the above models.

An additional approach has been to examine separately the phonological and phonetic components of prosody (Fox 2000), in an effort to better diagnose the nature of both the L2 prosodic production and the acquisition process. The impression that a prosodic production is non-native is the culmination of a combination of phonological and phonetic factors still anchored to the speaker's L1 prosody, which contribute to production that is perceived as non-target. The phonological properties associated with prosody include the choice and placement of pitch accents, and the phonetic properties are those associated with temporal alignment of pitch events. The term pitch accent refers to pitch movement on a stressed syllable, in addition to the other characteristics of stress (duration, amplitude, spectral quality of vowel). The distribution of pitch accents guides the patterns of prominence placement among words in a sentence that contribute to its meaning (Pierrehumbert & Hirschberg 1990, Xu & Xu 2005). See Ueyama & Jun (1998) and Jilka (2007) for examples of studies that examine both the phonological and phonetic properties of L2 prosody acquisition.

The current approach differs from both the "difficulty" arguments as well as from the separation approach by modeling the acquisition process dynamically in order to account for both the discrete and continuous aspects of complex systems (Gafos & Benus 2006). The availability of different prosodic patterns in English can be modeled as multiple discrete modes of a dynamical system that regulates a continuous variable: the relative prominence of adjacent constituents. While this relation is continuous, the modes of its distribution correspond to prosodic patterns that are qualitatively distinct from any other in terms of meaning, and it is this discrete structure that the dynamics also models. This shift in qualitative meaning along a continuum with variation figures prominently in modeling the differences between Spanish and English at the prosodic level, as well as the change in L2 behavior as a result of acquisition.

1.2 Dynamic systems theory and second language acquisition

The current work couches the L2 prosody acquisition process within a dynamic systems theory (DST) framework by positing that prosody is a complex system with self-organization, and that the L2 speaker's prosodic system is undergoing qualitative changes in the dynamics of its organization. We hypothesize that the reorganization process involves a quantitative shift in successive vowel duration coordination before a qualitative shift in prominence placement at the phrasal level occurs. By analyzing the relation between prominence placement and rhythm as part of the acquisition process within a complex systems framework the prosodic structure can be modeled and these implications tested empirically.

Complex systems are characterized by the non-linear interaction of a large body of components that demonstrate self-organization. Prosody exhibits self-organization due to the multiplicity of components that interact in a non-linear fashion, in a hierarchical schema that is, however, not unidirectional (Tilsen 2009). Prosody qualifies as a complex system with self-organization precisely because of this circular causality: a given prosodic pattern results from the coordination between its components parts, but also exerts an influence on the behavior of these components.

What has been observed (and experimentally reproduced) is a synergy among these events, such that quantitative differences in temporal events yield qualitative effects in meaning. Integral to any model of language is the quantitative-qualitative symmetry that speaks to a generalizability emerging from the structure, which in turns affords predictions about the behavior. Relating the qualitative to the quantitative in a principled and formal way is the domain of dynamics and differential equations that can be used to model complex self-organized systems. The qualitative aspects of the system have typically been studied in isolation from the quantitative ones as a grammar of prosody, such as in the case of the above-mentioned SLM, OPM, and the phonological/phonetic separation approaches. While separating out the qualitative aspects in this way when describing the system does no violence to the understanding of the structure of the system when it is in a stable state, in order to address the issue of change in a system it is necessary to simultaneously embrace the qualitative and the quantitative.

Behavioral patterns self-organize such that they emerge, stabilize and change in accordance with the given control parameters. A control parameter moves the system through the various patterns, and is identifiable because its variation causes qualitative change, for example the qualitative change in meaning associated with a given prosodic pattern. An order parameter, on the other hand, is a single dimension along which the state of the observable can be described, so as to reveal its stable state(s). Such states result from the coordination between the elements and units of the system, at the same time influence their behavior.

The various linguistic and contextual parameters that influence the placement of relative prominence include discourse context/information structure, verb type and syntactic structure, word order, and possibly speaking rate. The dynamics associated with each of these parameters can be modeled with a potential function that expresses the preferred state or states of relative prominence associated with that parameter. These preferred states are attractor values. In producing any given utterance, the composite preference for prominence placement can be modeled by superposition of the parameters that are relevant to this particular utterance, and determining the attractors of that composite potential function. As a result of this combination of potential functions, the main prominence (MP) can end up falling in a number of states, of which we will be restricting in this work to phrase-internal or phrase-final.

The prosodic pattern is an example of a stable state along the order parameter of relative prominence. This should be thought of as a nested subsystem with multiple parts: the rhythm, or rather vowel durations as a measurable indicator of rhythm, and the emergent rhythmic pattern that gives rise to the phrasal prosodic pattern. Attractors are stable states, in this case an example of a prosodic pattern associated with a particular combination of verb type, discourse context, etc. Each speaker's prosodic system exhibits attractors for prosodic production, with very little noise or fluctuation around a given stable state. However, in order for a system to move from one stable state to a new one as a result of a change to a particular control parameter, an attractor would have to undergo destabilization. A phase transition in the case of prosodic patterns would shift relative pitch accent placement as a function of verb type and information structure, and in the case of rhythm refers to shift in the preferred ratio of adjacent vowel durations as a function of the speech rate.

While both Spanish and English have the same factors governing relative prominence placement, the measurable difference along this order parameters is a result of the differential flexibility caused by self-organization of the systems, in particular as determined by the vowel durations and their interactions in the coordinative space. The prediction is that a change in certain component properties of the system (for instance, vowel duration) will result in broader/larger qualitative changes of the system (prosodic patterns). But the notion that acquisition changes the properties of the system makes no prediction about the direction of that change. The specific hypothesis advanced here is that changes in the coordination between vocalic units are the necessary precursors to promoting a qualitative change at the phrasal level. This hypothesis was tested with a series of experiments designed to probe both the differences in organization between English and Spanish and the relationship between rhythmic flexibility and prominence placement as evidenced by L1Spanish/L2English speakers of varying prosodic proficiency levels.

II. Experiment 1: Phrasal prominence events

2.1 Introduction

Within a given utterance, there is one word perceived as more prominent than the others, which is assumed to correspond to a more prominent production of that word with respect to its syntagmatic cohorts. Words are flagged as prominent in response to informational focus operations. Focus, in general terms, regulates the incorporation of information in to the discourse flow (Ladd 1996), with words under focus serving as the locus of prominence placement. “Wide focus” refers to a context where all the information is new, or no information is presupposed, as in example (1) below.

(1) a. What happened?
b. Jane made a shirt.
c. Jane made a shirt at home.

In the case of each response, the final word is analyzed as receiving main prominence for this transitive verb type and a wide-focus context. However, there are wide-focus contexts where MP placement can occur phrase-internally in English (Schmerling 1976, Gussenhoven 1984), specifically in the case of the unaccusative class of intransitive verbs, examples of which are given below in (2) (cited from Zubizarreta 1998: 68).

(2) a. The sun came out.
b. The mail arrived.

An additional source of phrase-internal prominence in English is object-verb (OV) compounds (Selkirk 1984):

(3) a. tree-eater

The insight that both phrase-final and phrase-internal prominence placement patterns are produced in English for wide-focus contexts has been tested and confirmed empirically in the current work (see results from Experiment 1 given in section 2.2).

In Spanish, however, it is widely noted that prominence placement is phrase-final across all verb types for wide-focus contexts, regardless of word order (Zubizarreta 1998, Sosa 1999, Hualde 2007).

(4) a. María hizo una camisa. (Maria made a shirt)
b. Salió el sol. (came out the sun) c. El sol salió. (the sun came out)
d. Llegó el correo. (arrived the mail) e. El correo llegó. (the mail arrived)

Furthermore, the corresponding compound construction does not exist in Spanish.

Under the complex systems analysis, English is analyzed as having both a phrase-final and a phrase-internal attractor for phrasal prominence, whereas Spanish has only a phrase-final attractor for the same discourse context. The hypothesis is that the L1Spanish/L2English acquirer will start out with the same attractor landscape for English as in their native Spanish, and that the development of a second attractor (for phrase-internal prominence) results from the reorganization of vowel duration coordination (micro-temporal flexibility).

2.2 Methodology

The first experiment designed to test this hypothesis was a scripted question and answer (Q&A) dialogue task between the experimenter and the participant. Thirty-five English native control speakers (ENC) and 45 L1Spanish/L2English (L2E) speakers participated in Experiment 1. Participants' responses were recorded and analyzed using PitchWorks software program.

2.2.1 Design

For the verb type and discourse context conditions syntactic structures were paired with different information structure contexts. The full experiment included multiple verb type - discourse context pairing, but only results from the unaccusative verbs and OV compounds with neutral, wide-focus contexts are discussed here (see Nava 2010 for a complete review of results).

2.2.2 Coding and statistics

Data were coded for the presence vs. absence of pitch accents (PA) and for the location of the nuclear PA, marking main phrasal prominence. Coding was done by one Tones and Breaks Indices (ToBI) (Silverman et al. 1992) trained and one naïve native English speaker to ensure inter-rater reliability. Inter-rater reliability was 94%, and any coding discrepancies were resolved by a third, expert ToBI labeler. For the statistical analysis the results were pooled across participants for a Chi-square analysis, where each participant contributed 12 observations in the case of the unaccusatives and compounds.

2.2 Experiment 1: Results

A representative token of MP production for unaccusative verbs is given in (5). The significant difference was observed between the ENC and L2 populations with the wide focus SV unaccusative structures; the ENC produce main prominence on the subject 97% of the time, and the L2E only 23% of the time ($\chi^2 = 137.45$, $p < .001$). These results add to existing experimental data that have also found a statistically significant main prominence placement on the subject for unaccusatives with native English speakers (Hoskins 1996).

(5) What happened?

a. A <u>glass</u> broke. (ENC)   b. A glass <u>broke</u>. (L2E)

Earlier studies on the realization of PAs in English compounds have shown that no PA is present on the second constituent for this type of OV compound

(Gussenhoven 2004: 18). A significant difference between the ENC and the L2E ($\chi^2 = 37.54$, $p < .001$) was likewise found in the case of the compounds, with the ENCs producing main prominence on the argument 96% of the time, and L2ers doing so 43% of the time.

(6) What will Tim do in Africa?
a. He will go lion-hunting. (ENC) b. He will go lion-hunting. (L2E)

The L2 population clearly demonstrates a difference in prominence realization, where a unimodal organization is seen as influenced by the system-specific organization of their L1 (Spanish) – despite the presence of the same additional variables that affect the behavior of the ENC speakers. Thus far, the results have been described as a categorical presence of positions for MP realization, but the behavioral indices (such as pitch and duration) that correspond to those distinct modes vary in a continuous fashion. Here, we attempt to understand the ability to make the transition to one of these MP modes as a function of changes in the more microscopic variables controlling speech production (rhythm).

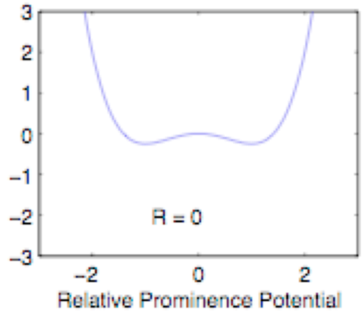
2.3 Modeling prominence

The position of prominence in the data of this experiment is modeled by means of an order parameter, which is the relative prominence of non-final vs. final position. This order parameter is in principle a continuum corresponding to the potential continuous physical parameters that give rise to the perception of prominence. The presence of an attractor causes the observed values to cluster in two discrete ranges in English and a single discrete range in Spanish. In English prominence can be placed either phrase-finally or phrase-internally, and the relevant point is that realizing prominence in either of those locations is a stable pattern in English. To model this, an equation known as the tilted anharmonic oscillator is used (Tuller et al., 1994, Gafos & Benus 2006).

(7) $V(x)= -Rx-x2/2+x4/4$

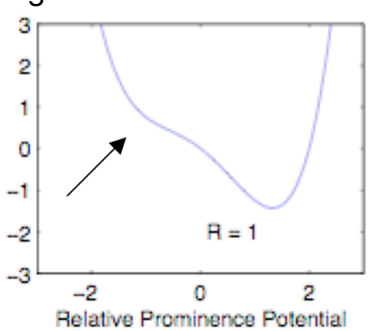
Spontaneous changes in category have been tested in the literature using this potential function, to probe the notion that such changes are only possible if there is more than one stable mode (Tuller et al. 1994, Gafos & Benus 2006). When the value of R (the (a)symmetry parameter) in equation (7) is set to zero, the potential function graphed below is obtained.

Figure 1. Potential function as plotted for English



The potential shows two attractors where relative prominence can fall; one with a positive value, and the other a negative value. If relative prominence is defined as a ratio of the prominence of the final word over the non-final word, then a positive value of the log of that ratio corresponds to more prominence on the final word, and a negative value corresponds to more prominence on the non-final word. Thus the two modes of this attractor correspond to final vs. non-final prominence. This function will be called the relative prominence potential (RPP). If English-speakers' prominence placement is guided by the dynamics of the anharmonic oscillator, with the asymmetry parameter R set to zero, the fact that relative prominence can occur either finally or non-finally can be modeled. In order to model the behavior of Spanish speakers, the identical anharmonic oscillator equation is used, but with the value of R set to 1. The result is the potential function shown in Figure 14 below.

Figure 2. Potential function as plotted for Spanish



There is a single attractor, with a positive value – indicating final prominence. The formal model of prominence proposed assumes that there is a fully-formed, non-dynamic representation of a sentence that includes parsing into constituents, ordering, and primitives like argument and head. The dynamical determination of prominence is thus a formally (modularly) separate process that does not interact with the syntax. For the current cases, in which the RPP only models relations between adjacent constituents, such as the short subject-verb phrases presented from Experiment 1, this approach could be adequate. However in future versions of the model, a dynamical recasting of some aspects of syntactic structure itself will be necessary in order to accomplish the relevant computations in a dynamical system.

In order for an L1Spanish speaker to move towards a different attractor space, a restructuring of the self-organizing components within that system must take place. It becomes increasingly more and more untenable to maintain the same modes of organization engaged when speaking the L1 while speaking the L2. The very identity of the L2 input (differences in vowel quality, syllable structures, etc.) forces an accommodation that brings in to question L1 modes of coordination whose functionality and efficiency decrease as exposure to and use of the L2 increase. An example of this – explored here – is that as vowel durations shift in the production of the L2 speaker speaking English, this is the necessary change to move the speaker's system into a different mode of organization, which has repercussions at the phrasal level.

The results presented above reflect group averages. A closer look at the individual data reveals that 9 of the 45 L2 speakers produced PAs phrase-internally at least 75% of the time. These speakers with English-like main prominence (+MP) in their speech have modified the organizational landscape such that a phrase-internal attractor is operative. Given the hypothesis that re-organization at the phrasal level is preceded by reorganization at the rhythmic level, we expect to find that these 9 +MP speakers would demonstrate native-like rhythmic properties in English. And given this directionality, we would also expect to find speakers with native-like rhythmic properties in English, who have not yet developed native-like phrasal main prominence. This hypothesis is tested with a forced repetition rhythm task.

III. Experiment 2: Rhythmic events

3.1 Introduction

Experiment 2 tests the hypothesis that the necessary condition for English-like prominence production is the coordination of adjacent durations into larger prosodic units. This hypothesis is tested with a repetition task experiment, where language-specific preferences for forming rhythmic units are expected to emerge in response to task demands.

Among the most characteristic distinctions of English and Spanish rhythm is vowel reduction in the former but not the latter (Dauer 1983, Hayes 1984). The pacing of vowels with varying durations yields a foot structure in English, a prosodic unit larger than the syllable, a structure that Spanish is argued not to have (Harris 1983, Roca 1997). The broad hypothesis is that there is a systematic relation between the flexibility to subdivide rhythmic intervals and the existence of an active foot oscillator. The organization into feet requires producing asymmetric syllables within a foot-sized unit, and the flexibility to subdivide the rhythmic interval can be the basis for this asymmetry.

Experiment 2 was designed to reveal and uncover the realization of the language-specific organization of monosyllabic words into larger rhythmic units. Specifically, a repetition task was used to probe the coordination of vowel durations given different task demands, a syllable condition and a foot condition, and a variation in a control parameter, speech rate.

3.2 Methodology

Participants for Experiment 2 included four ENCs from Experiment 1, eight L2ers from Experiment 1 (4+MP, 4-MP), and three monolingual Spanish speakers of Mexican Spanish.[1]

Three conditions were manipulated as part of the design: word pair, rhythmic unit, and speech rate. For word pair, four possible combinations of homophonous content and function words were used: do to, do two, due to, due two (English); de te, de té, dé te, dé té (Spanish). The word pair condition was designed to probe whether a difference in lexical category sequence across the pairs would result in a different relation between vowel durations such that for a function-content or content- function pair the durational difference between vowels is expected to be greater than in the case of pairs with words from the same lexical category. While it is expected that this might be observed in English, since vowels in function words are generally reduced, we do not expect to observe this in Spanish.

The rhythmic unit condition varied along whether participants were asked to align each word in the word pair with a single click, the syllable condition, or whether they were asked to align two words with one click, the foot condition.

The speech rate condition refers to the time between the clicks used to guide speakers' production. In the syllable condition, each click occurred 400 milliseconds (ms) apart for the first 40 repetitions (the "before" speech rate), after this point in the trial the time between clicks was decreased by 5 ms for every subsequent repetition (for a total of 80 repetitions), with the final repetitions occurring 200 ms apart (the "after" speech rate). In the foot condition, clicks occurred 600 ms apart for the first 40 repetitions in the before condition, and increased by 5 ms until reaching 400 ms apart in the after condition (a total of 80 repetitions).

3.3 Procedure

Participants were tested in two separate experiment sessions, with one week between sessions. Half of the participants participated first in the syllable condition, and half participated first in the foot condition.

There were 16 target trials, with four repetitions of each target word pair described above, and 16 filler trials for a total of 32 trials. Trials were presented in four blocks of eight trials each. There was a pause, whose length was controlled by the participant, between each block. The order of the target trials was randomized within block, and the experiment began with a filler trial, with fillers flanking each target trial.

In the syllable condition, participants were instructed that they would hear a repeating sequence of clicks and that they should align each word with a click.

[1] ENC speaker number 3 was excluded from the group data results due to deviation in behavior of when compared with the other speakers. Data from only two monolingual Spanish speakers are presented in both the individual and group results, as the data from one speaker were extremely variable, as this speaker did not seem to be reliably aligning productions with the metronome.

Both words appeared simultaneously on the screen, and a box that was synchronized with the click highlighted each word in succession. In the foot condition, participants were instructed that they should align both words with a click. The box that was synchronized with the click highlighted both words simultaneously.

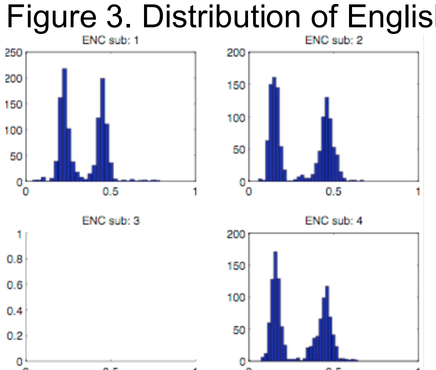
3.4 Measurements and results

Broad phonetic transcriptions were made of each trial for all participants. Forced Text Alignment was used to align these transcriptions to the speech recordings for the current experiment. Three measurements were then made on the vowels in the data set: the ratio of the duration of the first vowel in the pair (either one of "do" forms or one of the "de" forms) over the duration of the second vowel in the pair (either a "to" form or a "te" form), the duration of the interval between the release of the [d] consonant up to the release of the following [d] consonant ("DtoD"), and the duration of the interval between the release of the [d] consonant to the release of the following [t] consonant ("DtoT").

In order to examine the issue of the different ways – or lack thereof – of dividing durations into sub-intervals in English, the first step is to combine for any given subject the data from the foot and syllable conditions in order to reveal whether they have bimodal or unimodal distributions for these measures, as revealed by the proportion of a phrase cycle that is occupied by the first word. This measure is calculated with the ratio of DtoT over DtoD.

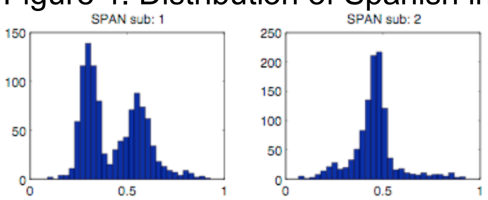
Results from the individual ENCs are shown below in Figure 3. Speakers 1, 2, and 4 clearly show a bimodal distribution, which speaks to differences between the foot and syllable conditions in terms of the proportion of first word as part of entire phrase cycle. These results show that speakers clearly can divide the interval either symmetrically (syllable condition), or asymmetrically (foot condition) by beginning the two-word interval a quarter of the way through the phrase cycle. Speakers can divide the interval between phrases either into two or into four parts. The results from these English speakers are consistent with the hypothesis that flexibility in the rhythmic domain is mirrored in flexibility in the phrasal domain.

Figure 3. Distribution of English individual speakers ratio of DtoT over DtoD



For the monolingual Spanish speakers, the prediction was that they would not exhibit the same degree of flexibility as the English speakers, and would divide the interval only into two parts. For Spanish speaker 2, the distribution is clearly unimodal, speaking to the fact they have one way of dividing the interval. However, the other Spanish speaker does appear to exhibit bimodal behavior. No speakers were pre-screened regarding musical ability, and the debriefing period after the second test session, Spanish speaker 1 mentioned that as a musician, he much enjoyed the task because of its similarity to musical entrainment techniques. It is hypothesized that in the case of this speaker his behavior was more a reflection of treating this as a musical task than as a speech task, leveraging their musical background as a strong ability to divide sub-intervals flexibility. Thus, it is reasonable to maintain that Spanish is linguistically inflexible with regards to this microscopic parameter.

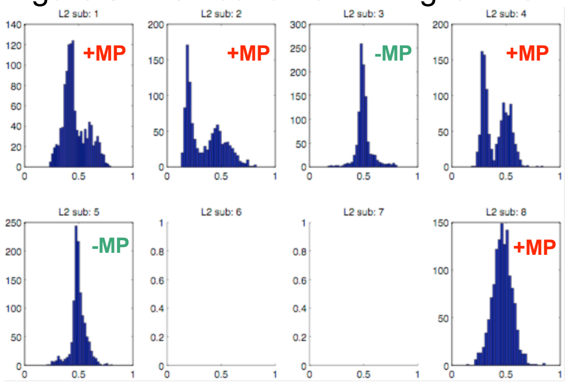
Figure 4. Distribution of Spanish individual speakers ratio of DtoT over DtoD



Results for the L2 speakers are shown in Figure 5 below.[2] Speakers 1, 2, 4 and 8 are the +MP speakers, and speakers 3 and 5 are the –MP speakers. Speakers 3 and 5 clearly exhibit unimodal distributions, while participants 1, 2, and 4 exhibit bimodal distributions. The hypothesis that acquiring this flexibility is a sufficient condition for acquiring flexibility at the phrasal prominence level receives some support here. Speakers 2 and 4 have clear bimodal distributions and are also speakers with flexible nuclear stress placement. Speakers 3 and 5 show the unimodal distribution and are among those speakers without flexible nuclear stress placement. Speaker 8 indeed demonstrated English-like MP placement but does not show a clear bimodal distribution. Even though this speaker does not exhibit bimodal distribution, the speaker clearly has flexibility in dividing the interval into a different number of ways, although not necessarily discretely in two ways. Note that the distribution of speaker 8 is wider than that of speakers 3 and 5 (which are narrowly distributed around .5), so that speaker 8 has a large proportion of observations that have values substantially less than .5. An additionally potentially relevant fact about speaker 8 is that they revealed only after the repetition task that they are a bilingual speaker of Catalan, but did not share this information at the time of the Q&A. It is possible that Catalan already has a system of prominence not based on the duration in the same way as English, which leads to a situation where it would be difficult to predict Catalan speakers' performance with regards to this task.

[2] The data for speakers 6 and 7 are not included here because in both cases there is only data from one of the tasks.

Figure 5. Distribution of L2English individual speakers ratio of DtoT over DtoD



IV. Discussion and conclusion

In English, there is flexibility at the microscopic level in dividing a repeating temporal interval in more than one way, and at the macroscopic level there is flexible phrasal prominence location. On the one hand, there seems to be a predictive relationship between these two flexibilities in acquisition, such that if you have flexibility at the microscopic level you should also have flexibility at the macroscopic level.

Relative prominence placement in English occurs both phrase-finally and phrase-internally, to the extent to which being able to place prominence in more than one location implies different ways of dividing the temporal ways where the prominence finds itself. The differential prominence placement requires dividing the temporal intervals associated with the words or syllables in different ways. The relative time at which the amount of the sentence cycle allocated to the prominent word when it is non- final is different than when it is final. Different ways of assigning prominence require dividing the sentence into temporal intervals in different ways. If the temporal interval can only be divided in two ways, as in Spanish, then it is not possible to get the patterns of discrepancy that allow for the non-final phrasal patterns.

In theory, the sentences from Experiment 1 could be divided into temporal intervals where different patterns of intervals in the final and non-final cases could be seen. This can be formalized by again returning to the value of R in the RPP parameter, which in English has a value of zero. One way of enforcing this relation between the microscopic and macroscopic phrasal potential function is to propose there is also a nonlinear dynamical attractor at the microscopic level which in turn is governed by a parameter like R, which we could call Rint (R interval). The hypothesis is that one of the determinants of RPP is the value of Rint for the potential function. There is a yoking of the control parameter in the phrasal parameter at the level of the durational interval division (the temporal subdivision) attractor. The quantitative part of working this out in detail is not pursued here. Additionally, it may be necessary to either assume that there are other variables upon which RPP can also depend, or whether the relevant parameter is not the modality but rather the percentage of observations substantially below .5.

References

Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, 51-62.

Flege, J. E. (1995). Second-language speech learning: Theory, findings and problems. In: *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*. Edited by W. Strange. Timonium, MD: York Press Inc, pp. 233-272.

Fox, A. (2000). *Prosodic Features and Prosodic Structure. The Phonology of Suprasegmentals*. Oxford University Press, Oxford.

Gafos, A. & S. Benus. (2006). The dynamics of phonological cognition. *Cognitive Science* 30, 5, 905-943.

Gussenhoven, C. (1984). *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.

Gut, U. (2003). Prosody in second language speech production: the role of the native language. *Fremdsprachen Lehren und Lernen* 32, 133-152.

Hayes, B. (1984). The phonology of rhythm in English. *Linguistic Inquiry* 15, 33–74.

Hoskins, S. (1996). A Phonetic Study of Broad and Narrow Focus in Intransitive Verb Sentences. In: *Proceedings of the Fourth International Conference on Spoken Language Processing*, October 3-6, Philadelphia, PA.

Hualde, J. I. (2007). Stress removal and stress addition in Spanish. *Journal of Portuguese Linguistics* 5.2/6.1, 59-89.

Jilka, M. (2007). Different Manifestations and Perceptions of Foreign Accent in Intonation. In: *Non-Native Prosody - Phonetic Description and Teaching Practice*. Edited by J. Trouvain & U. Gut. Mouton De Gruyter, Berlin, pp. 77 – 96.

Ladd, R. (1996). *Intonational Phonology*. University Press.

Larsen-Freeman, D., & M. H. Long. (1991) *An Introduction to Second Language Acquisition Research*. New York: Longman.

Major, R.C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mennen, I. (1999) The realisation of nucleus placement in second language intonation. *Proceedings of the InternationalCongress of Phonetic Sciences*. San Francisco, August.

Nava, E. (2010). *Connecting phrasal and rhythmic events: Evidence from second language speech*. Unpublished doctoral dissertation, University of Southern California.

Pierrehumbert, J. & J. Hirschberg (1990). The Meaning of Intonation in the Interpretation of Discourse. In: *Intentions in Communication*. Edited by P. Cohen, J. Morgan & M. Pollack. Cambridge, MA: MIT Press, pp. 271-311.

Roca, I. (1997). On the role of accent in stress systems: Spanish Evidence. In: *Issues in the Phonology and Morphology of the Major Iberian Languages.* Edited by F. Martínez-Gil & A. M. Front. Georgetown University Press, Washington D.C., pp. 618-663.

Schmerling, S. (1976). *Aspects of English sentence stress*. Austin: University of Texas Press.

Selkirk, E. (1984). *Phonology and Syntax. The Relation between Sound and Structure*. Cambridge, MA: MIT Press.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & J. Hirschberg. (1992). TOBI: A Standard for Labeling English Prosody. In: *Proceedings of the 1992 International Conference on Spoken Language Processing*. Banff, Canada, pp. 867-70.

Sosa, J. M. (1999). *La entonación del español*. Madrid: Ediciones Cátedra.

Tilsen, S. (2009). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science* 33, 839-879.

Tuller, B., Case, P., Ding, M., & , J. A. S. Kelso. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1, 3-16.

Ueyama , M. & S.-A. Jun. (1998). Focus realization in Japanese English and Korean English intonation. In: CSLI, vol. 7. Cambridge University Press, pp. 629-645.

Xu, Y., & C. X Xu. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33, 159-197.

Zubizarreta, M. L. (1998). *Prosody, focus, and word order*. Cambridge, Mass.: MIT Press.