UNIVERSIDADE FEDRAL DO RIO DE JANEIRO

Faculdade de Letras

Programa de Pós-graduação em Linguística

NATHACIA LUCENA RIBEIRO

THE PROSODY FACULTY:

a biolinguistic perspective on prosody processing

Rio de Janeiro

November, 2022

NATHACIA LUCENA RIBEIRO

THE PROSODY FACULTY:
a biolinguistic perspective on prosody processing

Dissertation presented to the Linguistics Graduate Program of Federal University of Rio de Janeiro as a requirement to PhD degree.
Advisor: Professor Aleria Cavalcante Lage
Co-advisor: Professor Andrew Ira Nevins

Rio de Janeiro
November, 2022

CIP - Catalogação na Publicação

L935t

Lucena Ribeiro, Nathacia
THE PROSODY FACULTY: a biolinguistic perspective on prosody processing / Nathacia Lucena Ribeiro. -- Rio de Janeiro, 2022.
164 f.

Orientadora: Aleria Cavalcante Lage.
Coorientador: Andrew Nevins.
Tese (doutorado) - Universidade Federal do Rio de Janeiro, Faculdade de Letras, Programa de Pós Graduação em Linguística, 2022.

1. Psycholinguistics. 2. Biolinguistics. 3. Prosody. 4. Evolution of Language. 5. Language Processing. I. Cavalcante Lage, Aleria, orient. II. Nevins, Andrew, coorient. III. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.

THE PROSODY FACULTY:

a biolinguistic perspective on prosody processing

Nathacia Lucena Ribeiro

Advisor: Professor Aleria Cavalcante Lage; Co-advisor: Professor Andrew Ira Nevins

Dissertation presented to the Linguistics Graduate Program of Federal University of Rio de Janeiro as a requirement to Ph.D. degree.

Examined by:

President, Professor Aleria Cavalcante Lage, advisor
Federal University of Rio de Janeiro

Professor Andrew Ira Nevins, co-advisor
Federal University of Rio de Janeiro

Professor Eva M. Fernández, supervisor during *Doutorado Sanduíche SWE*, at CUNY Queens College

Professor João Antonio de Moraes
Federal University of Rio de Janeiro

Professor Michele Calil dos Santos Alves
Federal University of Acre

Professor Aniela Improta França
Federal University of Rio de Janeiro

Professor Marije Soto
Federal University of Rio de Janeiro

Professor Leticia Rebollo Couto; alternate member
Federal University of Rio de Janeiro

Dr. Ana Luiza Henriques Tinoco Machado; alternate member
Federal University of Rio de Janeiro

Rio de Janeiro

November, 2022

ABSTRACT

RIBEIRO, Nathacia Lucena. **The prosody faculty**: a biolinguistic perspective on prosody processing. Rio de Janeiro, 2022. PhD dissertation, Languages College, Federal University of Rio de Janeiro, Rio de Janeiro.

The present work investigates prosody processing under the light of Generative Grammar Theory. As prosody is the melodical part of speech, its organization maps information of different kinds into intonational content. Thus, processing a sentence that we hear involves auditory input processing (see Fernández & Cairns, 2010), in which it will be determined if the sound is a speech sound; and later it involves prosody processing, in which the intonational contour will bootstrap language processing. When we refer to Language Processing, we are referring to both Faculty of Language Narrow - FLN and Faculty of Language Broad - FLB (Hauser, Chomsky, and Fitch, 2002), as Prosody maps Syntax, Pragmatics and extralinguistic content.

From a psycholinguistic perspective, we aim to understand and explain how Prosody Processing is and its role in Language Processing. A total of three experiments were conducted. The first one is intended to test syntax-prosody interface. It is a perception test with a forced choice task, in which the participants listened to delexicalized sentences and then had to decide which sentence they had listened to. This test we ran both in Brazilian Portuguese and American English. The results showed that the perception is successful and the prosodic mark is enough to perceive syntactic structure to coordination or to attachment. We also noticed that when we have the mark on the subject, it is harder to process, probably because of a combination of recency effect and subject-object asymmetry.

In the third and final experiment, we investigated the Pragmatics-Prosody interface, in order to understand the efficiency of prosody mapping deixis. The data shows that prosody markers are efficient in triggering the expected pragmatic interpretation.

Keywords: Prosody Processing; Prosody Domain; Generative Grammar; Psycholinguistics.

Rio de Janeiro

November, 2022

RESUMO

RIBEIRO, Nathacia Lucena. **A faculdade de prosódia**: uma perpectiva biolinguística do processamento prosódico. Rio de Janeiro, 2022. PhD dissertation, Languages College, Federal University of Rio de Janeiro, Rio de Janeiro.

O presente trabalho investiga o processamento prosódico à luz da Teoria Gerativa. Considerando a Prosódia como a parte melódica da fala, a forma como ela se estrutura mapeia diferentes tipos de informação em conteúdo entonacional. Dessa forma, processar uma sentença que ouvimos envolve processamento do sinal auditivo (ver Fernández & Cairns, 2010), durante o qual se determina se um som é sinal de fala ou não; e envolve posteriormente processamento prosódico, durante o qual a estrutura entonacional vai desencadear o processamento linguístico. Quando nos referimos a Processamento Linguístico, nos referimos tanto à Faculdade da Linguagem Estreita quanto à Faculdade da Linguagem Ampla (Hauser, Chomsky and Fitch, 2002), uma vez que a Prosódia mapeia estrutura sintática, conteúdo pragmático e informação extralinguística.

Da perspectiva da Psicolinguística, buscamos entender e explicar como é o Processamento Prosódico e qual o seu papel no Processamento de Linguagem. Um total de três experimentos foram implementados. O primeiro tem o objetivo de investigar a interface Sintaxe-Prosódia. Trata-se de um teste de percepção com tarefa de escolha forçada, no qual os participantes ouviram sentenças deslexicalizadas e foram instruídos a decidir qual sentença tinha sido ouvida. Este experimento foi rodado tanto em português do Brasil quanto em inglês Americano. Os resultados mostraram que a percepção foi bem sucedida e que a marcação prosódica para coordenação ou para encaixe é suficiente para a identificação da estrutura sintática. Além disso, é possível observar que quando a marcação está localizada no sujeito da sentença, o processamento é mais custoso, provavelmente devido a uma combinação de dois fatores: o efeito de recência e a assimetria sujeito-objeto.

O segundo experimento investiga com a interface Pragmática-Prosódia, com o objetivo de entender qual é a eficiência da Prosódia no mapeamento de informação dêitica. Os dados mostraram que as marcas prosódicas são eficientes em recuperar a informação pragmática.

Palavras-chave: Processamento Prosódico; Domínio Prosódico; Teoria Gerativa; Psicolinguística.

Rio de Janeiro

Novembro, 2022

I dedicate the present work to my sons, Ryan and Ethan, as they are the very reason why I kept going, with God's help, even when I had no strength left.

ACKNOLEDGEMENTS

I would like to first, and always, thank my God and Father, for He is the reason I made it this far.

I thank my husband, for his full support, and my sons, Ryan and Ethan, that encourage me to keep accomplishing life.

I thank my parents and my siblings, as they were always a great support in many ways,

I thank my advisor Professor Aleria Lage, for all these years of mentorship and guidance.

I thank my co-advisor, Professor Andrew Nevins, for so many enlightening talks that steward my research in the right direction so many times.

I thank Professor Eva Fernández for hosting me in the US and for her crucial mentorship during my stay.

I thank my American informants TP and RL, for helping me editing and recording my experimental sentences.

I thank my lab fellows, who helped me run my experiments in Brazil.

SUMARY

GRAPHS

TABLES

FIGURES

1 INTRODUCTION

Human beings have a highly complex language. And understanding what makes this capacity as unique as it is has been an endeavor for over 60 years now. The link between language and speech has been widely discussed in the field. However, the place of Prosody in human beings' language capacity is still very unclear. Whether Prosody is a speech domain or language domain, or even its own domain, is not settled. If we turn to evolutionary history in order to understand human beings' language capacity and cognition, pursuing a better understanding of Prosody, we are faced with a lot of challenges. One of the biggest challenges is that the organs of speech don't fossilize (Fink, 1963; Hauser, Chomsky and Fitch, 2002). Therefore, it is hard to pinpoint exactly where language, or at least speech, has started. Furthermore, it makes extremely hard to determine how it started, and what were the innovations involved. Uomini and Meyer (2013) propose that language would have emerged around the same time as the technology for Acheulean toolmaking, complex tools forged by striking stones against each other, 2.5 million years ago (Gibson, 1993 *apud* Uomini and Meyer, 2013; Roche *et al.,* 1999 *apud* Uomini and Meyer, 2013). According to them, language and toolmaking would both need structured and hierarchical action plans. These action plans would be the results of similar cognitive processes and overlapping neural circuits, meaning that to some extent, we would use partially the same neural circuits to both plan tool making and compute sentence structure. They found relevance in the overlapping neural activation during a functional transcranial Doppler ultrasonography (fTCD) test, where participants would engage in toolmaking and word generation. This would endorse their hypothesis that the network for complex action planning might be the innovation that led to both language and toolmaking. However, there is no evidence that helps us understand how language cognition was first structured, and what aspects of language first emerged. For example, have interjections emerged first as meaningless shouts that were later paired with lexical meaning such as "Behold!" "Food!" "Watch out!" etc., or have they had complex sentence structure from the start and what emerged later was our capacity to process and use the structure creatively?

The linguistic signal encodes highly complex linguistic structure (Kirby, 2007; Fernández and Cairns, 2010). The prosodic features of a sentence will dictate how the sentence will be interpreted, interfering with the recovery of its syntax structure (Wagner, 2004, 2015) and its propositional structure (Büring, 2016) by the listener. When we listen to an ambiguous sentence as (1), out of a context that helps semantic interpretation, the only clue we have about the syntactic structure is prosody organization:

(1) The girl who is a student at this college appeared on a TV show

If we have a pause (break) after *girl*, signalizing Intonational Phrase (IP) boundary, [the girl] is a topic followed by an appositive expression. If we have no strong boundary between [the girl] and the who-sentence, it is a relative clause embedded in a Determiner Phrase (DP).

This is a simple example of what many studies have shown: prosody seems to map syntax into intonational content (see Ladd, 1986, 1988; Christophe, 1997, 2008; Ribeiro, 2015a, 2015b; Ribeiro and Lage, 2015). Therefore, there is a prosody processing deeply linked to syntax and its distribution, hierarchy, relations, and operations.

We can also imply from prosodic realization many pragmatic information. By the way an intonational curve is performed, a sentence as (2) can be either a statement, or a request, or an order:

(2) I want more

The listener may distinguish which speech act is being performed by the IP organization, the tone targets, and the segments duration. This is an example of how prosody may carry pragmatic and extralinguistic information. Many studies show prosody processing strongly linked to pragmatic and extralinguistic factors, such as emotion, mental state, and empathy (Moraes & Collamarco, 2007).

The examples above reveal that prosody processing is a crucial and indispensable part of sentence processing. We assume that a better understanding of what was the role of prosody in the emergence of modern human language with the *homo sapiens*, how it was affected by this emergence, what is its role in sentence processing, and to what extent it is exclusively human is the key to have a better understanding of human language faculty itself. That is because we assume prosody is the exact interface between modern human complex language faculty and what remains from human primitive communication system. In other words, we assume that prosody is a module within language modular system. To investigate that claim, we are basing our studies in the generative approach to language theory (Chomsky, 1965, 1981a, 2014), within a cognitive perspective.

For the present work, our objective is to investigate the prosodic aspects of language cognition in order to understand its place in language architecture. Our prediction is that Prosody is a specific, independent domain, withing the language system, processed by its dedicated module.

The first part of this dissertation is dedicated to describing our perspective of prosody, and it is divided into five chapters. On 2.1 we will describe prosody as the innate capacity it is, alongside the faculty of language innateness. On 2.2 we will describe cognitively what it means to understand language as a modular system, and why prosody should be considered as one of its modules. On 2.3 we will look into the architecture of language and language design as proposed by Chomsky (1965, 1981a, 2014), observing what is prosody's role in language architecture. On 2.4 we make a review on acoustics and physical aspects of language signal that are relevant to prosody processing. On 2.5 we advocate for prosody being a module and its own architecture.

The second part of this dissertation is dedicated to reporting the experiments run in an effort to test some of our hypotheses.

2 WHAT IS PROSODY?

Many have attempted to define prosody, but, as it is with language studies in general, the task is doomed to fail if one doesn't take a theoretical approach and an adequate perspective. A complex system cannot be explained by the simple description of the properties of its elementary components, as they are described at their own level (see Simon, 1962). To describe a complex system, it is needed an understanding of different levels, in a sense that they belong cohesively to a whole, even if they can't be linked to each other in complete detail.

The problem seems to be the difficulty scientists must delimitate clearly, throughout history, what is prosody and what is not, when looking to a linguistic event. It is hard to define an object whose boundaries you can't see. Although this is an old and intrinsic problem in Linguistics (see Saussure, 2008), the difficulty with prosody gets worse because of the abstractness of prosodic units[1]. Despite how challenging it is, we have countless attempts to answer it. Ladd (2014) argues that one of the hardest tasks in defining prosody is finding an agreement among the many definitions available, a converging point, through which we would be able to see the object *Prosody*, *per se*, in raw. To reinforce the problem around the many definitions to prosody, he cites the revies of Crystal (Crystal, 1992, *apud* Ladd, 2014: 65-66):

> "Prosody: Variation in pitch, loudness, *tempo* and rhythm, as encountered in any use of spoken language (…); also called prosodic features, and in phonemics analyzed in terms of prosodemes. In generative phonology, prosodic features are one of the main dimensions of speech sound classification. In Metrical Phonology, one of the levels of structure in a metrical tree is the prosodic level. The canonical pattern of segments in a form is a prosodic template. In Prosodic Phonology, a prosody is a feature extending over a stretch of utterance (such as the syllable or sentence), contrasting with the segmental notion of phonematic units."

In face of such complexity, Wagner (2015) will say that it is probably easier to define prosody by listing what it is not. Following his direction, prosody is not the set of properties and rules that regulates phonemes; it is not part of syntactic structure; it is not organized in inventories as lexical entries are (cf. Xu, 2015); it is not something whose characteristics are specific to each society, as Pragmatics issues; it is not propositional structure only; it is also not bare articulation.

[1] Xu (2015) makes a point in pointing out a major problem with the impossibility of representing orthographically the prosodic units. That increases the volatility and harms any attempt of unified definition.

Prosody is a multidimensional object of study and, as such, it needs to be treated as a complex object under the classic understanding Linguistics inherited from Saussure (2008): it is not a given object; it is the viewpoint that creates the object.

As Fonseca (2012) points out, it was a necessity to prosody studies to have different approaches and points of view advancing separately in order to better understand and properly describe language and prosody, but it is time to cross the fields and approach an integrated way of doing research.

One of the possible problems we find in common among the many definitions to prosody, besides oversimplification or lack of integration, is that, in Linguistics, we look to Prosody through a language perspective. We deal with it as it is another level of language structure, such as Syntax or Phonology. It may be limiting the scientific observation and description of it. It is clear that Prosody is related to Phonetics, as it has acoustic and articulatory properties regulated by language structure, and to Phonology, as it has rules and inner structure dictated by language system. It is also largely sustained that Prosody is interacting with Syntax (see Wagner 2010, Fonseca 2012, Ribeiro 2012), to Semantics (Büring, 2016), and to Pragmatics (Moraes, 2011a, 2011b). But what are we left with if we take off all this strictly linguistic and propositional information from Prosody? If we go ahead and think of Prosody Processing, the definition of Prosody gets even blurrier since Prosody Processing seems to be in interface with Faculty of Language Narrow - FLN, with Faculty of Language Broad - FLB and its interfaces, with the limbic system (for emotional mapping and state of mind mapping), and with the motor and sensory cortices (for output and input, respectively). What are the characteristics of bare prosody?

When we hum or babble a song that we don't know the lyrics, or a sentence in a language we don't speak, or when we see kids that don't talk yet mimicking a conversation, all we have available is prosody. We are able to replicate part of what we hear by replicating the rhythm, the tunes (or more specifically, tone variation and contrast), its pace, and the quality with which it is produced. Accordingly, prosody resembles music. In fact, since the origin of the term *prosōdia* (see Ladd, 2014), prosody has been associated with music. When we think of the object Prosody itself in an opaque mode to the other levels of language structure, it looks like music, in a sense that it has Rhythm and Harmony, not to mention pace (*tempo*), quality (tone) and volume (loudness/ intensity). Therefore, we can say that Prosody is the Melody of Language. And by doing that we can relate it to both the structural (phonological) part of it (as a Music Sheet) and the actual performance (phonetic) of it (as in a concert).

Finally, it is imperative for us to understand better prosody processing, for we cannot forget that, evolutionarily, language comes from our ancestors' communication systems. Prosody being the closest structural system we probably have to non-human animal communication system; it presents strong relations with language and communication as well.

Chomsky's theory doesn't approach in detail what happens to linguistic information once syntax is processed, dealing only with idea that a link is needed between syntax structure and performance. Prosody seems to be that link.

2.1 AN INNATE CAPACITY

"I talk, therefore I am." – Lieberman (2012:620)

Humans have an innate capacity to develop language. Our language system, however, seems a lot more complex than any non-human animal's system of communication. What makes it different? And how?

In this chapter we are going to review and discuss what makes human language unique amidst all non-human animals' communication system. We are also going to review what we know on how our language evolved and how we develop this complex innate capacity. Additionally, we are going to discuss the role of prosody in enabling our language complexity.

2.1.1 The language-communication dichotomy

It is undeniable that human language and communication are deeply linked, to the point that people often take them to be the same thing. We argue against or in favor of something, we speak up when we want to communicate something, we text or call people we need to contact and communicate with. However, it doesn't take specialized observation to notice that we have communication beyond human language, and language goes beyond communication, meaning that, although they are intertwined, they are two different capacities.

Communication is the capacity to convey information[2] to others, in an expressive way (Hauser, 1996). However, as a system, communication may present alternate designs. For example, the way we dress ourselves makes an impression on people because it communicates about our interests and intentions. Our facial expression and body posture also communicate emotions and intentions. On the other hand, language is an innate capacity to develop a dedicated cognitive system to verbalize and manage thoughts and generate representative symbols and categories. For example, we can be doing math exercises, in silence, alone at home, and still use language in our minds ("two plus two equals four"), sometimes even speaking out loud. There's no communicative need in that, if we take into consideration that a communicative event involves exchange of information that the speaker has, and that the

[2] There is an interesting discussion in Buckland (1991) around three different definitions for *information*. He argues that *information-as-process* is the act of changing what someone else knows; *information-as-knowledge* is what is the knowledge conveyed when changing what someone else knows; and *information-as-thing* is a resource that is intrinsically informative such as data or a document. When we argue here that communication conveys information, we are leaning toward the *information-as-knowledge*.

speaker believes the listener doesn't have. In Grice's (1975) term, the maxim of relevance is usually respected in a conversation. If we can have one without the other, we may assume that language and communication are independent.

The discussion around the relationship between language[3] and communication[4] is extensive. Researchers (see Hauser, Chomsky, and Fitch, 2002) have questioned whether the ultimate reason for language is communication and whether it has evolved from and for it (see next section for further discussion). Although this dissertation assumes that language evolved for reasons other than communication needs, I'm far from attempting to answer to this matter. It is necessary, though, to make clear four premises in order to go forward with no misunderstanding:

i) Communication is a function of language

It is important to understand that, as communication and language are two independent capacities, language uses our communicative capacity as a functionality, one of many. In other words, language can be used to communicate something, but it can be used in other situations, motivated by other necessities. Also, language goes beyond mapping and output mechanisms that it uses when attempting to communicate.

ii) We can assess language characteristics through communication

Every time language is used to communicate something to someone, it carries information from the language system as well. Observing and analyzing this data can make information about syntax structure, lexical paradigm, morphology, etc., evident. However, in order to study the system, mere observation is not enough to fully describe and understand the abstract system, as learning about sentences the system doesn't license is more informative than counting on the positive evidence. For the generative system that language is, the number of sentences the system license is infinite.

iii) Communication can represent constraint to linguistic output

From Grice's Maxims to deixis indentation, communicative needs can veto the licensing of a perfectly grammatical sentence. Therefore, it is important to consider communication when studying certain aspects of language.

[3] From now forward, any time we use the term language, we mean human language.

[4] As the term before, any time we use the term communication, from now forward, we mean human communication.

iv) Prosody structure is regulated by both language system and communication

As we will see in 2.5, prosodic structure has many layers. Some layers are satisfying linguistic demands, others are fulfilling communicative needs. In all of them we also have acoustics restrictions operating.

We will be touching, then, both realms: language system and communication system, without confusing them with one another. That is imperative for us to understand better Prosody processing, for we cannot forget that, evolutionarily, language comes from our ancestors' communication systems. Prosody being the closest structural system we probably have to non-human animal communication system, it presents strong relations with language, and communication as well.

2.1.2 Prosody and the Evolution of Language

Communication evolved. Not in a figurative way, but it truly evolved alongside human beings' evolution throughout history. Each animal species has its own communication system, that is effective amongst their own, and it is, at least partially, opaque or meaningless to other species representatives. It seems obvious, and it is, to acknowledge that, if the species evolved, so did their communication systems, but this is far from being an exhausted issue when it comes to the study of human language. That simple first sentence carries so much more questions with it than answers: What did communication evolved for? What parts of it, if not all, are results of adaptation through natural selection? Are there similarities between human language and other systems of communication? Are those similarities evolutionary traits? When in the course of history things changed in such a way that what we know as the human language happens? I do not attempt to answer all the possible questions, but to discuss a little further on the evolutionary history and distinction of our human language system to animal's communication systems, and what role Prosody plays in it.

2.1.2.1 The uniqueness of human language

Among several theories in Biolinguistics field of research, one of the most extensively tested states that human language is a unique system different from all other animal's systems of communication, starting by the simple fact that, although it serves communication, as we have seen in the previous section, it is not just a system of communication (Hauser, Chomsky and Fitch, 2002; Fitch, Hauser, Chomsky, 2005). The hypothesis is that language didn't evolve to its current state because of, or for, communication. In the contrary, communication as we know was facilitated by language, a result from language complex system and its operation in the human mind. We do not need oral or sign language to communicate, but we most likely need language to think, and that is one of the key differences between human language and other animal systems of communication, according to this proposal.

It may seem that one of our greatest capacities in language is the ability to pair information (meaning) and representation (sound), or in Saussure's terms (Saussure, 2008), "signified" and "signifier", respectively, through categorization and mapping. When we listen to the word *dog*, we immediately associate it with a concept of *dog* – a four-legged pet, that barks –, that is, we have an abstract category in our mind that meets all the possible dogs in the world by matching the abstract features that make a dog distinct from all other animals, and an abstract[5] matching sound, that evokes this category. However, mapping and categorization are not something exclusive to our system; hence, not what makes our system unique. There is vast research on how non-human animals can establish and recognize categories, even categories of human language (see Kuhl and Miller, 1975; Savage-Rumbaugh, Rumbaugh, and Boysen, 1980; Berwick et al., 2011). Also, within their own communication systems, non-human animals have specific signals for specific information. It may not have human language's level of abstraction, but it still requires mapping between signal perception or production and information. The difference is in what we do with this ability and in how it works.

Human language appears to have three exclusive features that, once combined, creates a limitless complex system. The first one is what we call language creativity. It has its links with general creativity in a sense that we may formulate sentences on any issue we could ever think of. It doesn't matter if it is fantasy or real life, present, past or future; our system is able to project different worlds and times by using specific grammatical resources, and that originates from our cognitive capacity for abstraction. Language creativity refers to the capacity human language must create an infinite number of products (sentences) from a finite set of materials (phonemes, morphemes, and words). We can observe and list bird songs, but the list isn't endless, nor its use by the bird community. The same way, we can teach apes words (Savage-Rumbaugh, Rumbaugh, and McDonald, 1985), but we cannot teach them its

[5] Saussure (2008) refers to signifier as an abstract sound because he talked about of the idea of the sound we have in our mind before it is physically articulated, or after it has been received by our ear and translated into electric signal to our brain.

endless use. In other words, none known non-human animal's system of communication makes infinite use of its finite means.

What enables this kind of creativity is a mechanism called *merge*, the second exclusive characteristic of human language. Merge is a syntactic operation, a mechanism, that takes two objects and projects a third one. There are two types of recursive structures built by merge (Hollebrandse & Roeper, 2014). Direct recursion is simpler, has only one level, that creates coordination of terms:

(3) [I ate [[pie] and [a banana]]]

And the indirect recursion is more complex and creates embedding of terms:

(4) [I ate [banana [pie]]]

The latter appears to be exclusively human[6]. *Recursion* is a mechanism that allows multiple nesting embedding of propositions and/or similar structures:

(5) [[The aunt [of the friend [of the mayor [of the city [of the actor [of the [tv show [on ABC]] [I told you about]]]]]]] is my godmother]

Through recursion, a grammar can potentially built infinite sentences. However, recursive structure, especially indirect recursion, is constrained by working memory, both to production and to comprehension. This mechanism, despite some controversy (cf. Everett, 2005, 2009; Nevins, Pesetsky, Rodrigues, 2009a, 2009b), is found in every known human language and not found in any animal's system of communication.

All these unique features are relevant to study Prosody, as prosodic structure will map syntactic structure (Wagner, 2015). Observe (6):

(6) My sister, who is a biomedical researcher, is moving to the USA

When prosody processing receives as an input a recursive syntactic structure as (6), it generates a recursive prosodic structure as well (Ladd, 1986). Ribeiro (2015a) found out that partial declination reset (Ladd, 1986, 1988) marks Prepositional Phrases (PPs) attachment by coordination, i.e., direct recursion, and that the absence of partial declination reset marks embedding Prepositional Phrases (PPs) attachment, i.e., indirect recursion.

Key computational capacities such as merge seem to have evolved not specifically because of communication (Hauser, Chomsky and Fitch, 2002). The operation of merge is clearly present in many other aspects of cognition and in the general operation of the mind. But when these capacities served the new faculty of language in the new evolutionary state,

[6] For a discussion on recursive structures in birdsongs see Bolhuis, Okanoya, Scharff (2010) and Berwick *et al.* (2011a).

they may have been remodeled by constraints imposed by language computation and communication systems (Hauser, Chomsky and Fitch, 2002).

In order to understand human language's innate capacity and how Prosody contributes to it, it is productive to investigate to some extent the evolutionary history of language and communication. In order to do that, let's also review a few basic concepts and premises to evolutionary studies in Linguistics.

2.1.2.2 When did the vocal communication start?

It is important to understand language evolution on three fronts: the signal, or speech; the structure, or syntax; and the meaning, or semantics. This separation is the minimum necessary to properly investigate specific mechanisms that might have different phylogenetic and functional history (Hauser, Chomsky and Fitch, 2002). When we look to the whole language system we have now, we may be tricked into seeing the more superficial features functioning for communicative means and take the naïve step of assuming that it all evolved together for communicative purposes. However, a complex system will go through small specific changes that at some point come all together to function as a whole new thing. Therefore, by dividing fronts, we can investigate more effectively how the mechanisms underlying language capacity differ from those underlying animals' systems of communication. Identifying what is unique and what is shared between human language and non-human animal systems of communication is crucial to evolutionary studies (Hauser, Chomsky and Fitch, 2002), because the determination of those will help pinpoint key innovations[7] that have happened during evolution, and it will help us to understand what forces are playing a role in selecting the innovations to go forward in the evolutionary line (Fitch, 2002). As the object of investigation for this work is Prosody, we are going to focus on the evolution of speech.

Language is thought. And speech refers to a signaling system that is orally articulated, used for communication of human language. It seems that human speech system has two unique characteristics: vocal imitation and reconfigured vocal tract (Fitch, 2002). Speech is considered important as our default means of communication not only because of its

[7] Key innovations (Liem, 1973) are distinguishing qualities or characteristics that play crucial role in natural selection.

frequency of use, but also because our closest living ancestors in the evolutionary track, that is, primates, cannot produce it (Janik and Slater, 1997; Kako, 1999; Fitch, 2002, 2005).

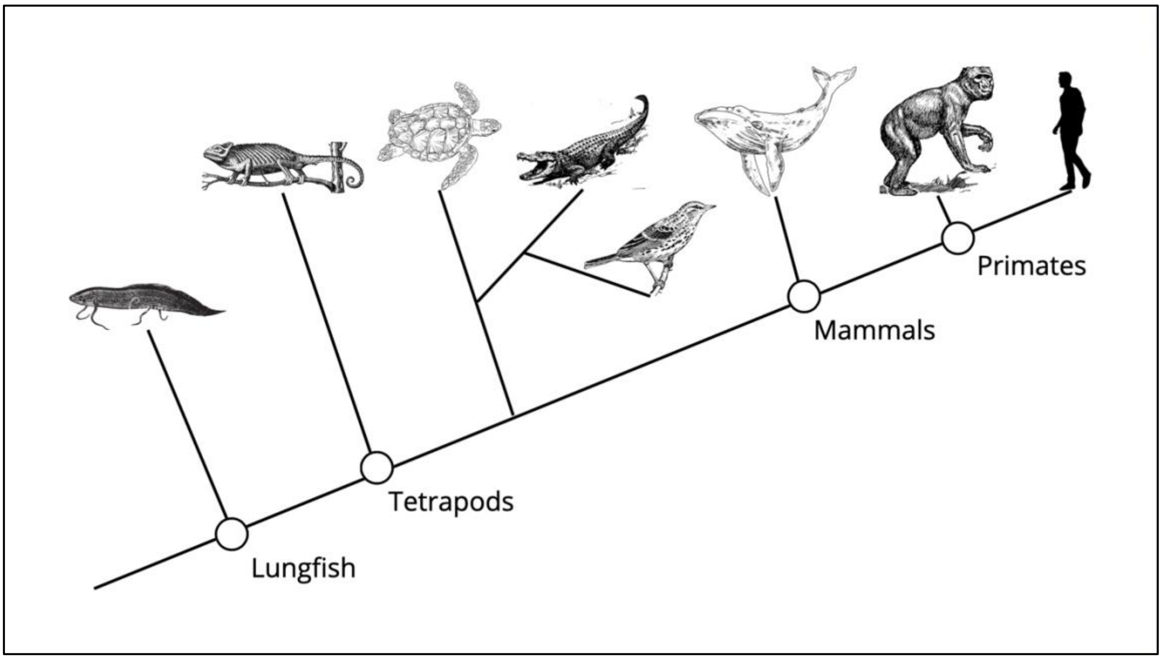


Figure 1: Simplified branch of clades important to speech evolution

In Figure 1, there's a simplified representation of some of the important *clades* to speech evolution. Clades are groups of living and extinct organism that are related by a common ancestor (Lawrenece, 2005). Observing clades distribution is important in order to understand if a trait[8] being observed evolved from a common ancestor or if it is an innovation. We call homologous trait if it is being observed in species of the same clade – without missing nodes (Darwin's, 1859 *apud* Boyden, 1947; Wagner, 1989). For example, all mammals nurse their children with milk. This trait is homologous in humans, chimpanzees, and dogs for example. A trait is analogous (or convergent) if it is observed in species of different clades (Darwin, 1859 *apud* Boyden, 1947). For example, musicality in birds, whales, and humans is an analogous trait as birds belong to a different clade than mammals, and as between whales and humans we have nodes in which this trait is not observed (e.g., primates). Homologous traits help us to look back in the evolutionary line and infer characteristics about common extinct ancestors. Analogous traits help us to understand selective forces and constraints that might have operated in the evolutionary line. And key innovations are traits that end up transforming the selective pressure, or their action, in an evolutionary lineage (Liem, 1973).

[8] Traits are specific characteristics of an organism, that can be determined genetically or environmentally (Lawrenece, 2005).

A large variety of species communicate through multiple types of signal, such as body language and facial expression, similarly to humans. Yet, vocal communication is what apparently links our language to non-human animals' system of (vocal) communication (Fitch, 2002).

When it comes to vertebrates, we have strong evidence that led us to believe that vocal communication goes back to over 400 million years ago, with the beginning of *bony fish*. The *Osteichthyes*, the bony fish, includes two classes: *Sarcopterygii* and *Actinopterygii*. The sarcopterygians, also known as lobe-finned fish, are bony fish that originated lungfish and tetrapods (later, amphibians, reptiles, birds, and mammals). Although there is record of lungfish's vocalization in the 19th century, it is an under investigated matter, with no robust evidence. *Coelacanth*s have been studied for long then two decades with no evidence of vocalization (Bass and Rice, 2018).

The *Actinopterygii*, the other aforementioned class within bony fishes, also known as ray-finned fishes, have fins that are webs of skin and by bony spines (rays). There is strong evidence of vocalization within the living ray-finned fish, mostly by using sonic muscles pressuring their swim bladders. We have register of both broad-band sounds and harmonic sounds, both used in social context, such as danger warning, and mating (Bass, Rice and Feng, 2019).

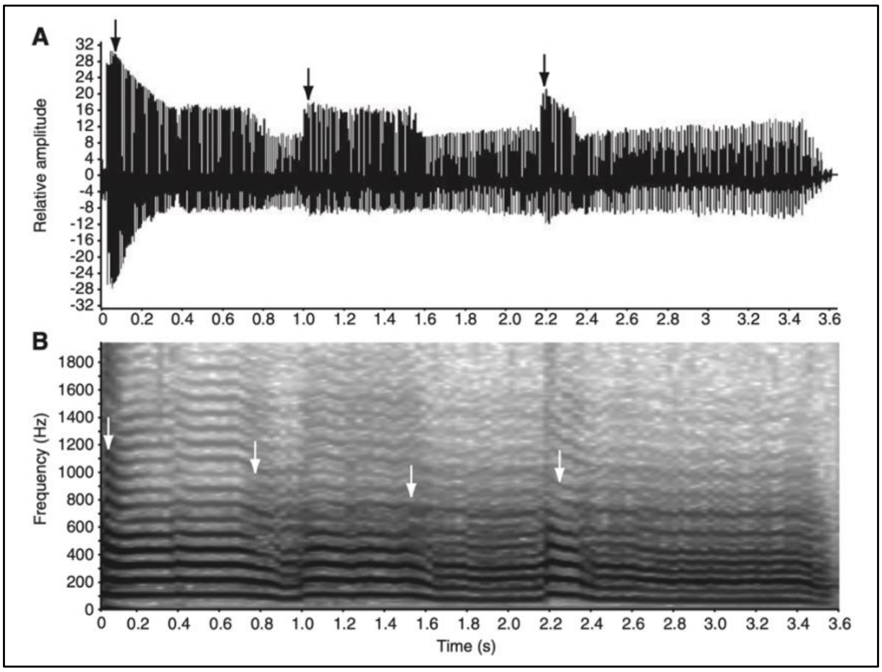


Figure 2: Patterns of modulation in the growl call of the Plainfin Midshipman fish, Porichthys notatus, an actinopterygian. Source: Bass, A.H., Rice, A.N., 2010. Vocal-acoustic communication in fishes: Neuroethology. In: Breed, M.D., Moore, J. (Eds.), Encyclopedia of Animal Behavior. Vol. 3, pp. 558–567.

There's also registered evidence of vocal communication in ancestral lineages of ray-finned fish (Bass, Rice and Feng, 2019), adding the irrefutable evidence of vocal communication in land vertebrates, and evidence in lungfish vocal communication, even though not vastly explored, that indicates that vocal communication is an ancient component of bony fishes as a clade. The strong hypothesis is that the development of a central nervous system was the key innovation necessary to bootstrap vocal communication (Bass, Rice and Feng, 2019).

Going back to the sarcopterygians, the subclade Tetrapodomorph is believed to host the first hybrid water-land vertebrate, that would evolve later to tetrapods' subclade (Clack, 2009). Regarding vocalization, the new propagation media, the air, seems to have potentialized the spread of the acoustic signal and consequently the continuous increase of vocal communication on land vertebrates. Besides that, the increasing use of vocal communication appears to be deeply linked to nocturnal activity, when visual signals are less efficient (Chen and Wiens, 2020).

Within land vertebrates, there's evidence of very early acoustic communication with *Anuras*, an amphibian order that contemplates adult frogs that don't have tails – among other physiological specificities (Chen and Wiens, 2020).

2.1.2.3 When did speech start?

We today know a lot about apes' capacity of learning our language system: They are great with learning words; they can put sentence-like together; they pair meaning and sound; they pair even meaning and visual symbols (Savage-Rumbaugh, Rumbaugh and McDonald, 1985; Lyn, Greenfield and Savage-Rumbaugh, 2011). However, it was never possible to teach them how to talk (Fitch, 2002). Speech is supposed to be the closest capacity we have to apes' original system of communication, and yet there is a gap that keeps them from it. We can teach some birds how to mimic speech (Patterson *et al.*, 1997), but even chimpanzees that were raised fully by human families weren't able to develop speech (Fitch, 2002). They have their own oral production that serves their communication system, yet that is not enough to enable them to produce, or imitate, speech. Vocal learning/ imitation is a behavior that we can find in a few mammals and birds (Tchernichovski *et al.*, 2001; Fitch, 2002; Studdert-Kennedy, 2002; Knörnschild *et al.*, 2010; Balsby, Momberg, Dabelsteen, 2012; Mercado, Mantell, Pfordresher, 2014; Roffman *et al.*, 2015). In humans, its crucial role seems to be enabling the extensive Lexicon every language has (Fitch, Huber, Bugnyar, 2010). In some extent, to birds it plays a similar role, as it enables song repertoire and tribe culturalization (Fitch, 2002, 2005; Mercado, Mantel, Pfordresher, 2014).

The distinction between language and speech comes a long way in Linguistic Science. The father of Linguistics, Saussure, was probably the first one to make this distinction relevant to the study (Saussure, 2008). When we say speech, we are talking about the auditory/ vocal medium that humans predominantly use to convey language. So, yes, it is, in some sense, completely distinct from language, our computational and processual system. That, though, does not mean that speech did not evolve alongside with everything else. That does not mean either that Prosody and speech are the same thing. We agree that the evolution of language has been independent of communicative needs (Hauser, Chomsky, Fitch, 2002; Fitch, Hauser, Chomsky, 2005), but the evolution of speech seems, at least partially, closely linked to sound production and perception, which are deeply linked to communication. A robust hypothesis is to consider Prosody as the domain at the language and speech interface. That is because what we see in Prosody is both restrictions and limitations, or even alterations and system predictions, based in the articulatory and perceptual systems, but also the complexity of its structure, coming from the language system. While the evolution of language brought forth a complex system of structures and concepts, the evolution of speech had to bring forward vocalizations that were adequate to such complexity, that involves our vocal learning capacity, our vocal tract phonetic range, and possibly our perception systems. We propose that prosody is the linking process between them.

Although finding similarities between human language and non-human animal's systems of communication is not an easy task – if it is even possible –, finding similarities between human speech and non-human vocalizations appears to be much more productive. We can find in our ancestors some abilities that were shown to be required to the current speech capacity later in the evolutionary line. One of the main distinctions we see from apes' vocalizations is human use of varied and numerous formants[9] in the vocal tract (Fitch, 2002). Lieberman (2012) proposes that this wide range of formants in the human vocal tract was enabled by the lowering of the larynx, alongside with sophisticated motor control.

The proposal that lowering of the larynx was central in evolution of speech has been challenged new data showing that some non-human mammals lower their larynx during vocalization, indicating that it is actually a primitive trait (Fitch & Reby, 2001). Also, some

[9] Formants are the resonant frequencies produce in the vocal tract. They are generated by resonating the harmonic produced by the vocal folds. Their frequencies are determined by articulation, that modifies the shape, length and height of the vocal tract. See 2.4 for further explanation.

other species have been discovered to have a permanently lowered larynx, as koalas (Sonntag, 1921 *apud* Fitch & Reby, 2001), indicating it as an analogous trait to human's, even as apes don't have it (Fitch, 2002, 2005). As these other species do not use their lower larynx to vocalize, it is very probable that other selective forces may have operated in the lower larynx trait. And that might have played as a preadaptation to the expanded phonetic ability explored later.

In regard of communication ability, highly trained apes have been successful in learning and using for communication (not orally) above 100 words. Bonobos exposed to heavy training and reinforcement has shown success in perceiving multiword spoken utterances, as well as some basic syntax, such as word order (Savage-Rumbaugh *et al.*, 1993). No more complex structure has been observed to have been achieved, however (Kako, 1999).

Apparently, primates do have some sort of primitive syntax in their own communication system, as they produce sequence of calls and the way these sequences are made relevant and perceived by the listeners (Bergman *et al.*, 2003; Seyfarth *et al.*, 2005). There is no evidence, though, of any sort of compositional structure; and most importantly no evidence of recursive structure. Those crucial aspects of human grammar, alongside the ability of perceiving anaphora, that are learned naturally, are very difficult or even impossible for other primates to master (Fitch, 2005). This might also play a role in perceiving complex prosody structure.

Speech has a complex organization as, although it is linear, it is composed of chunks with inner structure. Supposedly, the precursor of articulatory abilities that enable structures like a syllable is mandibular movement as chewing and sucking (Fitch, 2002). The movements involved would serve as a frame for the set of movements involved in the structure like this. This alone doesn't explain apes' inability to articulate our syllables, especially when compared to parrot's ability.

When we look for vocal imitative ability in the evolutionary lineage, the distribution of occurrence we find is odd and sparse, which indicates multiple convergent evolution (Nottebohm, 1976). Although non-human primates seem to struggle with this ability, we have other animals as birds, seals, and bats displaying high skills on it (Janik and Slater, 1997). As we discussed above, the number of hypotheses on what constrain non-human primates' ability to vocal imitation has risen. But despite strongly accepted in the past that traits like lower larynx would play a crucial role in this inability, the presence of lower larynx in other species have put this hypothesis in perspective. The strong hypothesis now relies on neural basis: vocal imitation would require high voluntary motor control of vocal tract and the ability to link auditory input to motor abstract representations of how to produce it (Jürgens, 1998; Lieberman, 2012). One point still up for debate, though, is whether vocal imitation is a domain-specific or domain-general ability (Donald, 1991; Moore, 1992).

MacLarnon and Hewitt (1999) *apud* Fitch (2000) measured the diameter of the thoracic vertebral canal in different primates, extinct hominids, and modern humans. According to their analysis, *homo sapiens* and *homo erectus* have larger thoracic spinal cord than other hominids and primates. Thoracic motor neurons from the spinal cord would be responsible to abdominal and intercostal muscle movements. These muscles are responsible for breath control. Therefore, later hominids would have greater breath control, enabling complex speech. The problem with this hypothesis is that there is no way to determine if this anatomical change has any speech reason or function at all, showing itself not of great help.

The relation between larynx height and formants has shown to be a productive investigation topic within the comparative method. Birds and primates are known to perceive formants with great accuracy (Fitch, 2002). That would have to do with the possible role it plays in individual discrimination and body size prediction, as formants are related to vocal tract length. Important to notice that formants are different than pitch (see chapter four for further explanation on pitch properties), that plays no role, apparently, in individual discrimination. This highly accurate formant perception would not facilitate our speech, though. When we investigate human perception of formants, we face some sort of formant normalization, where formants from different body sizes are normalized into categories, which enables human language phonetic accurate perception, independently of the size of who is speaking. The descend of larynx, as explained before, would have been crucial to the variety of formants we humans can produce. Evolution wise, the downsize would be more vulnerability to chocking, which was corrected with a change in nutritional behavior (Lieberman, 2012).

Within genetic studies, the discovery of the gene called FOXP2 (Forkhead-box P2) represents great advance in understanding the origins of human speech. This gene is part of a gene family responsible for coding transcription factors, that is, proteins that bind to DNA and regulates gene's expression (Fisher, 2006). This gene is the same to all humans and significantly different in function in chimpanzees (Fisher, 2006; Fisher & Marcus, 2006). It was discovered through a clinical study with a large British family that presented problems with oro-motor praxis, including both speech production and non-speech oral movements, as well as some perceptual cognitive deficits (Vargha-Khadem *et al*., 1998; Fisher, 2006; Fisher, 2016). Precipitated assumptions on what was known by the time made this gene famous as 'the grammar-gene' erroneously (Pinker, 1994; Fisher, 2006). The correction of this misconception, however, does not take out the importance to the field of this discovery, as FOXP2 is indeed strongly regulating some of our speech functions.

All the evidence and hypothesis discussed above lead to one simple conclusion: language as we know today did not evolved at once. Indeed, evolutionists believe that some sort of protolanguage took place at the beginning of language evolution history (Fitch, 2005). The point of discussion is what was the nature of this protolanguage. There are a few theories that can be categorized into two subgroups: those who hypothesize the protolanguage with a synthetic nature; and those who hypothesize language with an analytic nature (Fitch, 2005). The synthetic models represent more of a traditional understanding of the protolanguage. Fitch (2005) explains that synthetic models postulate that the protolanguage was composed by simple sentence structure (formed by simple concatenation of words, or simple merge), or even one-word structureless sentences. They follow the most primitive idea that the basilar language grammar is to pair representation (sound) and meaning, in a direct, non-compositional way. Then, the leap forward in language evolution would be adding syntax structure to its grammar.

Fitch (2005) also explains that analytic models, on the other hand, propose that, although the primitive relation of meaning and representation (or form) is true, meaning would be entire propositions that map a complex yet undecomposable structure, opaque to primitive speakers of protolanguage. In other words, the primitive syntactic unit was not a word, but a full sentence from which structure was yet opaque. That way, synthetic models propose an evolution bottom-up, while analytic models propose an evolution top-down (Fitch, 2005).

The analytical model is specially of our interest when we look back at Darwin proposal of a musical protolanguage (Darwin, 1871). He postulates that the protolanguage would be composed of complex learned vocalizations with no meaning or holophrastic meaning (e.g.: courtship song, war song, etc.). The two main properties of this protolanguage would be a focus on the auditory/vocal channel and complex structure culturally transmitted. If we go further in Darwin's hypothesis and add another stage for this protolanguage guided by kin selection,[10] where it is added of full propositions and intentional semantics (Fitch, 2005), we

[10] Evolutionary studies list three types of selection: natural, sexual and kin selection. Kin selection is a little different than natural selection because it doesn't focus on the continuation of one individual's genes, but in all individuals that share those genes, as a community or family. It is cooperative, in a sense that when individuals that share similar genes (similar alleles) share resources, they are amplifying the chances of that allele to continue (see Dawkins, 1979).

have a strong dual stage analytical hypothesis to the evolution of language. Modern music would be a kind of fossil to primitive language that would have evolved its own path (exaptation and then adaptation).

In conclusion, Prosody is the closest system we have to animal's system of communication, as it takes information from other cognitions or computations and encodes them into melodic content that will then be transduced into acoustic signal via vocalization and articulation. That being true, it endorses the hypothesis of a prosodic protolanguage, that, as we will see in the next section, is very consistent with the human language acquisition process. It even may be that our Prosody cognition is partially an old structure present in our ancestors that has been repurposed (exapted) and improved with language insurgence, after complex thought began.

2.1.3 How do we acquire language?

When a baby is born, the main contact it has with language is through its main caretaker. On its surface, communication seems to play great role in language acquisition. Once a child is placed within a community where people talk to her in certain language, that is the language she will eventually talk. For a long time, researchers believed that the communication of the mother (or caretaker) with her child was crucial to the child's language development (Gleitman *et al.*, 2019). *Motherese* refers to the way parents and caretakers typically talk to infants and young children (Newport, Gleitman, Gleitman, 2020). The simple sentences, the exaggerated contours, and the high pitch were believed to help, or even enable children language acquisition. However, it's been showed that *motherese* has little to no effect on the success of a child's language acquisition (Newport, Gleitman, Gleitman, 2020).

The Poverty of Stimulus Argument (POS) is a collection of claims that are related and endorse each other, although slightly different from each other (Laurence & Margolis, 2001; Berwick *et al.*, 2011b). They all agree that the knowledge one has (on any matter) is greater than what can be learned strictly from experience (Laurence & Margolis, 2001; Berwick *et al.*, 2011b). When it comes to language, most of the claims under POS suggests that there is genetic endowment to human language capacity (Chomsky, 2005; Berwick *et al.*, 2011b). This innate endowment would play a greater role to language acquisition than any circumstances of the empirical data, such as motherese (Newport, Gleitman, Gleitman, 2020). A child's knowledge of language may be shaped by experience, or even bootstrapped by it, but it does not depend on it (Gleitman *et al.*, 2019), as it is internally directed, similarly to the development of other cognitions. In other words, we are born cognitively prepared to develop language by engaging multiple capacities, such as memory, statistical predictions, and categorization.

Chomsky (1965, 1981a, 1981b, 1982, 1993, 2014) proposes that the initial stage of human's language capacity is some sort of raw grammar from which humans develop the grammar for any specific language. During the process of language acquisition, the Universal Grammar, as Chomsky calls this initial raw stage, will be modified by inferences and generalizations made by language cognition based on empirical data collected through experience (Chomsky, 1981a; 1981b). The theory (Chomsky, 1981a; 1981b, 1982, 1993, 2014) is that this raw grammar is composed of a set of principles and a set of parameters. The principles are rigid universal rules that every language follows. The parameters are unmarked variable rules that constrain the possible grammars to be generated, but allows variability among the languages of the world. The process of language acquisition would consist in marking down these parameters, generating, that way, the grammar of a specific language. In order to mark the parameters, the young mind would use the data to which they've been exposed through experience. We call this data that is intensively used to assess core system information *primary data*. Once in contact with it, the young children's mind will analyze the primary data, qualitatively and quantitatively, and, at some point, empirically, comparing the evidence with the possibilities of the set of parameters in the Universal Grammar. Once marked, the parameters marking will be put to test either by output or by further analysis of input data.

At the very beginning of language acquisition, we have the beginnings of Prosody acquisition. Prosody acquisition starts in the womb (Burnham, Kitamura, Lancuba, 1999; Gervain, 2018a, 2018b; Martinez-Alvares *et al.*, 2021), right after the fetus can properly hear and perceive sounds, sometime between 20 and 28 weeks of pregnancy (Ruben, 1992; McMahon, Wintermark, Lahav, 2012; Gervain, 2018a, 2018b). Surrounded by amniotic fluid, much of the complexity of the speech signal is lost, but most of its prosody is still perceivable, as (some) pitch, contour, breaks, intensity and duration (Gervain, 2018a, 2018b). There is evidence that an infant recognizes her mother's voice right after birth (DeCasper & Fifer, 1980). There is also evidence that the first cry of an infant shows characteristics of the prosody of the mother's language (Mampe *et al.*, 2009), indicating that at least some of it was accessed and acquired in uterus. If considering that the fluid and tissues of the womb will low-pass filter the sound a fetus hears (Gervain, 2018a), prosody is the only information available to assess speech in uterus.

Going back to what we've seen earlier in this section, "motherese" may be a result of (or means to) survival strategies of human as species, but for other reasons than language computation needs. There is evidence that infants have preference to listening to "motherese" than to regular adult speech (Fernald, 1985; cf. Cooper *et al.*, 1997). Although it's been showed that there's no linguistic advantage or preference to "motherese" when it comes to language acquisition (Newport, Gleitman, & Gleitman, 1977; Newport, 1975), these data do not disqualify the fact that children give more attention to "motherese" speech (Fernald, 1985; cf. Cooper *et al.*, 1997). In the child back end, "motherese" may play a role in developing an infant's attention through the oddness of its prosodic characteristics (Cooper *et al.*, 1997), as well as communicate affect (Saint-Georges *et al.*, 2013). The mother would then instinctive use motherese, in order to make sure her child is paying attention to the information she is passing down. Also, the high pitch, and the slow pace, accompanied by body language and facial expression, promote emotional bonding between them, what today is known as essential to a child to thrive (Le Bas *et al.*, 2020). Therefrom, motherese would be making use of prosodic sources to fulfill more primitive needs of survival, although using the already evolved and specialized structure and system of language.

In conclusion, it seems that language development starts with Prosody acquisition (Gervain, 2018b). Prosody plays a crucial role in language cognition development. That is because it bootstraps language acquisition, as it is the first structure to be acquired and decomposed (Fisher & Tokura, 2014; Fernald & McRoberts, 2014; Da Silva, 2019). Considering strong the analytical model hypothesis to language evolution, discussed in the previous section, it is possible that the same stages the model proposes human language cognition evolved to process language is similar to the current stages an infant brain will mature its language cognition. In other words, the data children are exposed to is an opaque complex structure with full proposition, that the developing system will analyze and learn to break down, similarly to the earlier men and their protolanguage. In that case, the very first layer of this opaque chunk of speech is prosodic. Also, prosody is the very first almost intact structure they hear in the womb, as discussed above, which endorses the hypothesis of prosody being first thing children acquire.

2.2 A MODULE IN THE MIND

"(...) understanding the process that designed the human mind will advance the discovery of its architecture"
Cosmides, Tooby, & Barkow, 1992

We've seen in the previous chapter that, although human language serves human communication, research shows language is not a functionality of human's communication capacity, rather, language is a capacity in its own. Arguments endorsing that, range from evolutionary trait to genetic encoding to ontogenetic development. What is implicit (or explicit yet unmentioned) in the concept of language as an innate capacity, is the concept of language as a cognition.

Mind is the executive control of all things behavioral and the home of thought. Being that speech is the behavioral outcome of language, and that thought and reasoning are done linguistically, language is not only a capacity, in a sense of genetically ingrained potential, nor simply a behavior, but what produces and regulates the speech behavior.

Mind can be described as an information-processing system that transforms input into data structures, such as mental representations, and behavior (Cosmides & Tooby, 1987, 1995). In order to do that, mind is a highly structured entity. To understand language cognition, it is necessary to understand its place and role within mind structure and architecture, besides investigating its own architecture and design.

We argue here that language, while a faculty and cognition, is a modular system, containing a selection of domain specific modules, being Prosody one of them. The nature of the domain specific modules will be briefly discussed in 2.3. Prosody as a module will be discussed in 2.5, where we argue in favor of its domain specificity, information encapsulation and autonomy during language processing.

2.2.1 How do our minds interact with the world?

Minds are some sort of symbol-manipulating devices, in a sense that what they deal with are representations of the real word and of thought itself. The awareness of this gap between what we know in our minds and what it is in the outside world can be track back to Plato. In his *Allegory of the Cave* (Plato, 2010), Plato proposes that what we know about the real world are actually shadows of it. In other words, what we know does not correspond exactly to what the real world is. What we know is a flat bidimensional *representation* of the real four-dimensional world.

In despite of that, the environment must be accounted for mind structure, as it shapes knowledge through experience (see Menari, 2010; Fodor, 1983, 2000) and bootstraps cognitive development to many capacities (see Chomsky, 1995; Fodor, 1983). Although the contact between mind and environment is done with mediation, minds are constantly being impacted by their contact with the world. Understanding this mediation is understanding the way we get the real-world information and turn it into thought.

All the contact we have with the world is made through our senses. We smell, taste, see, hear, and touch things, and that sensorial information is what our minds will deal with in order to make sense and memories of things. The first challenge to our minds is right there: the difference between the information our senses receive and the information our senses transmit to our nervous system. Aroma, taste, light, sound, and texture are what we call analogue signal. Analogue signal is any physical information that varies continuously within a given period[11]. Its variation is related to another time-based variable (analog of it). As a result, an analogue system allows an infinite number of values to be represented, as it can achieve any value within the parameters (the analog time-based variables) governing the system. For example, a smell is a continuous signal that is increased or decreased by other factors such as distance from the source, time in which the smell started, temperature of the environment, etc. Our minds operate in our brains, which are composed of neurons, cells that transmits electric pulses. (We are going to discuss a little further on that in the next sections).

Pulses are discrete values. In other words, the kind of signal our brains produce is the opposite of the continuous variables our senses receive in the analogue signal the environment provides them with. We call this signal composed by a sample of discrete values a digital signal[12]. Digital signals are binary, that means, they are sampled by using only two values: a logic *1* (high) and a logic *0* (low). Therefore, in order to read the world, our brains, and consequently our minds, need to have the analog signal transformed into digital signal. Electronic devices all have an analog-to-digital converter, that takes the continuous signal from the environment (i.e., electricity from the outlet) and transforms it into discrete binary values (i.e., electric pulses of high or low voltage). Our bodies have similar devices. Our

---

[11] analogue signal. In World Encyclopedia: Philip's. Retrieved 12 Mar. 2022, from https://www.oxfordreference.com/view/10.1093/acref/9780199546091.001.0001/acref-9780199546091-e-418.

[12] digital signal. In Ince, D. (Ed.), A Dictionary of the Internet: Oxford University Press. Retrieved 12 Mar. 2022, from https://www.oxfordreference.com/view/10.1093/acref/9780191884276.001.0001/acref-9780191884276-e-907.

senses are stimulated by an analog signal, triggering neurons in our nervous system, that will transmit the now digital signal to the brain. We call these systems that interface environment and brain transducers.

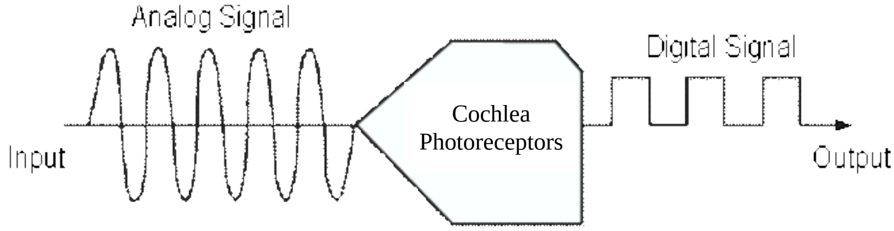


Figure 3: scheme of an analog to digital converter with addition of examples of sensorial transducers (font: https://wiki.analog.com/_media/university/courses/electronics/text/chptr20-f1.gif?w=650&tok=6b5353 ; accessed on April 8th, 2022 at 10:32PM).

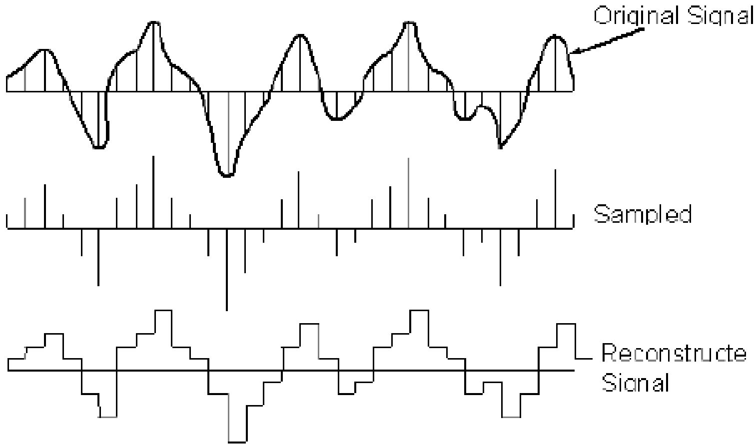


Figure 4: Scheme of an analog signal being converted to digital signal through sampling and reconstruction (font: http://www.azimadli.com/vibman/_aintroduction%20to%20machine%20vibration-52.png ; accessed on April 8th, 2022 at 10:32PM).

Once the information reaches the brain, it will be processed by our second interface system: the perception system. Fodor (1983) classifies transducer systems and perception systems slightly different then we do here, when making his argument in favor of what he calls input systems. He says that input systems are perception systems, or systems of interface, responsible for feeding central systems (thought) with information that is an interpreted representation of the world. To our opinion, that is not the exact function of transducer systems and perception systems. According to Fodor (1983), and we agree here, transduction process is simply converting real world information in information that the brain/mind can interpret. It happens by the origin of the sensorial information (ear, eyes, skin, nostrils, and tongue). Whereas perception systems are responsible for generating representations of the information provided from transduction that the mind can process. It happens in the brain. We don't fully disagree with that definition, but the main divergence point is in the understanding of how these representations are generated. To Fodor (1983), the perception process stays true to its input information, in a sense that a representation is as accurate as it can be to the real-world original information. In the other hand, we understand here that perception involves judgement of the input information interfered by the knowledge one already has of the world (cf. Munkong & Juang, 2008). To give a practical example, when a Brazilian adult is being exposed to English in the beginning of his studies, some phonemes, e.g., [θ] (/th/), will remain unperceivable by his mind, because he doesn't have it in the inventory of phonemes stored in his memory yet. In other words, he will not "listen" to that sound, even though the acoustic signal for it is present. That selective deafness will last until when, by frequency of exposure, attention, and active memorization, new information on that sound is stored in the phonemes inventory.

Language would be one input system, according to Fodor (1983). This categorization does not fit our view to the language faculty we argue here. First because we don't see language as simply a feeder of world information to thought. Also, we do believe, propositional content and will can interfere in language processing, which would be characteristics of central systems. Ross (1990) says that the core problem in Fodor's trichotomous functional taxonomy of psychological processes is in the hypothesis of a category as central systems. Our goal to this dissertation is to study and investigate language and Prosody, therefore we will not argue in favor or against central systems as it doesn't concern us directly.

2.2.2 Understanding the concepts of faculty, module, and domain

Beyond transducers and perception systems, it is important that we understand a few terms to move forward. The first of them is faculty, within psychology perspective. Faculty is an inherent power of the mind[13], as opposed to capacities acquired through experience, training, etc. It is defined by its functionality, its role in the thought process, and its effects.

[13] Gregory, R. (2004). faculty psychology. In The Oxford Companion to the Mind. Oxford University Press. Retrieved 13 Mar. 2022, from https://www.oxfordreference.com/view/10.1093/acref/9780198662242.001.0001/acref-9780198662242-e-331.

For example, memory is a faculty, as it is an innate capacity, we have to store information. Notice that we can *train* our memory to improve its *performance,* i.e., to make it function in a more productive and effective way, but that does not make it any less innate, as it doesn't require active training. Fodor (1983) brings up two types of faculties that has divided the field of faculty psychologist theoretically: horizontal faculties and vertical faculties. Horizontal faculties are faculties whose functionality is not limited to one type of content. They are accessed by different contents domains without being modified by them or being contained by any of them. Their operation is independent of whichever content domain is using it. Memory stores information independently of the nature of the information's content. It will operate always the same way, let's say, to store a smell, or a grocery list (Fodor, 1983). The problem with this theory is that someone who is good in certain cognitive function, such as remembering, should be good in every and each time this function would be applied, i.e., remembering every sort of information (e.g., remembering names, and faces, and numbers, etc.), as a faculty cannot be both strong and weak at the same time (Fodor, 1983).

Gall is one of the stronger proposers to the idea of vertical faculties[14]. His theory is that we have independent abilities that are not interactive, being that its operations are closed in themselves (Gall *apud* Fodor, 1983). Gall thought mental structure not by its functionalities, but by enabling and causing behavior. In that case, Gall looks to mind as a set of abilities, propensities, aptitudes, and functional powers. When looking through this perspective, he proposes that the skill set that causes or enables certain ability, e.g., cooking, is distinct from the skill set that causes or enables another ability, e.g., dancing, as they are distinct abilities. The biggest problem with this theory is that it presupposes a great number of faculties to be able to account for every human ability, since no psychological mechanism is cross domain (Fodor, 1983). It is then unbearable.

Indeed, complex behaviors such as cooking, or dancing, seem to involve a variety of capacities. What makes someone good at cooking, but bad in dancing, or vice-versa, is not a lack of ability or development in the whole set of skills and mechanisms needed to cook or to dance. At the same time, the degree of success in certain cognitive functionality doesn't mean succeeding every time, in every attempt, with every matter. Spearman (1927 *apud* Fodor, 1983) says that there must be a correlation between the operative (horizontal) faculty and what it is being used with. Fodor (1983) proposes a mixed model where mental structure is composed both by vertical faculties, which he calls cognitive domains, and horizontal

[14] Although Gall doesn't call them vertical faculties. Fodor is the one that proposes the term when referring to Gall's theory as it is a direct opposition to horizontal faculties theory.

faculties. To him, what determines high ability of someone is the skill set a person has, and how her mind operates and mixes these horizontal faculties and domains (Fodor, 1983). Now it is not a matter of having a specific faculty strong or weak only, but also how the mind combines faculties resulting in different degrees of abilities.

Performance, then, is not the result of the operation of one faculty, but the organism operating various faculties in order to execute a behavior. That also goes well with the idea of an optimum system, proposed by Chomsky (1993a, 1993b, 2008, 2013), as different capacities are the result of the interaction of faculties, being that we have several capacities arising from different combinations of a small number of mental faculties (Fodor, 1983).

Fodor (1983, 2000) proposes that the mind is partially modular. According to him, there's no evidence good enough to support that central systems are modular, but what he calls the *input systems*, including language, are. Fodorian modules [15] have a few characteristics. First and foremost a module is domain specific (Fodor, 1983). Domain is a class of information. The term *domain specific* entails that a cognitive mechanism operates with only specialized class of information. Take vision faculty as an example, we have a module for processing motion, a module for processing shape, a module for processing color and so on (Frankenhuis & Ploeger, 2007).

Modules are autonomous. Within our minds we have operations that are conscious and those that are unconscious (see Fodor, 1983). It is not about a will to think of something, but the level of conscious effort. Fodor (1983) argues that the operation of a module is mandatory. That means that one cannot refrain its mind from doing it. It is unavoidable. Taking the vision faculty again as an example, one that has healthy eyes cannot refrain one's mind from seeing what is in front of one's open eyes. Even when you're daydreaming, or just focusing on your thoughts, what shifts actively and somewhat willingly is your attention on the information being processed and output by your visual cortex, but it does not make your mind go blind for a second.

That leads us to the next characteristic: modules have limited access to central systems. Although we are not adopting or exploring Fodor's concept of central systems, we agree that modules have limited access to higher cognitions. By that, we mean that the intermediate/transitional representations an operation within a module may project is not accessible to consciousness. Still thinking of the vision faculty, there is evidence that object recognition goes through several different processes, within several different modules (cf.

[15] There is at least one other theory that proposes a modular mind with slightly different terms than Fodor's theory. To see more on Massive Modularity theory, see Frankenhuis & Ploeger (2007).

Hubel & Wiesel, 1962; Cohen & Dehaene, 2004; Dehaene, 2010), but the only representation we are conscious about is the final whole output of it. If we are to make a conscious reflective effort in analyzing more specific information, such as shape, texture, color, we are still unable to “unsee” the object as a whole.

Figure 5: diagram of a bottle as an object in the world; some of its perceivable features; the object and its features as a representation of incapacity to disassociate the features processed by the transducers and the whole object processed by the perception system.

This limited access is one main key to enable the processes within a module to be fast. Due to consciousness, reasoning and logical thought take a while to be processed. If every mid-process output were to be taken to reasoning, the processes would be slowed down a lot, not to mention the working memory overload. That basically means that the information being processed within a module cannot be affected by information flowing through the rest of the mind. Its processing cannot be interfered by outside information. Once again, taking vision as an example, in order for our brain to see an image being presented, it needs 13 ms (Potter *et al*., 2014). Now, in order for our brain to recognize an object, it takes 150 ms, more than ten times over (DiCarlo, Zoccolan and Rust, 2012). When we face something as the Müller-Lyer illusion, we can recognize the shapes and, even if we know the catch, the horizontal lines are the same size.

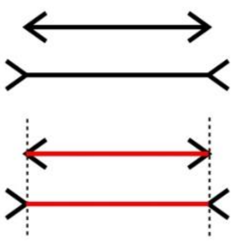


Figure 6: horizontal lines with inwards and outwards arrow tops, an illustration of Müller-Lyer illusion. (source: https://brainstuff.org/blog/muller-lyer-illusion-psychology-test , accessed on March 25th, 2022, at 12:08 PM).

It requires prolonged practice to grasp that they are same size (Lewis, 1908), otherwise, our knowledge about its actual measures doesn’t affect our perception of them (Frankenhuis & Ploeger, 2007).

On the other hand, a module is informationally encapsulated (Fodor, 1983, 2000). That means that the other cognitive processes have only access to a module output, but not its intermediate levels of representation during its processes. For example, face recognition involves analyzing virtual lines on the face (see Andrade & França, 2021a, 2021b), but this information is never processed by other module, not to mention being accessible to awareness.

Other modules are not able to even retrieve intermediate levels of representation, as a module projects only shallow outputs (Fodor, 1983, 2000). That means that the intermediate levels of representation of an output is not (completely) retrievable as the output is flattened into one opaque piece of information.

All of this is ratified by the fact that modules exhibit characteristic and specific breakdown patterns (Fodor, 1983, 2000). In other words, pathologies that are associated with one modular system does not affect directly other modules. A person who is brain-blind is not necessarily also brain-deaf. That points strongly in favor of the information encapsulation and autonomy of the modules, as what affects the process of one will not affect directly the processes in another.

This would have to do with the fact that a module is associated with fixed neural architecture (Fodor, 1983, 2000). Conditions derived from localized brain injuries are usually linked to impairment or defective functioning of one specific system, more specifically, one specific module. For example, vision is associated with parietal, temporal and occipital lobes in the brain, each region is related to different processes within vision faculty. The occipital lobe is known to house visual cortex, processing information sent by the transducer system in the eyes, to process visual recognition. A stroke on the occipital lobe is known to cause blindness (any degree), or hallucinations (Morenas-Rodríguez *et al.*, 2017), while a stroke on the right temporal lobe can cause prosopagnosia, which is the inability to recognize previously familiar faces (Evans *et al.*, 1995; Bentin, Deouell and Soroker, 1999; Joubert *et al.*, 2003).

Finally, a module develops based on genetic information as the innate capacity it is. Its development follows the same stages and sequencing, in every human being. The developmental stages are usually bootstrapped by environmental stimulation. Or, in Fodor's word: "(…) neural mechanisms subserving input analysis develop according to specific, endogenously determined patterns under the impact of environmental releasers." (Fodor, 1983: 100).

To give an example, a human fetus has in itself the capacity for vision. Once the eyes begin to be stimulated by light, the vision faculty starts its sequential stages of development.

It is known that at birth, a baby's vision is extremely way undeveloped when compared to an adult.

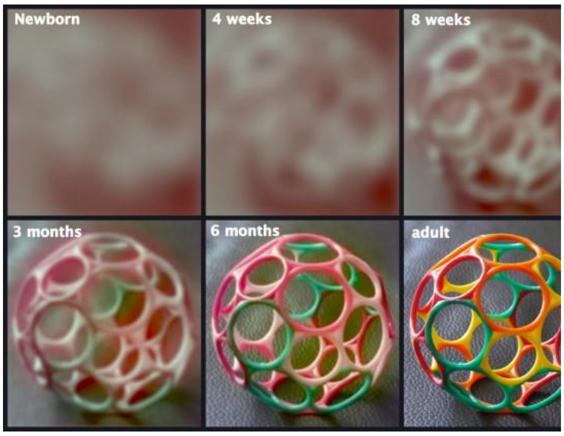


Figure 7: Simulation of an infant's vision. (source: https://lasermom.wordpress.com/2012/06/24/infant-vision-research/ ; accessed on March 25th, 2022, at 11:41 AM)

During its development, it improves different characteristics. For example, by 2 months old an infant start discriminating colors from white light (Teller, Peeples and Sekel, 1978); by 3 months, their perception of depth is improved significantly (Birch & Petrig, 1996); by 4.5 months an infant can individuate shapes of objects (Wilcox, 1999); while by 6 months infants can integrate shapes and objects to discriminate contour (Taylor *et al.*, 2014; Siu & Murphy, 2018).

In summary, cognitive modular systems are domain specific, innately specified, hardwired, autonomous, and not assembled. Let's take a look now on how all of this applies to the language faculty.

2.2.3 The Faculty of Language

When we think of human language, many different human behaviors can be associated with human communication capacity and be mistaken as language (e.g., body language). However, the perspective we assume here is centered on the hypothesis that human language is a faculty (Hauser, Chomsky & Fitch, 2002; Fitch, Hauser, Chomsky, 2005). By that we mean that we assume language is an innate, inherited capacity of the mind (see 2.1 for further discussion). Now, going further on the nature of this faculty and its properties, a lot has been discussed over the years, in an effort of better understand its origins, development and functionality (see Hauser, Chomsky & Fitch, 2002; Fitch, Hauser, Chomsky, 2005). In the next chapter we will look into Chomsky's proposal to faculty of language functionality and operation, but before getting there, it is important to understand its place within mind architecture. As mentioned before, Fodor (1983) indexes language as an input system. For reasons discussed previously, that classification does not fit in our proposal. However, we do see language as a modular system and here's the why.

a) As a modular system, language is composed by domain specific (sub)modules.

One way of looking at the Minimalist Program (Chomsky, 2014) is that it seeks to understand and explain which one of these modules is uniquely human and what information it computes. The idea of Syntax first (Chomsky, 1957), as the very core of language faculty would be speaking to one module of the language system. Would this module be exclusively human? Maybe. Or maybe the way it operates, as one capacity is accounted for a collection of modules and operations, and the way they combine (Fodor, 1983). In any case, plenty of research has shown that whatever the computation for syntax is, the computation for semantics, pragmatics, lexicon, phonology are independent (see França & Gomes, 2015; Kraljic, Samuel & Brennan, 2008; Ribeiro, 2015a, 2015b). For example, a sentence like (7), vastly used in memes as a linguistic illusion:

(7) I love Paris in the the springtime

the duplicate 'the', although present phonetically, even once we realize the illusion, does not affect syntax processing, or semantic processing. That can only be possible if we assume different types of linguistic information are being processed within different modules.

b) The operation of language modules is mandatory.

The same way we cannot willingly refrain our brains from hearing something someone is telling us, we cannot willingly refrain a sentence from being processed (Hartsuiker & Moors, 2017). Once we listen to (or read) a sentence, it will be processed no matter what[16]. More specifically, a sentence as the well-known coined by Chomsky (1957):

(8) Colorless green ideas sleep furiously

will inevitably be processed by syntax although it will be blocked by semantics. Another example is when we listen to a sentence in a language we are not fluent in. As the processing of phonological content seems to bootstrap syntax processing (Scott & Cutler, 1984), lexical access, and semantic processing, it is common to mistakenly hear words in our own languages even if the sentence doesn't make sense (see Kentner, 2015).

[16] Very few internal circumstances can interfere with that truth. High anxious mental state is one of them (cf. Eysenck, 1988, 2000), as it interferes with cognitions in general.

c) Language computation requires or entails limited central access.

As we've argued on the previous topic, language modules' operation is autonomous. That means that there is no will, attention, effort, reasoning necessarily involved in their operation. Take (9), for example:

(9) The dog devoured the shark

No context given, this kind of sentence provokes an aggravation on an event related potential (ERP)[17] signature called N400. This ERP is affected by unexpected roles and relations meaning wise. Dogs devouring things is not absurd. It is very common, depending on the breed we have in mind. Now, a shark, the fierce full ocean animal, being devoured by a dog, it is very unrealistic, or at least unexpected. However, some reasoning would easily imagine circumstances where this sentence would be totally fine. For example, the dog has a toy that is a shark. Or the dog's owner has a fish named shark that was sadly murdered by the dog. Or the dog earned some beach themed dog cookies, and his favorite was the shark one. In despite every possible scenario we can think of where this sentence would be okay, the N400 will consistently show aggravation. That shows us that, while processing language, reasoning is not accessed until further later. If it were, there would be no place for the N400, as reason would immediately play its role.

d) Language processing is fast.

Intuitively, we can easily become aware that it is faster to produce an automatic sentence when we're freely thinking of something than it is to write a speech about something, where intentionality, style, social rules and more extralinguistic factors play a huge role on it. Furthermore, the advance of linguistic experiments within the fields of psycholinguistics and neurobiology has shown that we process a word in 350ms (Pylkkänen & Marantz, 2003), and a sentence as soon as 600ms (Schmidt-Kassow & Kotz, 2009).

e) Language processing is informationally encapsulated.

As discussed before, when we say a module is informationally encapsulated, we mean that it has autonomy to process its information within, without sharing all the steps with other modules. Linguistic modules process their specific information without sharing inner layers of its structure. In a way, that is the concept behind the old-fashioned concept of Deep

[17] When neurons fire, they produce and transmit electro-magnetic pulses. They are not aleatory and unique. They follow a pattern in order to transmit information thorough the circuit. An ERP is a result of the sum of pulses of certain area of the brain, during certain time, in response to certain stimuli. N400 is a pulse that happens within 400 milliseconds after stimulation. For further information see Kutas & Federmeier (2011).

Structure (Chomsky, 1965). Although we now believe it is counter-efficient to have a step such as a Deep Structure to a sentence, the idea that the mid-process representation is accessible only by the syntax module is still current (Chomsky, 1957, Fodor, 1983, 2000). That is why concepts as phases, and spell-out, to be discussed further in the next chapter, are still incorporated to the newest versions of Chomsky's theory (Chomsky, 2014). Another example is that Semantic Theory describes what they call covert movement. It is a movement of an item in the structure to another place within the sentence structure, which is not accessible to syntax and phonology (Kratzer & Heim, 1998).

f) Language computation projects shallow outputs.

A consequence of information encapsulation is that linguistic output from a module is always shallow. That means that the mid-process representations not only is not accessible to other modules while being processed, it is also non-retrievable once the final output is shared with other modules. An example would be the output from syntax module. Although syntax builds up sentence structure when processing a sentence (see Lage, 2005; Lage *et al*., 2008), this structure is flattened when it takes phonological form, in order to be sent to the motor cortex to articulation (see Chomsky, 2014). This "flattening" process is called linearization, and it is a mandatory process that undergoes within the phonology modules.

g) There is a neural architecture dedicated to language processing.

There is plenty evidence on neural architecture dedicated to language processing. Although, investigate the neurobasis for language is not within the scope of this dissertation, we strongly assume the importance of it to understanding mind structure and language processing. See further in Hickok & Poeppel (2000, 2004, 2007); Giraud & Poeppel (2012); Friederici (2011, 2012); Friederici & Gierhan (2013).

h) Language faculty seems to have specific breakdown patterns.

Linguistic impairments associated with brain damage was actually the evidence that led to uncovering brain regions dedicated to language processing specific domains. The different aphasias are associated with different brain regions (Damasio, 1992). Also, cases like Genie (Curtiss, 2014) show us that vocabulary processing is independent than syntax processing, which points to a modular system.

i) Language faculty development is ontogenetic.

We've seen on Chapter 1 that, although language acquisition is bootstrapped by stimulation of empirical data, language capacity is innate. Because of that, its development is ontogenetic, that is, the development of the faculty of language is motivated and carried by the system itself. All the information the mind/brain needs to develop language is already coded in the DNA (see Graham & Fisher, 2013; Rice, 2013), leaving to the empirical data only the content for parameterization (see Chomsky, 2014). That is why language acquisition is considered natural and infallible, because it utilizes empirical data, but does not depend on it (see Chomsky, 1980).

By all the discussed above, we have sufficient evidence to assume that the faculty of language is a modular system. In the next chapters, we will discuss a little further on the nature of the modules that compose the language faculty. For now, it is important to point out that the language's modular system is in direct relationship to the perception systems, from which it receives input. The final output of language processing depends on the mind's demand, as it can be the motor cortex, for articulation, or one of the higher cognitive functions, for thinking.

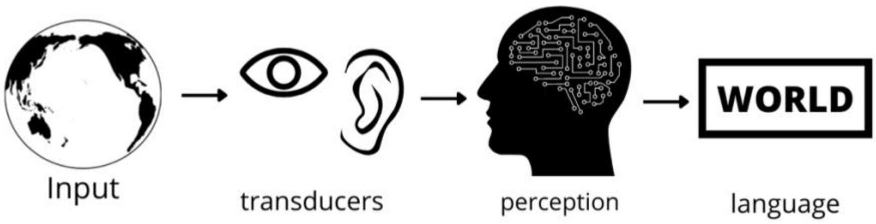


Figure 8: Diagram of language placement in mind architecture

2.3 A COMPUTATIONAL PROCESSING

In Chapter 1, we've seen that Chomsky's innateness hypothesis to language has been accepted and endorsed by much evidence in evolutionary studies. As part of his argumentation, he proposed how we develop this innate capacity into first language, and, afterwards, how language is processed in the competent mind, through a computational approach. The generative theory (Chomsky, 1965; 1993a) is one of the most explored versions of his theory to language computation, but now we have a more updated version of it, the Minimalist Program (Chomsky, 2014). The Minimalist Program is an effort to answer questions as of what is exclusively linguistic and/or human in our language capacity (Chomsky, 2005). Although he doesn't touch directly Prosody processing in his model, we believe that our ideas are very complementary to his proposal and that is the main reason we are going to briefly review it here.

2.3.1 Core concepts to a cognitive approach

Cognitive psychology is the study of knowledge representation and use in human beings. Consequently, a cognitive approach to psycholinguistics will be interested in understanding what are the representations for linguistic information and how they are processed. To think is to manipulate representations using the operations of computational procedures (Thagard, 2005). Hence, language is the collection of computational procedures and representations, and how they operate.

Representation is one of the fundamental concepts in a cognitive approach to language. A representation is a system for producing symbolic entities through formal schemes in the mind. Its content must be grounded to its referent in the world in a way that it is interpretable (Marr, 1982; Friedenberg, Silverman & Spivey, 2021). That means that the description, the how the entity is being represented, is constitutional part of a representation (Marr, 1982).

Representations are static mental entities. The dynamic part of the mind is called computations. Computations are responsible for information processing, that means that they generate, manipulate and transform representations (Friedenberg, Silverman & Spivey, 2021).

There are an endless number of computations of different granularity operating the mind. Therefore, in order to study them, they are usually categorized by two criteria: i) type of operation (what do they do); and ii) type of information they act upon (what do they manipulate). In that sense, it is considered language computation the (collection of) system(s) that process language information through linguistic operations (i.e., merge, linearization, etc.).

The computational theory, as proposed by Marr (1982) postulates three levels to information processing: the computational level, the algorithmic level, and the implementational level.

Within the computational level, a process contains separate arguments about what is computed and why. In other words, a process is defined by the type of information it processes and the purpose of it. When we say ‘why’ and ‘purpose’, it is important to make it clear that it doesn’t mean some sort of belief or elaborate proposition. The purpose of a process points to its functionality – what is it for? For example, sentence processing is the process of building up sentences, it computes sentences.

A process involves two components: i) a representation to be manipulated as an input and output, for example, lexical entries as an input and a sentence as an output; and ii) an algorithm by which transformation may be accomplished, for example, merge for building up phrases. Therefore, the choice of algorithm depends on the specific representation it is being processed, as processing an output depends on the constraints it must satisfy (Marr, 1982).

Before moving forward, it is important to explain that transformation is different than computation, because the former talks about the difference between input and output representations (i.e. what changed when comparing input and output) and the latter talks about the process by which transformation is achieved (i.e. what are the operations involved in the process applied to the input in order to make that output).

2.3.2 The architecture of language

Following the tri-level hypothesis proposed by Marr (1982) – see previous section –, describing the language faculty involves describing it computationally, algorithmically and implementationally. Although briefly touched in Chapter 1, implementational aspects of the faculty of language, in what it concerns to neurophysiology, will not be discussed by us, as it is not our point of interest for this present work. Regarding the computational and algorithmic descriptions, here is our understanding of the matter.

The faculty of language is a faculty dedicated to process linguistic information. Its final product will serve higher cognitive systems, such as reasoning, horizontal faculties, such as memory, and communication needs, such as speaking (see Fodor, 1983, 2000; Chomsky, 2014; Hauser, Chomsky & Fitch, 2002; Fitch, Hauser & Chomsky, 2005). As we've discussed in the previous chapter, the faculty of language is a modular system, which means that it is composed by a collection of domain specific modules, each a specialized computational system (see Fodor, 1983, 2000). The computational systems are responsible for creating new and transforming existing linguistic representations, each within the scope of their specific domains.

Therefore, to understand the architecture of the faculty of language is, in part, to understand its modular distribution of computational systems and the way those modules interact among themselves before dispatching information to other faculties. In the case of the Minimalist Program, as mentioned before, it aims to understand which mechanisms and characteristics in language processing are exclusive to human's linguistic capacity. That implies that some of the mechanisms and characteristics involved in language processing are not exclusive either to our language capacity, being some sort of domain general, or exclusive to humans, being present in animal's system of communication. Yet, non-exclusivity doesn't fully disqualify a mechanism as linguistic. Therefore, Hauser, Chomsky, and Fitch (2002) propose that there is the broad approach to language faculty, called the Faculty of Language in the Broad Sense (FLB), and the narrow approach to language faculty, the Faculty of Language in the Narrow Sense (FLN). While FLB would include every aspect and mechanism involved in language processing, FLN includes only what is exclusive to humans' linguistic capacity. That doesn't necessarily mean some modules within language modular system belongs only to FLB and some to both FLB and FLN (as FLN is content to FLB – Hauser, Chomsky & Fitch, 2002). We may find the same modules in animals and humans, but they may be different in the algorithms they employ. Or we may find some disparities in the modules' distribution, yet some similarities in the algorithms applied. We are, however, a long shot from having clarity in this matter, as the computational and cognitive research with animals is way behind when compared to human's (Hauser, Chomsky, and Fitch, 2002; Fitch, 2002, 2005).

Hauser's, Chomsky's, and Fitch's hypothesis will be that FLN foresees the mechanism of merge, with its recursive property, being the only component of it. As this is the main mechanism to building sentence structure, they argument that FLN contemplate syntax (narrow syntax; Hauser, Chomsky & Fitch, 2002; Fitch, Hauser & Chomsky, 2005). In other words, the syntactic module is in the core of language's modular system.

That's the reason Chomsky describes the faculty of language as composed by a lexicon and a computational system responsible to generate structural descriptions (SDs). Each structural description is an expression, that is, a sequence of representations, contemplating a pair (π, λ), to be checked on the Phonological Form (PF) and the Logical Form (LF), final representations to be sent out of language processing modules.

> "A language consists of a lexicon and a computational system. The computational system draws from the lexicon to form derivations, presenting items from the lexicon in the format of x-bar theory. Each derivation determines a linguistic expression, an SD, which contains a pair (π, λ) meeting the interface conditions." (Chomsky, 1995:186)

The computational system Chomsky (2014) mentions, in our view, refers mainly to the syntactic computational system, which deals with sentence structure. The interface systems, called Sensory-Motor (S-M) interface and Conceptual-Intentional (C-I) interface, we believe them to be, in part other modules within language faculty, such as segmental-phonology module, Prosodic module, (formal) semantic module, etc., and in part the "performance systems" Chomsky mentions language is "embedded in" (perception systems, articulatory systems, inference systems, etc.), where the linguistic information will serve a function – express, communicate, think, reason, etc. (Chomsky, 2014).

The sequence of operations employed to process a sentence, that is, to produce the pair (π, λ), is called *derivation*. Chomsky (2014) proposes that a derivation starts by accessing and selecting *lexical items* to be called into the sentence structure. These lexical items are organized into a list called *numeration*, that pairs lexical items, and indexes that represent the number of times the selected lexical items are to be called into an operation. By this process, the system is able to map π and λ properly matching, in what regards the lexical items selection and summons. Theoretical frameworks such as Distributed Morphology (Morris & Marantz, 1993) and Exoskeletal Model (Borer, 2005a, 2005b, 2013) propose that at the beginning of syntactic derivation, lexical access is not done as proposed by Chomsky (2014), as words are to be constructed within syntactic derivation as well. We will not explore this topic here, as it doesn't affect our point of interest, which is Prosody Processing. Independently to what is the nature of the representation accessed and selected to be part of sentence processing, some level of arbitrary symbols to concepts are accessed in memory and used within the derivation, and that is all that concerns us here.

Syntactic computational system will then construct syntactic structure. To do that, the system takes items in the numeration list and merge them together, forming a phrase. The X-Bar theory is a way of representing this syntactic representation, the phrase structure. Chomsky (2014) prefers the version that uses only binary branching because it supposedly avoids ambiguous paths to mapping (π, λ). The computational system will exhaustively use all items in numeration list, and recursively the syntactic objects projected during derivation to build complex sentence structure.

Through the algorithms of *merge*, that takes two items and transform them into one (see Hauser, Chomsky & Fitch, 2002), *label,* that identifies the head of the syntactic phrase (see Chomsky, 2013), and *move*, that copies an item and inserts it somewhere higher in the structure[18] (see Chomsky 2014), the syntactic computation builds up syntactic structure into *phases,* syntactic objects that are transferred to the interfaces S-M and C-I. The operation of transfer to the interfaces is called *Spell-out*.

Chomsky (2008) argues that there are two phases at which Spell-out applies, one at v*P and one at CP. The Spell-out operations map the syntactic structure to Phonological Form (PF) and Logical Form (LF), generating the pair (π, λ), to be interpreted by the performance systems (Chomsky, 2008, 2013, 2014). A derivation will either converge, if it satisfies the conditions imposed by the interfaces, or crash, if otherwise.

It is not very clear, in Chomsky's theory, the details of what happens with language information between projected sentence structure and final output to performance systems (i.e., articulation). We believe that what he calls PF is in fact a collection of processes made by other modules then syntax, responsible for phonological aspects of language. When it comes to Prosody, the theoretical approach is unclear about its place in the language architecture and the functioning of its processing. What we have is somewhat extensive research on the interfaces between Prosody and other levels of linguistic representation, to be reviewed bellow. In summary, we know that prosodic structure carries syntactic, semantic, pragmatic, and emotional information, but we don't know how this structure is processed in face of the language processing as a whole. As each representational level belongs to a symbolic system (Fodor, 1983, 2000), we argue that segmental phonology and prosody are processed by different modules, being that prosodic modules are the last ones to transform linguistic information within language system. That makes their output to be the final version of π, the set of instructions to the performance systems (Chomsky, 2014).

[18] There are other syntactic operations proposed throughout the years of research in the field. We are only citing here the most relevant ones to prosodic structure.

As matters of θ-role application apparently doesn't affect prosody, and other reasons to be discussed in Chapter 5, we believe the information is not transferred to prosodic modules at phase v*P, but only phase CP.

Finally, all mid-process representations are discarded at the end of language processing, lasting only the pair (π, λ), to be transferred to the performance systems (Chomsky, 2014). That endorses Fodor (1983, 2000) hypothesis that a module output is shallow, as the final pair is nothing, but the complex sum of representations and transformations in one whole object to be put to use.

2.3.3 The (un)placement of prosody

Within the Generative Approach, Prosody studies have been under explored. We see a massive amount of work in (syntactic) sentence processing, as this is the main lane Chomsky and his collaborators take in the core work on the field (Chomsky, 1965, 1993a, 1993b, 2003, 2005, 2013, 2014; Hauser, Chomsky & Fitch, 2002; Fitch, Hauser & Chomsky, 2005). Therefore, the most representative work with successful attempts of exploring Prosody Processing from a Generative Grammar point of view focus on investigating the interface between Syntax and Prosody. Prosody, as we see it here, interfaces with few other modules, that have been explored to different extents, and not necessarily under a generative/ minimalist perspective.

We believe that in order to understand the place and role of the Prosody Module in the architecture of the faculty of language, we must understand better the dynamics of its interfaces.

2.3.3.1 The Syntax-Prosody Interface

Let's take a look at the sentence below:

(10a) The thief hit the lady with the cane

This sentence contemplates what we call syntactic ambiguity. At a first glance, because of the Minimal Attachment Principle (Frazier, 1978), we are most likely to interpret the sentence as a synonym to the following sentence:

(10b) The lady that is using the cane was hit by the thief

However, (10a) could be mapping another meaning. Something similar

(10c) The thief used the cane to hit the lady

This structural ambiguity happens because the Prepositional Phrase (PP) could be attached either to the Determiner Phrase (DP) *the lady* or to the Verbal Phrase(vP) *hit the lady*.

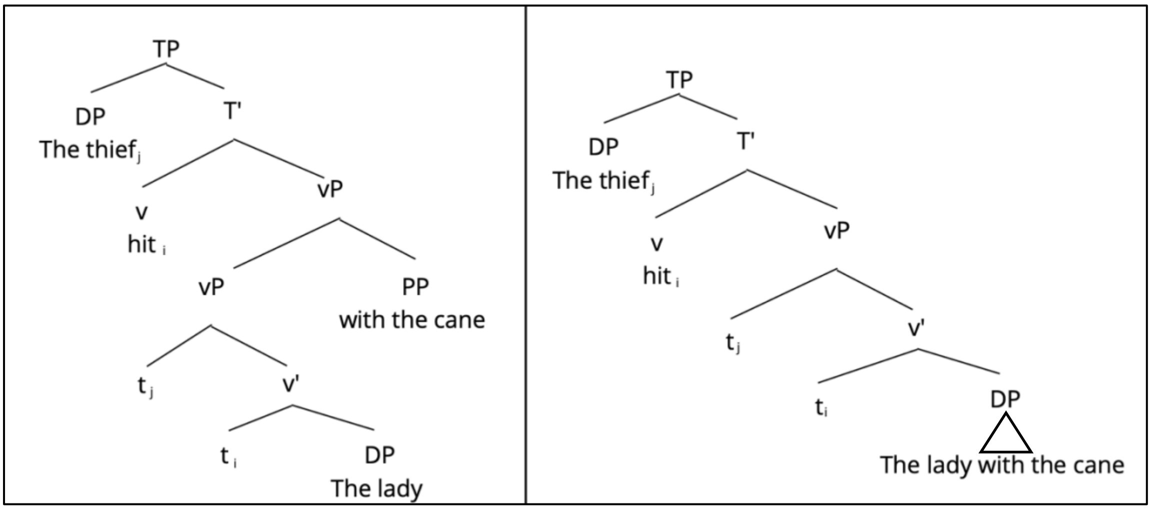


Figure 9: Syntactic trees representing the two possible structures for the sentence in (10a). On the right, the similar in meaning to (1b), and on the left, the similar in meaning to (10c).

The point here is that, if this sentence was spoken instead of written, there would be no ambiguity. Several studies have shown that prosodic structure will cue the sentence processing to which syntactic structure it should be mapped (see Marslen-Wilson *et al.*, 1992; Gravina & Fernandes-Svartman, 2013). In the case of (10a), the difference is in the prosodic domains: to map the structure in (10d),

(10d) [The thief] [hit the lady with the cane]

the PP *with the cane* should belong to the same prosodic domain as the DP *the lady*; and to map the structure in (10e),

(10e) [The thief] [hit the lady] [with the cane]

the PP *with the cane* should belong to a different prosodic domain then the one the DP *the lady* belongs. This would be phonologically signalized by the presence of a boundary between *the lady* and *with the cane*, which would change the configuration of the intonation.

The example explained above is evidence to the argument that prosodic structure maps syntax structure (see Wagner, 2004, 2007, 2015; Ribeiro, 2015b; Ribeiro & Lage, 2015; Elfner, 2018). When we say here *prosodic structure*, we are referring more specifically to *prosodic phrasing*.

Prosodic phrasing is the part of prosodic structure that determines what are the prosodic constituents. So, taking the example we used above, there is a distinction in the prosodic phrasing of the two syntactically ambiguous sentences:

There is no ambiguity to this sentence in prosodic terms because the formation of the prosodic constituents is based on the formation and distribution of syntactic constituents. This can be observed if we play with the placement of the [$_{PP}$ with the cane].

The sentence (10f)

(10f) [The thief with the cane] [hit the lady]

represents the only normal prosodic constituent's distribution option for a sentence where *with the cane* is placed right after *the thief*. That is due to syntactic restriction where when the PP is not attached to the [$_{DP}$ the thief], it should either remain at the end of the sentence, or be moved to the very front of it (to Spec of CP). That's the reason (10g)

(10g) *[The thief] [with the cane] [hit the lady]

is ungrammatical, as it maps a syntactic structure that is ungrammatical.

There are plenty of research that investigates the Syntax-Prosody interface and the way prosodic phrasing maps different syntactic structure in different languages. Ribeiro (2015a) investigated the prosodic phrasing in Brazilian Portuguese (BP) and Hebrew to sentences with three PPs attached either by conjunction or adjunction. They found different prosodic phrasing to different syntactic structures. The sentences with coordination of PPs had a different prosodic constituent to each attached PP. The sentences with embedding of PPs had a single prosodic constituent to the three PPs. This difference was attested in both BP and Hebrew.

2.3.3.2 The Semantics-Prosody Interface

Not extensively investigated, the interface semantics-prosody seems to be gaining some ground recently with the arise of informational semantics. It is known that the lexical items are the most important units to construct meaning, specially within the formal semantics framework. However, when it comes to information in the discourse, it seems that prosody plays a role in mapping critical information. Ladd (2014) points out that researchers haven't given much importance to the fact that there is crucial distinction between propositional content and any other information conveyed by speech. Yet, Büring (2016) argues that different prosodic realizations systematically correspond with a difference in interpretation.

While the Syntax-Prosody interface will result in prosodic phrasing, in what regards its constituents, mapping syntactic structure, the Semantics-Prosody interface will result in pitch configurations mapping informational structure.

Consider a sentence as follows:

(11a) Your cellphone is ringing

(11b) Your cellphone is ringing?

The sentences in (11a) and (11b) have different intonations, in spite of having similar syntactic structure. The question in (11b) is not a typical question that requires an unknown information. Rather, it is a confirmation question that is stated only to validate or falsify a proposition's truth value. This intention is conveyed by pitch alone, as the prosodic phrasing (its constituents) are the same in (11a) and (11b).

Now consider the following question, from Pruitt and Roelofsen (2013):

(12) Is Marcia allergic to dairy or soy?

This sentence can be answered two ways: either by choosing *dairy*, or *soy*, or by stating lack of knowledge on the matter; or by saying *yes*, or *no*, or again something like *I don't know*. The type of answer being required is conveyed by a distinction in the intonation of the two questions.

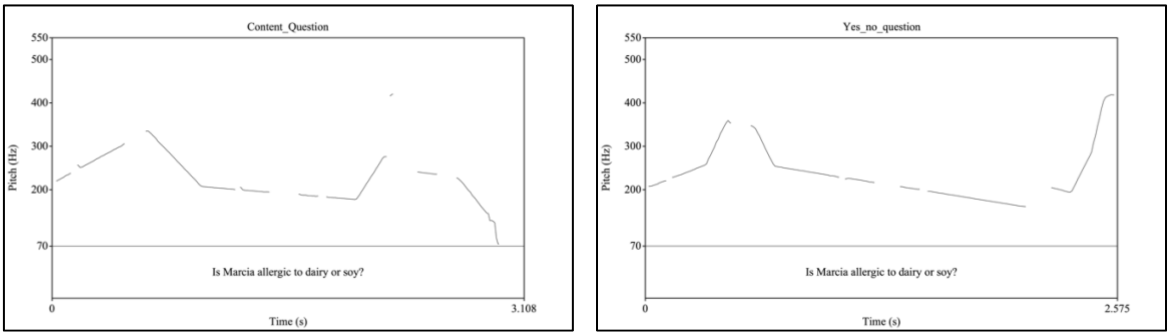


Figure 10: Stylized pitch tracks of the two types of questions for (12)

The examples in (11) and in (12) have a similarity that they don't share with the examples in (10): they're variation is made through what we call Pitch Accent (PA). A PA is a distinctive prominence in the prosodic domain, more specifically, in the intonational phrase, that is characterized by an increasing shift in the pitch (either a peak or a valley) and, often times by a lengthening on the duration of the melodic segment to which it is applied (usually a vowel) as well.

The sentences in (12) demonstrate that often times there are more than one way of PA placement to a sentence, but they are not informationally equivalent (see Bolinger, 1972; Büring, 2016). That does not mean that there is no structural relevance in PA placement. The way an utterance is phrased, which is, as we saw on the previous chapter, directly related to syntactic structure, will play a role in the PA placement, as will the grammar determination for nucleus placement. However, the structural constraints presented by the prosodic phrasing does not explain the arbitrary choice of pitch curves amidst several possible intonational contours. Bolinger (1972:633) will say that “accent should be viewed as independent, directly reflecting the speakers intent and only indirectly the syntax”.

2.3.3.3 The Pragmatics-Prosody Interface

Pragmatic information carries meaning beyond linguistic level. The propositional content of a sentence is deeply linked to the sentence, and to it alone, in a sense that building meaning is based and constrained by the lexical items and syntactic structure. Pragmatic content, on the other hand, is constrained by extralinguistic factors, only relying on linguistic information as a mean of communication.

Let's observe the sentence (13):

(13) You are awesome!

The propositional content tells us that we are associating the quality of being *awesome* with this entity *you,* equating one with the other. The lexical content adds that the quality of being awesome is the quality of being extremely impressive or daunting, someone or something that leaves us in awe, and that *you* is the entity to which the sentence is being directed. Moreover, if I say (13) for example to my son, pragmatic. content will tell us that *you* is my son. The restriction of the scope and the specialization on what *you* means in (13), in the case just described, are there because the sentence has a place and a time of occurrence and actants to perform the enunciation act. These restrictions are due to extralinguistic factors, as the language system, as seen here, does not process reference the same way deixis does.

Not all pragmatic content is mapped to prosodic content, but there a few iconic interactions, like the speech acts. Moraes (2011b) argues that not all speech acts will have a distinct characteristic intonation, but that there are a few speech acts with consistent specific melodic contours.

Take a sentence as the following.

(14) Unlock the door

Because of the lexical choices and the syntactic structure of the sentence, when an audio of the sentence being spoken is not provided, we will probably take this sentence as an order. However, depending only on the way it is spoken, the same sentence can be either an order, or a request, or an appeal, or a challenge, or a suggestion, or an advice (Searle, 1969; Moraes, 2011b).

The interface between pragmatic processing and prosody processing results~~,~~ then~~,~~ in alteration in the intonation configuration to convey information regarding speech acts, deixis, and the sort.

2.3.3.4 The Affective Prosody

Speech conveys emotions and other affective states. There is no argue about that. In fact, humans are supposedly better at perceiving emotions through speech than through facial expression (Scherer, 1981 *apud* Chuenwattanapranithi *et al*., 2008)[19]. Therefore, it is necessary to mention that the Prosody System seems to interface with some of the affective systems.

The main point of investigation on this interface has been identify what acoustic and systematic characteristics are crucial to perceive a specific emotion of affective state from prosody cues (see Moraes & Rilliard, 2016). An emotional state[20] is known to have longer duration than a sentence, and its influence lasts until the affect event is ended (see Scherer, 2005). Because it affects the vocal tract as a whole, the prosody will be affected at different aspects such as pitch, pace, intensity and vocal quality (Scherer, 1981 *apud* Chuenwattanapranithi *et al.,* 2008; Moraes & Rilliard, 2016).

It is not our goal here to investigate this matter further, but to raise awareness on the role of Prosody Processing in regard to affective states communication.

2.3.3.5 The Performance Systems Interface

Prosody Processing, as seen here, predictably interface with performance systems of perception and articulation. The reason for that is related to Chomsky's proposal of the pair (π, λ) – Chomsky (2014). Taking into consideration that Prosody maps syntactic, semantic, pragmatic, and affective information, it has to be receiving input from all these different systems and providing the output for that. In other words, prosody processing is done later than all these other processes. Therefore, we here argue that it is possible, if not probable, that the prosodic information is the last one added to the pair (π, λ), right before it is sent to be articulated. Similarly, during comprehension, prosody would be the first to process as it cues other processes on its structures.

---

[19] For a counter argument see Moraes & Rilliard (2016).

[20] An affective state can have even longer duration as it can be behavioral linked to one's personality (cf. Chuenwattanapranithi *et al*., 2008).

This matter has been under investigated within psycholinguistic framework, being more successful in neurosciences framework. Although, we don’t investigate it further in this work, it was also worth mentioning to delineate properly the Prosody Modular System.

2.4 A PHYSICAL PHENOMENON

When we talk about Prosody, people usually first think of it as a representation of the way we speak. And although we strongly believe here that Prosody is closely related to speech, as we believe its processing is the last one within faculty of language (see our argumentation in Chapter 5), speech data is nothing more, but a product of Prosody processing (and other linguistic modules processing), and not Prosody itself. However, in order to assess the properties of prosodic computation, the data it provides is of extreme relevance. Therefore, we will look in this chapter further on what forms the physical component of prosodic data as well as how it is produced by human beings, and what are the relevant measurements to prosodic analyses.

2.4.1 The acoustic features of Prosody

Acoustics is the area of Physics that studies sound. In physical terms, sound is a mechanical longitudinal wave composed of compression and rarefaction that propagates transmitting energy. A wave is a disturbance, that is, the interruption of a settled condition, a change in equilibrium. We have different types of waves, that are directly related to the type of energy they are caused by (i.e., electromagnetic). Mechanical waves involve mechanical energy that is generated by the vibration of matter. To Physics, matter can be defined as anything that can have its mass measured (e.g., bones, air, water, etc.). Thus, a mechanical wave is a local deformation in matter that propagates (travels) from particles to particles by creating local stresses (disturbance) that cause deformation in neighboring particles as well (see Sueur, 2018).

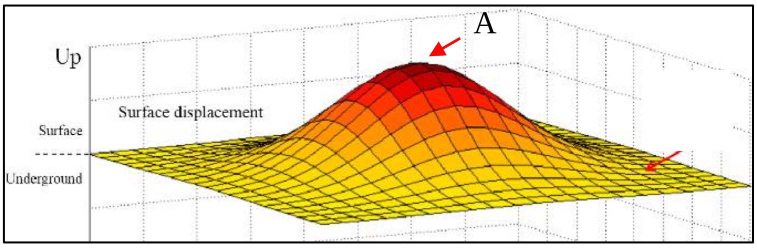


Figure 11: Graphic representation of a surface deformation. A disturbance caused to point A affecting neighboring quadrants. (source: Hu et al., 2017 [modified])

In order to better understand this, let’s take a guitar as an example. When we hit a string of a guitar with a guitar pick, we generate a disturbance at one point of the string.

Figure 12: Top string vibrating in response to disturbance caused by guitar pick. (source: photo taken by me of a volunteer L.R. at April 23rd, 2022)

That disturbance causes that one point to oscillate and consequently, it vibrates the neighboring points of the point that was hit. This vibrating movement, that was caused by the force applied to the string and facilitated by the elasticity of the string matter, generates mechanical energy, that causes the body to keep vibrating for a while. The consequent vibration of the neighboring points of the original disturbance, causes the next neighboring points to vibrate as well, resulting in the whole string vibrating (not to mention the air that resonates through or inside the guitar body, the air close to the strings, etc.). This phenomenon is called resonance (see Sueur, 2018). Therefore, this effect of one punctual disturbance causing the disturbance of an area (or many points in a string) by making the neighboring points vibrate in resonance to the initial disturbance is called propagation (see Sueur, 2018) As we will see in the next section, this property of the sound is crucial to articulate different vowels, as the way resonance and propagation go within vocal and nasal tracts will determine the nature of different vowels.

Propagation, then, is the trip of the wave from the original point to other points through matter. Its trip can be multidirectional, but the disturbance it causes to the matter is unidirectional. There are two possible types of direction to the disturbance a wave causes: vertical, when the disturbance has a perpendicular direction to the dislocation direction, or horizontal, when the disturbance direction is in parallel to the dislocation direction (see Sueur, 2018). Imagine a rock being dropped in a water surface. The waves generated by this event would propagate radially, but the original disturbance the rock causes to the surface is vertical, consequently, all the disturbances of the waves will be vertical as well.

Figure 13: Disturbance caused by a rock on the water surface. (source: Math is fun[21])

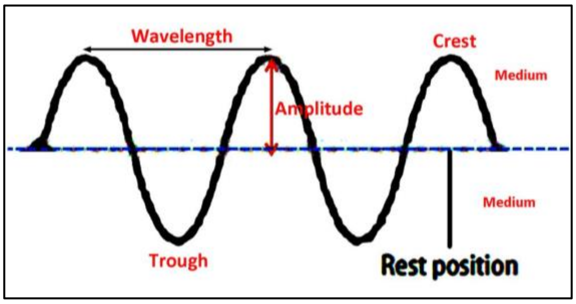


Figure 14: Graphic of vertical waves, as the ones caused by the rock on the previous picture. (source: Through Physics[22])

The sound wave has a different type of disturbance. Its oscillation (a single event of vibration) occurs in parallel to the direction of the dislocation, causing the propagation to go further (as the dislocation rate is aided by the disturbance's energy), that's why we call it longitudinal, as it usually goes lengthwise, compressing and decompressing matter. Let's take a speaker as an example. It receives electronic information from the player device, that causes it to oscillate back and forth.

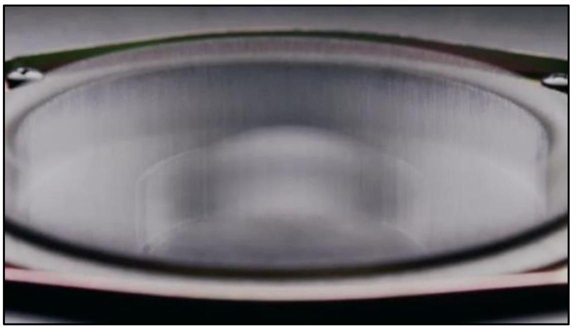


Figure 15: Picture of a speaker moving (source: Shutterstock[23])

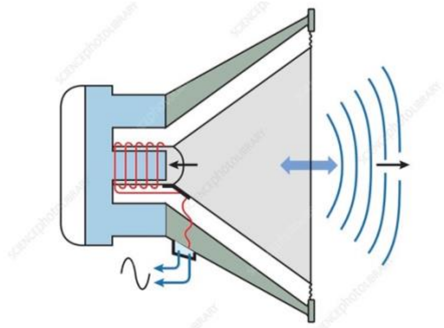


Figure 16: Scheme of a speaker moving. (source: Science Photo Library[24])

The speaker, then, causes the air particles in its cone to move forth every time it goes forward, and to move back (by vacuum), every time it goes backward. When the air particles go forward with such power and speed, we call it a pulse, a peak of released energy. The pulses cause the neighboring air particles to move forward as well, but before that the initially in motion particles get very close to the neighboring particles, making that section of air very

[21] https://www.mathsisfun.com/physics/waves-introduction.html

[22]https://d2nchlq0f2u6vy.cloudfront.net/17/07/10/e8bbde98e050403af6e96884d123103c/9d844625e837eaa1e2049725480d864c/image_scan.jpg

[23] https://ak.picdn.net/shutterstock/videos/6838240/thumb/1.jpg

[24] https://media.sciencephoto.com/c0/50/81/98/c0508198-800px-wm.jpg

dense. That is what we call compression (see Sueur, 2018). When we are close to speakers in a concert or something alike, we can feel the compression, every pulse the speaker emits: it is that punch feeling we usually feel in our chest. When air particles get neighboring air particles to move, as a push, through the pulse, the next thing that happens is a quick lowering in air density at that section of air, that is called rarefaction.

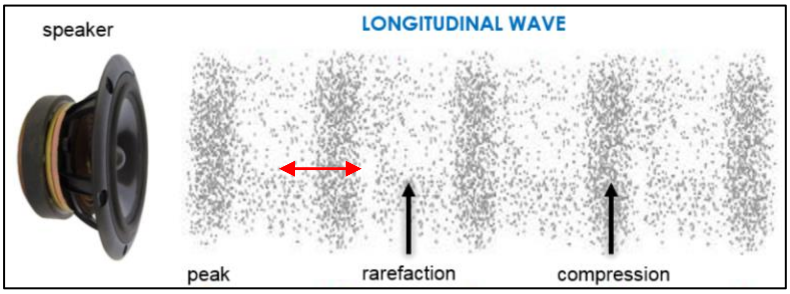


Figure 17: Scheme of a speaker transmitting sound waves through rarefaction/ compression of air particles. (source: Science in the News[25] [modified])

Once the neighboring air particles hit the next neighboring air particles, they bounce back, restoring the air density until the next pulse comes through.

The ears work as a receptor for these pulses. The ear canal will capture the sound waves, intensifying its frequency by providing an acoustic tunnel for resonance. The ear drums, vibrate as a response to the pulses, causing the bones of the year to move and consequently cause vibration inside of the cochlea, that will then transduce the signal to the nervous system.

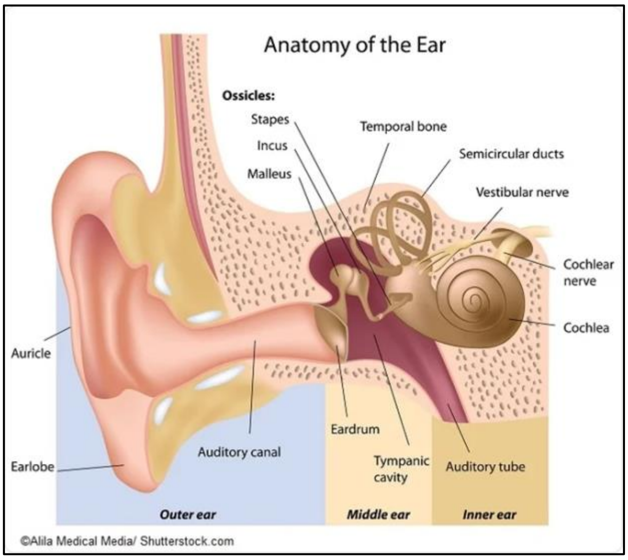


Figure 18: Anatomy of the ear, displaying the auditory canal, eardrum, middle ear bones and cochlea. (Source: Shutterstock[26])

Hence, sound is a wave that propagates mechanical energy through a medium by oscillating in the same direction of its dislocation. Its pulses cause the matter to compress and

[25] https://i0.wp.com/sitn.hms.harvard.edu/wp-content/uploads/2015/12/Figure-2.png

[26] https://image.shutterstock.com/z/stock-photo-ear-anatomy-79593499.jpg

decompress. Basically, any matter can transmit sound but, its elasticity will dictate how this transmission will be done.

2.4.2 The articulatory properties of Prosody

Having established that sound is a mechanical wave that needs an energy source, an elastic matter, and a media of propagation to be produced, it is now imperative that we understand how human beings produce the sounds of speech.

Human beings, as most of the mammals do, use their respiratory system, with its oral and vocal air ways, to vocalize sounds.

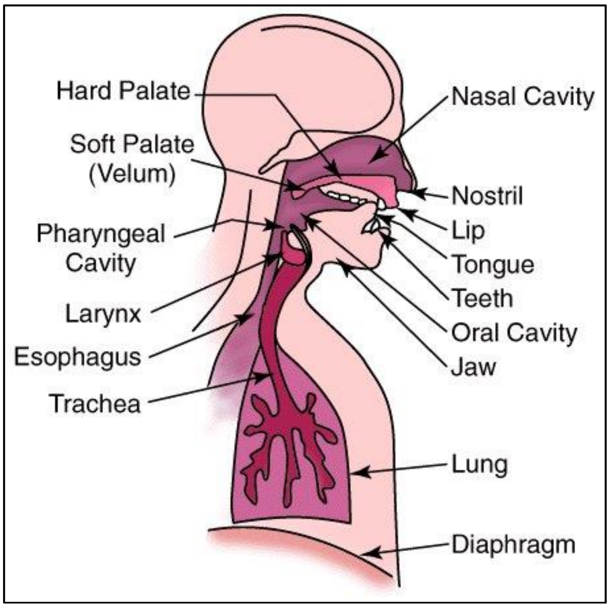


Figure 19: The human vocalization system. (Source: O'Cinneide, Dorran & Gainza, 2008).

The lungs and the diaphragm function as the source of power. They pressure air up through the system in a voluntary controlled movement. In the larynx, also known as voice box, the pressured air encounters the vocal folds. In a vocalization effort, the person applies some degree of tension to the vocal folds, the elastic matter needed to sound production. The tension of the vocal folds will act against the pressured air, causing the vocal folds to vibrate as the air passes through. This vibration of the vocal folds is the source of the sound of human voice. In order to produce the sounds of speech, we need both voiced and voiceless sounds. The movement of the pressured air against the tracts is what makes noise when the vocal folds remain relaxed to produce voiceless consonants. If, alongside the noise, the vocal folds are tensioned, we have voiced consonants. The consonants are, therefore, a result of the pressured air passing through obstructed oral and nasal tracts. On the other hand, the vowels are product of the vocal folds periodic[27] vibration, and the resonating waves (formants) produced by the different textiles in the tracts, the position of tongue and height and wideness of the configuration of the oral tract (see Cristófaro-Silva, 1999; Mateus, Falé & Freitas, 2016).

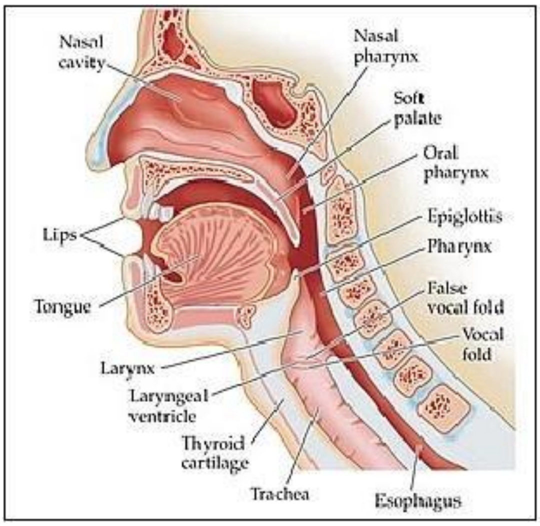


Figure 20: Vocal tract anatomy (source: Dalva, 2012).

Fundamental Frequency (F0) is the frequency of the vibration of the vocal folds. When these waves travel through the vocal tract, they resonate at many walls, generating multiple harmonics, of which the most prominent ones are called formants[28].

F0 oscillates throughout an utterance, going up and down in Hertz. That variation is due to the tension in the vocal folds by their stretching or loosening, and by the air pressure that comes out of the lungs.

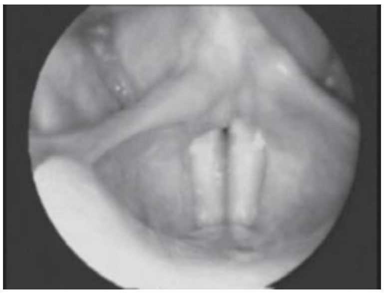


Figure 21: Tensioned vocal folds producing sounds (Source: Esling et al., 2019).

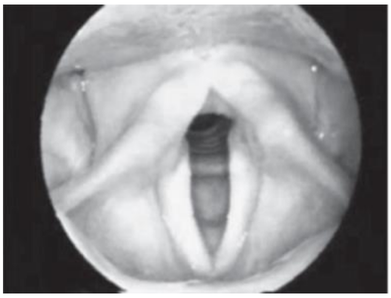


Figure 22: Loose vocal folds producing sounds (Source: Esling et al., 2019).

[27] Periodic waves are waves that have identical cycles continuously.

[28] Vowels are categories perceived from the superposition of frequencies, the F0 and the first four formants.

How wide the range is and how low and high a F0 event can go depend on some variants. The immutable one is the timbre. Every physical body has its natural frequency, that means, every physical body has a specific constitution of matter that can only resonate with a specific frequency. The human body has multiple parts, multiple tissues, that resonate with multiple frequencies, but cannot resonate with others. Also, the individual's anatomical format of the vocal tract, thoracic cavity, and the anatomy of one’s body as a whole, as well as the control and power of diaphragm regular use, determine the maximum and minimum frequency somebody can perform.

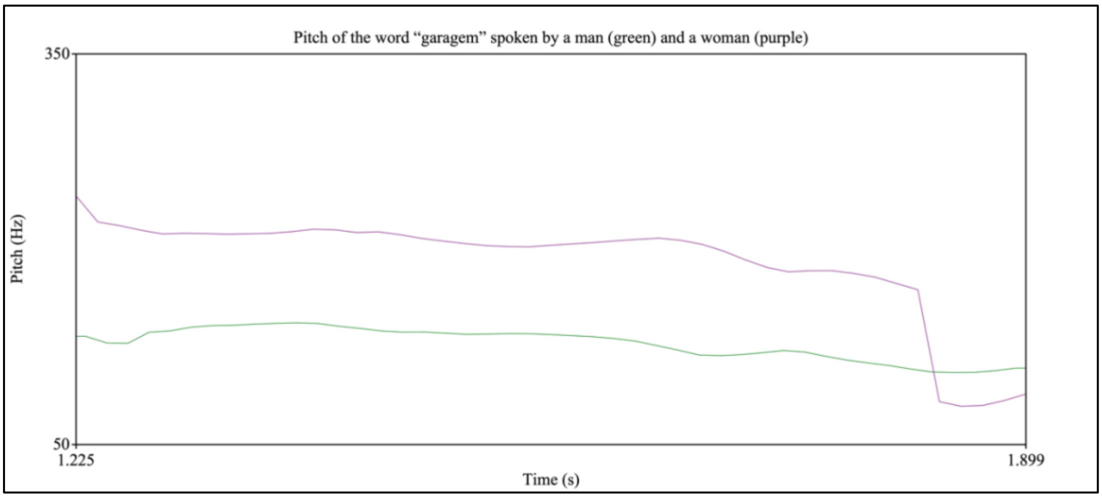


Figure 23: Pitch tracks of the word ***garagem*** *([ga.'ɾa. ʒɛ̃j̃] – garage) recorded, analyzed and drawn with Praat.*

Now, within that natural range, the range can variate due to other factors such as emotion and mental state, that alters the tension of the vocal tract and the control of diaphragm.

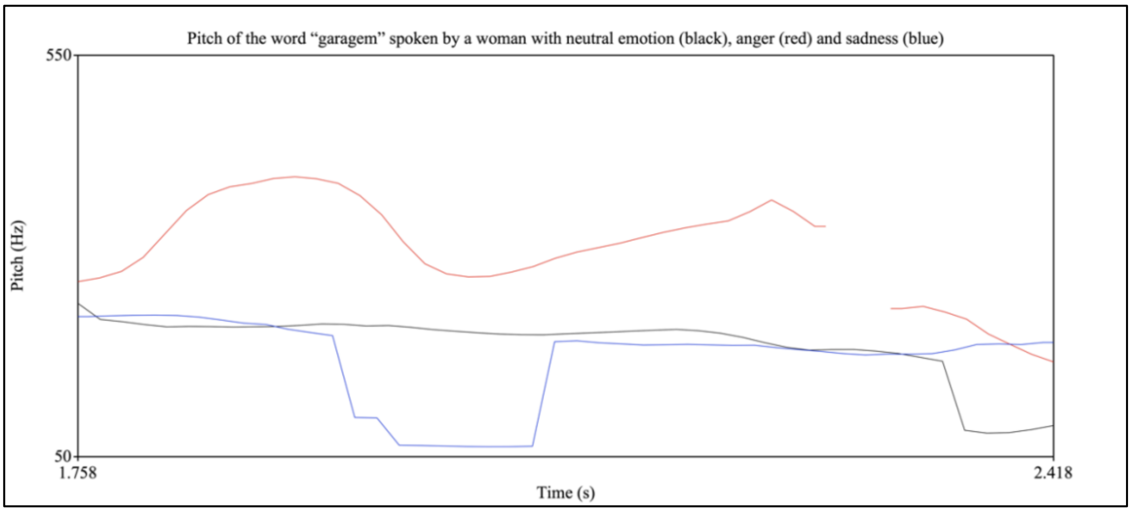


Figure 24: Pitch tracks of the word garagem *([ga.'ɾa. ʒɛ̃j̃] – “garage”) recorded, analyzed and drawn with Praat.*

2.4.3 Analyzing acoustic data to investigate Prosody

Pitch is intrinsically related to F0. That is because while F0 is the physical event of the speaker, Pitch is the perception of F0 by the hearer (or acoustic analysis software as Praat[29]). Pitch is the perception of periodicity within a time frame[30]. In other words, it is equivalent to how many times a pulse of that amplitude and length would happen in one minute. Its variance is caused by a number of factors as it depends on vocal folds vibration, but much of its variance is abstracted by the perception system (i.e., event variation of pitch due to differences between control of the muscles that tension the vocal folds) or is not relevant for the phonological system (i.e., female pitch range vs. male pitch range).

Although it seems redundant having two different concepts to deal with the same physical acoustic events (i.e., the sound waves produced by the vibration of the vocal folds), what the speaker actually produces does not correspond perfectly to what the hearer (‘s brain) hears. As discussed in Chapter 2, perception does not correspond to the signal being received by the sensory system (ears). Hearing is far beyond simply capturing sound waves through one’s ear. The actual responsible organs for listening are the auditory cortices. And, as any cognition, they deal with representations, hence, categorization of digital information. Therefore, there is a disparity between the analog nature of the sound of speech, more specifically here, F0, and the digital nature of pitch categories. This is endorsed by the finding that human pitch awareness is not nearly as high as it is often assumed, especially when it comes to recognizing melodic events in prosody (Dankovicova et al., 2007 *apud* Xu, 2015:187).

When dealing with prosodic data, as discussed in Chapter 3, the external layer of prosodic structure, that envelopes the rest of the structure, is the intonational phrase. Intonation is a suprasegmental phenomenon, and as such, it is defined by its F0, intensity and duration. If we want to define from the hearer point of view, intonation is defined by its pitch, loudness and quantity, the psychological (digital) correspondent to the physical aspects of intonation (Ladd, 2008).

Intensity is the amount of energy employed when producing the sound. It can be altered by non-linguistic factors, as emotion, or by linguistic factors, as the nature of each sound unit, that require different amount of energy.

---

[29] Software: Boersma & Weenink, (2017).

[30] See https://www.fon.hum.uva.nl/praat/manual/Pitch.html

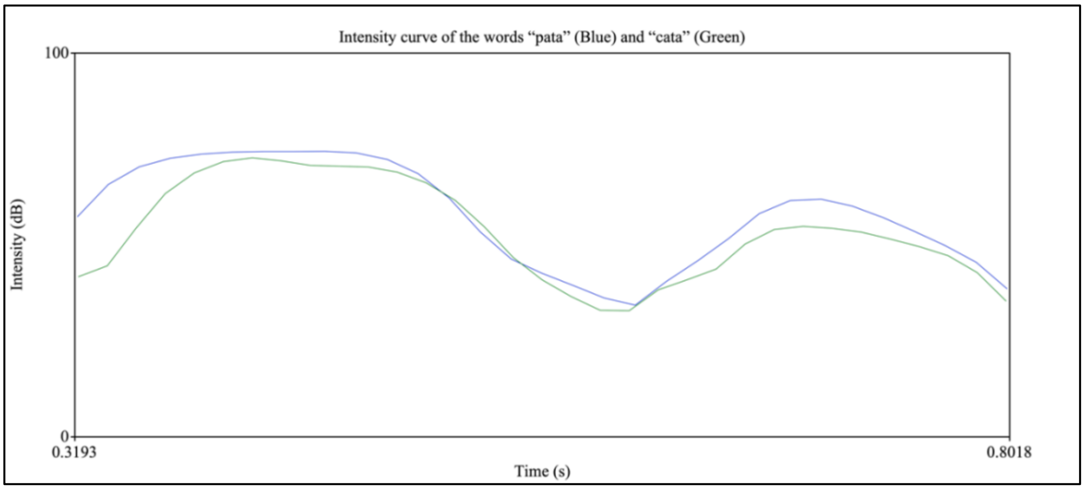


Figure 25: intensity curves of pata *(['pa.tɐ] – “paw”) and* cata *(['ka.tɐ] – “to pick something up of a place”). Praat calculates the mean amplitude for the signal (RMS) and generates a synodal curve that can relate to our perception of loudness.*

Duration is the pace of the speech. It is set by the duration of each sound unit and breaks. It can be altered by several non-linguistic factors, such as sleepiness and personal style, or by linguistic factors, like stressed syllable.

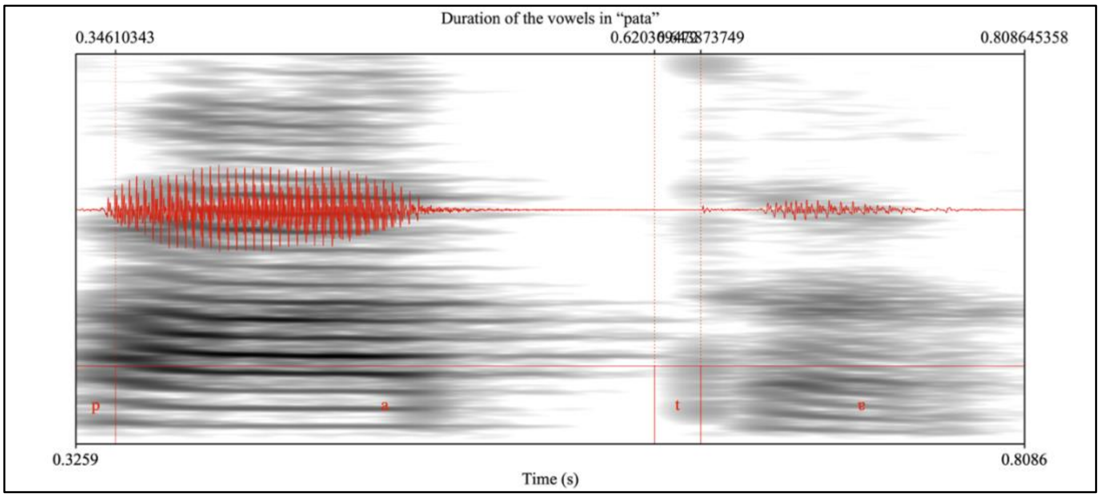


Figure 26: Spectrum and sound waves per segment showing the different ranges in vowel duration, between tonic and post-tonic vowels.

Although not accounted for in the definition of intonation, there is another factor that relates to prosody information, as we believe here: voice quality. Voice quality is, as the name suggests, the quality of the sound produced, in the sense of smoothness, clarity, and cleanness of the sound. It can be altered, as far as we know, only by non-linguistic factors, as style, timbre, and emphasis (see Esling, 2012).

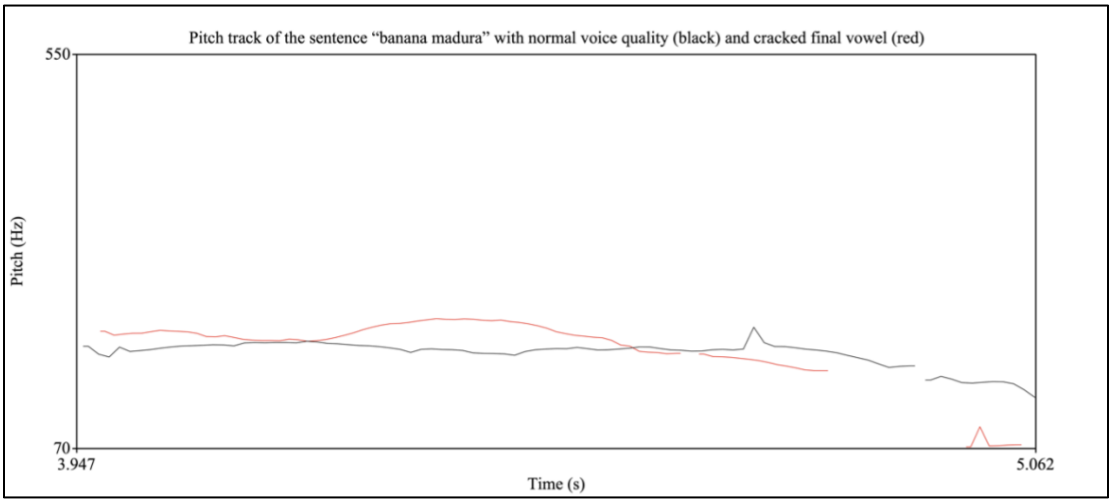


Figure 27: Pitch track of the sentence banana madura *([ba.'nã.nɐ#ma.'du.ɾɐ] – "ripe banana") with different voice qualities at the final vowel. In black we have regular intonation, and in red we have creaky voice caused by abnormal voicing.*

As discussed in the beginning of this chapter, although the physical phenomenon of speech itself is not prosody, it is the basis for the abstract representation prosody has as its units. As with all linguistic studies within the psycholinguistics approach, we can only infer the computations of prosody processing by looking onto actual data, gathered either by elicitation or experimentation. Therefore, understand the physical properties of the sounds of speech that are relevant to intonation, as well as how to analyze it and/or manipulate it are unavoidable. Furthermore, as we will see on Chapter 6, an important part of the methodology applied here is experimental. Thus, going through technical details as the one exposed in this chapter was of extreme relevance to understanding our work.

2.5 THE PROSODY MODULE: A PROPOSAL

We've been discussing in the previous chapters some of the characteristics of language and prosody that are relevant to this research, in an effort to delineate what prosody processing is. Based on what has been discussed, in this chapter we are going to present what we understand on the matter.

We've seen in Chapter 1 that language is an innate capacity (Hauser, Chomsky & Fitch, 2002). That means that we are genetically predisposed to develop the faculty of language and its cognitions (Fitch, 2005; Fisher, 2006). Part of this capacity is inherited from ancestors in our evolutionary line (Hauser, Chomsky & Fitch, 2002; Fitch, 2005; Fitch, Hauser & Chomsky, 2005). Prosody, as a domain of language, seems to be the link between what is new in the evolutionary line when it comes to *homo sapiens* language capacity and what is inherited from ancestors (Darwin, 1871; Fitch, 2005). That is, when we look closely to prosody, we see marks of both modern human language complexity, such as recursion (see Wagner, 2007), and similarities with animals' oral system of communication, such as melodic structure.

In Chapter 2, we've assumed that language is a modular system (Fodor, 1983). What we propose here is that prosody is one independent module within the language system. Prosody is domain specific in a sense that it deals with specialized information. Its information assigns rhythm, harmony, and some other melodic suprasegmental aspects to speech, such as pace and loudness. This assignment is the result of mandatory processing. Although the output information of prosodic module is strictly melodic, therefore directly related to speech, we have prosodic information being processed even in the absence of articulation/externalization of speech (Fodor, 1998; Fodor, 2002a). Prosody processing will happen even when the speaker has no intention of speaking as, similarly to other modules of language, it entails limited central access, making most of its operation unconscious. Prosody processing being mandatory and unconscious is directly related to three other aspects that argues in favor of its modularity: It is a fast process; its information is encapsulated even to other modules of language; and it projects shallow outputs. Prosody processing is fast as its processing is required to full sentence processing (see Fodor, 2002b). This is partially possible, as Fodor (1983) argues, because of information encapsulation. As we have mentioned above, prosody deals with melodic information, and we believe here that its information can be categorized into three distinct categories: rhythmic, harmonic, and dynamic information. When we analyze qualitatively prosodic data, we see certain independence among rhythmic structure, that is, phrase structure in what it regards to boundary placement; harmonic structure, that is, manipulation of pitch to design intonation; and style assigned, that is, melodic variations that change execution of melody. However, there is no sentence with rhythmic structure, but no harmonic structure; or with only harmonic structure, but no rhythmic structure or style assignment. The processing of its layered structure seems to be, then, closed in itself, in a sense that the distinction among structures is only salient to prosodic module itself. The result of its encapsulation is the third aspect mentioned above: the projection of shallow outputs. As we know, speech is a linear apparent one-level string of sounds (Fernández & Cairns, 2010). Thus, similarly to syntax structure, that goes through a linearization process, prosodic structure goes through a flattening process, making the distinction of its structures (rhythm, harmony, and style) opaque to other cognitions.

Hence, prosody processing is autonomous and informationally independent, yet it is in close interaction with other modules of language system. We've seen in Chapter 3 that language is a (collection of) cognition(s) that processes linguistic information, building up sentences by pairing sound structure (π) with propositions (λ) – Chomsky (1995). The prosody module would have the role of processing the melodic information related to language, mapping linguistic information to and from other modules of language system into melodic structure, and getting information to and from the S-M interface (see Chomsky, 1995). As discussed in a forementioned Chapter 3, prosody module seems to interface with syntax, semantics, pragmatics, limbic system, and performance systems. From those interactions, we believe that the prosody module processes different parts of its structure, as different linguistic information (syntax, semantics, etc.) will be mapped into different prosodic structure. Some of the syntactic information will be mapped into rhythmic structure, mainly in the way prosody phrasing will be distributed (see Wagner, 2004, 2015). Some of the propositional content and pragmatic content will determine the intonational curve and pitch accent assignments (see Moraes & Collamarco, 2007; Moraes, 2011b; Büring, 2016). And some of the affective information will be mapped through style assignment (see Moraes & Rilliard, 2008).

The effectiveness of these mappings has been endorsed by different studies (see Ladd, 1986, 1988; Vigário & Fernandes-Svartman, 2010; Fernandes-Svartman, 2012; Ribeiro, 2015a; Ribeiro, 2015b; Ribeiro & Lage, 2015, to cite some) and is being tested in this study through different experiments (see next chapter). We believe here that prosodic information carries sufficient information into sentence processing to retrieve sentence structure and propositional particularities, as well as contrastive information relevant for pragmatic context.

To summarize, prosody is a module of the language modular system responsible for suprasegmental melodic information processing. It is composed by three submodules: rhythm, harmony, style. Each one is responsible for processing one specific type of prosodic structure. Rhythm module, responsible for forging rhythmic structure, maps syntax. Harmony module is responsible for forging intonation. It will modulate pitch and apply PA. It maps semantic and pragmatic information when needed. Style transforms output from Harmony and adequates it to emotional and mental state and style. The output from prosody modular system is the final one into pi, being ready for articulation.

3 METHODOLOGY

To better understand prosody processing as I am proposing, I implemented two psycholinguistic experiments in Brazilian Portuguese (BP), and one in American English (AE). In the first one, it was investigated how prosody processing is mapping syntax and how efficient the perception of the prosodic cues to syntax structure is in BP and AE. In the second experiment, it was investigated how prosody processing is mapping pragmatic information and how efficient the perception of the prosodic cues to pragmatic information is in BP.

We have chosen English for three reasons. The first one is the fact that English is from a different language family than Portuguese, Anglo-Germanic and Romance language respectively, and that will give us an ampler picture of the phenomena to be analyzed. In fact, English will be a good comparison to Portuguese, as they have different prosody properties and prosody phrasing. Also, not only English is a language in which I have proficiency, making it convenient to work with, but most importantly, I went on a one-year research at CUNY Queens College in NYC, through the SWE program of CNPq.

Going into my one-year research trip, I had expectations on finding as many English speakers subjects available as I needed. Indeed, the college environment provided me with that. What I haven't foreseen before coming to NYC is that New York is an extremely bilingual community. NYC and northeast NJ have been a big immigration target for generations. They have many different cultural communities, as Jewish, Indian, Nigerian, Polish, Hispanic, to name some, and, therefore, many bilingual households. There's no problem with bilingualism, however, bilinguals are not target subjects to my research, as I need them to rely on language intuition when going through my tests. Because of bilingualism, not only their brains work differently (Fabbro, 2001; Mechelli, et al, 2004), they also have more than one grammar to rely their intuition on, making my data not accurate to the target grammar. That is the reason why I decided to carry on with only one experiment in AE, so I could concentrate on it my efforts in gathering monolingual subjects.

3.1 EXPERIMENT 1 – SYNTAX AND PROSODY: PROSODY PHRASING PERCEPTION IN DELEXICALIZED SENTENCES IN BRAZILIAN PORTUGUESE (BP)

When we listen to a common sentence, we use prosodic cues as well as syntactic information to process that sentence. Wagner (2015) says that prosody maps syntax through organization and distribution of its constituents. He also claims that this mapping is useful during sentence processing. That hypothesis has been largely tested and seems to be true (Ladd, 1986, 1988; Christophe, Guasti & Nespor, 1997; Christophe et al. 2008; Ribeiro, 2015a, 2015b; Ribeiro & Lage, 2015, to mention some studies). One of the most popular tests involves coordination. When we coordinate phrases, there is a strong boundary, composed of stop and declination reset, in intonation contour between the two coordinate segments, marking the separation between two Major Phrases – MP (Ladd, 1986, 1988). That boundary is salient and associated to the distribution of Pitch Accents (PA) throughout the utterance, one to each MP. It indicates a syntactic coordination when the boundary is present, or a syntactic attachment when the boundary is absent.

So, when listening to a sentence as

(15a) Tomato, sauce, and meat are ingredients to this recipe

The listener understands that three ingredients were mentioned: i) *tomato*, ii) *sauce*, *and* iii) *meat*, instead of just two, *tomato sauce and meat*, by the prosodic information of the intonational contour, which have strong boundaries (||) between *tomato* and *sauce*, and between *sauce* and *meat*:

(15b) ||Tomato, || sauce, || and meat | are | ingredients | to this recipe ||

Those boundaries (||) are acoustically marked either by pitch fall, as the intonational curve presents final down steps and it is reset at the next Intonational Phrase (IP), as by stop. Besides that, each MP will receive its own PA, which serves as a clue as well.

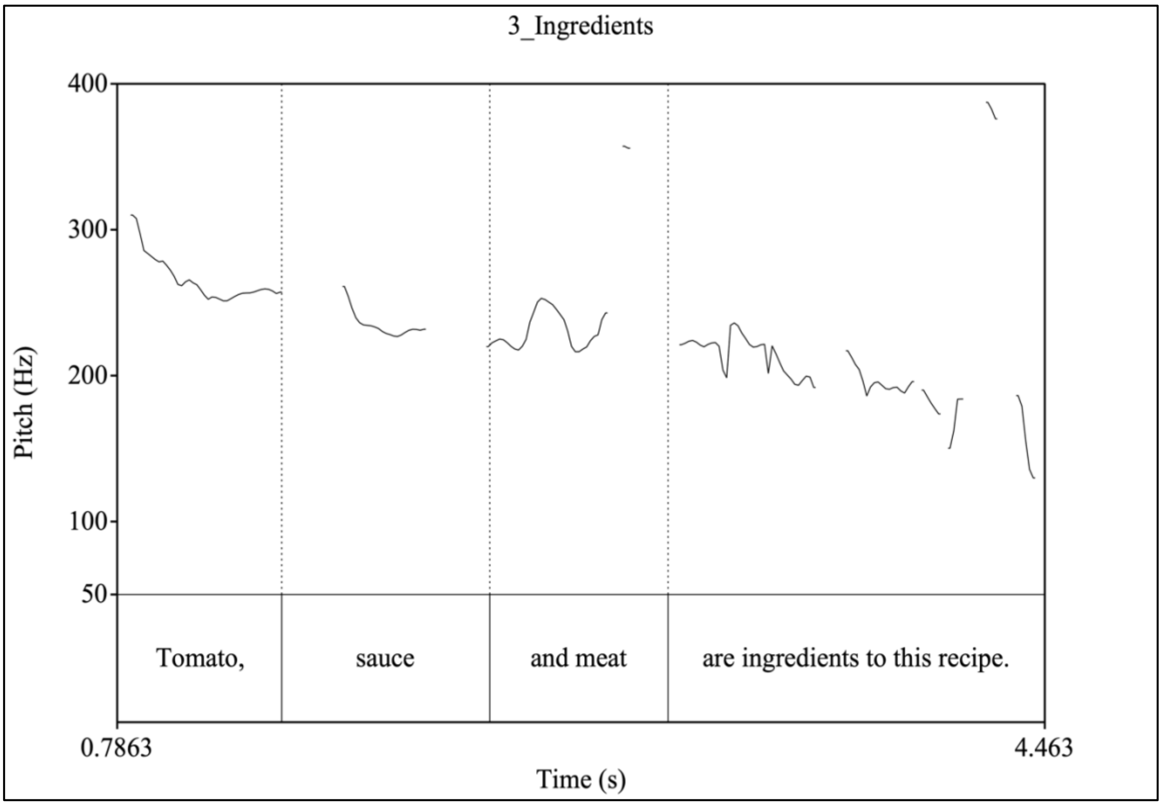


Figure 28: Pitch track of the sentence listing three ingredients.

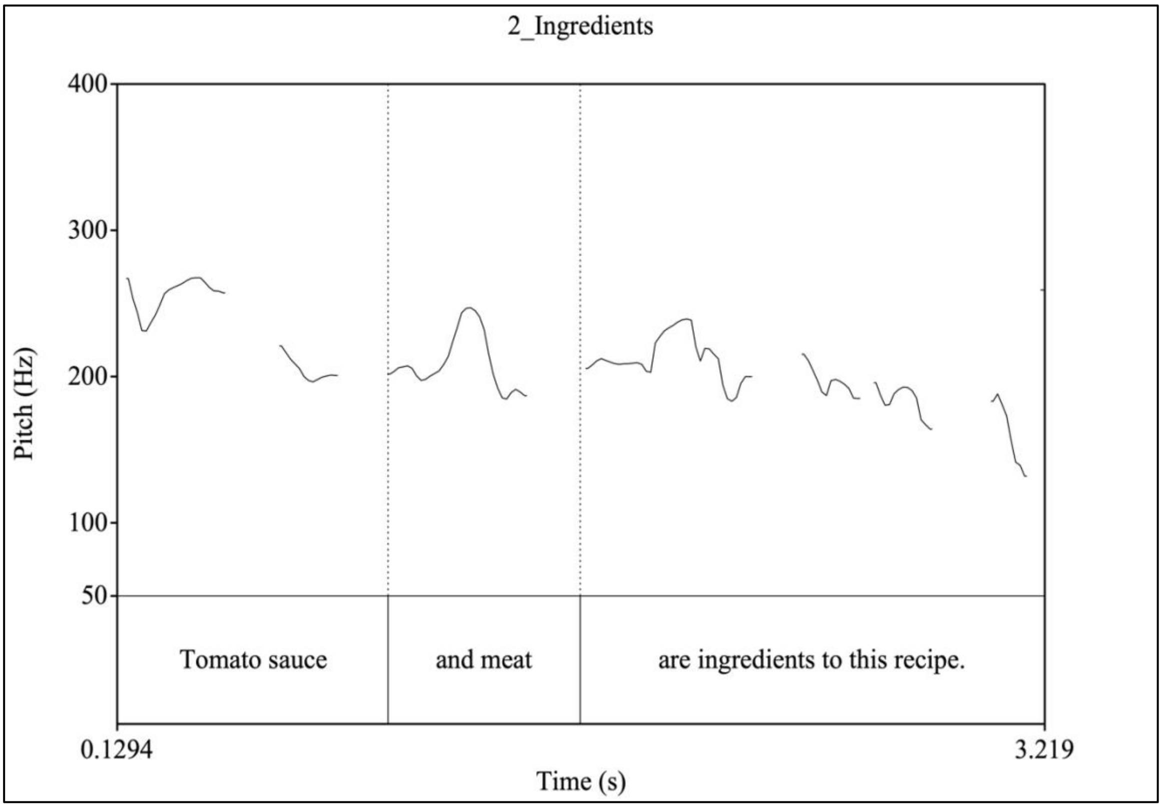


Figure 29: Pitch track of the sentence listing two ingredients.

Nevertheless, the prosodic cues are undoubtedly present and active in language processing, more specifically, syntax processing. The question to be raised is how independent the prosodic information is from other cues when mapping syntax. Are they informative enough to enable the listener to recover the syntax structure? Lexical entry plays probably the most important role in recovering syntax structure. How does the prosodic structure perform in giving the cues to syntax structure when there isn’t lexical basis that the listener can hold on to?

The major aim of this experiment is to observe the perception degree of the strong boundaries, and how independent that perception is from the lexical item recognition and propositional comprehension.

3.1.1 Hypotheses

Our hypothesis is that there are different levels of prosody processing. One of those levels would be in interface with syntax processing. At this level, prosody would use prosodic phrasing to properly map syntactic structure and relations into prosodic content – from now on, we will be calling this level *syntax-prosody level*.

At the syntax-prosody level, prosodic structure marks the syntactic structure for attachment or for coordination with different strategies (see Ladd, 1986, 1988; Ribeiro, 2015a, 2015b). The prosodic marking bootstraps syntax processing for coordination mechanisms (see Christophe, Guasti & Nespor, 1997; Name, 2008). The coordination mechanism is marked by a strong boundary (Ladd, 1988). We assume here that the information of syntax-prosody level is perceived as language information and, because of that, relevant information to syntactic processing.

3.1.2 Independent variables

Within the experimental approach, we understand that variables are the aspects that are being observed in a phenomenon. Independent variables are the aspects that we are manipulating when we design our stimuli. We usually consider independent variable those aspects that we assume are the most relevant to testing our hypothesis. In the case of this study, as the hypothesis to be tested is on the interaction between prosody and syntax, more specifically, in the interaction between prosody phrasing and syntax phrasing, we are interested in observing the perception of strong boundaries and how they relate to syntactic constituency. In order to observe that, we are manipulating two independent variables, described below.

3.1.2.1 Syntactic function of the target phrase

Given that prosody perception cannot escape phonetic and performance aspects as it is, to some extent, inseparable of its acoustic realization (see chapter 2.4 for further explanation on it), we are considering that the position of the boundary might interfere with the perception of it due to its acoustical saliency. Its saliency might change either due to acoustic realization, as the breath control changes throughout an utterance, or due to working memory restrictions. Therefore, we decided to test the difference of perception when a boundary is placed within i) *the subject, at the beginning of the sentence*:

(16) <u>The attendant and the client</u> understood the new rules of the store.

Or ii) *the object, at the end of the sentence*:

(17) The attendant disappointed <u>the manager and the colleagues.</u>

3.1.2.2 Manipulation of the intonational contour

The next aspect we are manipulating in our design targets the observation of how well a prosodic structure triggers the projection of a synctatic structure. More specifically, how well the cues for a intonational phrase is successful in mapping the proper synctatic structure. To do that, we are using complex constituents with either an embedded structure or a coordinated structure within it. The embedded structure is mapped into one single intonational phrase. The coordinated structure is mapped into two intonational phrases, one for each coordinated constituent. Therefore, when listening, we expect that the presence or absence of an intonational boundary will trigger a coordinated structure or an embedded structure, respectively. So the experimental sentences either have:

i) *presence of strong boundary* (||), generating a curve with two prominences, referent to a coordination of DPs:

(18a) <u>The attendant and the client</u> understood the new rules of the store *ii) absence of the strong boundary*, generating a curve with one PA, referent to a DP without coordination:

(18b) <u>The impolite attendant</u> understood the new rules of the store

3.1.3 Dependent variable

We understand that under the experimental approach, dependent variables are aspects that we are measuring, that generates our data. Within psycholinguistic field of research, the dependent variables are related to a behavior and, therefore, usually come from either an action or a reaction, elicited or not. To this experiment we had a task for each trial that will be explained below. For that task, the participant would have to press a key on a keyboard. We are measuring what key is pressed, the response index, and how long it takes for the participant to press a response key, the response time

Design 2x2	Subject (S)	Object (O)
With boundary (F)	SF	OF
Without boundary (N)	SN	ON

Table 1: Experimental design of experiment 1

3.1.4 Procedures and task

The experiment was a perception test, in which the participants were told to identify as much sentences as possible. To do that, they listened to delexicalized sentences.

The subjects were comfortably sat in front of a *core i7*, 17” Dell laptop. The experimental program was coded and run through *Open Sesame* platform. First, a screen with the consent form was displayed, where we explained that the experiment would represent no harm to the participant and that their personal information would be private. Then, several instruction screens were displayed, explaining the subject task. During instructions, I helped the participants to adjust the volume on the headphone, that was a *Kadosh, KH-800*. After making sure they understood what they were supposed to do, the program would move forward to a *Practice loop* with 5 (five) trials. After the Practice, if the participant was comfortable enough with their task, the program would move forward to the actual testing loop.

Both the Practice loop and the *Testing loop* consisted of a preparation screen, a blank screen with audio and a task screen to each trial.

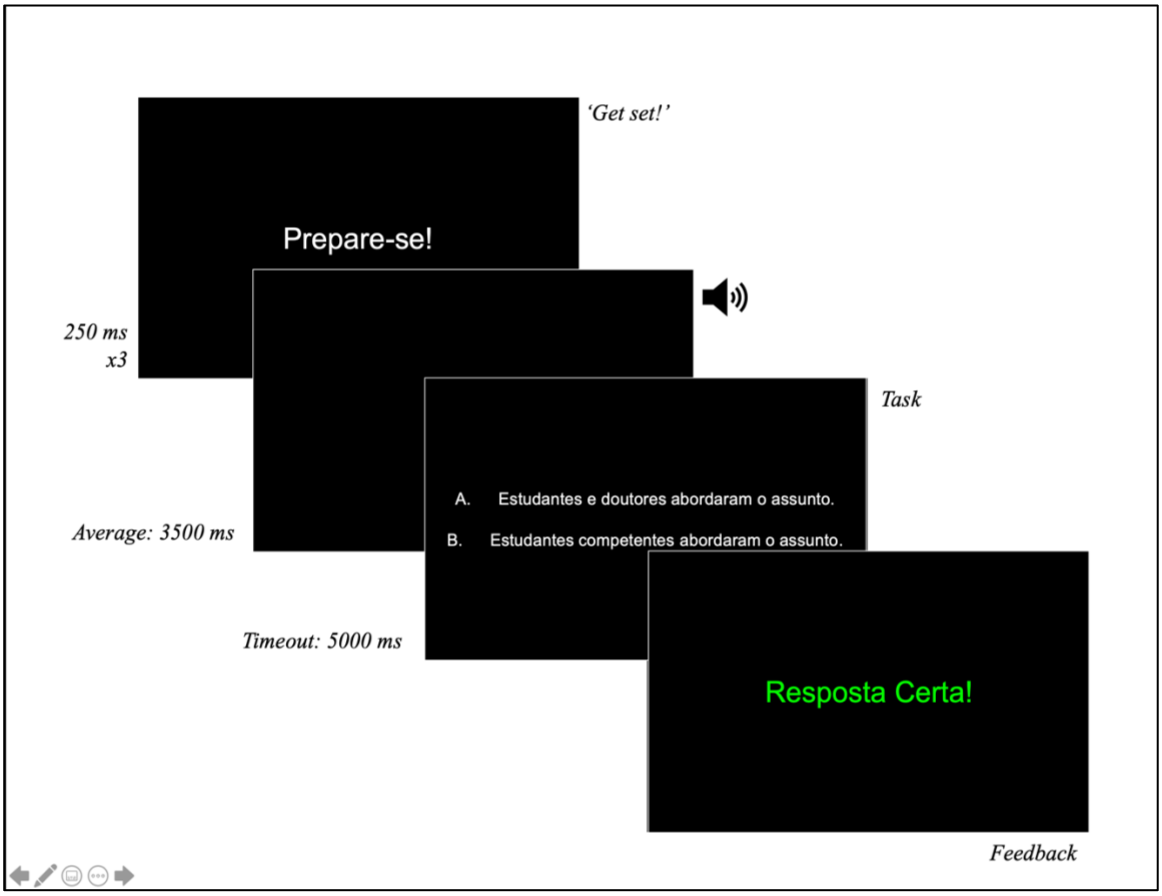


Figure 30: Testing loop trial sample

The preparation screen displayed the message *Prepare-se!* (*Get set!*), which would blink three times, each of it been displayed for 250 milliseconds (ms). Then, automatically, a blank screen was displayed and the audio file corresponding to the recorded sentence of that trial would play. At the end of each listening act, one screen with two options of sentences was shown. At this point, the participant must have chosen, between the two options, the sentence that they thought it was the closest with the one they had just listened to, which was delexicalized, that is, the high frequencies had been taken off of the audio file, maintaining the pitch contour, but making it impossible to understand the lexical item. The participant monitored only the transition from the task screen to the next screen, by answering the task. After the task screen, a feedback screen was displayed with the message *Resposta certa!*

(*Correct answer!*) or *Resposta errada!* (*Wrong answer!*), accompanied by a tune: 440Hz *sine waveform* to positive feedback, and 440Hz *saw waveform* to negative feedback.

Figure 31: "Correct answer!" feedback screen

Figure 32: "Wrong answer!" feedback screen

The whole experiment program was formatted to display black background with white foreground, with the exception of feedback screens that had green/red foreground.

3.1.5 Materials

Twenty sets of four stimuli (20 x 4), one to each condition (1 x 4), and forty (40) fillers/distractors were created. The 80 experimental sentences were then recorded and treated. Each of them was manipulated, based on a delexicalization technique: *low pass filter*, that is, sound manipulation in which high frequencies are taken off, maintaining the pitch contour, but making it impossible to understand the lexical item. We used *Praat*® software to filter the sentences, establishing a band from 0Hz to 450Hz.

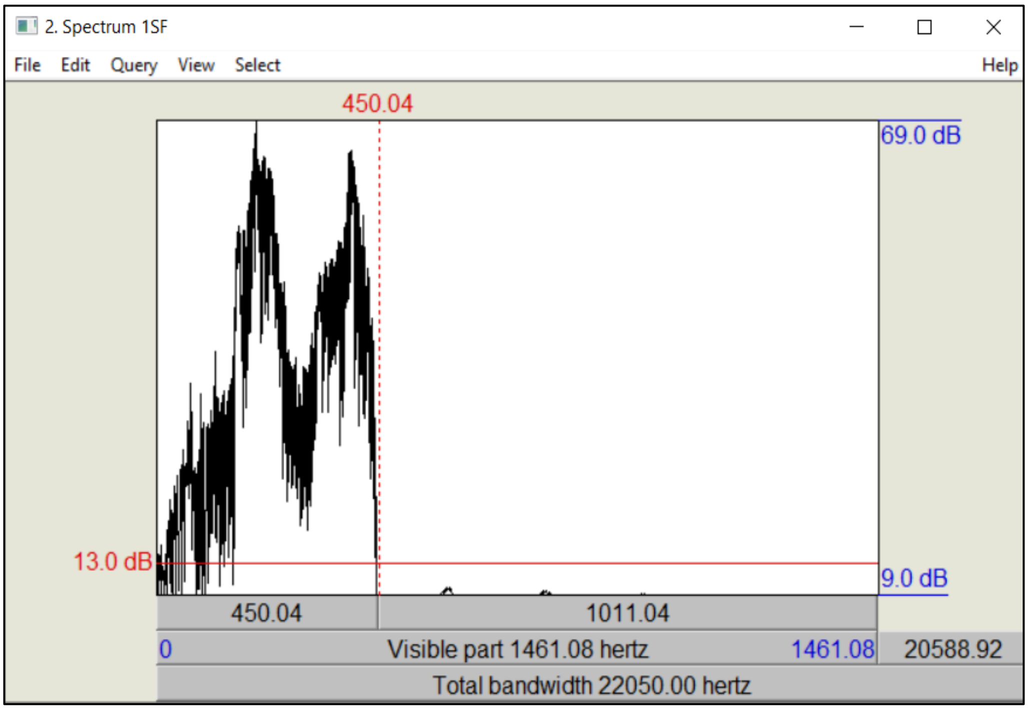


Figure 33: Spectrum of stimuli 1SF, showing maximum frequency at 450Hz

The 40 distractor sentences were controlled in a way that there wouldn't be strong boundaries inside the phrases. All the sentences were pseudorandomized, controlling the number and order of the experimental sentences. There were four experimental lists of stimuli, generated by *Latin-square*, so that the experiment didn't take too much time, taking about 15 minutes.

3.1.5.1 Set 1 – SF condition

(1Sfa) Estudantes e doutores abordaram o assunto

Students and PhDs addressed the issue

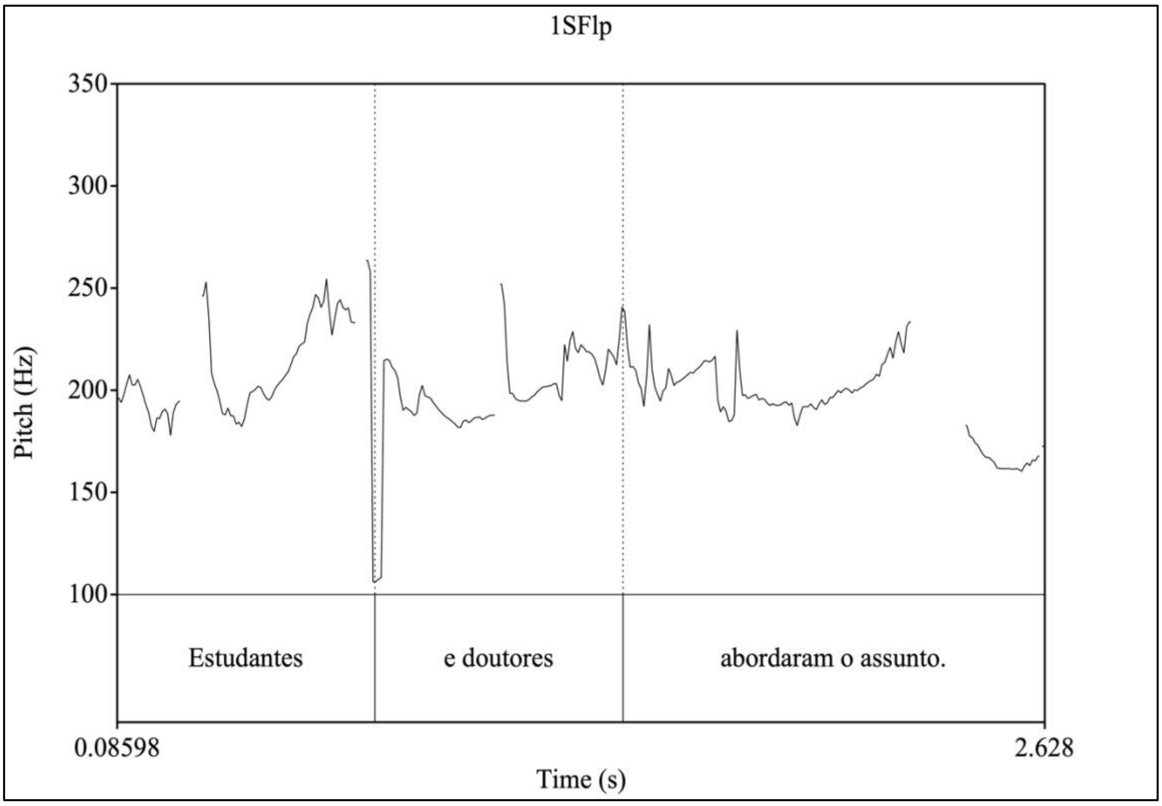


Figure 34: (1SFb) – experimental sentence represented by diagram

Task – (A) Estudantes e doutores abordaram o assunto

Students and PhDs addressed the issue

(B) Estudantes competentes abordaram o assunto

Competent students addressed the issue

3.1.5.2 Set 1 – SN condition

(1Sna) Estudantes competentes abordaram o assunto

Competent students addressed the issue

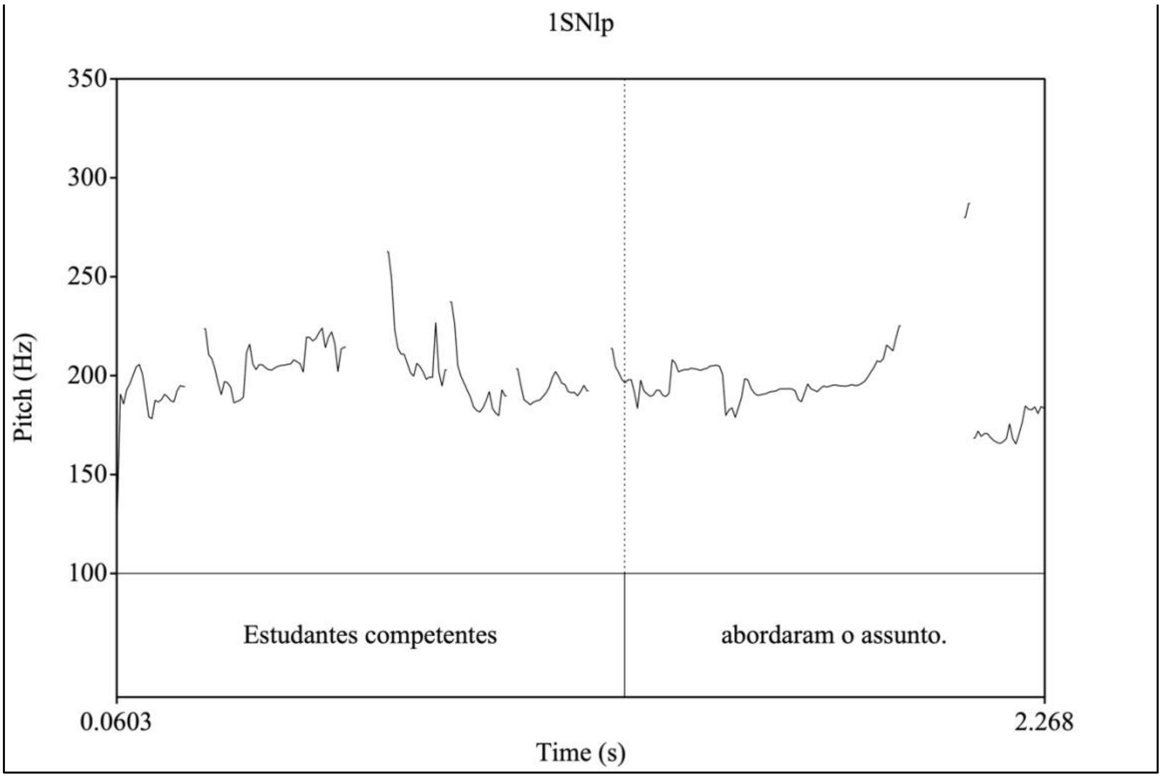


Figure 35: (1SNb) – experimental sentence represented by diagram

Task – (A) Estudantes competentes abordaram o assunto

Competent students addressed the issue

(B) Estudantes e doutores abordaram o assunto

Students and PhDs addressed the issue

3.1.5.3 Set 1 – OF condition

(1Ofa) Estudantes abordaram o assunto e a agenda

Students addressed the issue and the agenda

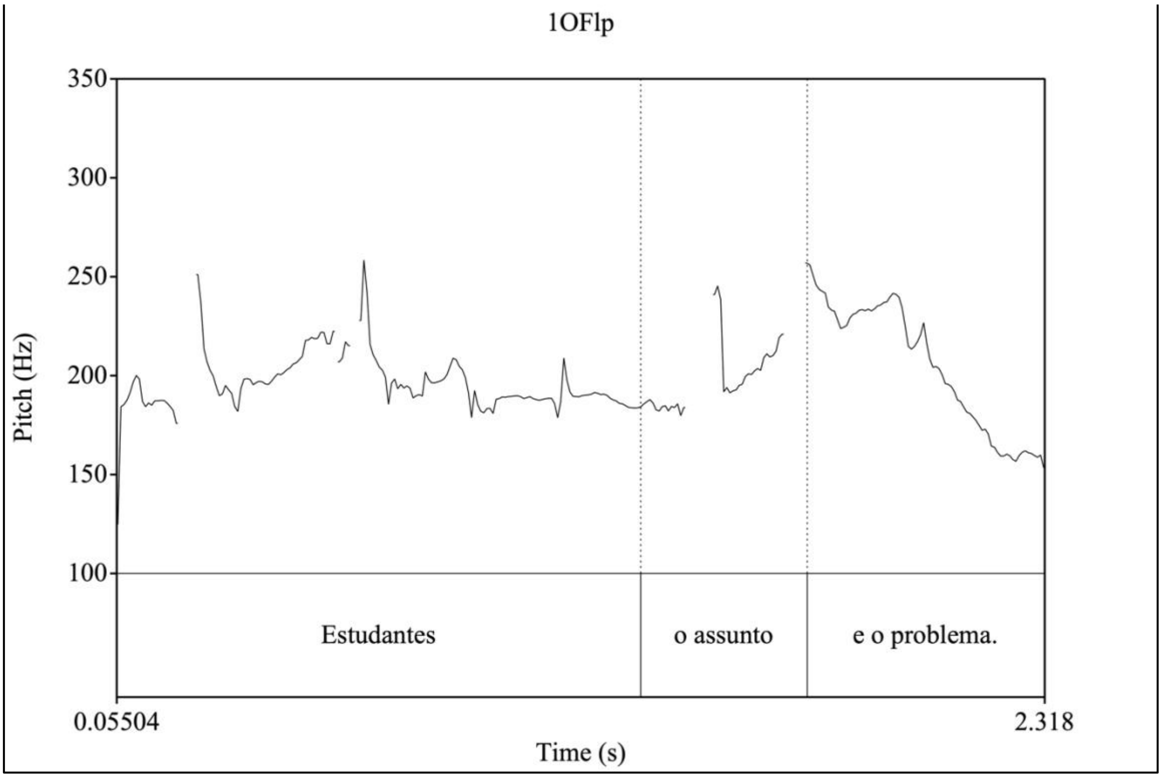


Figure 36: (1Ofb) – experimental sentence represented by diagram

Task – (A) Estudantes abordaram o assunto e a agenda
Students addressed the issue and the agenda
(B) Estudantes abordaram o assunto importante
Students addressed the important issue

3.1.5.4 Set 1 – ON condition

(1Ona) Estudantes abordaram o assunto importante
Students addressed the important issue

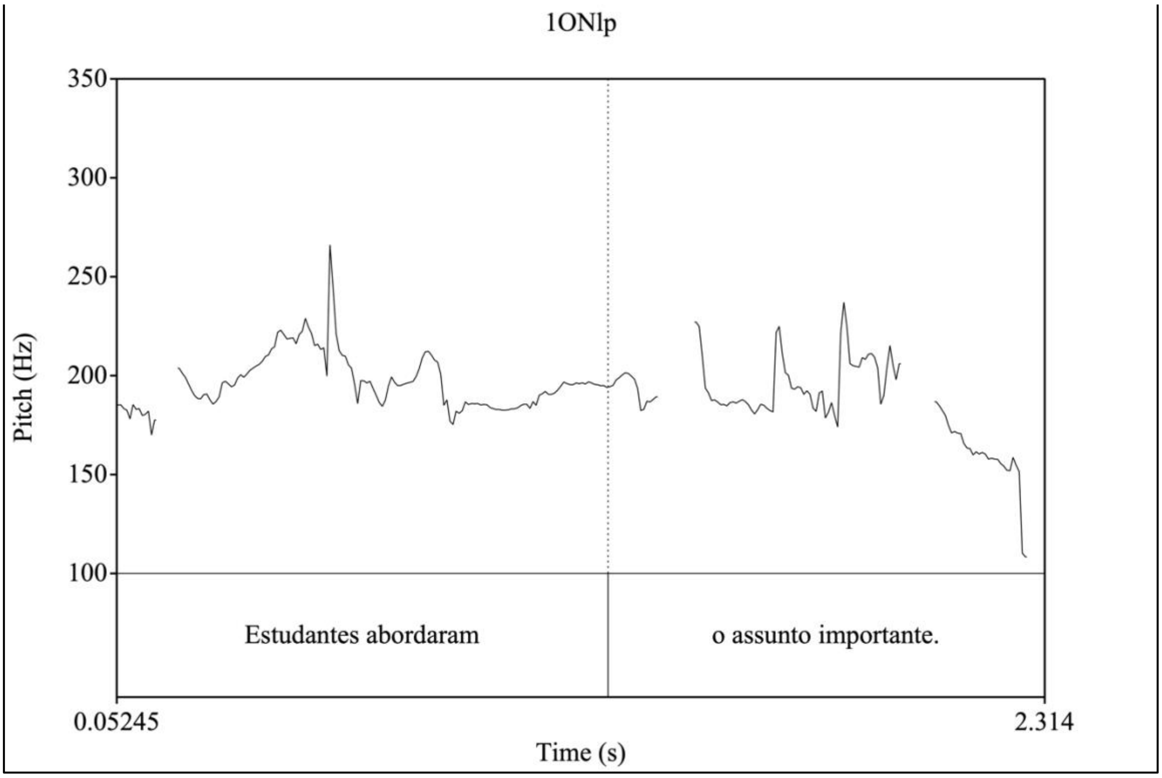


Figure 37: (1Onb) – experimental sentence represented by diagram

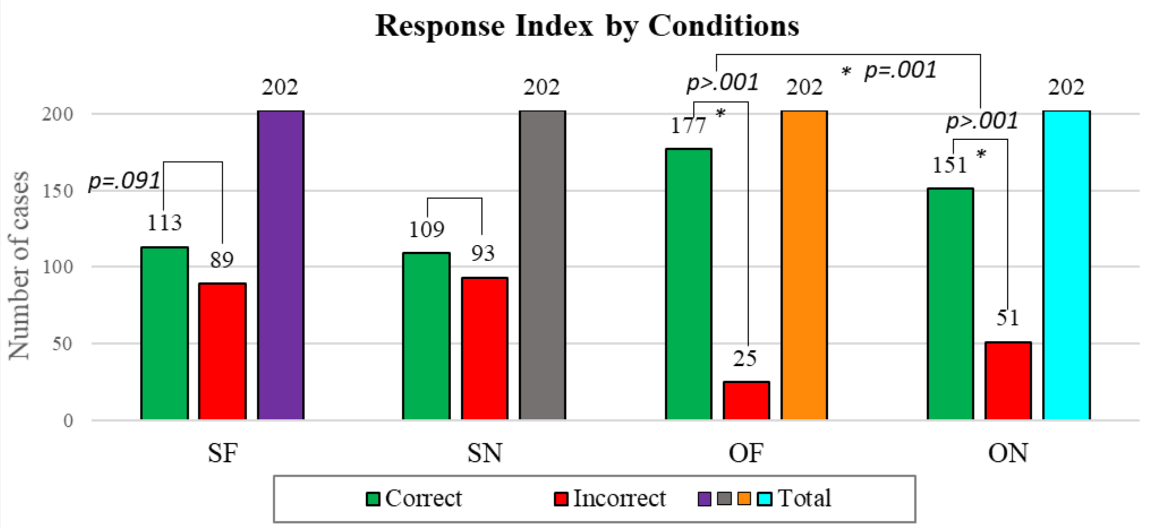
Task – (A) Estudantes abordaram o assunto importante
Students addressed the important issue
(B) Estudantes abordaram o assunto e a agenda
Students addressed the issue and the agenda

3.1.6 Predictions

The answer indexes will be satisfactory (more successful answers than mistakes) to all conditions. The response time will be larger to conditions SN and ON, due to the comparison complexity between the possible sentences and the absence of a prosodic cue to perceive the sentence structure.

3.1.7 Results

The response indexes and response times were measured and compared to pursue the perception degree and processing cost. All the results were statistically analyzed using *IBM® SPSS® Statistics, version 20*. An alpha level of .05 was used for all statistical tests. We gathered a total of 840 cases for study. Due to low engagement and some issues during the conduction of the experiment, one participant's data were discarded, leaving us with 820 cases. Once the data were filtered, and the outliers were taken out, there was a total of 808 cases to study. Here's their distribution:

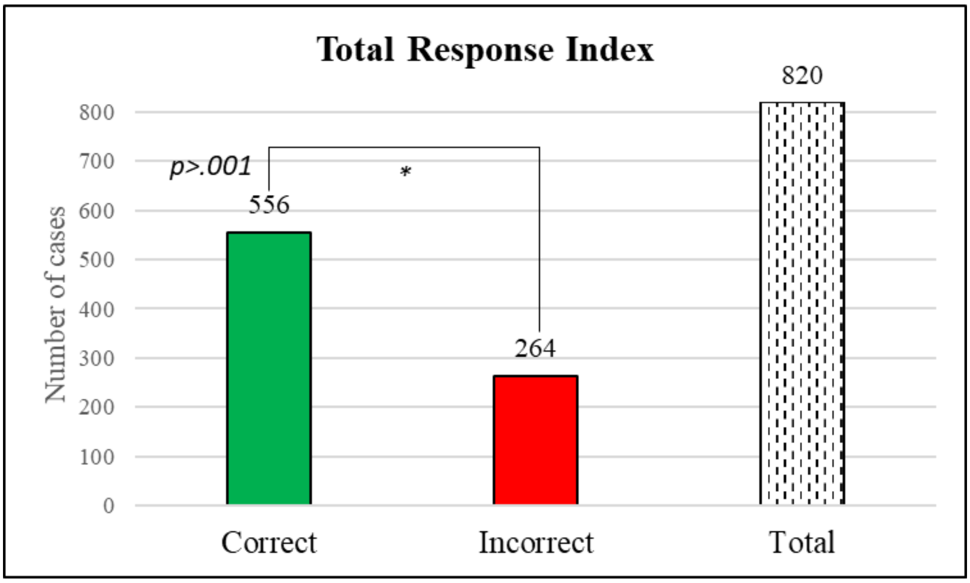


SF	**Complex Subject with Prosodic Boundary Marking Coordination of Nouns**
	Estudantes e prefeito divulgaram a proposta.
	student-PL COP mayor disclose-PERF.PAST.3.PL ART.DEF.F proposal
	"Students and Mayor disclosed the proposal"
SN	**Complex Subject without Prosodic Boundary Signalizing Attachment of Nouns**
	Estudantes engajados divulgaram a proposta.
	student-PL commit.PART.PAST.M-PL disclose-PERF.PAST.3.PL ART.DEF.F proposal
	"Committed students disclosed the proposal"
OF	**Complex Object with Prosodic Boundary Marking Coordination of Nouns**
	Estudantes divulgaram a proposta e o problema.
	student-PL disclose-PERF.PAST.3.PL ART.DEF.F proposal COP ART.DEF.M problem
	"Students disclosed the proposal and the problem"
ON	**Complex Object without Prosodic Boundary Signalizing Attachment of Nouns**
	Estudantes divulgaram a proposta rejeitada.
	student-PL disclose-PERF.PAST.3.PL ART.DEF.F proposal reject-PART.PAST.F
	"Students disclosed the rejected proposal"

Graph 1: Response indexes counted by condition and quality (correct or incorrect), sided by the total number of response indexes to that condition

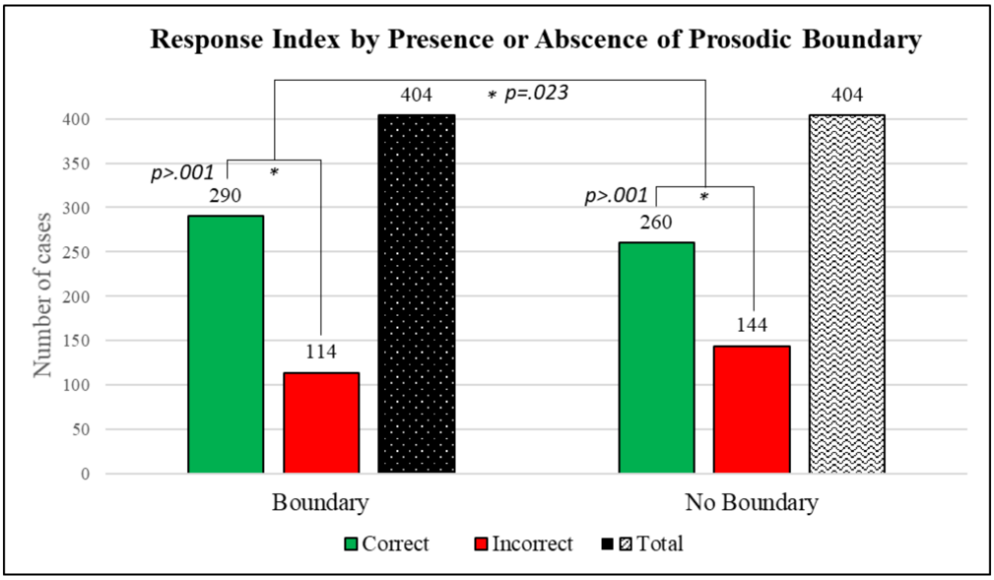
Graph 1 shows the response indexes distributed by condition. The samples for each condition are composed of 202 cases. For all conditions, the number of correct answers is larger than the number of incorrect answers. Condition SF shows 24 more cases of correct answers than incorrect ones. For condition SN, we have a difference of 16 cases. Condition OF shows a difference of 152 cases, and condition ON, 100 cases.

A *Chi-square goodness-of-fit* was conducted for each condition, in which an equal distribution of correct and incorrect cases (chance level) was established as null hypothesis. The test showed a statistically significant *relevance* in rejecting the null hypothesis for conditions OF ($X^2(1) = 114{,}376$, $p<.001$) and ON ($X^2(1) = 49{,}505$, $p<.001$). A *tendency* in rejecting the null hypothesis can be observed for condition SF ($X^2(1) = 2{,}851$, $p = .091$). The null hypothesis prevails for condition SN ($X^2(1) = 0{,}970$, $p = .325$). A *Hypothesis test* was conducted, between pairs of conditions, in which the null hypothesis was that the distribution of different values across both conditions were equally likely. The test showed a statistical significance of $p<.001$ in rejecting the null hypothesis when comparing the conditions OF with ON. The other combinations did not reach statistical significance.



Graph 2: Total response indexes sorted by quality (correct or incorrect).

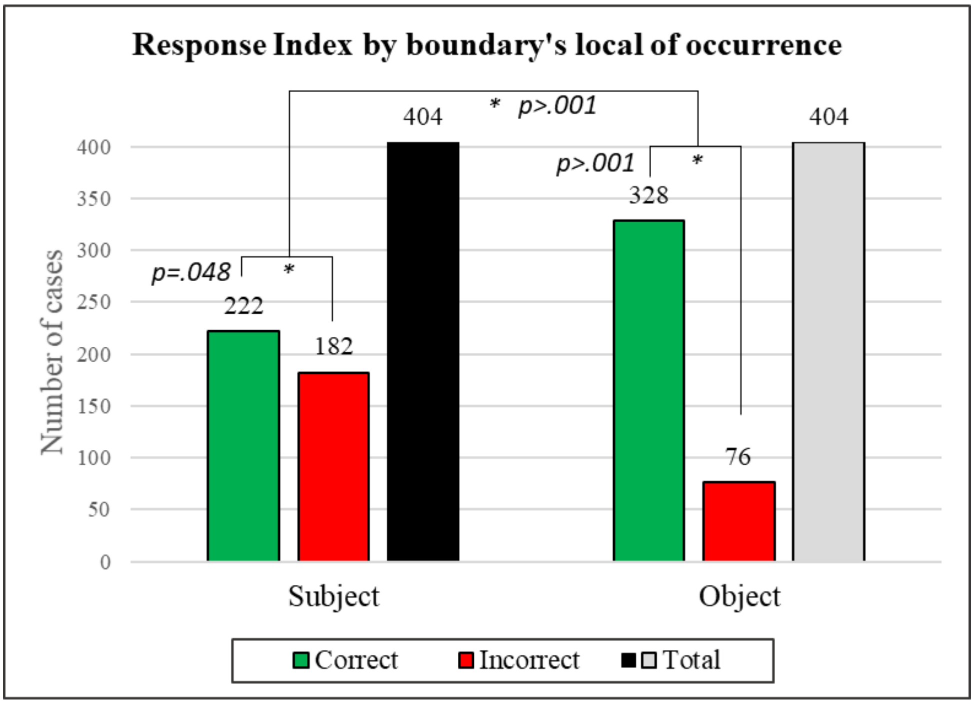
To test if the overall data show a tendency for correct answers cases, we combined the samples. Note that we have a total of 820 cases (and not 808, expected from the sum of previous graph's totals), as we didn't need to equal the number of cases on each condition. Another Chi-square goodness-of-fit test was conducted and showed a statistically significant relevance in rejecting the null hypothesis ($X^2(1) = 103{,}98$, $p<.001$), i.e., a statistically significant relevance in rejecting that the distribution between correct and incorrect cases is chance leveled.



Graph 3: Response indexes sorted by presence or absence of prosodic boundary.

To isolate the *Prosodic Boundary factor*, we combined SF with OF samples and SN with ON samples. The results show that, to *Boundary conditions*, Correct answer cases outnumbered Incorrect answer cases by 176. Similarly, to *No Boundary conditions*, Correct answer cases outnumbered Incorrect number cases by 116.

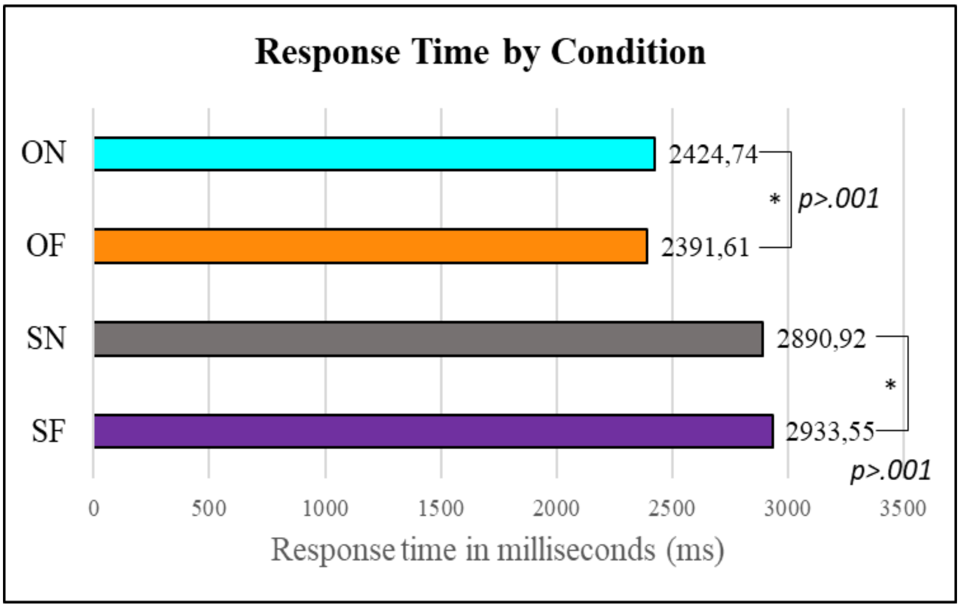
Again, a Chi-square goodness-of-fit test were conducted, showing a statistically significant relevance in rejecting the null hypothesis for both conditions (Boundary: $X^2(1) = 76{,}673$, $p<.001$; No Boundary: $X^2(1) = 32{,}346$, $p <.001$). In addition, a Hypothesis test was conducted, in which the null hypothesis was that the distribution of different values across Boundary condition and No Boundary condition was equally likely. The test showed a significance of $p = .023$ in rejecting the null hypothesis.



Graph 4: Response indexes sorted by local of occurrence of the prosodic boundary, i.e., phrase with coordination or attached item.

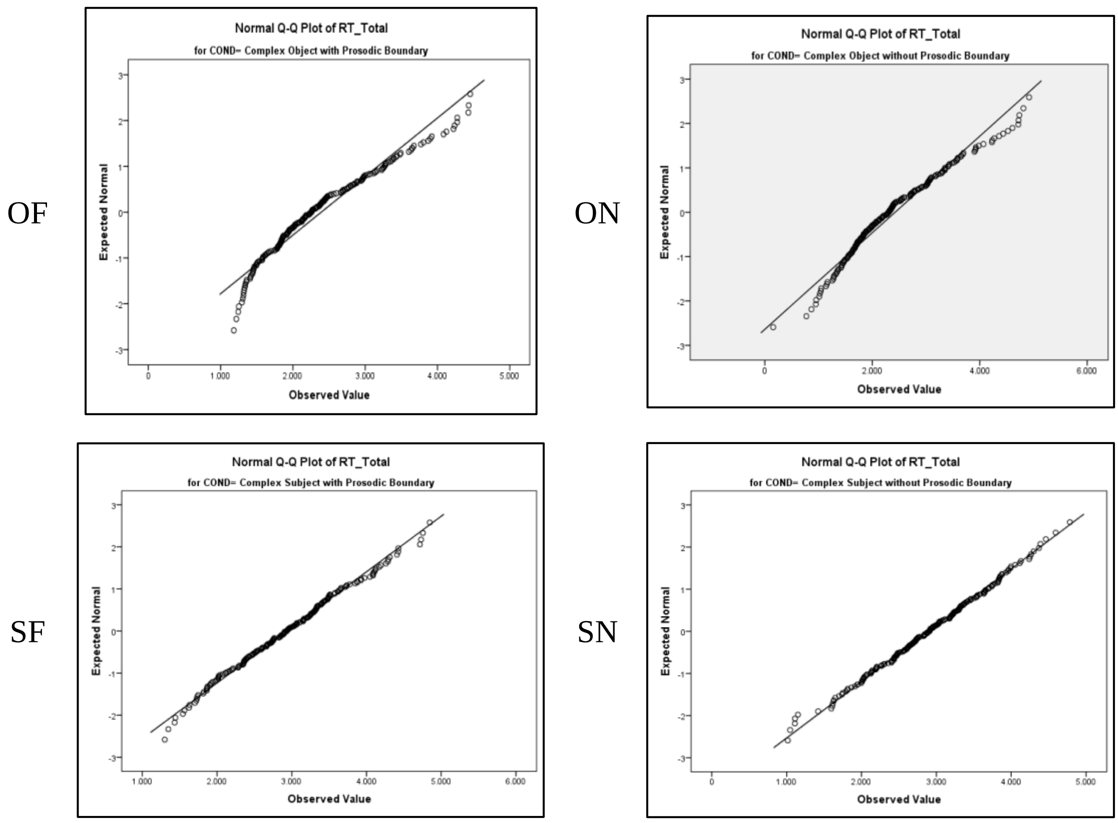
To isolate the Local-of-occurrence factor, we combined SF sample with SN sample and OF sample with ON sample. The results show 40 more cases of correct answers to Subject condition; and 252 more cases of Correct answers than Incorrect answers to Object condition. We can notice a big difference between Subject and Object samples. Object condition shows a greater distance between the number of Correct answers and Incorrect answers.

Once more, a Chi-square goodness-of-fit test was conducted. In both cases, it showed a statistically significant relevance in rejecting the null hypothesis (Subject: $X^2(1) = 3{,}902$, p=.048; Object: $X^2(1) = 154{,}888$, p <.001). Besides, a Hypothesis test was conducted and showed a statistically significant relevance in rejecting the null hypothesis (p<.001), i.e., in rejecting that the distribution of different values across Subject condition and Object condition was equally likely.



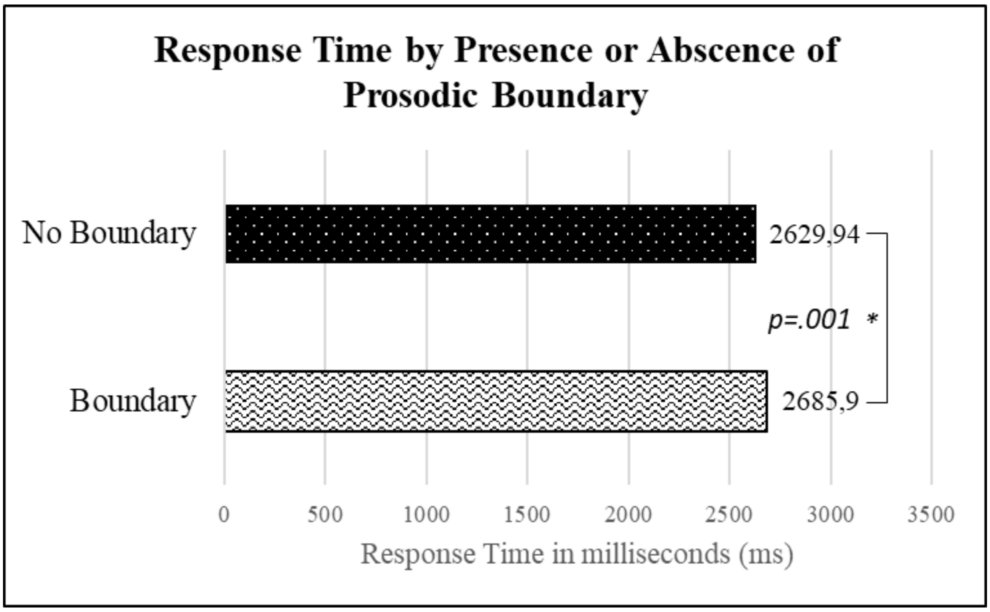
Graph 5: Mean of response times sorted by condition

The Response time results show the following averages (from the highest to the lowest): SF: 2933,55ms; SN: 2890,92ms; ON: 2424,74ms; and OF: 2391,61ms. *A Shapiro-Wilk test of normality* was conducted. OF and ON samples were shown not normal (OF: *p*<.001; ON: *p*<.001). SF and SN samples showed no statistical significance in rejecting the null hypothesis, i.e. the samples were normally distributed (SF: *p* =.263; SN: *p* =.805).



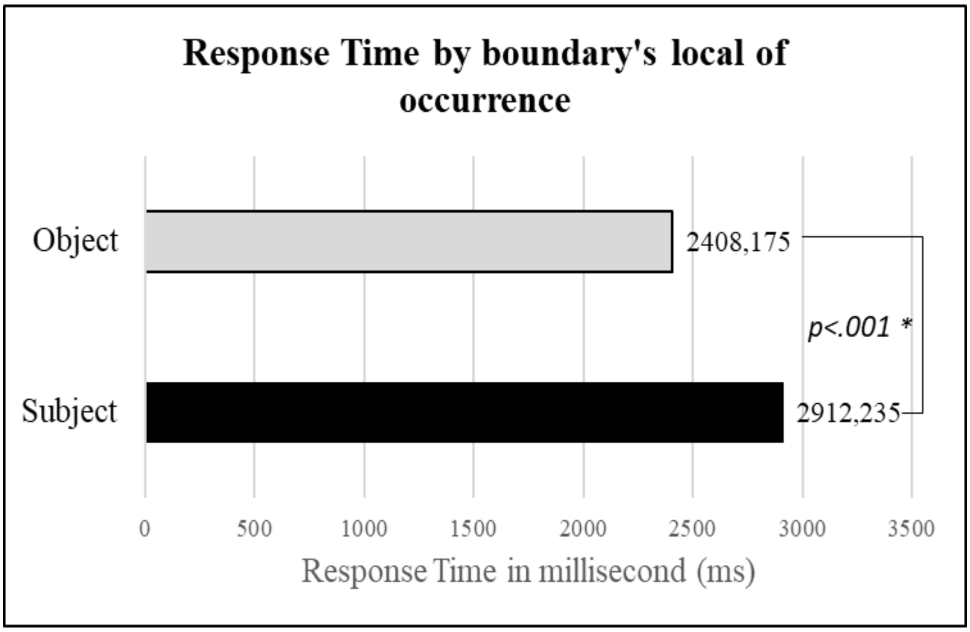
Graph 6: *Normal Q-Q plot* of response times by condition

A *Wilcoxon Signed test* was conducted to the differences in occurrence distribution between groups. The test showed statistically significant relevance in rejecting the null hypothesis, i.e., the sample distribution wasn't likely, when comparing SF with SN (Z = -11,604, p<.001) and OF with ON (Z = -3,770, p<.001).



Graph 7: Average of response times sorted by presence or absence of prosodic boundary

To isolate the Prosodic Boundary factor, we combined SF sample with OF sample and SN sample with ON sample. The results show a mean of 2685,9ms to Boundary condition and 2629,94ms to No Boundary condition. A *Wilcoxon Signed-Rank test* showed that, although the means seems very close, the samples are significantly distinct in statistical terms (Z = -3,285, p=.001).



Graph 8: Average of response times sorted by local of occurrence of the prosodic boundary

To isolate the *Local-of-occurrence factor,* we combined SF sample with SN sample and OF sample with ON sample. The graph shows a mean of 2408,175ms to Object condition and a mean of 2912,235ms to Subject condition. Once more, a Wilcoxon Signed test was conducted. The test showed statistical significance in rejecting the null hypothesis (Z = -17,394, p<.001), i.e., the difference in value distribution between the samples is relevant.

3.1.8 Linguistic analysis

The data show that the total number of correct answers is greater than the total number of incorrect answers (Graph 2), pointing toward a successful perception of intonational curve. As we had delexicalized sentences, it seems that the prosodic information is enough to trigger syntax processing. Although we don’t need lexical information to perceive syntax structure through prosody, lexical item support seems to be important to memory. We can say that, by observing the indexes to Subject conditions versus Object conditions (Graph 4). Object conditions have greater number of correct answers. The difference in Subject rates and Object rates can be caused by two factors: recency effect or subject-object asymmetry. Recency effect would cause the subject acoustic information to be harder to retrieve for task purposes than the object acoustic information as the subject information was presented earlier than the object information. According to the theory (Baddeley & Hitch, 1993), our short-term memory has a short capacity to hold information. Once the information is displaced, the retrieval is harder. Subject-object asymmetry would cause the mapping of the object and the mapping of the subject to be different. That would be due to the basic structure of a sentence being asymmetric, as the verb and object of a sentence form a constituent that exclude the subject (Baker, 1991). Also, the response times show that Subject sentences were harder to be identified, as the task took longer on those cases (Graph 8). Because of that, it seems very possible that both factors, recency effect and asymmetry, are contributing to it.

Despite the differences in indexes rate and response times between Subject and Object conditions, when we isolate the Boundary factor, we can see that the presence of a boundary, i. e., the realization of a prosodic mark, facilitates the identification of coordination structure (Graph 3). We see in Graph 3 that the absence of the boundary was also enough to exclude the coordination structure option, as we have larger rate of correct answers to both matters. One can say that it is the Object condition that is carrying those rates, and s/he will not be totally wrong about it. We can see in Graph 1 that the OF and ON conditions are the only conditions with statistical significance to difference in index rates. We can only see a tendency to SF and absolutely no statistical significance to SN. Once again, as argued before, it seems to be a memory issue, aside with a processing asymmetry. But although we don't have significant difference in index rates to SF and SN, when we combine the samples, we can see that participants made more correct choices than incorrect ones, thus supporting the hypothesis.

3.2 EXPERIMENT 2 – SYNTAX AND PROSODY: PROSODY PHRASING PERCEPTION IN DELEXICALIZED SENTENCES IN AMERICAN ENGLISH (AE)

The experiment that was ran in Brazilian Portuguese (BP) was supposed to be replicated in American English (AE), but there were some challenges, and some changes had to take place. Prosody phrasing rules vary from language to language and the prosodic structure generated by the stimuli to BP experiment would not be similar to the prosodic structure generated by AE sentences with same syntactic structure. Coordination of phrases in English do not map a strong boundary in prosody when it is placed inside subject/object domain, or, at least, it is not perceived as strong. Therefore, there was a need to make it more salient. In order to do that, the verbs chosen to the stimulation are indirect transitive verbs, and the coordination was placed inside the indirect object domain. Due to preposition attachment, the boundary would become more salient and properly perceived.

Also, after Experiment 1 results, we decided not to test subject in comparison to object position due to recency effect

3.2.1 Hypotheses

The Hypotheses for this experiment are the same as the hypothesis for the experiment described on item 3.1. Hence, we are assuming that prosody processing has different levels of processing and representation, and that one of those levels maps syntax structure into prosodic content. We assume that different syntactic structure will be mapped into different prosodic structures, and that the difference among prosodic structures is salient enough to trigger the proper syntax structure during sentence comprehension.

3.2.2 Independent variables

To this study, as explained before, we will not be observing different constituents as on the previous experiment. Therefore, only one independent variable was selected.

3.2.2.1 Manipulation of the intonational contour

Similarly to the previous experiment, we are testing if the perception of a strong boundary will trigger the syntax structure for coordination of items when placed within an object, and if the absence of the boundary will trigger an embedded constituent. Therefore, the experimental sentences were distributed into two groups, based on:

a) *presence of strong boundary (||)*, generating a curve with two prominences, referent to a coordination of PPs (*to Jason and to Ashley*); or
b) *absence of strong boundary*, generating a curve with one prominence, referent to a PP without coordination (*to Jason Parreno*).

3.2.3 Dependent variable

Two off-line measures, i.e., response time and response index.

Design 1x2

Strong boundary	No boundary
SB	NB

Table 2: Experimental Design 1x2

3.2.4 Procedures and task

The experiment was a perception test similar to the Experiment 1 in Brazilian Portuguese. The participant was instructed to listen to a (delexicalized) sentence and then choose, between two options, the one that fits better to the sentence listened.

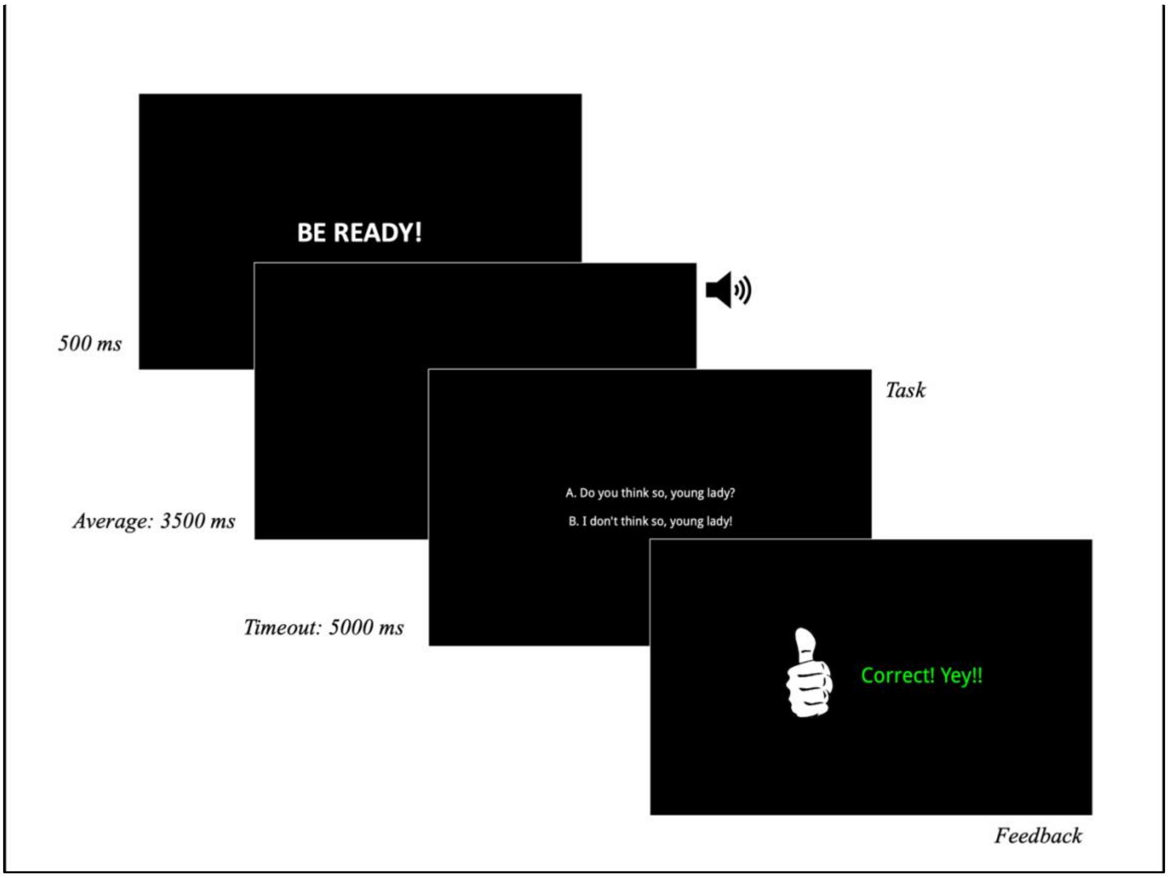


Figure 38: Trial sequence to experiment 2

After responding to it, a feedback screen would show a *right* or *wrong* image with different sound cues, accompanied by synthesized tunes.

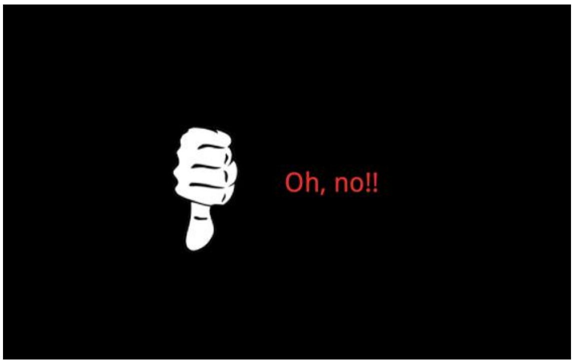


Figure 39: Negative feedback

Figure 40: Positive feedback

3.2.5 Materials

To test the hypothesis, it was created 10 sets of two (10 x 2) stimuli sentences, one to each experimental condition, and 20 fillers/distractors. The sentences were controlled in length and stress placement. The structure was nominal subject + verb + direct object + indirect object + adverbial phrase. The indirect object would be duplicate by coordination (i.e., *to Jason and to Ashley*) or it would have a complex Noun Phrase (i.e., *to Jason Parreno*). Each condition would generate a different pitch contour as can be seen bellow.

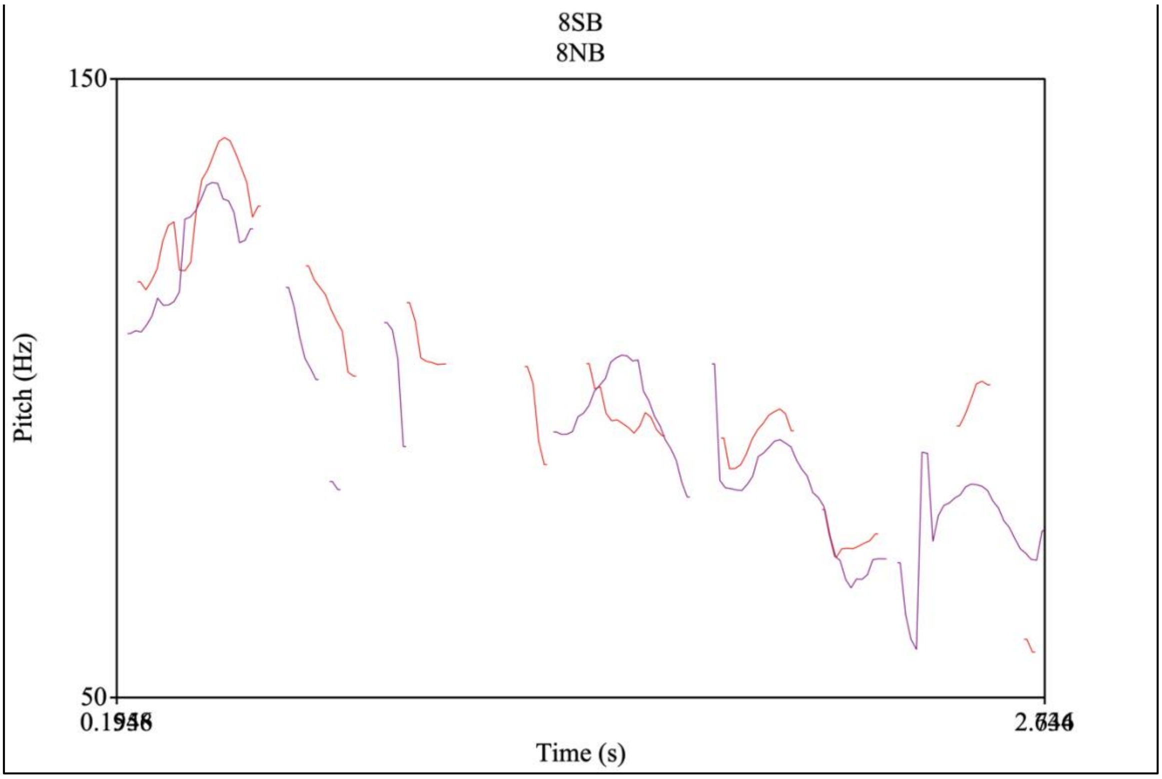


Figure 41: Pitch tracks of two experimental sentences
(8SB – in purple): Gregory showed pics to Caleb and to Mary at the meeting
(8NB – in red): Gregory showed pics to Caleb Gallagher at the meeting

All experimental sentences were revised by my supervisor Eva Fernández and my American informant TP[31]. The sentences were recorded in a sound proof lab, with professional microphone brand Blue Yeti. They were then manipulated using the low pass filter using *Praat*® software, establishing a band from 0Hz to 450Hz.

[31] His name is not displayed for privacy.

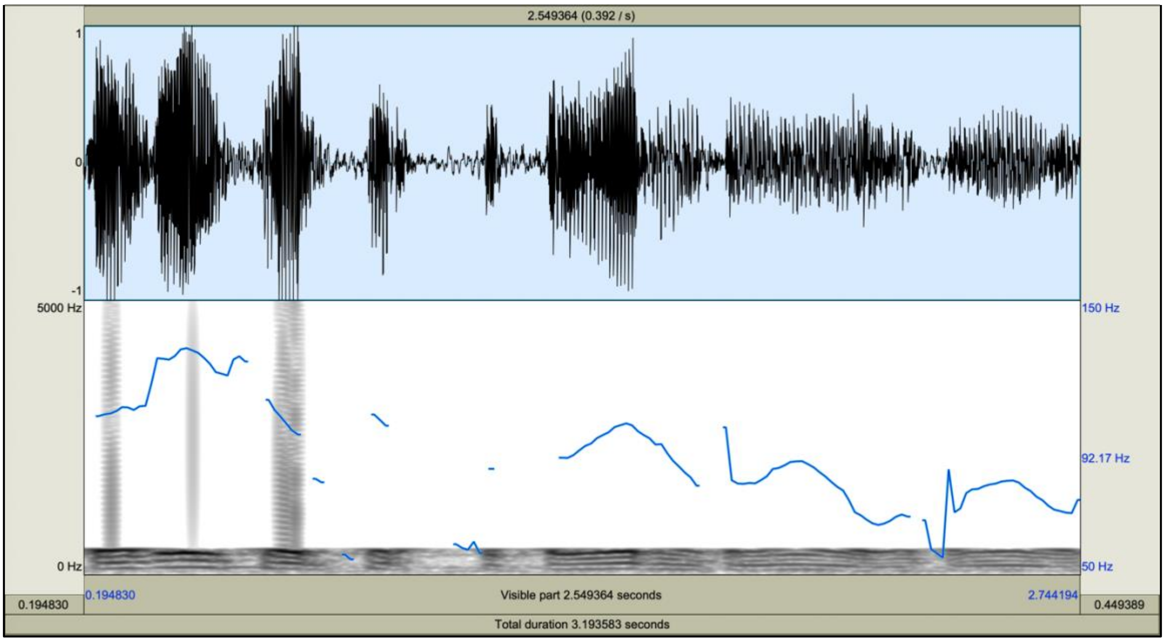


Figure 42: Spectrogram and pulses view of Praat® window displaying 8SB sentence pitch track

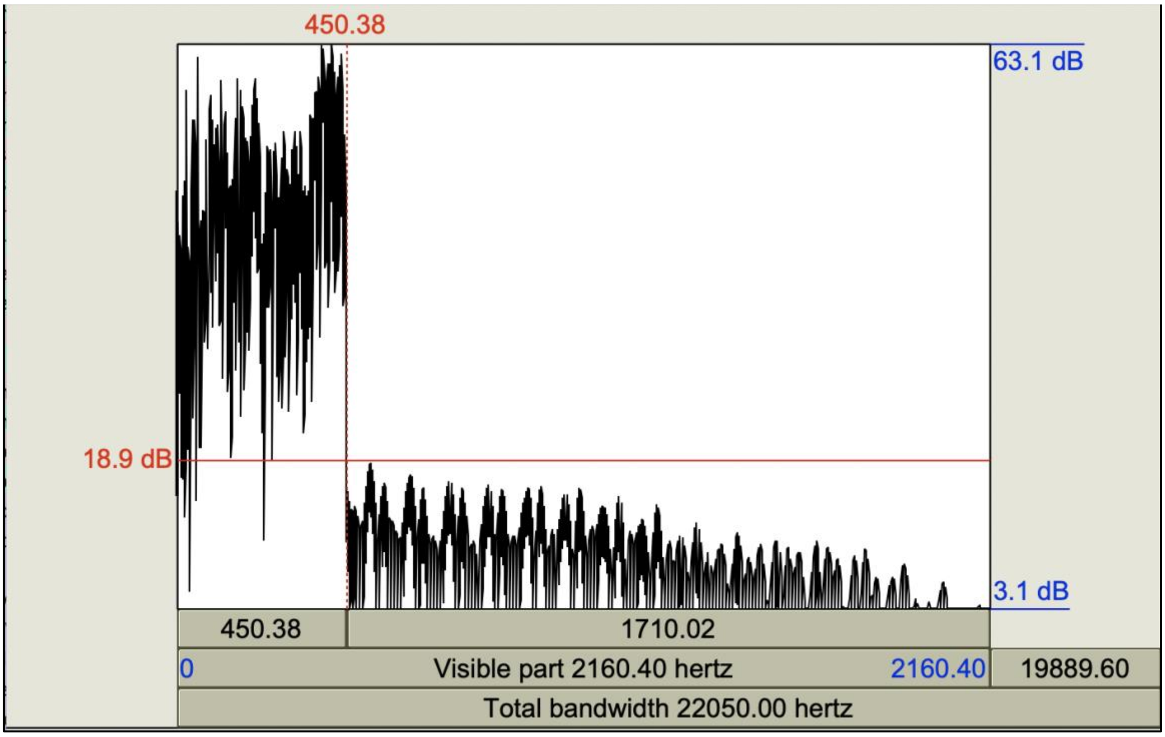


Figure 43: Spectrum window view of sentence 8SB

3.2.5.1 Set 1 – SB condition

(1Sba) *Gregory showed pics to Caleb and to Mary at the meeting*

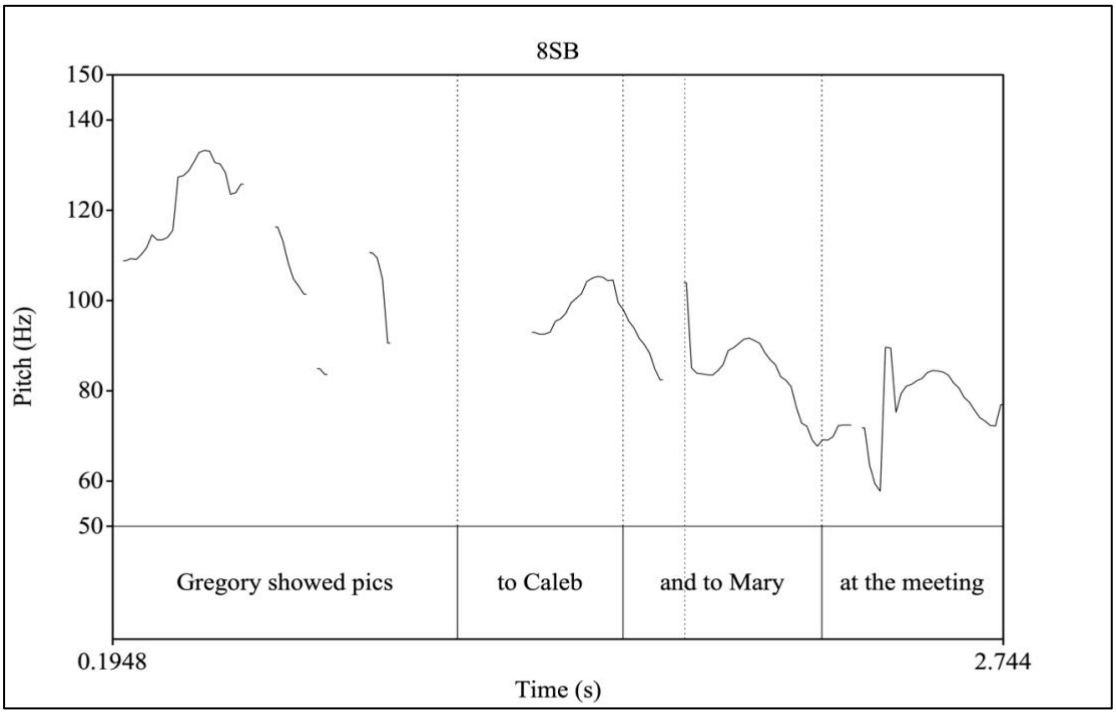


Figure 45: (1SBb) – experimental sentence pitch drawing

Task – (A) *Gregory showed pics to Caleb Gallagher at the meeting*

(B) *Gregory showed pics to Caleb and to Mary at the meeting*

3.2.5.2 Set 1 – NB condition

(1Nba) *Gregory showed pics to Caleb Gallagher at the meeting*

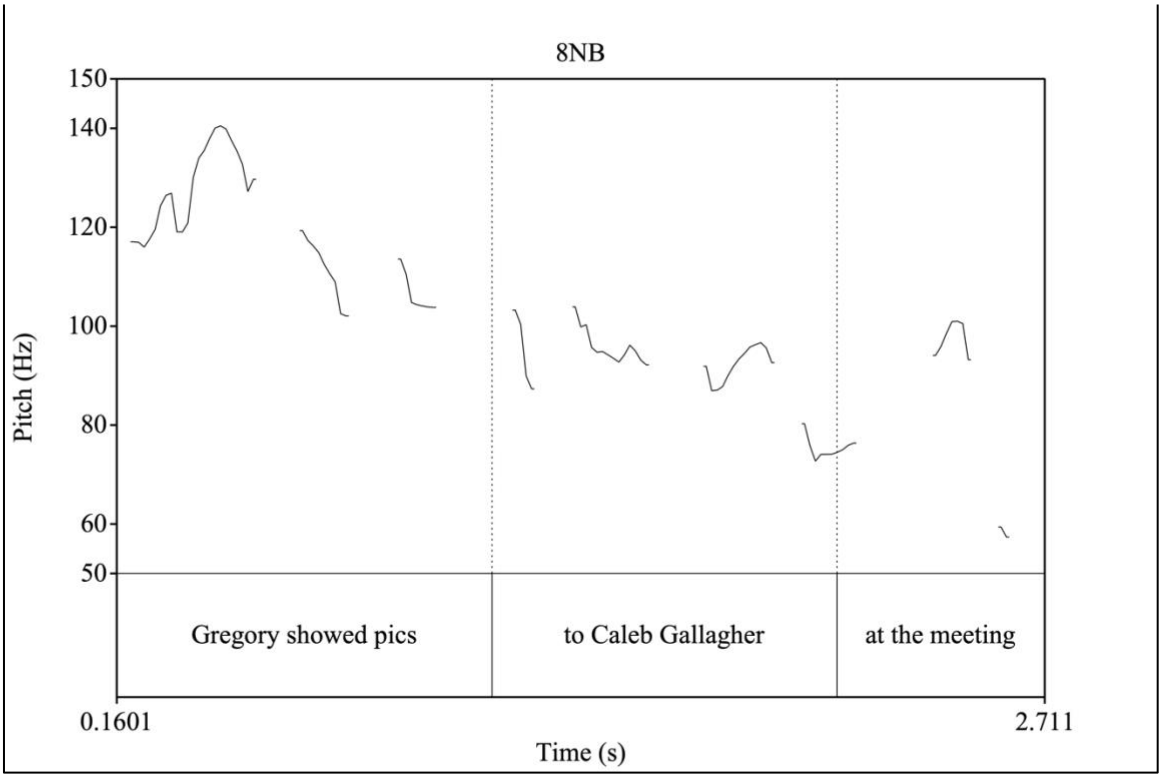


Figure 46: (1Onb) – experimental sentence pitch drawing

Task – (A) *Gregory showed pics to Caleb and to Mary at the meeting*

(B) *Gregory showed pics to Caleb Gallagher at the meeting*

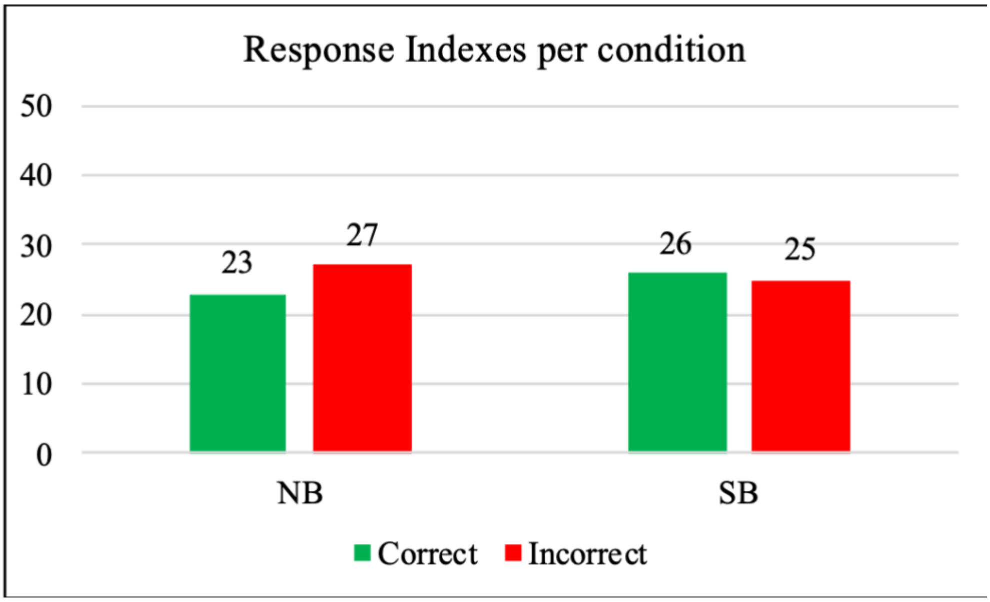
3.2.6 Predictions

The answer indexes will be satisfactory (more successful answers than mistakes) to both conditions. The response time will be larger to condition NB, due to the comparison complexity between the possible sentences and the absence of a prosodic cue to perceive the sentence structure.

3.2.7 Results

The response indexes and response times were measured and compared to pursue the perception degree and processing cost. All the results were statistically analyzed using *IBM®*

SPSS® Statistics, version 20. An alpha level of .05 was used for all statistical tests. The data were filtered and the outliers were taken out.

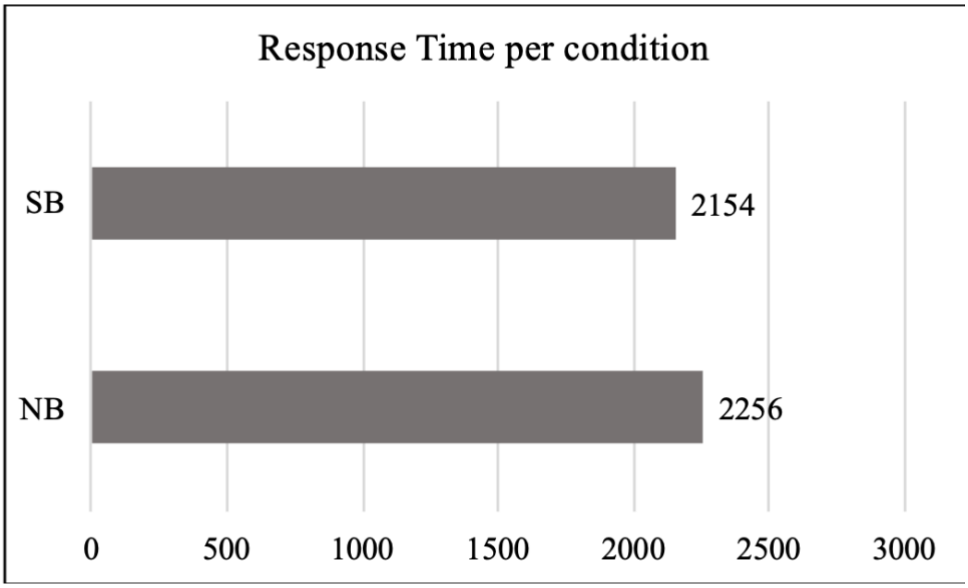


Graph 9: Total of Response Indexes per condition distributed by Accuracy

Graph 9 shows the response indexes grouped according to correctness of answers, distributed by presence/absence of strong boundary. The samples for each condition are composed of approximately 50 cases. For all conditions, the number of answers corresponding to the expected (correct) is very close to the number of answers that go against of the expected, lying statistically on the level of the chance. Condition NB shows four more cases of incorrect answers. For condition SB, we have a difference of one case.

A *Chi-square goodness-of-fit* was conducted for each group, in which an equal distribution of single referent and multiple referents cases (chance level) were established as null hypothesis. The test showed no statistically significant relevance in rejecting the null hypothesis for all conditions.

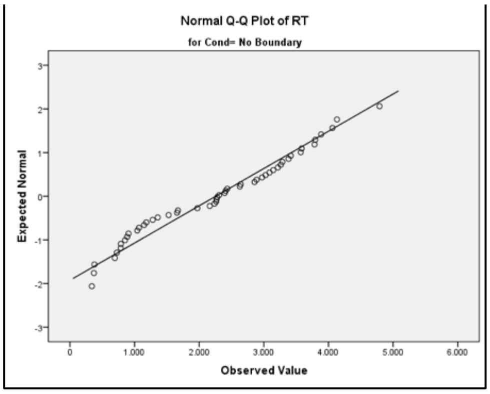
A *hypothesis test* was conducted between the pair of conditions. To this test, the null hypothesis was that the distribution of different values across both conditions were equally likely. In other words, the test examines if there is one condition is stronger than the other. The test showed no statistical significance in rejecting the null hypothesis when comparing the conditions, which means that there was no difference in distribution between conditions.



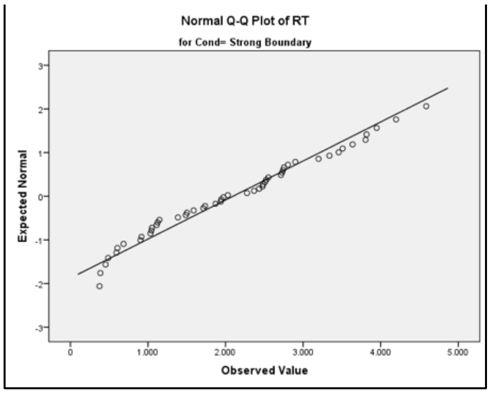
Graph 10: Average of Response Time per Condition

The Response time results show a higher response time to NB condition than SB condition in 102ms. A *Shapiro-Wilk test of normality* was conducted. Both samples showed no statistical significance in rejecting the null hypothesis, i.e., the samples were normally distributed (NB: *p=.089*; SB: *p=.146*).

NB

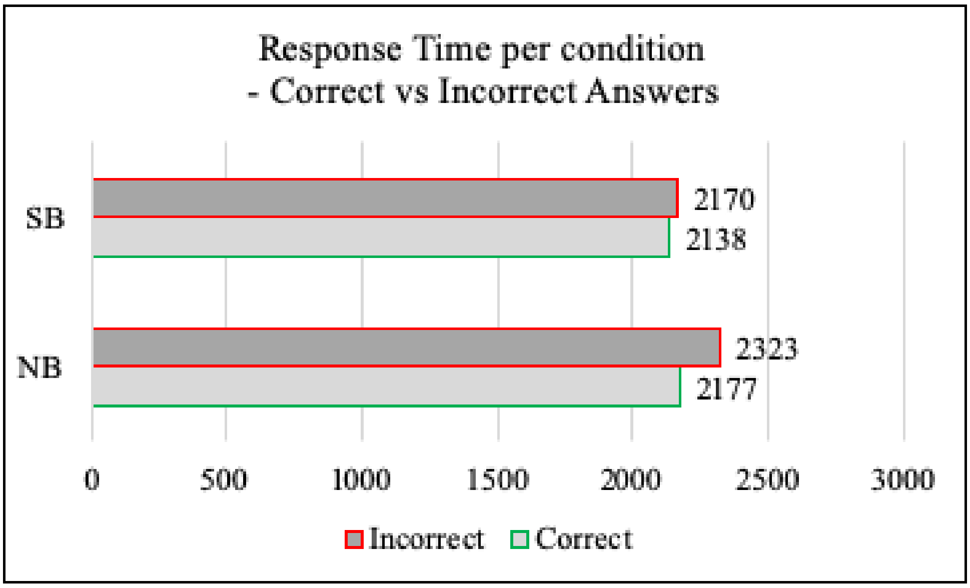


SB



Graph 11: Normal Q-Q plots of Response Times per condition

An *One-Way AnoVA* test was conducted to evaluate the difference in occurrence distribution between groups. No statistical significance was found when comparing the conditions.



Graph 12: Average of Response Time per condition distributed by accuracy

When comparing the response time grouped by success of the answer, incorrect answers have a higher average of RT than correct answer, but no statistical relevance was found.

3.2.8 Linguistic Analysis

The results show almost no interference of the prosodic cues to syntax processing, as the indexes remained in the level of chance. We can see that the boundary is acoustically there, but it is not perceived as such, or it is not salient enough. The response times show slightly greater cost to language processing when the boundary is absent, but the distribution is not strong enough to make assumptions. We may have to test further to understand if this result reflects a linguistic behavior in American English or if there are artifacts in play to our design.

The artifacts that might have interfered in the results involve the circumstances in which the experiment was run. Some of the participants were tested remotely, due to the pandemic, and some were tested in person. It was hard, then, to determine the engagement of the participants during the trials. Also, we were able to gather only a small group of participants that spoke only English within the community I was. A larger group of participants in a more controlled environment may give us in the future better understanding of the phenomenon in American English.

3.3 EXPERIMENT 3 – PROSODY AND PRAGMATICS: PROSODY FOCUS PERCEPTION IN SENTENCES CORREFERENCE AMBIGUITY IN BRAZILIAN PORTUGUESE (BP)

Prosody structure is not only composed of chunks and boundaries, but also by tones. The tones will play around with pitch range, giving movement to speech melody. Different information can be carried by the tone arrangement of a sentence. And pragmatics seems to use tone arrangement to map its information.

One of the high importance information in Pragmatics realm is deixis. The information regarding the enunciative act serves as core reference to full comprehension of a sentence contextualized in the world. And when it comes to contextualizing the actants of the discourse, things can get a little complex.

We know that syntax-semantics structure uses prominence as a resource to map focus (Ribeiro, 2015b). But Deixis seems to use prominence manipulation to point out special features of its contextualization. So, when we listen to a sentence:

(19a) You do the job while you think of your girlfriend

There is a structural ambiguity since the pronouns can either be co-referents or not:

(19b) You_i do the job while you_i think of your girlfriend

(19c) You_i do the job while you_j think of your girlfriend

This ambiguity seems to be overruled by prominence. When we make the pitch prominent on both pronouns, we are pointing out that they both need attention because they both refer to different individuals, in contrast with not making it prominent.

Büring (2016:1) says that "prominence is about the perception of the hearer regarding the strength in an utterance". The aim of this experiment is then to observe the perception of prominence due to pronoun focus, and how effective they are in mapping the number of referents to a sentence.

3.3.1 Hypothesis

Prosody is processed through a specific domain, and it interacts with the Faculty of Language. One interface is with narrow syntax. At this interface, prosody uses the needed means to map properly syntactic structure and relations into prosodic content (Syntax-

Prosody Interface). Another interface is with pragmatics. At this interface, prosody uses the needed means to map properly pragmatic conditions, such as enunciation conditions, and extralinguistic information. This interface of prosody processing marks the pragmatic intentions of the speaker during the enunciation (Moraes & Colamarco, 2007; Rilliard *et al*., 2013).

3.3.2 Independent variables

Two independent variables were selected.

3.3.2.1 Intonation Bias

a) Intonation with Bias to one discursive referent (no prominent prosodic focus);

b) Intonation with Bias to two discursive referents at same distance from speaker (prominent prosodic focus on both pronouns of the utterances);

c) Intonation with Bias to two discursive referents at different distances from speaker (prominent prosodic focus on both pronouns of the utterances and increased intensity at the second sentence.);

d) Unbiased Intonation: flattened intonation via Praat.

3.3.2.2 Pronoun

a) You (singular);

b) he.

3.3.3 Dependent variables

Response Indexes and Response Time.

Design 4x2	Bias to 1 Ref (B1)	Bias to 2 Ref n' same dist (B2)	Bias to 2 Ref n' Diff dist (B3)	Unbiased (U)
You (V)	B1V	B2V	B3V	UV
He (E)	B1E	B2E	B3E	UE

Table 3: Experimental design of experiment 3

3.3.4 Procedures and task

The experiment was programmed using OpenSesame platform. It starts collecting social data (name, age, and gender) just for register, but that were never revealed. It then proceeds to instructions, where it instructs the participant on the test to be performed giving an example of the auditioning and the task. The participants were instructed to sit in front of the computer and read carefully the instructions. Before the real test starts, it runs a sequence of four training trials, in order to make sure the participant learned his/her role. Once the participant is comfortable with the task, it proceeds to test trials. The test consisted of the audition of sentences followed by a task on each trial.

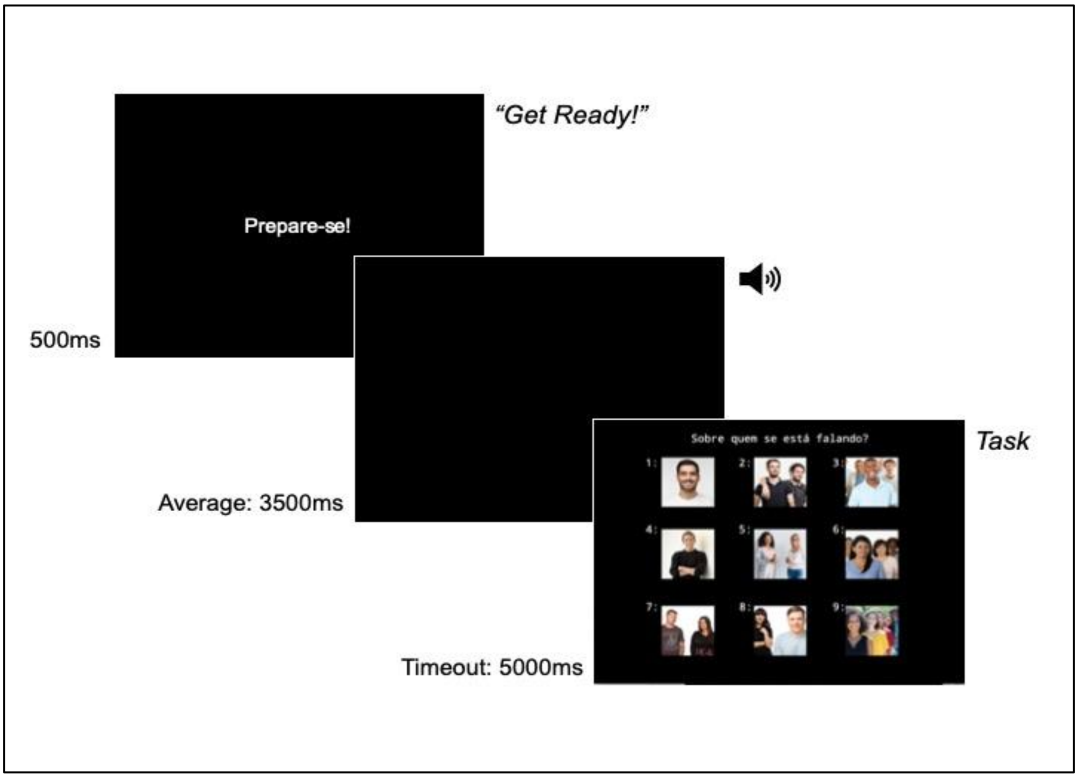


Figure 47: Trial sequence of experiment 3

The task of the subject was to answer to the question: *The sentence you just listened to talks about whom?* (*Sobre quem se está falando?*). To answer to that question, they had nine alternatives/options in form of photos, organized in a matrix as the one bellow. They answered to it by typing in the keyboard the number corresponding to the chosen picture.

Figure 48: Task screen with options

The experiment took place at the building of Languages College of UFRJ. Twenty participants were tested.

3.3.5 Materials

(24 sets of stimuli + 48 sets of distractors):

To test the hypothesis, I manipulated the way sentences with two deixis pronouns are intonated. The specific hypothesis here is that the prominence of the pitch accent will play as a contrastive prosodic focus. The sentences follow the model Filter + pronoun 1 + verb 1 + object 1 + conjunction of simultaneity + pronoun 2 + verb 2 + object 2, as in:

(20) É simples: Você faz o jantar enquanto você lava a louça

It is simple: You cook dinner while you do the dishes

In presence of two prominences, one in each pronoun, the contrastive focuses map the pragmatic interpretation of two distinct referents. Therefore, the test has four different intonations to each stimulus. B1 conditions have one prosodic focus on the first pronoun, hence they are mapping one referent.

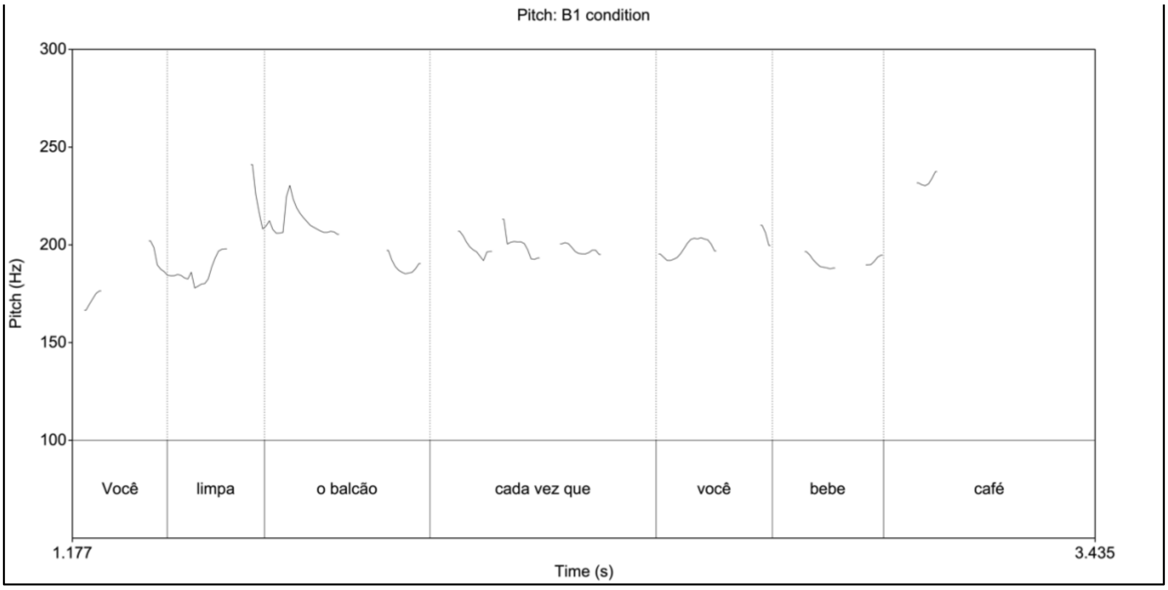


Figure 49: Pitch track of a B1 condition sentence: You clean the counter every time that you drink coffee

B2 and B3 conditions have two prosodic foci, one in each pronoun, hence they are mapping two referents.

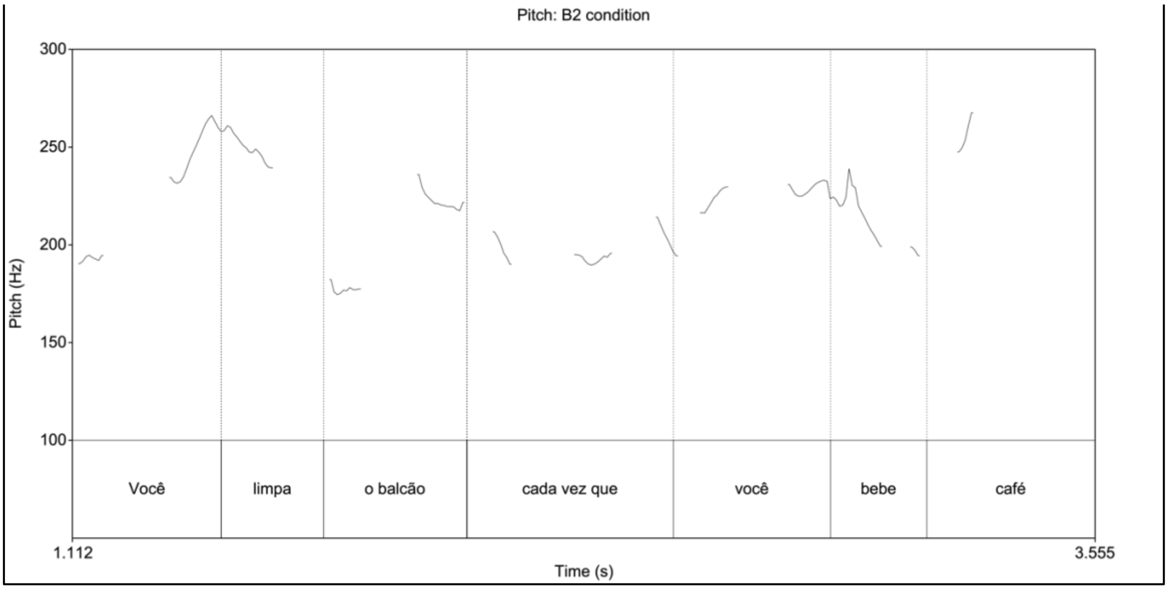


Figure 50: Pitch track of a B2 condition sentence: You clean the counter every time that you drink coffee

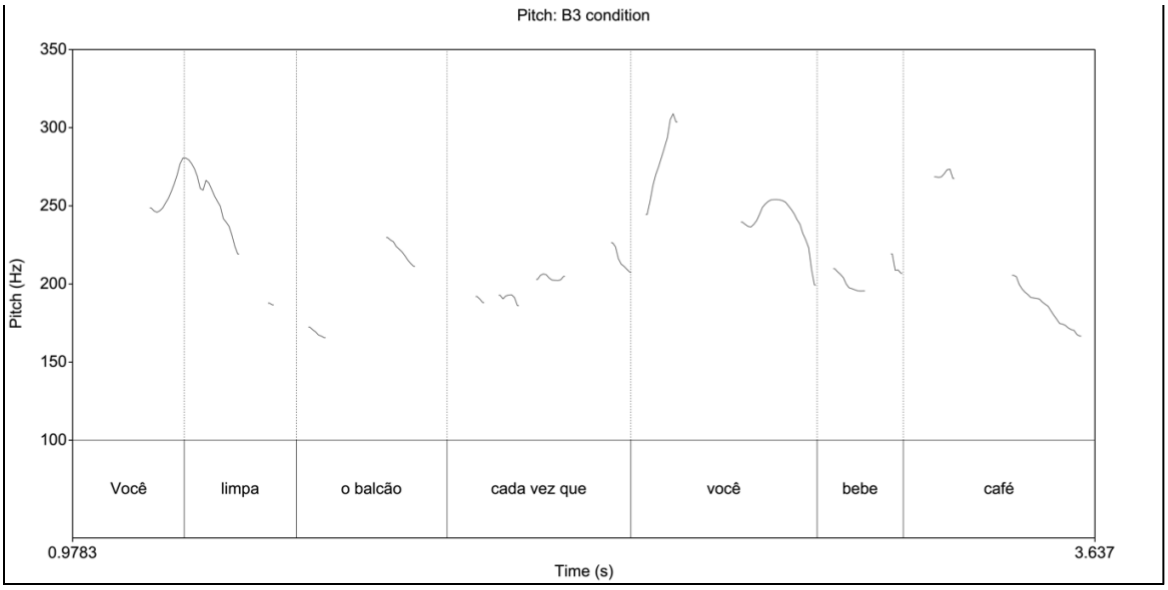


Figure 51: Pitch track of a B3 condition sentence
"You clean the counter every time that you drink coffee"

During the conceptualization of the experiment, an intuition about the intensity of the prosodic focus was raised. Apparently, the difference in intensity between two foci may map different distances between the speaker and the addressee of the message.

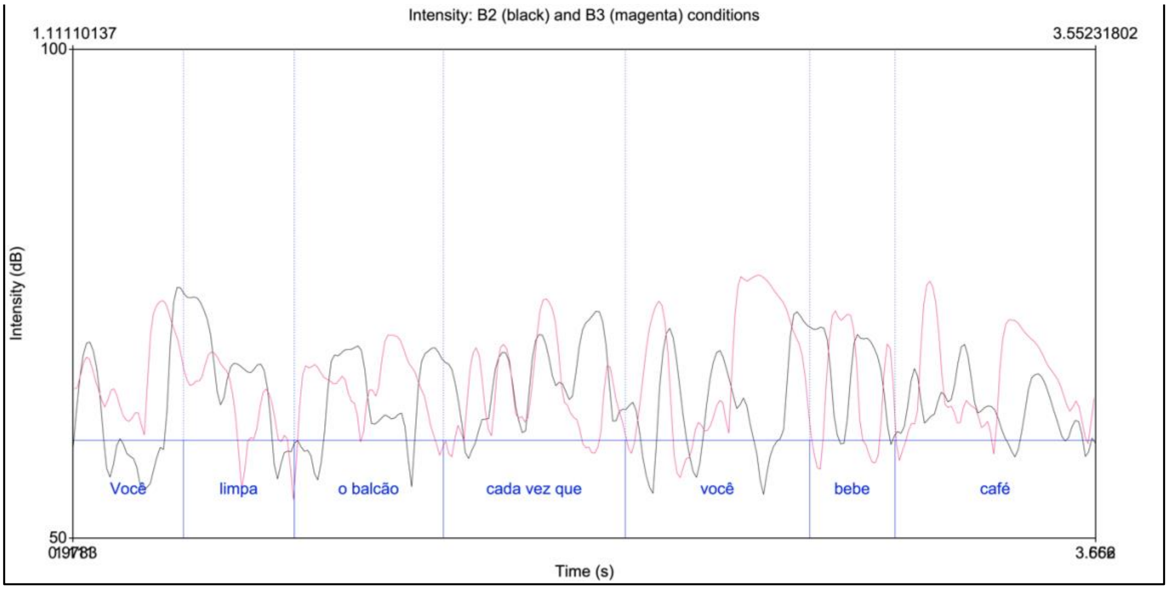


Figure 52: Intensity track of a B2 sentence (in black) and a B3 sentence (in magenta)

U conditions have a synthesized flattened intonation; hence they are unbiased in regard to the number of referents.

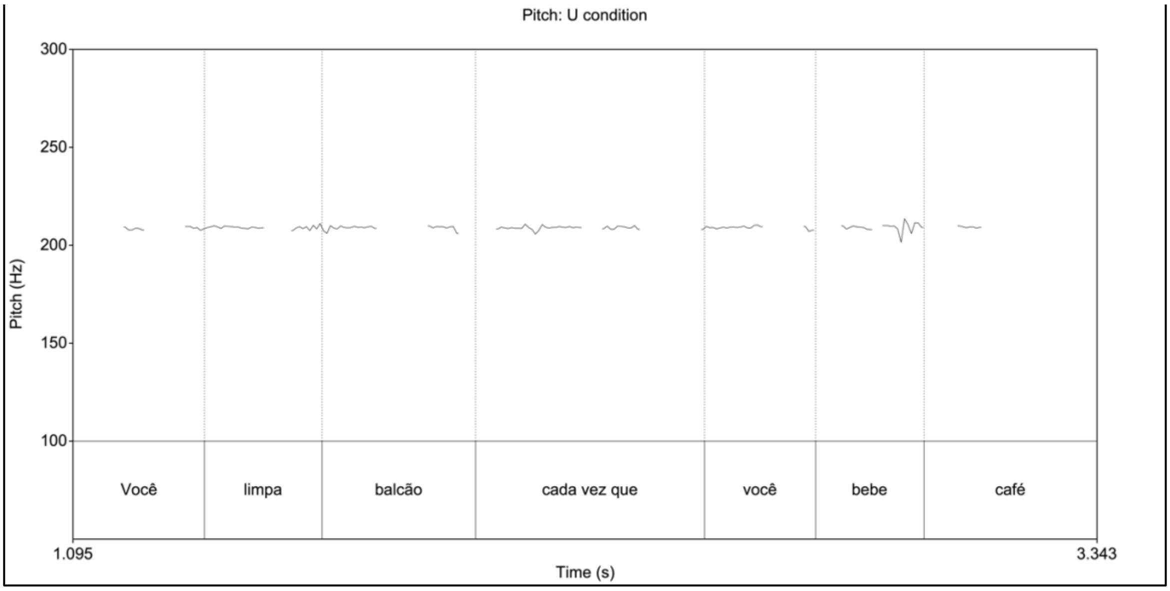
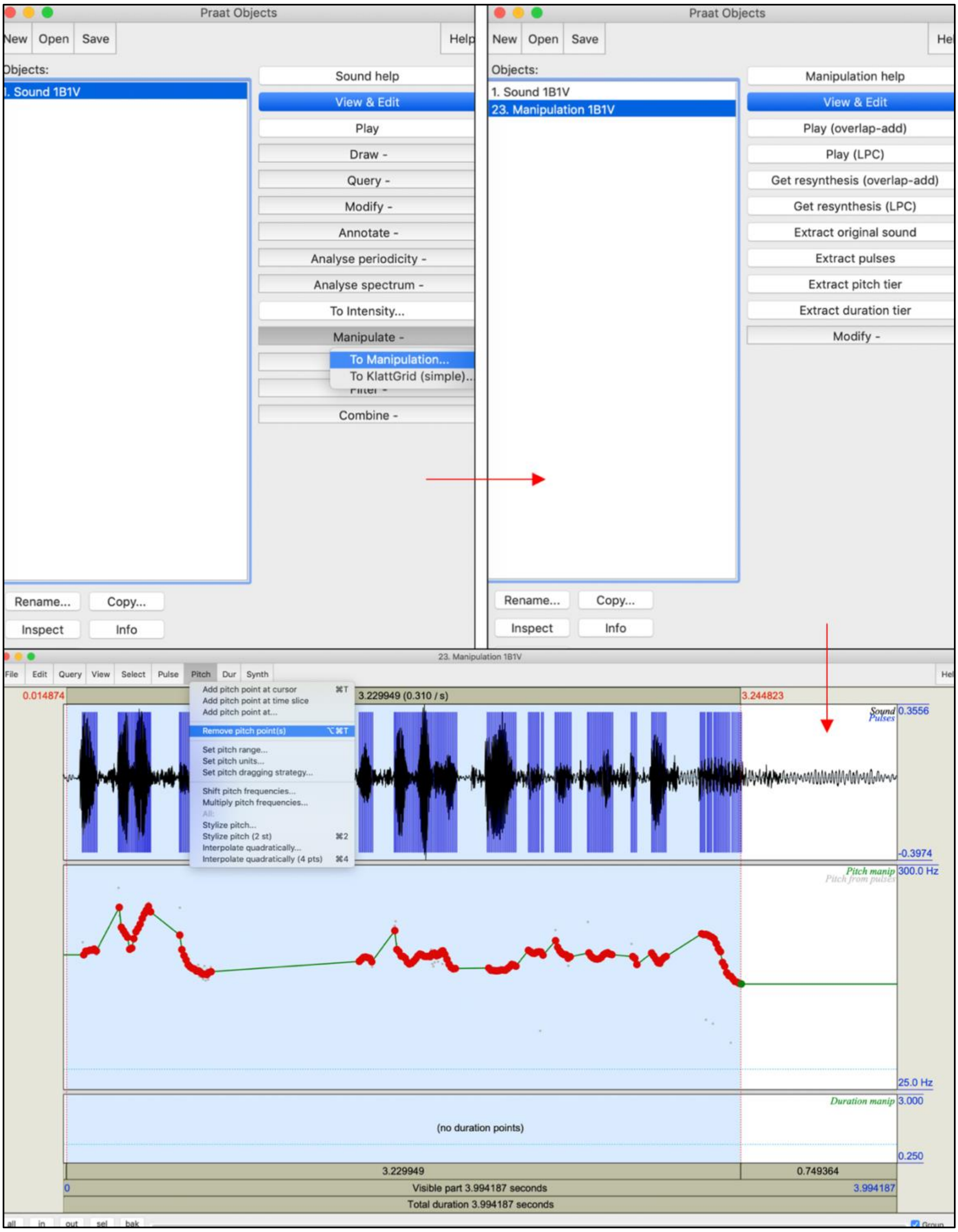


Figure 53: Pitch track of a U condition sentence
"You clean the counter every time that you drink coffee"

To synthesize the flattened intonation, I used the *Praat*® platform. Each sentence was converted to Pitch Manipulation and then had its pitches dots erased, except one. The remaining dot leveled the flattened intonation to a pitch that was closer to midrange. Each sentence was then published and saved as a WAVE file.


Praat Objects
New
Open
Save
Help
Objects:
1. Sound 1B1V
Sound help
View & Edit
Play
Draw -
Query -
Modify -
Annotate -
Analyse periodicity -
Analyse spectrum -
To Intensity...
Manipulate -
To Manipulation...
To KlattGrid (simple)...
Combine -
Rename...
Copy...
Inspect
Info
Praat Objects
New
Open
Save
Objects:
1. Sound 1B1V
23. Manipulation 1B1V
Manipulation help
View & Edit
Play (overlap-add)
Play (LPC)
Get resynthesis (overlap-add)
Get resynthesis (LPC)
Extract original sound
Extract pulses
Extract pitch tier
Extract duration tier
Modify -
Rename...
Copy...
Inspect
Info
23. Manipulation 1B1V
File
Edit
Query
View
Select
Pulse
Pitch
Dur
Synth
Add pitch point at cursor
Add pitch point at time slice
Add pitch point at...
Remove pitch point(s)
Set pitch range...
Set pitch units...
Set pitch dragging strategy...
Shift pitch frequencies...
Multiply pitch frequencies...
Stylize pitch...
Stylize pitch (2 st)
Interpolate quadratically...
Interpolate quadratically (4 pts)
0.014874
3.229949 (0.310 / s)
3.244823
Sound
Pulses
0.3556
-0.3974
Pitch manip
Pitch from pulses
300.0 Hz
25.0 Hz
Duration manip
3.000
(no duration points)
0.250
3.229949
0.749364
0
Visible part 3.994187 seconds
3.994187
Total duration 3.994187 seconds

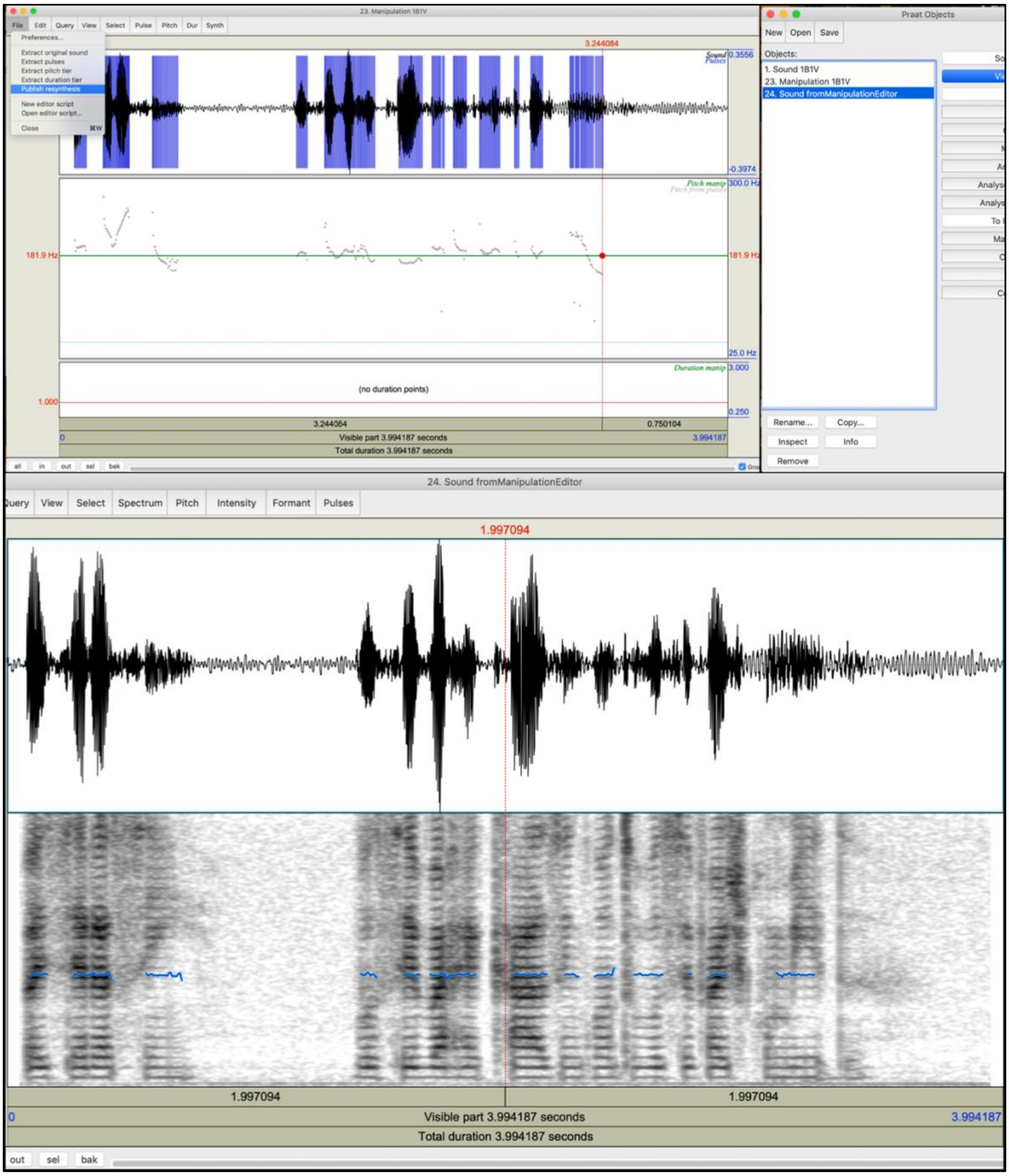


Figure 54: Steps to Pitch Manipulation and flattening technique

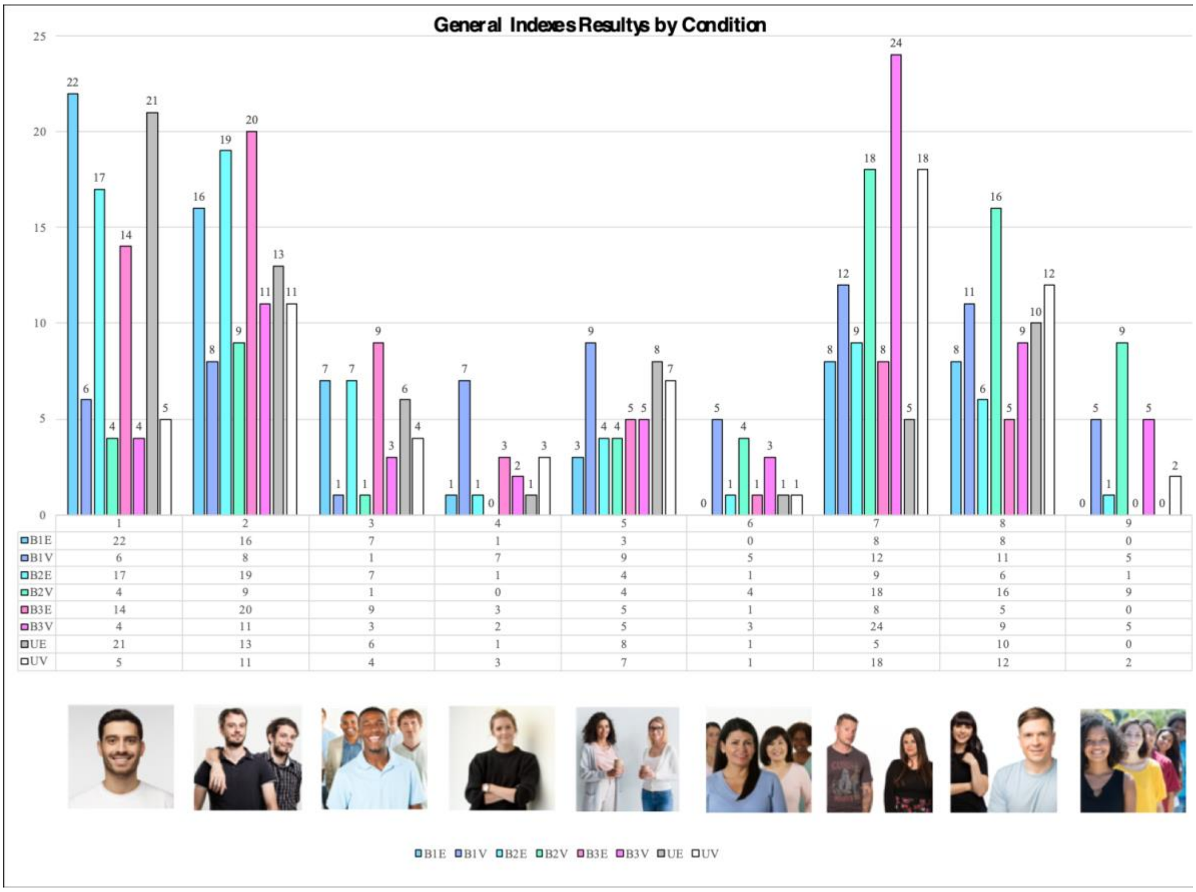
The test consisted of the audition of 74 experimental sentences, being 28 stimuli and 46 fillers. In total, 28 sets of stimuli (per Intonation variable, n=112) and 96 fillers (48 unfiltered and 48 filtered) were made. The fillers were distributed between two of the four experimental lists and copied on the other two. Four of the fillers were used in the training trials.

3.3.6 Predictions

The predictions were that the answers on B1 and B2 trials would be satisfactory, meaning that the perception of the contrastive prosodic foci on B2 and the absence of contrastive focus on B1 would be successful in mapping the pragmatic referents. The answers on B3 trials would show some tendency to satisfaction, meaning that the majority of the time, the perception of the raised intensity would successfully map the pragmatic notion of greater distance on the second referent. The answers would stay at the level of chance to U trials. Regarding the Response Time, the rates would be as follow: B1 = U < B2 < B3.

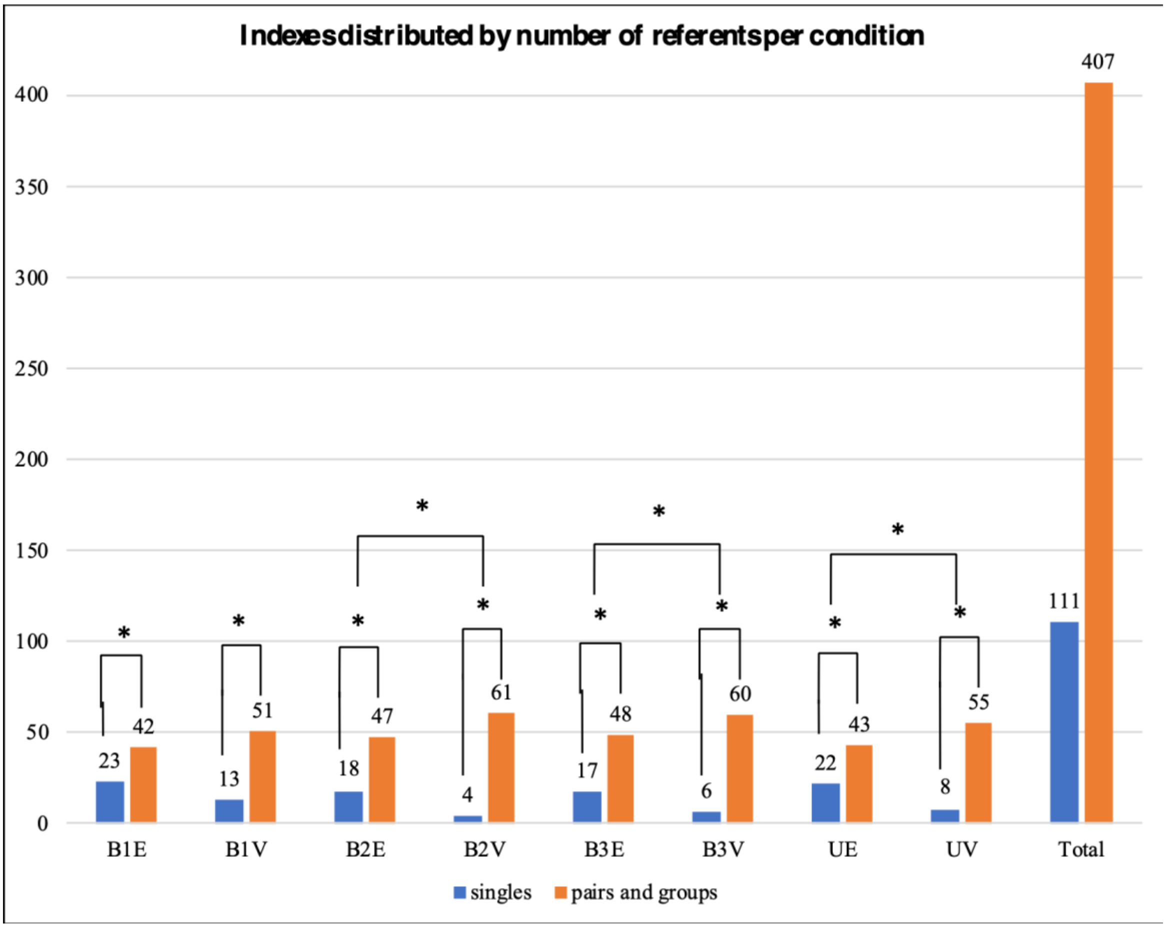
3.3.7 Results

The response indexes and response times were measured and compared to pursue the perception degree and processing cost. All the results were statistically analyzed using *IBM® SPSS® Statistics, version 20.* An alpha level of .05 was used for all statistical tests. The data were filtered and the outliers were taken out.



	1	2	3	4	5	6	7	8	9
B1E	22	16	7	1	3	0	8	8	0
B1V	6	8	1	7	9	5	12	11	5
B2E	17	19	7	1	4	1	9	6	1
B2V	4	9	1	0	4	4	18	16	9
B3E	14	20	9	3	5	1	8	5	0
B3V	4	11	3	2	5	3	24	9	5
UE	21	13	6	1	8	1	5	10	0
UV	5	11	4	3	7	1	18	12	2

Graph 13: Total response indexes distributed by option chosen per conditions

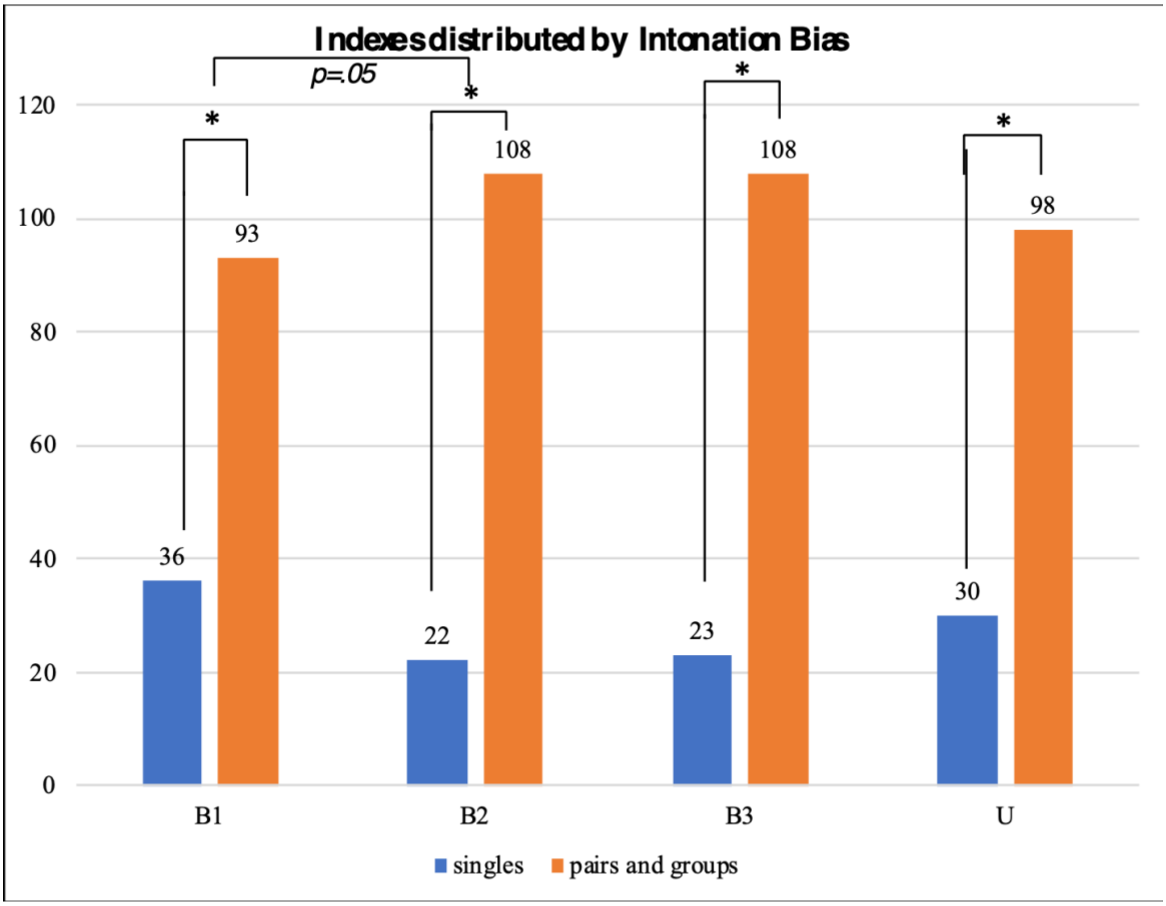


Graph 14: Result Indexes distributed by number of people in the chosen picture per condition.

Graph 14 shows the response indexes grouped according to the number of candidates to referents, distributed by condition. The samples for each condition are composed of approximately 65 cases. For all conditions, the number of answers revealing a preference for multiple referents is statistically larger than the number of answers revealing a preference for a single referent. Condition B1E shows 19 more cases of answers showing a preference for multiple referents than single ones. For condition B1V, we have a difference of 38 cases. Condition B2E shows a difference of 29 cases; condition B2V, 57 cases; condition B3E shows a difference of 31 cases; condition B3V, 54 cases; condition UE shows a difference of 21 cases; and condition UV, 47 cases. In total, the indexes show a preference for multiple referents 296 cases greater than single referent.

A *Chi-square goodness-of-fit* was conducted for each condition, in which an equal distribution of single referent and multiple referents cases (chance level) were established as null hypothesis. The test showed a statistically significant relevance in rejecting the null hypothesis for all conditions: B1E ($X^2(1) = 5{,}554$, p=.018), B1V ($X^2(1) = 22{,}563$, p<.001),

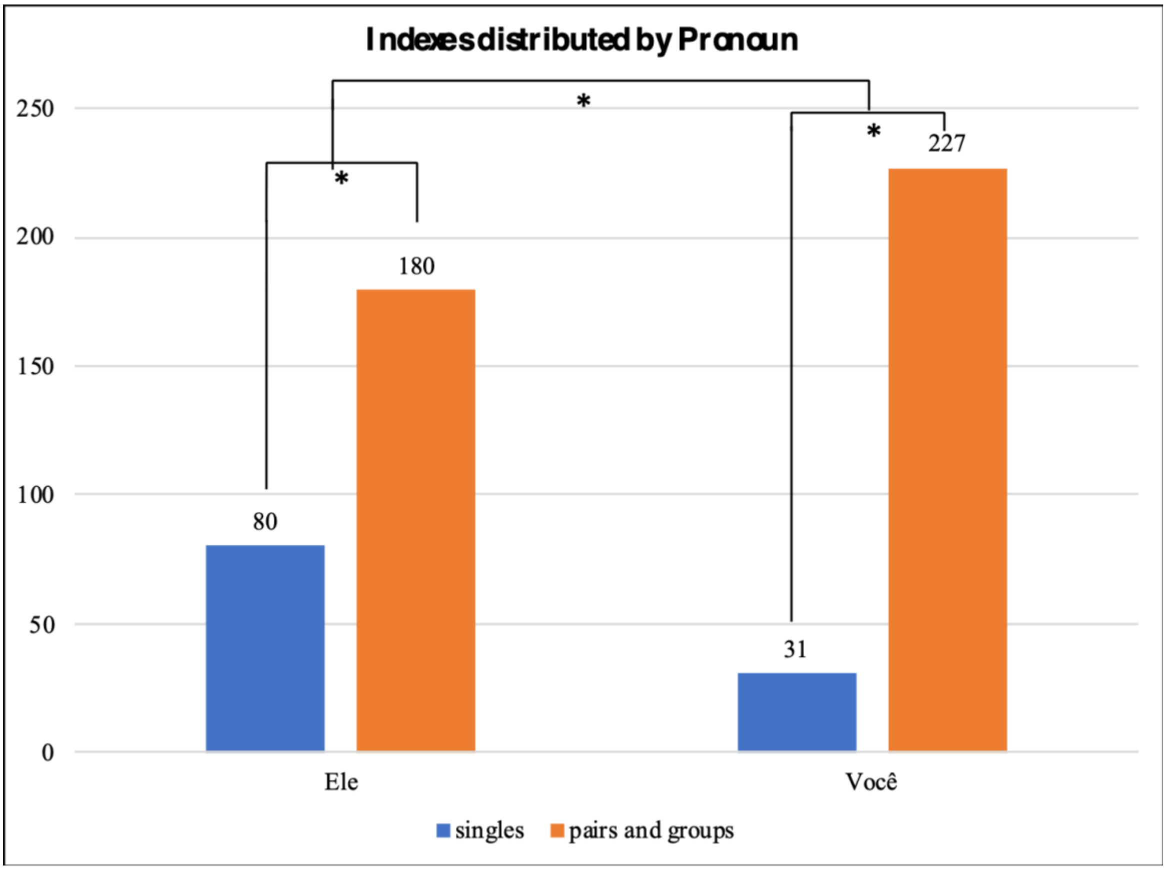
B2E ($X^2(1) = 12{,}938$, $p<.001$), B2V ($X^2(1) = 49{,}985$, $p<.001$), B3E ($X^2(1) = 14{,}785$, $p<.001$), B3V ($X^2(1) = 44{,}182$, $p<.001$), UE ($X^2(1) = 6{,}785$, $p=.009$), UV ($X^2(1) = 35{,}063$, $p<.001$). A Hypothesis test was conducted, between pairs of conditions, in which the null hypothesis was that the distribution of different values across both conditions were equally likely. The test showed a statistical significance in rejecting the null hypothesis when comparing B2E with B2V (*p=.001*), B3E with B3V (*p=.019*), and UE with UV (*p=.015*). B1E with B1V did not reach statistical significance.



Graph 15: Result Indexes distributed Intonation bias

Graph 15 shows the response indexes grouped according to the number of candidates to referents, distributed by intonation bias. The samples for each condition are composed of approximately 130 cases. For all conditions, the number of answers revealing a preference for multiple referents is larger than the number of answers revealing a preference for a single referent. Condition B1 shows 57 more cases of answers showing a preference for multiple referents than single ones. For condition B2, we have a difference of 86 cases. Condition B3 also shows a difference of 86 cases; and condition U, 58 cases.

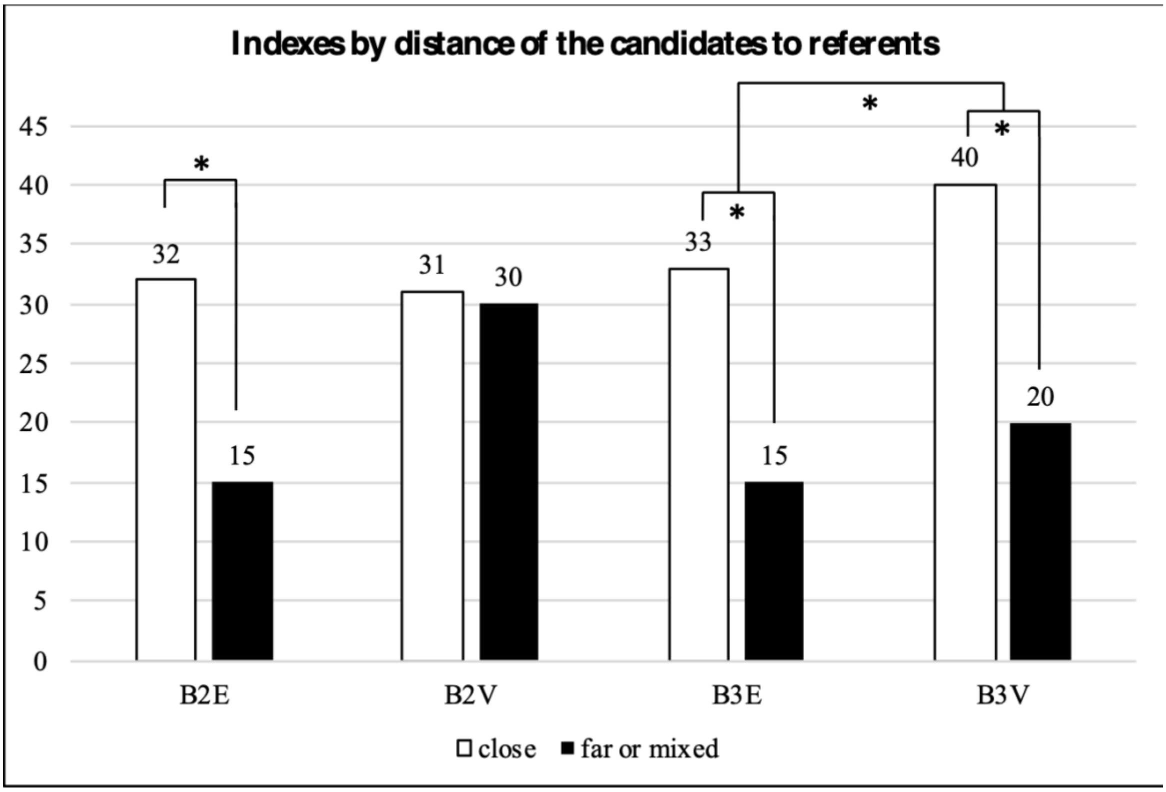
A *Chi-square goodness-of-fit* was conducted for each group, in which an equal distribution of single referent and multiple referents cases (chance level) were established as null hypothesis. The test showed a statistically significant relevance in rejecting the null hypothesis for all conditions: B1 ($X^2(1) = 25{,}186$, $p<.001$), B2 ($X^2(1) = 56{,}892$, $p<.001$), B3 ($X^2(1) = 55{,}153$, $p<.001$), and U ($X^2(1) = 36{,}125$, $p<.001$). A Hypothesis test was conducted, between pairs of conditions, in which the null hypothesis was that the distribution of different values across both conditions were equally likely. The test showed a statistical tendence in rejecting the null hypothesis when comparing B1 with B2 ($p=.05$). The other pairs did not reach statistical significance.



Graph 16: Result Indexes distributed by pronoun in the stimuli

Graph 16 shows the response indexes grouped according to the number of candidates to referents, distributed by pronoun used in the stimuli, either *ele* (*he*) or *você* (*you*, singular). The samples for each condition are composed of approximately 260 cases. For all conditions, the number of answers revealing a preference for multiple referents is larger than the number of answers revealing a preference for a single referent. Condition *Ele* shows 100 more cases of answers showing a preference for multiple referents than single ones. For condition *Você*, we have a difference of 196 cases.

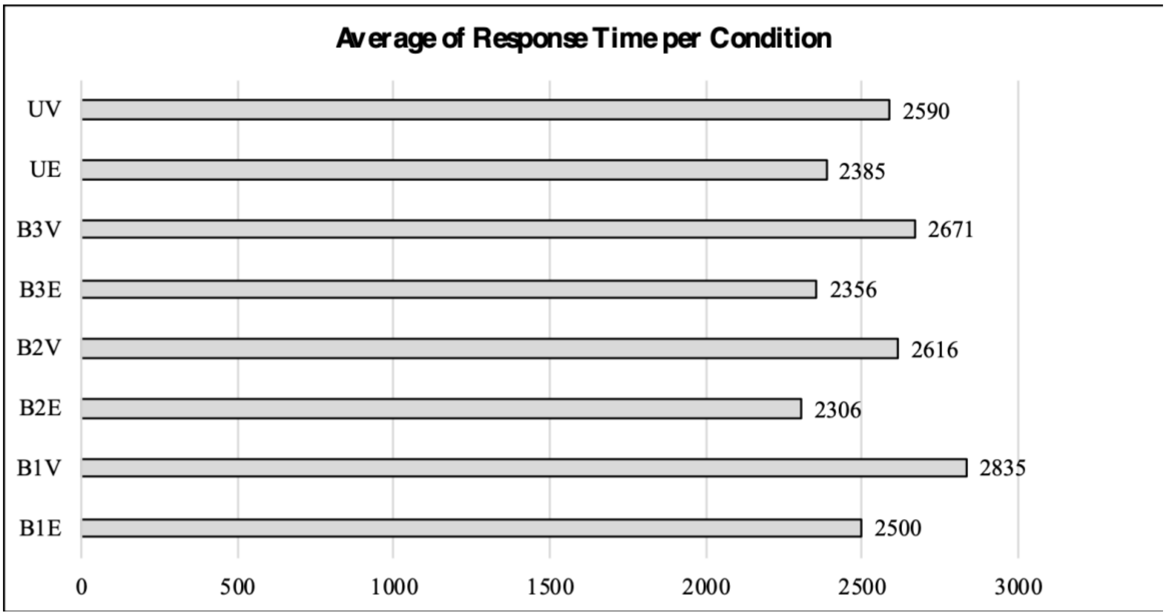
A *Chi-square goodness-of-fit* was conducted for each group, in which an equal distribution of single referent and multiple referents cases (chance level) were established as null hypothesis. The test showed a statistically significant relevance in rejecting the null hypothesis for both conditions: *Ele* ($X^2(1) = 38,462$, $p<.001$) and *Você* ($X^2(1) = 148,899$, $p<.001$). A *Hypothesis test* was conducted between the pair of conditions, in which the null hypothesis was that the distribution of different values across both conditions were equally likely. The test showed a statistical significance in rejecting the null hypothesis ($p<.001$).



Graph 17: Result Indexes distributed by distance of the candidates to referents

Graph 17 shows the response indexes grouped according to the potential distance of candidates to referents, distributed by relevant conditions, that is, conditions predicted to be perceived as making reference to multiple referents. The samples for each condition are composed of 47 cases to condition B2E, 61 cases to condition B2V, 48 cases to condition B3E, and 60 cases to condition B3V. For all conditions, the number of answers revealing a preference for same distance referents is larger than the number of answers revealing a preference for different distance referents, but to condition B2V, the distribution is almost even. Condition B2E shows 17 more cases of answers showing a preference for same distance referents than different distance ones. For condition B2V, we have a difference of only one case. Condition B3E shows a difference of 18 cases; and condition B3V, 20 cases.

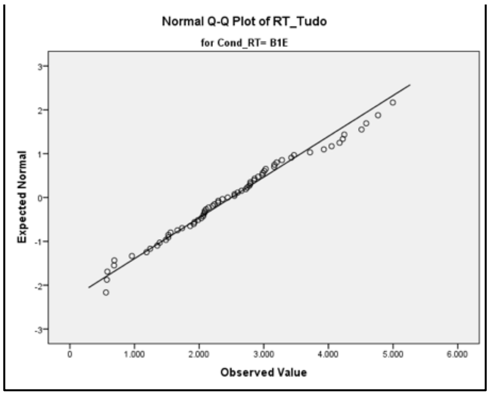
A Chi-square goodness-of-fit was conducted for each group, in which an equal distribution of same distance referents and different distance referents cases (chance level) were established as null hypothesis. The test showed a statistically significant relevance in rejecting the null hypothesis for conditions: B2E ($X^2(1) = 6,149$, p=.013), B3E ($X^2(1) = 6,750$, p=.009), and B3V ($X^2(1) = 6,667$, p=.010). A Hypothesis test was conducted, between pairs of conditions, in which the null hypothesis was that the distribution of different values across both conditions were equally likely. The test showed a statistical tendency in rejecting the null hypothesis when comparing B3E with B3V (*p=.016*). The pair B2E and B2V did not reach statistical significance.



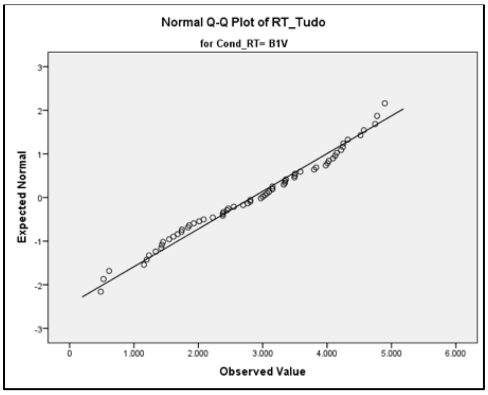
Graph 18: Average Response Time per condition

The Response time results show the following averages, from the highest to the lowest: B1V: 2835.14ms; B3V: 2670.86ms; B2V: 2616.38ms; UV: 2590.13ms; B1E: 2499.78ms; UE: 2385.37ms; B3E: 2356.38ms; B2E: 2306.06ms. A *Shapiro-Wilk test of normality* was conducted. B2E sample was shown not normal (*p=.012*). All other samples showed no statistical significance in rejecting the null hypothesis, i.e., the samples were normally distributed (B1E: *p=.350*; B1V: *p=.154*; B2V: *p=.337*; B3E: *p=.307*; B3V: *p=.453*; UE: *p=.144*; UV: *p=.061*).

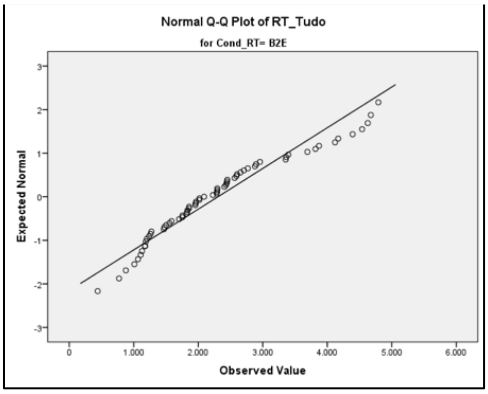
B1E



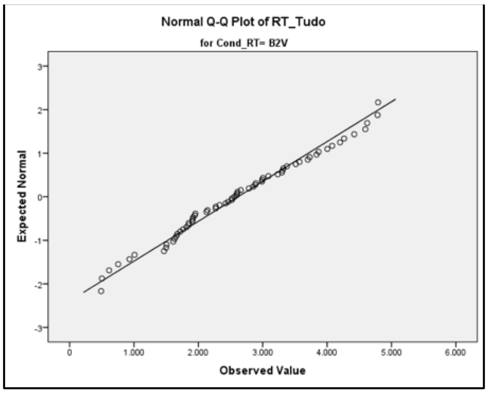
B1V



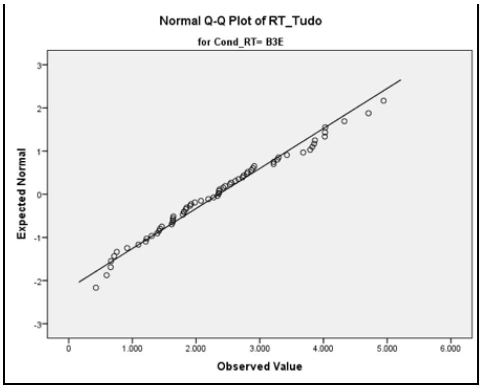
B2E



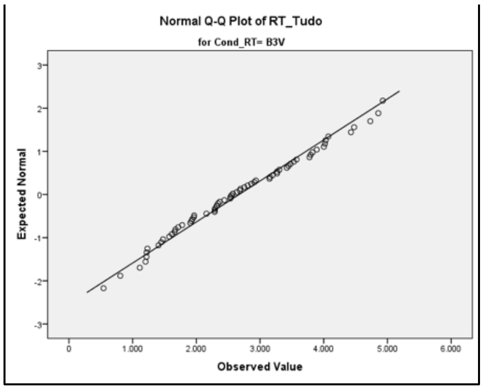
B2V



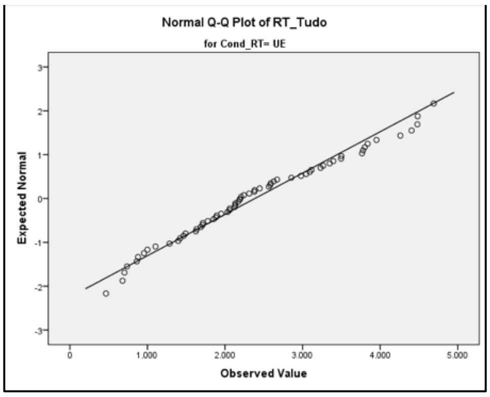
B3E



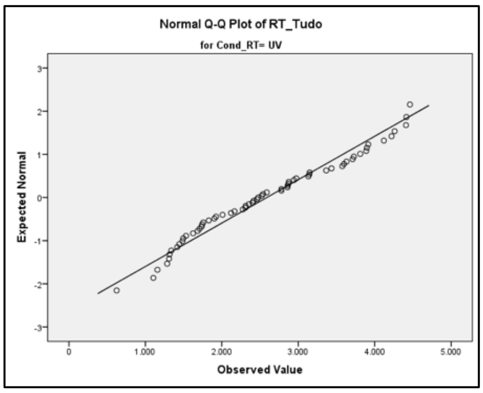
B3V



UE

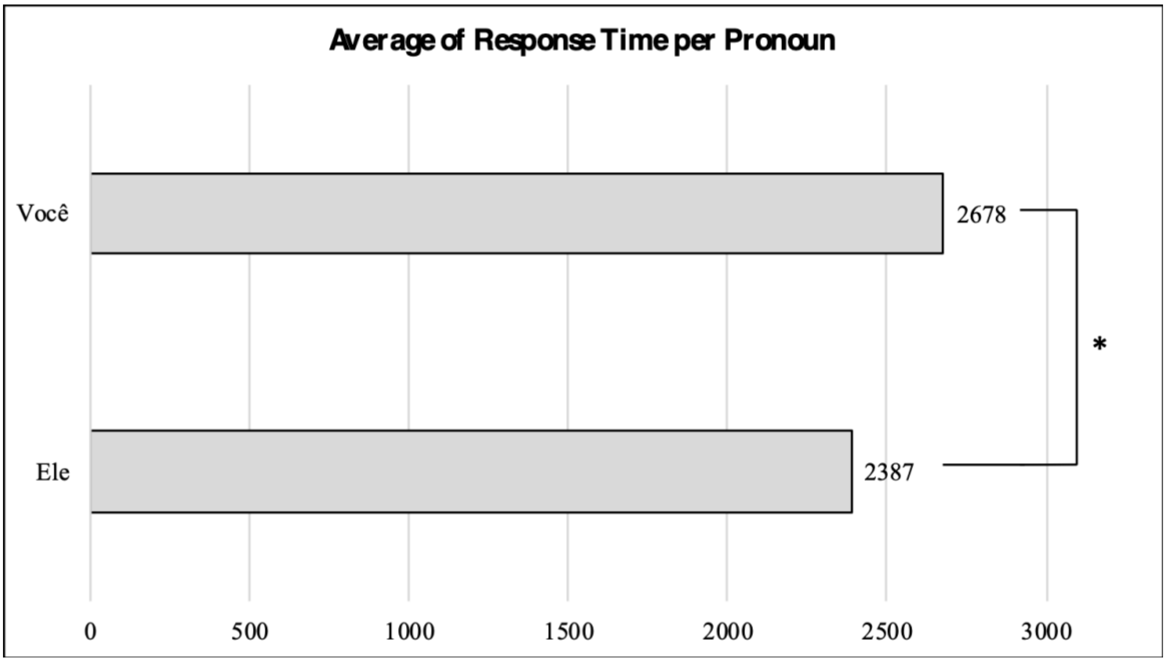


UV



Graph 19: Normal Q-Q Plots of Response Time data per condition

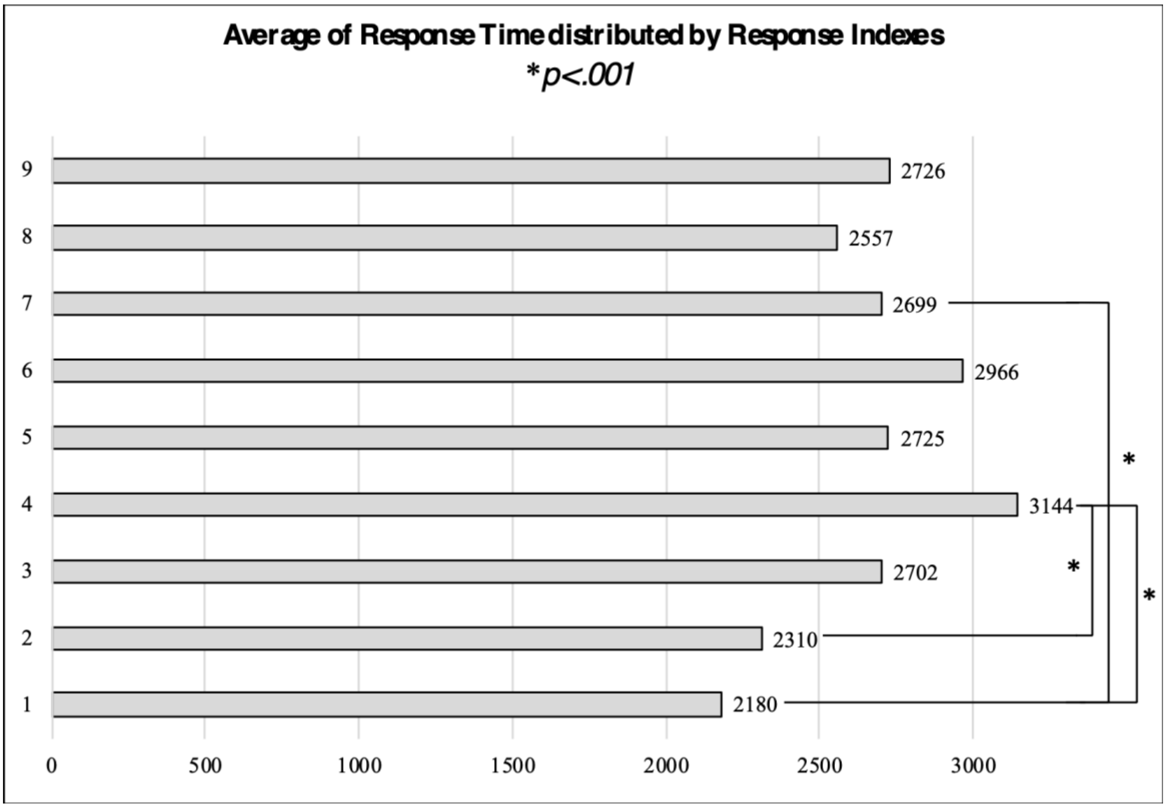
Several *One-Way AnoVA* tests were conducted to evaluate the differences in occurrence distribution between groups. No statistical significance was found when comparing the conditions.



Graph 20: Average of Response Time distributed by pronoun used in the stimuli

However, the tests showed statistically significant difference between groups, when comparing conditions using the pronoun *Você* (*you*, singular) and *Ele* (*he*), revealing that the 291ms response time of the task to *Você* conditions was statistically significantly higher than *Ele* conditions ($F(1,516) = 9.586$, p=.002).

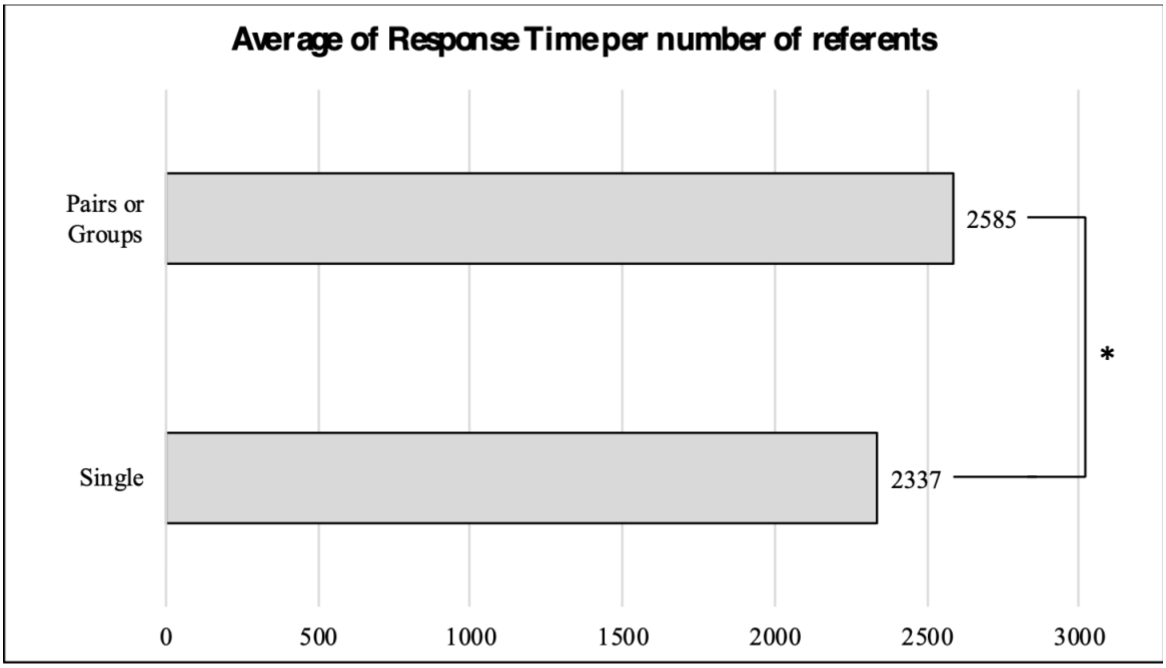
When comparing the response time distributed by answers, we have a statically significant difference between groups ($F(8,509) = 3.688$, $p<.001$).



Graph 21: Average of Response Time distributed by option chosen in the task.

A *Tukey post hoc test* revealed that the response time to complete the task was statistically significantly lower when the option choice was picture number 1 (single male) than when the choice was picture number 4 (single female; *p=.013*), and then when the choice was picture number 7 (couple at same distance; *p=.019*). The test also revealed a statistical tendency in the response time being lower when the option choice was picture number 2 (pair of men at same distance) than when the choice was picture number 4 (*p=.053*).

When grouping the answers into choices that reveal single referent and choices that reveal multiple referents, answers considering a single referent took significantly less time than answers considering multiple referents (*F(1,516)* = 4.653, *p*=.031).



Graph 22: Average of Response Time distributed by number of referents perceived

3.3.8 Linguistic Analysis

The data show that there is a clear preference, statistically attested, for interpreting the stimuli as making reference to two referents, in detriment of the single referent possibility, independently from what the intonational bias was, if there is any. That preference goes against our prediction in some way, since B1 conditions and U conditions reveal it as well.

However, B1 conditions show the preference for two referents in a less extent than B2 conditions, pointing to an interference of the intonational bias to one referent on the grammar preference. That interference can also be noted when comparing B1 with B3, although it didn't reach statistical relevance (*p=.061*).

We understand that this preference for pictures of multiple people in detriment of pictures of a single person is an artifact due to the design of the experiment: We have many more options with multiple people in the picture than with a single person. The number of options in the response grid was also elevated and might have interfered with the results, as it could have been overwhelming for the participants. The goal with the multiple options was to avoid gender bias, however, due to the number of independent variables we had for this design, the grid ended up having too many options. Another thing to be considered is the number of independent variables. Splitting this experiment in two, one testing single x multiple, and other testing close x far conditions might be more interesting in future research.

When comparing the indexes in regard of the stimuli' pronoun, the difference in number between answers showing a preference for single referent and answers showing a preference for two referents is statistically significantly smaller to stimuli with the pronoun 'he' (*ele*) than to stimuli with the pronoun 'you' (*você*). Two possibilities have arisen. One is in respect to the syntactic-semantic property of Brazilian Portuguese. Our hypothesis says that in sentences with two pronouns as subjects of two possibly concomitant events or actions, there is an ambiguity of Syntax-Semantic structure in respect to the control over the pronoun. Either both pronouns would be controlled by co-indexation or they would be controlled by different index (different denotations). In face of this ambiguity, the intonational bias would determine which structure should prevail. The overall data show that there is a preference for different controls to the two pronouns in one sentence, but when comparing 'you' (*você*) data with 'he' (*ele*) data, this preference is flattened a little in the 'he' (*ele*) data. Brazilian Portuguese still being a *pro-drop* language, it is understandable that there is a preference for choosing the different-control candidate over the same-control candidate when we reinforce the subject by realizing phonetically the pronoun instead of applying a *pro*. On the other hand, 'he' (*ele*) seems to be one of the pronouns that its necessity of realization is still high, due to the weak differentiation in the verb morphology, making it less likely that the realization of two of it in the same sentence is due to a structural necessity of marking difference in control. Therefore, the same-control candidate prevails a little bit more than to stimuli using 'you' (*você*), since the structure is slightly more ambiguous, and the bias becomes more salient. If it is not about the *pro-drop* property of the language, the hypothesis of Landau (2004) about non-control structures seems to be the one ruling more strongly for 'you' than for 'he'.

The other possibility is that the difference in range must be attributed to some pragmatic property of the pronoun in reference to its discursive persona (3rd person of singular). Being that 'I' and 'you' are the personas directly involved in an enunciative act, their deictic referents are strongly determined, making it less needed the reinforcement of referencing. Having that as a pragmatic implicit, and considering the conversational maxims of quantity, relevance and manner (Grice, 1975), the processor would avoid the interpretation of two referents in a sentence with realization of two 'you's, giving that BP is still a *pro-drop* language. 'He' is a discursive persona to which the speaker refers, thus it is a persona outside the enunciative act. In that case, it is not implicit its deictic referent, making it completely acceptable the double realization with no conversational fault. Hence the same-referent candidate is not quickly discarded, remaining available to the interference of the intonational bias.

Regarding the distance, the results do not meet the predictions, showing that the increased intensity is not mapping a greater distance between speaker and referent, or, at least, it is not efficient enough.

4 DISCUSSION: EXPLORING PROSODY DOMAIN

Taking the presented data and the literature reviewed in this dissertation into consideration, we want to propose that prosody processing is operated in a dedicated module, in articulation with the other modules of the faculty of language. As such, prosody is domain specific and may be activated multiple times during speech and language processing.

We assume there is a dedicated module for prosody because of its domain specific information being processed. When we have a structurally ambiguous sentence that is not ambiguous given its prosodic content. Likewise, the experiment in 3.1 conducted for this dissertation has shown the efficiency of the prosodic structure in retrieving syntax structure for a sentence, even when the lexical information is not available, and therefore, the syntactic labels to properly mapping the sentence structure are missing. In spite of that, participants were able to successfully retrieve the sentence structure solely based on prosodic structure. That shows that prosodic structure is processed in some formal way that it triggers certain syntactic operations and structures and not others.

As we see prosody processing here, its module is in operation multiple times during speech and language processing because we can distinguish at least three layers of semi-independent prosodic structure. The first and most inner structure would be the rhythmic structure, which will determine the lengthening of the prosodic elements (prosodic words, prosodic phrases, intonational phrases). This level of representation is the one mapping syntactic phrasing. This level was tested in experiments 3.1 and 3.2, as the manipulation of the experimental sentences focused on boundaries distribution, therefore, changing the lengthening of prosodic phrases and altering the rhythm of the prosodic structure. The next structure is the harmonic structure, which will determine the pitch contour (including, but not limited to the pitch accents). The operations of this level of representation will be restricted by the rhythmic structure. Respecting those restrictions, this level will be mapping semantic (informational) structure, as focus and topic, and pragmatic information. This level was tested on experiment 3.3, as the manipulation of the experimental sentences lead to different intonational contours by placement and manipulation of pitch accents within similar prosodic phrases. The final structure is the dynamic structure, which, in opposition to the other two structural levels, that are completely formal, is functional and will be mapping extralinguistic information, as mental state and emotions. The operation of this level is restricted by the other two levels, however, there is certain liberty as its operation is the final one before the final output of the Faculty of Language (Broad). That means its input is the sentence structure completely formed and, in some sense, closed as one linear object. So, the manipulation of this level is completely suprasegmental, and it is applied to the sentence as a unit, instead of to sections of it.

5 CONCLUDING REMARKS

Prosody is part of our Language Faculty. And as such, it has a unique complexity that we don't find anywhere else in the animal kingdom. However, it seems possible that it wasn't always like that. In section 2.1.2, we argued that there is a proximity between speech and animal systems of communication, and that that proximity might be a consequence of the way language emerged in human being's evolutionary history. The strong hypothesis is that prosody was the main feature of human being's protolanguage, the link between animal system of communication and human language. With the emergence of language with Human beings' evolution, prosody also developed into a more complex system, possibly facilitated by the emergency of background capacities, such as structured and hierarchical planning. The new, evolved, complex Prosody system would be an independent module within the Language system, that would be in close interaction with other language's modules. We proposed that the Prosody module would output three categories of representations: a rhythm structure, a harmonic structure, and a dynamic structure, where the rhythm structure would be mapping syntax structure into prosodic phrasing, the harmonic structure would be mapping semantic and pragmatic content into pitch intonation, and the dynamic structure would be mapping emotions and mental state into voice quality. To test the rhythm structure, we conducted experiments in Portuguese and English, where we tested the sufficiency of the salience of a prosodic boundary. The results of the experiment in Portuguese consistently reinforce our hypothesis of prosodic rhythmic structure mapping syntax structure, as significantly most of the time, the participants were able to retrieve the proper syntactic structure based on our delexicalized experimental sentences audios. On the other hand, our results for the experiment in English are not consistent with our hypothesis. We strongly argue that the results are under the influence of artifacts and that it is a matter of design and properly conducting the experiment. There's plenty of literature that shows successful mapping between syntax and prosody.

To test the harmonic structure, we conducted an experiment in Portuguese where we research the threshold for pitch accent salience. The data don't fully meet our predictions of proper interpretation of actants of the enunciation act. However, several studies, as reviewed on sections 2.3.3 and 3.3, showed consistent interaction between intonation and pragmatic information. We strongly argue that the number of variables and the amount of options in the task posed an extra layer of difficulty to the participant's mind when processing the sentence, acting as artifacts. A design with fewer options, splitting the variables to be tested in two or three separate experiments might display different results. In despite of the lack of supporting experimental data, the theories reviewed show consistency with our hypothesis of Prosody being and independent domain, withing the language system, processed by its dedicated module – a hypothesis that we aim to explore further in. the near future.

Going forward, the hypotheses for the three types of prosodic representation need further testing in order to understand its operations, representations, and interactions with other modules within the language system.

6 REFERENCES

ANDRADE, I. R. de; FRANÇA, A. I. **A influência da alfabetização nas cognições de reconhecimento de faces e dos sinais da fala:** um estudo psicolinguístico the influence of literacy on face recognition and speech sign cognitions: a psycholinguistic study. *Revista Prolíngua*, 1983(9979), p.101-126. 2021a.

________. **Ultrapassando os limites entre o natural e o cultural:** Uma revisão neuropsicolinguística sobre a aquisição da leitura. *Letrônica*, [S. l.], v. 14, n. 2, p. e38747, 2021b.

CHRISTOPHE, A.; GUASTI, T.; NESPOR, M.; DUPOX, E.; VAN OOYEN, B. **Reflections on phonological bootstrapping: its role for lexical and syntactic acquisition.** *Language and Cognitive Processes*, v. 12, n 5-6, p. 585-612. 1997.

BADDELEY, A. D.; HITCH, G. **The recency effect:** Implicit learning with explicit retrieval?. *Memory & Cognition*, 21(2), 146-155. 1993.

BAKER, M. C. **On some subject/object non-asymmetries in Mohawk.** *Natural Language & Linguistic Theory*, 9(4), 537-576. 1991.

BALSBY, T. J.; MOMBERG, J. V.; DABELSTEEN, T. **Vocal imitation in parrots allows addressing of specific individuals in a dynamic communication network**. *PLoS One*, 7(11), e49747. 2012.

BARRETT, M. J. **The holophrastic hypothesis:** Conceptual and empirical issues. *Cognition*, 11(1), 47-76. 1982.

BASS, A.H., RICE, A.N. **Vocal-acoustic communication in fishes**: Neuroethology. In: Breed, M.D., Moore, J. (Eds.), *Encyclopedia of Animal Behavior*. Vol. 3, pp. 558–567. 2010.

BASS, A.H.; RICE, A.N.; FENG, N.Y. **Singing behavior in fishes:** hormones, neurons, and evolution. Choe, J.C. (ed.): *Encyclopedia of Animal Behavior* (2nd Edition). Academic Press, pp. 340-351. 2019.

BENTIN, S.; DEOUELL, L. Y.; SOROKER, N. **Selective visual streaming in face recognition**, *NeuroReport*:, 19–9 - Volume –0 - Issue–4 - p 823-827. 1999.

BERGMAN, T.J., et al. **Hierarchical classification by rank and kinship in baboons.** *Science*, 302(5648), pp.1234-1236. 2003.

BERWICK, R. C. et al. **Songs to syntax:** the linguistics of birdsong. *Trends in cognitive sciences*, 15(3), 113-121. 2011a.

________. **Poverty of the stimulus revisited**. *Cognitive Science*, 35(7), 1207-1242. 2011b.

BIRCH, E.; PETRIG, B. **FPL and VEP measures of fusion, stereopsis and stereoacuity in normal infants.** *Vision Res.* 36(9):1321–1327. 1996.

BLOOM, A.; KIRSCH, A. **The republic of Plato**. New York: Basic Books, v.2. 1968.

BOERSMA, P.; WEENINK, D. **Praat:** doing phonetics by computer [Computer program]. Version 6.0.33, *retrieved 17 September 2017* from http://www.praat.org/

BOLHUIS, J. J.; OKANOYA, K.; SCHARFF, C. **Twitter evolution:** converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience*, 11(11), 747-759. 2010.

BOLINGER, D. **Accent is predictable** (if you're a mind-reader). *Language*, pp.633-644. 1972.

BORER, H. **Exo-skeletal vs. Endo-skeletal Explanations:** Syntactic Projections and the Lexicon. In: POLINSKY, M.; MOORE, J. (eds.) *The Nature of Explanation*. Chicago: Chicago University Press. 2003.

________. **In Name Only**. Structuring Sense, Volume I. Oxford: Oxford University Press. 2005a.

_______ . **The Normal Course of Events**. Structuring Sense, Volume II. Oxford: Oxford University Press. 2005b.

_______ . **Taking Form:** Structuring Sense Volume III. Oxford: Oxford University Press. 2013.

BOYDEN, A. **Homology and analogy.** A critical review of the meanings and implications of these concepts in biology. *American Midland Naturalist*, 648-669. 1947.

BUCKLAND, M. K. **Information as thing**. *Journal of the American Society for information science*, 42(5), 351-360. 1991.

BÜRING, D. **Intonation and meaning.** Oxford University Press. 2016.

BURNHAM, D.; KITAMURA, C.; LANCUBA, V. **The development of linguistic attention in early infancy:** The role of prosodic and phonetic information. In *Proceedings of the 14th International Congress of Phonetic Sciences*. Berkeley: University of California, pp. 1197-1200. 1999.

CHEN, Z. AND WIENS, J.J. **The origins of acoustic communication in vertebrates**. *Nature communications*, 11(1), p.369. 2020.

CHOMSKY N. **Syntactic Structures**. Mouton and co.: The Hague. 1957.

_________. **Aspects of the Theory of Syntax**. Cambridge, Mass.: MIT Press. 1965.

_________. **On Cognitive Structures and their Development:** A reply to Piaget. In M. Piattelli-Palmarini, ed. *Language and Learning:* The debate between Jean Piaget and Noam Chomsky. Harvard University Press. 1980.

_________. **Lectures on Government and Binding.** Dordrecht: Foris. 1981a.

_________. **Principles and parameters in syntactic theory.** In *Hornstein and Lightfoot*. 1981b.

_________. **Some Concepts and Consequences of the Theory of Government and Binding.** Cambridge, Mass.: MIT Press. 1982.

_________. **Lectures on government and binding:** The Pisa lectures (No. 9). Walter de Gruyter: 1993a.

________. **A minimalist program for linguistic theory.** In Hale, K.; Keyser, S.J. (eds.), *The View From Building 20:* Essays in Linguistics in Honor of Sylvain Bromberger. MIT Press. 1993b.

_________. **Three factors in language design.** *Linguistic inquiry*, 36(1), 1-22. 2005.

_________. **On phases.** *Current Studies in Linguistics Series*, 45, 133. 2008.

_________. **Problems of projection**. *Lingua*, 130, 33-49. 2013.

_________. **The minimalist program.** MIT press. 2014.

CHUENWATTANAPRANITHI, S. et al. **Encoding emotions in speech with the size code.** *Phonetica,* 65(4), pp.210-230. 2008.

CLACK, J.A. **The fin to limb transition:** new data, interpretations, and hypotheses from paleontology and developmental biology. *Annual Review of Earth and Planetary Sciences*, 37, pp.163-179. 2009.

CHRISTOPHE, A.; GUASTI, T.; NESPOR, M., 1997. **Reflections on phonological bootstrapping:** Its role for lexical and syntactic acquisition. *Language and cognitive processes*, 12(5-6), pp.585-612.

CHRISTOPHE, A. et al. **Bootstrapping lexical and syntactic acquisition.** *Language and speech*, 51(1-2), pp.61-75. 2008.

COHEN, L; DEHAENE, S. Specialization within the ventral stream: the case for the visual word form area, Neuroimage, n. 22, p. 466-476, 2004.

COOPER, R. P. et al. **The development of infants' preference for motherese**. *Infant behavior and development*, 20(4), 477-488. 1997.

COSMIDES, L.; TOOBY, J. **From evolution to behavior:** Evolutionary psychology as the missing link. In DUPRÉ, J. (ed.) *The latest on the best:* Essays on evolution and optimality. Cambridge, MA: MIT Press, pp. 276–306. 1987.

________. **From function to structure:** The role of evolutionary biology and computational theories in cognitive neuroscience. In GAZZANIGA, M. (ed.) *The cognitive neurosciences*. Cambridge, MA: MIT Press, pp. 1199–1210. 1995.

COSMIDES, L.; TOOBY, J.; BARKOW, J.H. **Introduction:** Evolutionary psychology and conceptual integration. In BARKOW, J.H.; COSMIDES, L.; TOOBY, J. (ed.) *The adapted* *mind:* Evolutionary psychology and the generation of culture. New York: Oxford University Press, pp. 3–15. 1992.

CRISTÓFARO-SILVA, T. **Fonética e fonologia do português:** roteiro de estudos e guia de exercícios. São Paulo: Contexto. 1999.

CRYSTAL, D. **Prosodic systems and intonation in English.** Cambridge: Cambridge University Press. 1969.

CURTISS, S. **Genie:** a psycholinguistic study of a modern-day wild child. Academic Press. 2014.

DA SILVA, H. P. N. **O bootstrapping prosódico na aquisição da linguagem e as funções da prosódia na compreensão de enunciados linguísticos.** *Anais do IX SAPPIL-Estudos de Linguagem*. 2019.

DALVA, D. **Automatic speech recognition system for Turkish spoken language**. Master's thesis, *Işık Üniversitesi*. 2012.

DAMASIO, A. R. **Aphasia.** *New England Journal of Medicine*, 326(8), 531-539. 1992.

DANKOVICOVA, J. et al. **The relationship between musical skills, music training, and intonation analysis skills**. *Language & Speech* 50: 177-225. 2007.

DARWIN, G. **The Origin of Species by Natural Selection:** Or, The Preservation of Favored Races in the Struggle for Life. 1859.

________. **The descent of man and selection in relation to sex**. London: John Murray. 1871.

DAWKINS, R. **Twelve misunderstandings of kin selection**. *Zeitschrift für Tierpsychologie*, 51(2), 184-200. 1979.

DE ANDRADE, I. R., & FRANÇA, A. I. **A influência da alfabetização nas cognições de reconhecimento de faces e dos sinais da fala:** um estudo psicolinguístico. *PROLÍNGUA*, 15(2), 101-126. 2020.

DECASPER, A.; FIFER, W. **Of human bonding:** newborns prefer their mothers' voices. *Science*, 208(4448), 1174–1176. 1980.

DEHAENE, S. et al. **How learning to read changes the cortical networks for vision andlanguage.** *Science*, v. 330, n. 6009, p. 1359-1364, 2010.

DICARLO, J. J.; ZOCCOLAN, D.; RUST, N. C. **How does the brain solve visual object recognition?** *Neuron*, 73(3), 415–434. 2012.

DONALD, M. **Origins of the modern mind:** Three stages in the evolution of culture and cognition. Harvard University Press. 1991.

ELFNER, E. **The syntax-prosody interface:** Current theoretical approaches and outstanding questions. *Linguistics Vanguard*, 4(1). 2018.

ESLING, J.H. **Voice Quality.** In**:** Chapelle, C.A. (ed.) *The Encyclopedia of Applied Linguistics*, 2012.

ESLING, J.H., et al. **Voice Quality**: The Laryngeal Articulator Model. (No. 1). Cambridge University Press. 2019.

EVANS, J.J. et al. **Progressive prosopagnosia associated with selective right temporal lobe atrophy:** A new syndrome?. *Brain*, 118(1), pp.1-13. 1995.

EVERETT, D. **Cultural constraints on grammar and cognition in Pirahã:** Another look at the design features of human language. *Current anthropology*, 46(4), 621-646. 2005.

________. **Pirahã culture and grammar:** A response to some criticisms. *Language*, 405-442. 2009.

EYSENCK, M. W. **Anxiety and attention.** *Anxiety Research*: An International Journal, 1:1, pp.9-15. 1988.

________. **A cognitive approach to trait anxiety**. *European Journal of Personality*, 14(5), 463–476. 2000.

FABBRO, F. **The bilingual brain:** Cerebral representation of languages. *Brain and language*, 79(2), 211-222. 2001.

FERNALD, A. **Four-month-old infants prefer to listen to motherese**. *Infant behavior and development*, 8(2), 181-195. 1985.

FERNALD, A.; MCROBERTS, G. **Prosodic bootstrapping:** A critical analysis of the argument and the evidence. In *Signal to syntax,* pp. 377-400. Psychology Press. 2014.

FERNANDES-SVARTMAN, F. R. **A entoação das sentenças clivadas em português brasileiro e a interface sintaxe-fonologia.** *Filologia e linguística portuguesa,* 14(1), 37-56. 2012.

FERNÁNDEZ, E. M.; CAIRNS, H. S. **Fundamentals of psycholinguistics.** John Wiley & Sons: 2010.

FINK, B. R. **Larynx and Speech as Determinants in the Evolution of Man**. *Perspectives in Biology and Medicine*, 7(1), pp.85-93. 1963.

FISHER, S. E. **Tangled webs:** Tracing the connections between genes and cognition. *Cognition*, 101(2), 270-297. 2006.

_________. **A molecular genetic perspective on speech and language**. In *Neurobiology of language,* pp. 13-24. Academic Press. 2016.

FISHER, S. E.; MARCUS, G. F. **The eloquent ape:** genes, brains and the evolution of language. *Nature Reviews Genetics*, 7(1), 9-20. 2006.

FISHER, C.; TOKURA, H. **Prosody in speech to infants:** Direct and indirect acoustic cues to syntactic structure. In *Signal to syntax,* pp. 355-376. Psychology Press. 2014.

FITCH, W. **The evolution of speech:** a comparative review. *Trends in cognitive sciences*, 4(7), pp.258-267. 2000.

________. **Comparative vocal production and the evolution of speech:** reinterpreting the descent of the larynx. In: Wray A. (ed.), *The Transition to Language*. Oxford University Press, Oxford, pp. 21 – 45. 2002.

________. **The evolution of language:** a comparative review. *Biology and philosophy*, 20(2), 193-203. 2005.

FITCH, W.T.; HAUSER, M.D.; CHOMSKY, N. **The evolution of the language faculty:** Clarifications and implications. *Cognition*, 97(2), pp.179-210. 2005.

FITCH, W.T.; HUBER, L.; BUGNYAR, T. **Social cognition and the evolution of language:** constructing cognitive phylogenies. *Neuron*, 65(6), pp.795-814. 2010.

FITCH, W.T.; REBY, D. **The descended larynx is not uniquely human.** *Proceedings of the Royal Society of London*. Series B: Biological Sciences, 268(1477), pp.1669-1675. 2001.

FODOR, J.A. **The modularity of mind.** MIT press. 1983.

________. **Précis of The Modularity of Mind.** *Behavioral and Brain Sciences*, 8(1), 1-5. 1985.

FODOR, J.D. **Learning to parse?** *Journal of psycholinguistic research*, 27(2), 285-319. 1998.

_______. **Prosodic disambiguation in silent reading**. In North East Linguistics Society (Vol. 32, No. 1, p. 8). 2002a.

_______. **Psycholinguistics cannot escape prosody.** In *Speech Prosody*, International Conference. 2002b.

FONSECA, A. A. **A prosódia no parsing:** evidências experimentais do acesso à informação prosódica no input linguístico. 2012. Pp. 208 (Letras – Linguística). Universidade Federal de Minas Gerais. 2012.

FRANÇA, A. I.; GOMES, J. N. **A técnica de ERP:** investigando a assimetria sujeito-objeto na interface sintaxe-semântica com EEG. *Letras de Hoje*, 50(3), 360-370. 2015.

FRANKENHUIS, W.E.; PLOEGER, A. **Evolutionary psychology versus Fodor:** Arguments for and against the massive modularity hypothesis. *Philosophical Psychology*, 20(6), pp.687-710. 2007.

FRAZIER, L. **On comprehending sentences:** Syntactic parsing strategies. Unpublished doctoral dissertation, *University of Connecticut*. 1978.

FRIEDENBERG, J.; SILVERMAN, G.; SPIVEY, M. J. **Cognitive science:** an introduction to the study of mind. Sage Publications. 2021.

FRIEDERICI, A. D. **The brain basis of language processing:** from structure to function. *Physiological reviews*, 91(4), 1357-1392. 2011.

________. (Ed.). **Language comprehension:** A biological perspective. Springer Science & Business Media. 2012.

FRIEDERICI, A. D.; GIERHAN, S. M. The language network. Current opinion in neurobiology, 23(2), 250-254. 2013.

GAZDAR, G. Pragmatics: Implicature, Presupposition and Logical Form. New York: Academic Press, 1979.

GERVAIN, J. **The role of prenatal experience in language development**. *Current Opinion in Behavioral Sciences*, 21, 62-67. 2018a.

________. **Gateway to language:** The perception of prosody at birth. In *Boundaries crossed, at the interfaces of morphosyntax, phonology, pragmatics and semantics*. Springer, Cham, pp. 373-384. 2018b.

GIBSON, J. J. **The senses considered as perceptual systems**. Boston, MA: Houghton Mifflin. 1966.

GIBSON, K.R. **Tool use, language and social behavior in relationship to information processing capacities.** In: Gibson, K.R.; Ingold, T., editors. *Tools, language and cognition in human evolution.* Oxford: Oxford University Press. 251–269. 1993.

GIRAUD, A. L.; POEPPEL, D. **Cortical oscillations and speech processing:** emerging computational principles and operations. *Nature neuroscience*, 15(4), 511-517. 2012.

GLEITMAN, L. R. et al. **The impossibility of language acquisition** (and how they do it). *Annual Review of Linguistics*, 5, 1-24. 2019.

GORDON, P.C., GROSZ, B.J., & GILLIOM, L.A. **Pronouns, names, and the centering of attention in discourse.** *Cognitive Science*, 17, 311–347. 1993.

GRAHAM, S. A., & FISHER, S. E. **Decoding the genetics of speech and language.** *Current opinion in neurobiology*, 23(1), 43-51. 2013.

GRAVINA, A. P.; FERNANDES-SVARTMAN, F. **Interface sintaxe-fonologia:** desambiguação pela estrutura prosódica no português brasileiro. Alfa: *Revista de Linguística,* 57(2), 639-668. 2013.

GRICE, H. P. **Logic and conversation**. In *Speech acts* (pp. 41-58). Brill. 1975.

HARTSUIKER, R. J., & MOORS, A. On the automaticity of language processing. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning:* How we reorganize and adapt linguistic knowledge, pp. 201–225. 2017.

HAUSER, M. D. **The evolution of communication**. MIT press. 1996.

HAUSER, M.D.; CHOMSKY, N.; FITCH, W.T. **The faculty of language:** what is it, who has it, and how did it evolve?. *Science*, 298(5598), pp.1569-1579. 2002.

HICKOK, G.; POEPPEL, D. **Towards a functional neuroanatomy of speech perception.** *Trends in cognitive sciences*, 4(4), 131-138. 2000

_______. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2), 67-99. 2004.

_______. The cortical organization of speech processing. *Nature reviews neuroscience*, 8(5), 393-402. 2007.

HOLLEBRANDSE, B.; and ROEPER, T. **Empirical results and formal approaches to recursion in acquisition**. In *Recursion:* complexity in cognition (pp. 179-219). Springer, Cham. 2014.

HU, J. et al. **Estimation of 3-D Surface Displacement Based on InSAR and Deformation Modeling**. *IEEE Transactions on Geoscience and Remote Sensing*, 55, pp. 2007-2016. 2017.

HUBEL, D. H.; WIESEL, T. N. **Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.** *The Journal of physiology*, 160(1), 106. 1962.

IBM Corp. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp. Released 2011.

JAMINI, D. **Harmony and Composition:** Basics to Intermediate. Trafford on Demand Pub. 2005.

JANIK, V.M.; SLATER, P.J. **Vocal learning in mammals**. *Advances in the Study of Behaviour*, 26(1), pp.59-100. 1997.

JOUBERT, S. et al. **Impaired configurational processing in a case of progressive prosopagnosia associated with predominant right temporal lobe atrophy**. *Brain*, 126(11), pp. 2537–2550. 2003.

JÜRGENS, U. **Neuronal control of mammalian vocalization, with special reference to the squirrel monkey**. *Die Naturwissenschaften*, 85(8), pp.376-388. 1998.

KAKO, E. **Elements of syntax in the systems of three language-trained animals**. *Animal Learning & Behavior*, 27(1), pp.1-14. 1999.

KENTNER, G. **Rhythmic segmentation in auditory illusions**-Evidence from cross-linguistic mondegreens. *ICPhS*. 2015.

KIRBY, S. **The evolution of language**. *Oxford handbook of evolutionary psychology*, pp.669-681. 2007.

KNÖRNSCHILD, M., et al. **Complex vocal imitation during ontogeny in a bat**. *Biology letters*, 6(2), 156-159. 2010.

KRALJIC, T.; SAMUEL, A. G.; BRENNAN, S. E. **First impressions and last resorts:** How listeners adjust to speaker variability. *Psychological science*, 19(4), 332-338. 2008.

KRATZER, A., & HEIM, I. **Semantics in generative grammar.** Oxford: Blackwell. 1998.

KUHL, P.K. and MILLER, J.D. **Speech perception by the chinchilla:** Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190(4209), pp.69-72. 1975.

KUTAS, M.; FEDERMEIER, K. D. **Thirty years and counting:** finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, 62, 621-647. 2011.

LADD, D. R. **Intonational phrasing: the case for recursive prosodic structure.** *Phonology*, Cambridge: Cambridge University Press, v. 3, p. 311-340, 1986.

______. **Declination reset and the hierarchical organization of utterances**. *Journal of the Acoustical Society of America*, v. 84, p. 530-544, 1988.

_______. **Intonational phonology**. Cambridge University Press. 2008.

_______. **Simultaneous structure in phonology.** Vol. 28. OUP Oxford. 2014.

LAGE, A.C. **Aspectos Neurofisiológicos de Concatenação e Idiomaticidade em Português do Brasil: Um Estudo de Potenciais Bioelétricos Relacionados a Eventos Lingüísticos (ERPs)**. 2005. 185. (Letras – Linguística). *Universidade Federal do Rio de Janeiro*, Rio de Janeiro, 2005.

LAGE, A. C. et al. **Potenciais Relacionados a Eventos (ERP) revelam o curso da derivação sintática e a dinâmica da integração entre micromódulos de cognição de linguagem.** *Ciências & Cognição*, 13(2), 03-13. 2008.

LAITZ, S. G.; LAITZ, S. G. **The complete musician**: An integrated approach to tonal theory, analysis, and listening. Oxford University Press, USA. 2008.

LANDAU, I. **The scale of finiteness and the calculus of control.** *Natural Language & Linguistic Theory*, 22(4), 811-877. 2004.

LAURENCE, S.; MARGOLIS, E. **The poverty of the stimulus argument**. *British Journal for the Philosophy of Science*, 52(2), 217-276. 2001.

LAWRENCE, E. (Ed.). **Henderson's dictionary of biology**. Pearson education. 2005.

LE BAS, G.A. et al**. The role of antenatal and postnatal maternal bonding in infant development:** A systematic review and meta-analysis. *Social development*, 29(1), 3-20. 2020.

LENT, R**. Cem bilhões de neurônios: conceitos fundamentais de neurociência.** 2004.

LEVELT, W. J. **A history of psycholinguistics:** The pre-Chomskyan era. Oxford University Press. 2013.

LEWIS, E.O. **The effect of practice on the perception of the Müller-Lyer illusion**. *British journal of psychology*, 2(3), p.294. 1908.

LIEBERMAN, P. **Vocal tract anatomy and the neural bases of talking**. *Journal of Phonetics*, 40(4), pp.608-622. 2012.

LIEM, K.F. **Evolutionary strategies and morphological innovations:** cichlid pharyngeal jaws. *Systematic zoology*, 22(4), pp.425-441. 1973.

LYN, H.; GREENFIELD, P.M.; SAVAGE-RUMBAUGH, E.S. **Semiotic combinations in Pan:** A comparison of communication in a chimpanzee and two bonobos. *First Language*, 31(3), pp.300-325. 2011.

MACLARNON, A.; HEWITT, G. **The evolution of human speech:** the role of enhanced breathing control. *Am. J. Phys. Anthropol*. 109, 341–363. 1999.

MAMPE, B. et.al. **Newborns' cry melody is shaped by their native language.** *Current biology*, 19(23), 1994-1997. 2009.

MARR, D. **Vision:** A computational investigation into the human representation and processing of visual information. San Francisco: W. H. Freeman. 1982.

MARSLEN-WILSON, W. D. et al. **Prosodic effects in minimal attachment**. *The Quarterly Journal of experimental psychology*, 45(1), 73-87. 1992.

MARTINEZ-ALVAREZ, A.; BENAVIDES-VARELA, S.; GERVAIN, J. **The newborns' brain detects utterance-level prosodic contours**. *34th Annual CUNY Conference on Human Sentence Processing.* 2021.

MATEUS, M. H. M.; FALÉ, I.; FREITAS, M. J. **Fonética e fonologia do português.** Universidade Aberta. 2016.

MCMAHON, E.; WINTERMARK, P.; LAHAV, **A. Auditory brain development in premature infants:** the importance of early experience. *Annals of the New York Academy of Sciences*, 1252(1), 17-24. 2012.

MECHELLI, A. et al. **Structural plasticity in the bilingual brain**. *Nature*, 431(7010), 757-757. 2004.

MENARY, R. **Dimensions of mind**. *Phenomenology and the Cognitive Sciences*, 9(4), 561-578. 2010.

MERCADO, E. III,;MANTELL, J. T.; PFORDRESHER, P. Q. **Imitating sounds:** A cognitive approach to understanding vocal imitation. *Comparative Cognition & Behavior Reviews*, 9, 1–57. 2014.

MICHAEL W. E. **Anxiety and attention**, *Anxiety Research*, 1:1, 9-15, 1988.

MOORE, B.R. **Avian movement imitation and a new form of mimicry:** tracing the evolution of a complex form of learning. *Behaviour* 12: 231 – 263. 1992.

MORAES, J. A. **From a prosodic point of view**: remarks on attitudinal meaning. In Heliana Mello, Alessandro Panunzi, Tommaso Raso (ed.): *Pragmatics and Prosody*: Illocution, Modality, Attitude, Information Patterning and Speech Annotation. Firenze: Firenze University Press, 2011a.

_______. **A entoação de atos de fala diretivos no PB**. *Resumos do III Colóquio Brasileiro de Prosódia da Fala.* 2011b.

MORAES, J. A.; COLAMARCO, M..**Você Está Pedindo Ou Perguntando?** Uma análise entonacional de pedidos e perguntas no português do brasil. *Revista de estudos da linguagem*, V. 15, P. 113-126, 2007.

MORAES, J. A.; RILLIARD, A. **Prosody and Emotion in Brazilian Portuguese**. In *Intonational Grammar in Ibero-Romance*. John Benjamins, pp. 135-152. 2016.

MORENAS-RODRÍGUEZ, E. ET Al. **Visual hallucinations in patients with acute stroke:** a prospective exploratory study. *European journal of neurology*, 24(5), 734–740. 2017.

MORRIS, H.; MARANTZ, A. **Distributed Morphology and the Pieces of Inflection.** In Hale, K.; Keyser, S.J. (ed.) *The View from Building 20*. MIT Press, Cambridge, 111–176. 1993.

MUNKONG, R.; JUANG, B.H. **Auditory perception and cognition**. *IEEE signal processing magazine*, 25(3), pp.98-117. 2008.

NAME, M. C. **Pistas prosódicas, sintáticas e semânticas facilitadoras da identificação dos elementos das categorias N e ADJ**. *Veredas*, v. 2, p. 1-15, 2008.

NEVINS, A.; PESETSKY, D.; RODRIGUES, C. **Pirahã exceptionality:** A reassessment. *Language*, 355-404. 2009a.

________. **Evidence and argumentation:** A reply to Everett. *Language*, 85(3), 671-681. 2009b.

NEWPORT, E. L. **Motherese:** The speech of mothers to young children. *University of Pennsylvania* doctoral dissertation. 1975.

NEWPORT, E.; GLEITMAN, H.; GLEITMAN, L. **Mother, I'd rather do it myself:** Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children* (pp. 109– 149). Cambridge: Cambridge University Press. 1977.

_______. **Mother, I'd rather do it myself.** Sentence First, Arguments Afterward: *Essays in Language and Learning*, 141. 2020.

NOTTEBOHM, F. Discussion paper:* **Vocal tract and brain:** a search for evolutionary bottlenecks. *Annals of the New York Academy of Sciences*, 280(1), pp.643-649. 1976.

O'CINNEIDE, A.; DORRAN, D.; GAINZA, M. **Linear Prediction:** The Problem, its Solution and Application to Speech. Conference Paper. *Dublin Institute of Technology*. 2008.

PANNEKAMP, Ann et al. **Prosody-driven sentence processing:** An event-related brain potential study. *Journal of cognitive neuroscience*, v. 17, n. 3, p. 407-421, 2005.

PATTERSON, D.K. et al. **How parrots talk:** Insights based on CT scans, image processing, and mathematical models. In *Medical Imaging 199*7: Physiology and Function from Multidimensional Images, Vol. 3033, pp. 14-24. 1997.

PINKER, S. **The language instinct**. William Morrow & Co. 1994.

PLATO, B. **The allegory of the cave.** P & L Publication. 2010.

POTTER, M. C. et al. **Detecting meaning in RSVP at 13 ms per picture**. *Attention, perception & psychophysics*, 76(2), 270–279. 2014.

PRUITT, K.; ROELOFSEN, F. **The interpretation of prosody in disjunctive questions**. *Linguistic inquiry*, 44(4), pp.632-650. 2013.

PYLKKÄNEN, L.; MARANTZ, A. **Tracking the time course of word recognition with MEG**. *Trends in cognitive sciences*, 7(5), 187-189. 2003.

RIBEIRO, Nathacia L. **Recursividade, coordenação e prosódia:** estudos psicolinguísticos com PPs em português e hebraico. 2015. Pp. 78. (Letras, Linguística) – *Universidade Federal do Rio de Janeiro*, Rio de Janeiro: 2015a.

_______. **Processamento do foco prosódico em clivadas invertidas reduzidas.** *ReVEL*, special edition n. 10, 2015b.

RIBEIRO, N. L.; LAGE, A.C..**Recursividade e coordenação em português a partir de sintagmas preposicionais:** da prosódia à sintaxe em estudos psicolinguísticos. *Letras de Hoje* (Online), v. 50, p. 112-119, 2015.

RICE, M. L. **Toward a genetics of language**. Psychology Press, 2013.

RILLIARD, A.; MORAES, J. A.; ERICKSON, D.; SHOCHI, T.. **Social affect production and perception across languages and cultures?** The role of prosody. *Leitura* (Ufal), V. 52, P. 15, 2013.

ROCHE, H., et al**. Early hominid stone tool production and technical skill 2.34 Myr ago in West Turkana, Kenya**. *Nature* 399: pp. 57-60. 1999.

ROFFMAN, I. et al. **Meaningful Informational Exchange and Pantomime in Chimpanzees and Bonobos:** Implications for Proto-Language in Hominins. *Human Evolution*, 30. 2015.

ROSS, D. **Against Positing Central Systems in the Mind**. *Philosophy of Science*, 57(2), 297–312. 1990.

RUBEN, R.J. **The ontogeny of human hearing**. *Acta Otolaryngol*. 112: 192–196. 1992.

SAINT-GEORGES, C. et al. **Motherese in interaction:** at the cross-road of emotion and cognition?(A systematic review). *PloS one*, 8(10), e78103. 2013.

SAUSSURE, F. de. **Curso de linguística geral**. Editora Cultrix. 2008.

SAVAGE-RUMBAUGH, E.S., RUMBAUGH, D.M. and BOYSEN, S. **Do Apes Use Language?** One research group considers the evidence for representational ability in apes. *American Scientist*, 68(1), pp.49-61. 1980.

SAVAGE-RUMBAUGH, E.S., RUMBAUGH, D.M. and MCDONALD, K. **Language learning in two species of apes.** *Neuroscience & Biobehavioral Reviews*, 9(4), pp.653-665. 1985.

SAVAGE-RUMBAUGH, E.S. et al. **Language comprehension in ape and child**. *Monographs of the society for research in child development,* pp.i-252. 1993.

SCHERER, K. R. **Speech and emotional states**. In Darby, J. (ed.) *Speech evaluation in psychiatry*. New York: Grune & Stratton, pp. 189-220. 1981.

________. **What are emotions?** And how can they be measured?. *Social science information*, 44(4), 695-729. 2005.

SCHMIDT-KASSOW, M.; KOTZ, S.A. **Event-related Brain Potentials Suggest a Late Interaction of Meter and Syntax in the P600.** *Journal of Cognitive Neuroscience*, 21(9), 1693–1708. 2009.

SCOTT, D.R.; CUTLER, A. **Segmental phonology and the perception of syntactic structure.** *Journal of Verbal Learning and Verbal Behavior*, 23(4), pp.450-466. 1984.

SEARLE, J. **Speech Acts**. An Essay in the Philosophy of Language. Cambridge: Cambridge University Press. 1969.

SEYFARTH, R.M.; CHENEY, D.L.; BERGMAN, T.J. **Primate social cognition and the origins of language.** *Trends in cognitive sciences*, 9(6), pp.264-266. 2005.

SIMON, H. A. **The Architecture of Complexity**. *Proceedings of the American Philosophical Society*, 106(6), 467–482. 1962.

SIU, C. R.; MURPHY, K. M. **The development of human visual cortex and clinical implications**. *Eye and brain*, 10, 25–36. 2018.

SONNTAG, C.F. **The comparative anatomy of the koala (*Phascolartos cinereus*) and vulpine phalanger *(Trichosurus vulpecula*).** In *Proceedings of the Zoological Society of London*, Vol. 91, No. 3, pp. 547-577. Oxford, UK: Blackwell Publishing Ltd. 1921.

SPEARMAN, C. **The Measurement of Intelligence**. *Nature* 120, 577–578. 1927.

STUDDERT-KENNEDY, M. **Mirror neurons, vocal imitation, and the evolution of particulate speech.** *Mirror neurons and the evolution of brain and language*, 207-227. 2002.

SUEUR, J. **What Is Sound?.** In *Sound Analysis and Synthesis with R* (pp. 7-36). Springer, Cham. 2018.

TAYLOR, G., et al.. **The development of contour processing:** evidence from physiology and psychophysics. *Frontiers in psychology,* 5, 719. 2014.

TCHERNICHOVSKI, O. **Dynamics of the vocal imitation process:** how a zebra finch learns its song. *Science*, 291(5513), 2564-2569. 2001.

TELLER, D.Y.; PEEPLES, D.R.; SEKEL, M. **Discrimination of chromatic from white light by two-month-old human infants.** *Vision Res.* 18(1):41–48. 1978.

THAGARD, P. **Mind: Introduction to cognitive science.** MIT press. 2005.

UOMINI, N. T.; MEYER, G. F. **Shared brain lateralization patterns in language and Acheulean stone tool production:** a functional transcranial Doppler ultrasound study. *PloS one*, 8(8), e72693. 2013.

VARGHA-KHADEM, F. et al. **Neural basis of an inherited speech and language disorder.** *Proceedings of the National Academy of Sciences*, 95(21), pp.12695-12700. 1998.

VIGÁRIO, M.; FERNANDES-SVARTMAN, F. **A atribuição de acentos tonais em compostos no português do Brasil. Textos Seleccionados**. *XXV Encontro Nacional da Associação Portuguesa de Linguística,* pp.769-786. 2010.

WAGNER, G. P**. The biological homology concept**. *Annual review of Ecology and Systematics*, 20(1), 51-69. 1989.

WAGNER, M. **Prosody as a diagonalization of syntax. evidence from complex predicates.** In *PROCEEDINGS-NELS*, Vol. 34, No. 2, pp. 587-602. 2004.

________. **Prosodic Evidence for Recursion?** Posted on *lingbuzz* in December 2007.

________. **Phonological Evidence in Syntax.** Tibor Kiss and Artemis Alexiadou (Eds.): *Syntax—Theory and Analysis*. An International Handbook. Handbooks of Linguistics and Communication Science. 42.1-3, Berlin: Mouton de Gruyte. 2015.

Wikipedia. **Melody**. *Wikipedia*, 2002. Available at: https://en.wikipedia.org/wiki/Melody . Accessed in May, 2017.

WILCOX, T. **Object individuation:** infants' use of shape, size, pattern, and color. *Cognition.* 72(2):125–166. 1999.

FRANKENHUIS, W.E.; PLOEGER, A. **Evolutionary psychology versus Fodor:** arguments for and against the massive modularity hypothesis. *Philosophical Psychology*, 20:6, 687-710. 2007.

XU, Y. **Speech Prosody**: Theories, models and analysis. In A. R. Meireles (ed.): *Courses on Speech Prosody.* Cambridge Scholars Publishing. Pp. 146-177. 2015.

APPENDIX

1) Experimental sentences list for Experiment 1

NT	LISTA	CONDICAO	SENTENCA
1	1	SN	Esta manhã, estudantes competentes divulgaram a proposta.
2	2	SF	Esta manhã, estudantes e tutores divulgaram a proposta.
3	3	ON	Esta manhã, estudantes divulgaram a proposta rejeitada.
4	4	OF	Esta manhã, estudantes divulgaram a proposta e o problema.
5	4	SN	Mês passado, promotores desordeiros delataram o prefeito.
6	1	SF	Mês passado, promotores e juízes delataram o prefeito.
7	2	ON	Mês passado, promotores delataram o prefeito desonesto.
8	3	OF	Mês passado, promotores delataram o prefeito e o ministro.
9	3	SN	Este ano, redatores criativos compuseram o roteiro.
10	4	SF	Este ano, redatores e amadores compuseram o roteiro.
11	1	ON	Este ano, redatores compuseram o roteiro rebuscado.
12	2	OF	Este ano, redatores compuseram o roteiro e o romance.
13	2	SN	Na estréia, dançarinos talentosos conquistaram a plateia.
14	3	SF	Na estréia, dançarinos e cantores conquistaram a plateia.
15	4	ON	Na estréia, dançarinos conquistaram a plateia desatenta.
16	1	OF	Na estréia, dançarinos conquistaram a plateia e o jurado
17	1	SN	Esta noite, seguranças contratados protegeram o cliente.
18	2	SF	Esta noite, seguranças e gerentes protegeram o cliente.
19	3	ON	Esta noite, seguranças protegeram o cliente furioso.
20	4	OF	Esta noite, seguranças protegeram o cliente e o gerente.
21	4	SN	No fim do dia, engenheiros responsáveis despediram o pedreiro.
22	1	SF	No fim do dia, engenheiros e chefia despediram o pedreiro.
23	2	ON	No fim do dia, engenheiros despediram o pedreiro desatento.
24	3	OF	No fim do dia, engenheiros despediram o pedreiro e o servente.
25	3	SN	Esta tarde, professores concursados condenaram o programa.
26	4	SF	Esta tarde, professores e conselhos condenaram o programa.
27	1	ON	Esta tarde, professores condenaram o programa disponível.
28	2	OF	Esta tarde, professores condenaram o programa e a palestra.
29	2	SN	No sábado, torcedores premiados conheceram o goleiro.
30	3	SF	No sábado, torcedores e blogueiros conheceram o goleiro.
31	4	ON	No sábado, torcedores conheceram o goleiros canadense.
32	1	OF	No sábado, torcedores conheceram o goleiros e o zagueiro.
33	1	SN	Até hoje, empregados cuidadosos conservaram a limpeza.
34	2	SF	Até hoje, empregados e ajudantes conservaram a limpeza.
35	3	ON	Até hoje, empregados conservaram a limpeza contratada.
36	4	OF	Até hoje, empregados conservaram a limpeza e a pintura.
37	4	SN	Recentemente, graduandos preocupados convidaram o patrono.
38	1	SF	Recentemente, graduandos e mestrandos convidaram o patrono.

39	2	ON	Recentemente, graduandos convidaram o patrono definido.
40	3	OF	Recentemente, graduandos convidaram o patrono e o padrinho.
41	3	SN	Hoje cedo, jogadores destemidos defenderam a vitória.
42	4	SF	Hoje cedo, jogadores e peritos defenderam a vitória.
43	1	ON	Hoje cedo, jogadores defenderam a vitória recebida.
44	2	OF	Hoje cedo, jogadores defenderam a vitória e o renome.
45	2	SN	Ontem cedo, senadores transgressores decretaram a mudança.
46	3	SF	Ontem cedo, senadores e juristas decretaram a mudança.
47	4	ON	Ontem cedo, senadores decretaram a mudança temporária.
48	1	OF	Ontem cedo, senadores decretaram a mudança e o recesso.
49	1	SN	No programa, colunistas rigorosos comentaram a filmagem
50	2	SF	No programa, colunistas e famosos comentaram a filmagem.
51	3	ON	No programa, colunistas comentaram a filmagem produzida.
52	4	OF	No programa, colunistas comentaram a filmagem e o fiasco.
53	4	SN	No serviço, motoristas descuidados dirigiram a carreta.
54	1	SF	No serviço, motoristas e aprendizes dirigiram a carreta.
55	2	ON	No serviço, motoristas dirigiram a carreta reforçada.
56	3	OF	No serviço, motoristas dirigiram a carreta e os tratores.
57	3	SN	Durante o ato, funcionários motivados sustentaram o boicote
58	4	SF	Durante o ato, funcionários e mercados sustentaram o boicote
59	1	ON	Durante o ato, funcionários sustentaram o boicote decidido
60	2	OF	Durante o ato, funcionários sustentaram o boicotes e a parada.
61	2	SN	No encontro, cineastas vanguardistas revisaram a tragédia.
62	3	SF	No encontro, cineastas e editores revisaram a tragédia.
63	4	ON	No encontro, cineastas revisaram a tragédia realista.
64	1	OF	No encontro, cineastas revisaram a tragédia e a comédia.
65	1	SN	Na época, atendentes carrancudos conferiram o trabalho.
66	2	SF	Na época, atendentes e fregueses conferiram o trabalho.
67	3	ON	Na época, atendentes conferiram o trabalho requerido.
68	4	OF	Na época, atendentes conferiram o trabalho e o dinheiro.
69	4	SN	Ainda agora, jornalistas renomados criticaram a novela.
70	1	SF	Ainda agora, jornalistas e cronistas criticaram a novela.
71	2	ON	Ainda agora, jornalistas criticaram a novela destacada.
72	3	OF	Ainda agora, jornalistas criticaram a novela e a matéria.
73	3	SN	No Domingo, cozinheiras talentosas prepararam o banquete.
74	4	SF	No Domingo, cozinheiras e doceiras prepararam o banquete.
75	1	ON	No Domingo, cozinheiras prepararam o banquete combinado
76	2	OF	No Domingo, cozinheiras prepararam o banquete e a bebida.
77	2	SN	Na madrugada, enfermeiros perspicazes socorreram o ferido.
78	3	SF	Na madrugada, enfermeiros e bombeiros socorreram o ferido.
79	4	ON	Na madrugada, enfermeiros socorreram o ferido desmaiado.
80	1	OF	Na madrugada, enfermeiros socorreram o ferido e o parente.
81	5	DS	No passado, a senhora que conheço fabricava sabonetes.
82	5	DS	Dessa vez, os idosos que ajudo precisavam de carinho.
83	5	DS	Anteontem, a farmácia que falaram anunciou vagas novas.

84	5	DS	Antigamente, o carteiro que entregou conversava todo dia.
85	5	DS	Em outro tempo, o sobrinho que repetiu foi expulso da escola.
86	5	DS	No fim da tarde, padaria que se preze oferece cafezinho.
87	5	DS	Pela manhã, a cortina que eu comprei escurece toda sala.
88	5	DS	Da outra vez, o técnico que gritava enfureceu quem assistia.
89	5	DS	Com frequência, as janelas que quebraram balançavam fortemente.
90	5	DS	No caminho, a sacola que pegaram arrebentou as laterais.
91	5	DO	Sobre o balcão, açougueiros afiaram suas facas de cerâmica.
92	5	DO	Na parede, prateleiras sustentavam as bandejas de madeira.
93	5	DO	No teatro, o violão ressoava a música de Toquinho.
94	5	DO	No avião, o celular funciona sem problemas de leitura.
95	5	DO	Na verdade, o terreno comportou o sobrado de três quartos.
96	5	DO	Na fronteira, militares avançavam territórios de guerrilha.
97	5	DO	Na balada, o sapato apertava o calcanhar de aquiles.
98	5	DO	Na cozinha, a abelha rodeava a bebida de chá mate.
99	5	DO	Na viagem, a camisa repelia a quentura do deserto.
100	5	DO	Na clínica, o treinador alimentou o filhote de gorila
101	5	DOF	Receosa, Tatiana comunicou o episódio ao João e ao Vitor
102	5	DOF	Amoroso, Alexandre ofereceu as doações à Bruna e à Carla
103	5	DOF	Prontamente, Janaína informou o problema ao José e ao Carlos
104	5	DOF	Prestativo, Leonardo emprestou o dinheiro pra Ana e pra Clara.
105	5	DOF	Sorridente, Liliane agradeceu o presente ao Pedro e ao Lucas
106	5	DOF	Animado, Adriano comemorou a conquista com a Célia e com a Maria.
107	5	DOF	Responsável, Isadora entregou as revistas pro Luiz e pro Miguel.
108	5	DOF	Chateado, Cristiano devolveu os pacotes pra Vera e pra Lúcia
109	5	DOF	Amigável, Eduarda recomendou a música pro Paulo e pro César.
110	5	DOF	Dedicado, Juliano ensinou a tarefa pra Lara e pra Bela.
111	5	DON	Exaltada, Eliane discutiu o assunto com o Bruno Lopes.
112	5	DON	Atencioso, Agostinho recebeu os pedidos da Lívia Seixas.
113	5	DON	Competente, Angélica reservou os lugares pro Caio Perez.
114	5	DON	Descansado, Edivaldo escreveu o relatório pra Carla Glória.
115	5	DON	Revoltada, Estéfane arrumou a estante pro Ciro Roger.
116	5	DON	Pacífico, Ezequiel negociou os valores com a Rita Silva.
117	5	DON	Preocupada, Fabiana acertou os detalhes com o Marco Túlio.
118	5	DON	Interessado, Frederico combinou o passeio com a Sara Castro.
119	5	DON	Pró-ativa, Filomena arquivou os artigos pro André Lélis.
120	5	DON	Resiliente, Gerônimo acatou o conselho do Luan Sales.

2) Experimental sentences list for Experiment 2

TN	List	COND	Sentence
1	1	NB	Jessica gave keys to Jason Parreno last Saturday.

2	2	SB	Jessica gave keys to Jason and to Ashley last Saturday.
3	2	NB	Gabriel brought books to Mathew Patterson an hour ago.
4	1	SB	Gabriel brought books to Mathew and to Kaylee an hour ago.
5	1	NB	Emily lent pens to Sophie Washington during the test.
6	2	SB	Emily lent pens to Sophie and to Andrew during the test.
7	2	NB	Jonathan made bags for Emma Robinson at his art class.
8	1	SB	Jonathan made bags for Emma and for Taylor at his art class.
9	1	NB	Natalie sent cards to Sarah Leventhal in February.
10	2	SB	Natalie sent cards to Sarah and to Dylan in February.
11	2	NB	Edward told lies to Ethan Flanagan for a long time.
12	1	SB	Edward told lies to Ethan and to Sydney for a long time.
13	1	NB	Margaret wished luck to Jacob Carrington before the play.
14	2	SB	Margaret wished luck to Jacob and to Molly before the play.
15	2	NB	Gregory showed pics to Caleb Gallagher at the meeting.
16	1	SB	Gregory showed pics to Caleb and to Mary at the meeting.
17	1	NB	Madison shared gum with Brenda Moreno this afternoon.
18	2	SB	Madison shared gum with Brenda and with Daniel this afternoon.
19	2	NB	Christopher bought cups for Lily Velazquez Tuesday morning.
20	1	SB	Christopher bought cups for Lily and for Michael Tuesday morning.
300	50	Filler	Do customers choose their bonus before the bill is paid?
301	50	Filler	Does security leave the building while the inspection takes place?
302	50	Filler	Will the professor post the grades after the term finishes?
303	50	Filler	Would your father babysit until the project is over?
304	50	Filler	Can the VIPs buy their tickets as soon as the list is released?
305	50	Filler	Might players bet on their game during final decisions?
306	50	Filler	What items come with the kit on the second subscription month?
307	50	Filler	Which researchers say birds can live up to forty-nine years?
308	50	Filler	Were the labels tied to the bags by the time the bride arrived?
309	50	Filler	Was the sweatshirt hung in the closet when Daddy went to sleep?
400	50	Filler	All guests are begging like there's no other room available!
401	50	Filler	The whole school is staring like there's something wrong with my clothes!
402	50	Filler	The class is acting like there's no teacther in the room today!
403	50	Filler	There are students talking as if there's no test going on right now!
404	50	Filler	The girls are crying as if there's no other day left to live!
405	50	Filler	The audience is laughing like there's a clown performing here!
406	50	Filler	There's no one smiling as if we were at someone's funeral!
407	50	Filler	The children are screaming as if they were being attacked by a ghost!
408	50	Filler	The guys are getting drunk as if there were no tomorrow!
409	50	Filler	The dogs are barking as if there's someone trying to break in the house!

3) Experimental sentences for Experiment 3

TN	LN	COND	SENTENCE
1	4	B1V	Vai ser assim: você varre a casa enquanto você faz comida.
2	4	B2V	Vai ser assim: você varre a casa enquanto você faz comida.
3	4	B3V	Vai ser assim: você varre a casa enquanto você faz comida.
4	4	UV	Vai ser assim: você varre a casa enquanto você faz comida.
5	20	B1E	Vamos combinar: ele estuda o livro sempre que ele vai pra escola.
6	20	B2E	Vamos combinar: ele estuda o livro sempre que ele vai pra escola.
7	20	B3E	Vamos combinar: ele estuda o livro sempre que ele vai pra escola.
8	20	UE	Vamos combinar: ele estuda o livro sempre que ele vai pra escola.
9	30	B1V	Estamos de acordo: você limpa o balcão cada vez que você bebe café.
10	30	B2V	Estamos de acordo: você limpa o balcão cada vez que você bebe café.
11	30	B3V	Estamos de acordo: você limpa o balcão cada vez que você bebe café.
12	30	UV	Estamos de acordo: você limpa o balcão cada vez que você bebe café.
13	41	B1E	Ficamos assim: ele revê o conteúdo toda a vez que ele faz a lição.
14	41	B2E	Ficamos assim: ele revê o conteúdo toda a vez que ele faz a lição.
15	41	B3E	Ficamos assim: ele revê o conteúdo toda a vez que ele faz a lição.
16	41	UE	Ficamos assim: ele revê o conteúdo toda a vez que ele faz a lição.
17	38	B1V	Fingir pra quê: você confere a nota sempre que você paga a conta.
18	38	B2V	Fingir pra quê: você confere a nota sempre que você paga a conta.
19	38	B3V	Fingir pra quê: você confere a nota sempre que você paga a conta.
20	38	UV	Fingir pra quê: você confere a nota sempre que você paga a conta.
21	7	B1E	Pode acreditar: ele toma sorvete cada vez que ele visita o parque.
22	7	B2E	Pode acreditar: ele toma sorvete cada vez que ele visita o parque.
23	7	B3E	Pode acreditar: ele toma sorvete cada vez que ele visita o parque.
24	7	UE	Pode acreditar: ele toma sorvete cada vez que ele visita o parque.
25	61	B1V	Nada novo: você tem problemas toda a vez que você conta mentira.
26	61	B2V	Nada novo: você tem problemas toda a vez que você conta mentira.
27	61	B3V	Nada novo: você tem problemas toda a vez que você conta mentira.
28	61	UV	Nada novo: você tem problemas toda a vez que você conta mentira.
29	71	B1E	Tá resolvido: ele serve as bebidas enquanto ele serve os pratos.
30	71	B2E	Tá resolvido: ele serve as bebidas enquanto ele serve os pratos.
31	71	B3E	Tá resolvido: ele serve as bebidas enquanto ele serve os pratos.
32	71	UE	Tá resolvido: ele serve as bebidas enquanto ele serve os pratos.
33	64	B1V	Não é difícil: você recolhe tudo cada vez que você arruma o quarto.
34	64	B2V	Não é difícil: você recolhe tudo cada vez que você arruma o quarto.
35	64	B3V	Não é difícil: você recolhe tudo cada vez que você arruma o quarto.
36	64	UV	Não é difícil: você recolhe tudo cada vez que você arruma o quarto.
37	10	B1E	É sempre assim: ele guarda a louça toda a vez que ele limpa a mesa.
38	10	B2E	É sempre assim: ele guarda a louça toda a vez que ele limpa a mesa.
39	10	B3E	É sempre assim: ele guarda a louça toda a vez que ele limpa a mesa.
40	10	UE	É sempre assim: ele guarda a louça toda a vez que ele limpa a mesa.

41	35	B1V	Então tá certo: você come o bolo enquanto você lava a fôrma.
42	35	B2V	Então tá certo: você come o bolo enquanto você lava a fôrma.
43	35	B3V	Então tá certo: você come o bolo enquanto você lava a fôrma.
44	35	UV	Então tá certo: você come o bolo enquanto você lava a fôrma.
45	48	B1E	Já virou tradição: ele toma cerveja sempre que ele joga bola.
46	48	B2E	Já virou tradição: ele toma cerveja sempre que ele joga bola.
47	48	B3E	Já virou tradição: ele toma cerveja sempre que ele joga bola.
48	48	UE	Já virou tradição: ele toma cerveja sempre que ele joga bola.
49	23	B1V	Tá decidido: você tira o lixo toda a vez que você cozinha a janta.
50	23	B2V	Tá decidido: você tira o lixo toda a vez que você cozinha a janta.
51	23	B3V	Tá decidido: você tira o lixo toda a vez que você cozinha a janta.
52	23	UV	Tá decidido: você tira o lixo toda a vez que você cozinha a janta.
53	67	B1E	É impressionante: ele tira foto enquanto ele faz as poses.
54	67	B2E	É impressionante: ele tira foto enquanto ele faz as poses.
55	67	B3E	É impressionante: ele tira foto enquanto ele faz as poses.
56	67	UE	É impressionante: ele tira foto enquanto ele faz as poses.
57	43	B1V	Sem briga: você escreve a carta sempre que você lê o e-mail.
58	43	B2V	Sem briga: você escreve a carta sempre que você lê o e-mail.
59	43	B3V	Sem briga: você escreve a carta sempre que você lê o e-mail.
60	43	UV	Sem briga: você escreve a carta sempre que você lê o e-mail.
61	16	B1E	Fácil: ele checa os resultados cada vez que ele calcula o valor.
62	16	B2E	Fácil: ele checa os resultados cada vez que ele calcula o valor.
63	16	B3E	Fácil: ele checa os resultados cada vez que ele calcula o valor.
64	16	UE	Fácil: ele checa os resultados cada vez que ele calcula o valor.
65	27	B1V	Faz assim: você pendura um quadro enquanto você apoia o outro.
66	27	B2V	Faz assim: você pendura um quadro enquanto você apoia o outro.
67	27	B3V	Faz assim: você pendura um quadro enquanto você apoia o outro.
68	27	UV	Faz assim: você pendura um quadro enquanto você apoia o outro.
69	52	B1E	Parece brincadeira: ele dá uma festa sempre que ele passa de ano.
70	52	B2E	Parece brincadeira: ele dá uma festa sempre que ele passa de ano.
71	52	B3E	Parece brincadeira: ele dá uma festa sempre que ele passa de ano.
72	52	UE	Parece brincadeira: ele dá uma festa sempre que ele passa de ano.
73	69	B1V	Foi combinado: você escolhe a janta toda a vez que você chega tarde.
74	69	B2V	Foi combinado: você escolhe a janta toda a vez que você chega tarde.
75	69	B3V	Foi combinado: você escolhe a janta toda a vez que você chega tarde.
76	69	UV	Foi combinado: você escolhe a janta toda a vez que você chega tarde.
77	50	B1E	Escuta isso: ele ouve música toda a vez que ele quer sossego.
78	50	B2E	Escuta isso: ele ouve música toda a vez que ele quer sossego.
79	50	B3E	Escuta isso: ele ouve música toda a vez que ele quer sossego.
80	50	UE	Escuta isso: ele ouve música toda a vez que ele quer sossego.
81	13	B1V	É simples: você ganha dinheiro sempre que você bate as metas
82	13	B2V	É simples: você ganha dinheiro sempre que você bate as metas
83	13	B3V	É simples: você ganha dinheiro sempre que você bate as metas

84	13	UV	É simples: você ganha dinheiro sempre que você bate as metas
85	54	B1E	Olha ali: ele descansa as pernas cada vez que ele malha os braços.
86	54	B2E	Olha ali: ele descansa as pernas cada vez que ele malha os braços.
87	54	B3E	Olha ali: ele descansa as pernas cada vez que ele malha os braços.
88	54	UE	Olha ali: ele descansa as pernas cada vez que ele malha os braços.
89	25	B1V	Presta atenção: você leva bronca cada vez que você chama a mãe.
90	25	B2V	Presta atenção: você leva bronca cada vez que você chama a mãe.
91	25	B3V	Presta atenção: você leva bronca cada vez que você chama a mãe.
92	25	UV	Presta atenção: você leva bronca cada vez que você chama a mãe.
93	58	B1E	É errado: ele aposta dinheiro cada vez que ele joga baralho.
94	58	B2E	É errado: ele aposta dinheiro cada vez que ele joga baralho.
95	58	B3E	É errado: ele aposta dinheiro cada vez que ele joga baralho.
96	58	UE	É errado: ele aposta dinheiro cada vez que ele joga baralho.
97	45	B1V	Sem erro: você compra o suco enquanto você decide o doce.
98	45	B2V	Sem erro: você compra o suco enquanto você decide o doce.
99	45	B3V	Sem erro: você compra o suco enquanto você decide o doce.
100	45	UV	Sem erro: você compra o suco enquanto você decide o doce.
101	32	B1E	É normal: ele faz a barba toda a vez que ele corta o cabelo.
102	32	B2E	É normal: ele faz a barba toda a vez que ele corta o cabelo.
103	32	B3E	É normal: ele faz a barba toda a vez que ele corta o cabelo.
104	32	UE	É normal: ele faz a barba toda a vez que ele corta o cabelo.
105	56	B1V	É mais prático: você pinta a tela enquanto você mistura a tinta.
106	56	B2V	É mais prático: você pinta a tela enquanto você mistura a tinta.
107	56	B3V	É mais prático: você pinta a tela enquanto você mistura a tinta.
108	56	UV	É mais prático: você pinta a tela enquanto você mistura a tinta.
109	73	B1E	Sem demora: ele apaga o quadro sempre que ele explica a lição.
110	73	B2E	Sem demora: ele apaga o quadro sempre que ele explica a lição.
111	73	B3E	Sem demora: ele apaga o quadro sempre que ele explica a lição.
112	73	UE	Sem demora: ele apaga o quadro sempre que ele explica a lição.
113	36	FillerB	Todo dia, eu ouço música enquanto eu arrumo o quarto.
114	36	FillerB	Pra melhorar, eu toco guitarra sempre que eu volto da aula.
115	28	FillerB	Meu chefe acha que eu me estresso cada vez que eu dirijo pro trabalho.
116	28	FillerB	Eu já sei que eu emagreço à medida que eu corto os doces.
117	9	FillerB	Você sabe que eu planejo meu dia toda a vez que eu vou dormir.
118	9	FillerB	Eles pensam que eu gosto de brincar ao mesmo tempo que eu gosto de me isolar.
119	49	FillerB	Toda tarde, eu caminho bastante enquanto ele fica em casa.
120	49	FillerB	O ruim é que eu almoço tarde sempre que ela faxina a cozinha.
121	57	FillerB	A verdade é que eu perco a razão cada vez que eles discutem assim.
122	1	Treino	Já notou que eu fiquei mais calma à medida que elas foram crescendo.
123	3	FillerB	Você não sabe mas eu quero te bater toda a vez que você fala assim.
124	57	FillerB	No verão, eu tenho treino ao mesmo tempo que vocês tem aula de dança.
125	14	FillerB	Vamos juntos já que ele anda de skate enquanto eu ando de patins.

126	3	FillerB	Em um debate, ela revira os olhos sempre que eu tenho razão.
127	5	FillerB	O legal é que eles dão risada cada vez que eu conto uma piada.
128	14	FillerB	No treinamento, elas entendem melhor à medida que eu mostro como faz.
129	55	FillerB	Não sei por que você aumenta a voz toda a vez que eu discordo de você.
130	5	FillerB	Pra agilizar: vocês cortam a carne ao mesmo tempo que eu coloco a mesa.
131	1	FillerB	No trabalho, eu desenho o modelo enquanto eu planejo o projeto.
132	55	FillerB	Por causa da alergia, eu espirro muito sempre que eu dobro roupa.
133	70	FillerB	Pra não errar, eu penso duas vezes cada vez que eu tomo uma decisão.
134	1	FillerB	Depois do parto, eu perdia os quilos à medida que eu amamentava.
135	62	FillerB	No início, eu consultava o chefe toda a vez que eu tinha dúvida.
136	70	FillerB	Pode falar: eu converso com você ao mesmo tempo que eu digito.
137	17	FillerB	Eu vi! nós passamos de carro enquanto ele entrava em casa.
138	62	FillerB	No Facebook, nós ignoramos sempre que ela posta alguma coisa.
139	39	FillerB	Ontem, nós comemorávamos cada vez que eles acertavam o gol.
140	17	FillerB	Sobre a briga, nós esquecemos à medida que elas se distanciaram.
141	24	FillerB	O jogo é: nós tomamos um gole toda a vez que você mexe no cabelo.
142	39	FillerB	Na sexta, nós saímos do bar ao mesmo tempo que vocês saíram de casa.
143	72	FillerB	Hoje cedo, ele cuidou do bebê enquanto nós fomos às compras.
144	24	FillerB	É alergia: ela tosse muito sempre que nós usamos perfume.
145	66	FillerB	Pura manha: eles fazem pirraça cada vez que nós olhamos.
146	72	FillerB	Na internet, elas ficam famosas à medida que nós divulgamos os vídeos.
147	59	FillerB	Quando criança, você chorava muito toda a vez que nós íamos embora.
148	66	FillerB	Hoje, vocês precisam de nós ao mesmo tempo que nós precisamos de vocês.
149	12	FillerB	Durante a tarde, nós damos atendimento enquanto nós estivermos aqui.
150	59	FillerB	É de costume: nós jantamos no Plaza sempre que nós vamos à São Paulo.
151	33	FillerB	Sobre a vovó, nós sentimos saudade cada vez que nós mexemos nas cartas.
152	12	FillerB	Na escola, nós avaliamos o aluno à medida que nós corrigimos a lição.
153	46	FillerB	No calor, nós trocamos de roupa toda a vez que nós suamos.
154	33	FillerB	Idealmente, nós fazemos dieta ao mesmo tempo que nós comemos doce.
155	3	Treino	Vou confessar que nós lavamos roupa enquanto nós tomamos vinho.
156	46	FillerB	Todos sabem que nós discutimos política sempre que nós nos encontramos.
157	21	FillerB	Na verdade, nós repensamos a vida cada vez que nós sofremos uma perda.
158	21	FillerB	É sempre assm: nós somos motivados à medida que nós atingimos as metas.
159	19	FillerB	Adoro que nós nos divertimos toda a vez que nós saímos juntos.
160	19	FillerB	Já reparou que nós lemos notiícias ao mesmo tempo que nós conversamos.
161	44	FillerU	Todo dia, eu ouço música enquanto eu arrumo o quarto.
162	44	FillerU	Pra melhorar, eu toco guitarra sempre que eu volto da aula.
163	6	FillerU	Meu chefe acha que eu me estresso cada vez que eu dirijo pro trabalho.
164	6	FillerU	Eu já sei que eu emagreço à medida que eu corto os doces.
165	29	FillerU	Você sabe que eu planejo meu dia toda a vez que eu vou dormir.
166	29	FillerU	Eles pensam que eu gosto de brincar ao mesmo tempo que eu gosto de me isolar.
167	42	FillerU	Toda tarde, eu caminho bastante enquanto ele fica em casa.

168	42	FillerU	O ruim é que eu almoço tarde sempre que ela faxina a cozinha.
169	65	FillerU	A verdade é que eu perco a razão cada vez que eles discutem assim.
170	65	FillerU	Já notou que eu fiquei mais calma à medida que elas foram crescendo.
171	37	FillerU	Você não sabe mas eu quero te bater toda a vez que você fala assim.
172	37	FillerU	No verão, eu tenho treino ao mesmo tempo que vocês tem aula de dança.
173	8	FillerU	Vamos juntos já que ele anda de skate enquanto eu ando de patins.
174	8	FillerU	Em um debate, ela revira os olhos sempre que eu tenho razão.
175	40	FillerU	O legal é que eles dão risada cada vez que eu conto uma piada.
176	40	FillerU	No treinamento, elas entendem melhor à medida que eu mostro como faz.
177	26	FillerU	Não sei por que você aumenta a voz toda a vez que eu discordo de você.
178	26	FillerU	Pra agilizar: vocês cortam a carne ao mesmo tempo que eu coloco a mesa.
179	22	FillerU	No trabalho, eu desenho o modelo enquanto eu planejo o projeto.
180	22	FillerU	Por causa da alergia, eu espirro muito sempre que eu dobro roupa.
181	34	FillerU	Pra não errar, eu penso duas vezes cada vez que eu tomo uma decisão.
182	34	FillerU	Depois do parto, eu perdia os quilos à medida que eu amamentava.
183	60	FillerU	No início, eu consultava o chefe toda a vez que eu tinha dúvida.
184	60	FillerU	Pode falar: eu converso com você ao mesmo tempo que eu digito.
185	2	Treino	Eu vi! nós passamos de carro enquanto ele entrava em casa.
186	74	FillerU	No Facebook, nós ignoramos sempre que ela posta alguma coisa.
187	74	FillerU	Ontem, nós comemorávamos cada vez que eles acertavam o gol.
188	18	FillerU	Sobre a briga, nós esquecemos à medida que elas se distanciaram.
189	18	FillerU	O jogo é: nós tomamos um gole toda a vez que você mexe no cabelo.
190	15	FillerU	Na sexta, nós saímos do bar ao mesmo tempo que vocês saíram de casa.
191	15	FillerU	Hoje cedo, ele cuidou do bebê enquanto nós fomos às compras.
192	53	FillerU	É alergia: ela tosse muito sempre que nós usamos perfume.
193	53	FillerU	Pura manha: eles fazem pirraça cada vez que nós olhamos.
194	68	FillerU	Na internet, elas ficam famosas à medida que nós divulgamos os vídeos.
195	68	FillerU	Quando criança, você chorava muito toda a vez que nós íamos embora.
196	31	FillerU	Hoje, vocês precisam de nós ao mesmo tempo que nós precisamos de vocês.
197	31	FillerU	Durante a tarde, nós damos atendimento enquanto nós estivermos aqui.
198	47	FillerU	É de costume: nós jantamos no Plaza sempre que nós vamos à São Paulo.
199	47	FillerU	Sobre a vovó, nós sentimos saudade cada vez que nós mexemos nas cartas.
200	63	FillerU	Na escola, nós avaliamos o aluno à medida que nós corrigimos a lição.
201	63	FillerU	No calor, nós trocamos de roupa toda a vez que nós suamos.
202	2	FillerU	Idealmente, nós fazemos dieta ao mesmo tempo que nós comemos doce.
203	2	FillerU	Vou confessar que nós lavamos roupa enquanto nós tomamos vinho.
204	4	Treino	Todos sabem que nós discutimos política sempre que nós nos encontramos.
205	11	FillerU	Na verdade, nós repensamos a vida cada vez que nós sofremos uma perda.
206	11	FillerU	É sempre assm: nós somos motivados à medida que nós atingimos as metas.
207	51	FillerU	Adoro que nós nos divertimos toda a vez que nós saímos juntos.
208	51	FillerU	Já reparou que nós lemos notiícias ao mesmo tempo que nós conversamos.