Same difference: literal and idiomatic interpretations of verb-object constituents

By

Erwin Lares

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

(Spanish)

at the

UNIVERSITY OF WISCONSIN-MADISON

2022

Date of final oral examination: 05/06/2022

The dissertation is approved by the following members of the Final Oral Committee:

Grant Armstrong, Professor, Spanish

Rajiv Rao, Professor, Spanish

Cathy Stafford, Professor, Spanish

Eric Raimy, Professor, English

Acknowledgements

I come from a family of *refraneros*[1]. My mother, my father, and my grandfather have always dispensed their nuggets of wisdom in the form of a proverb or a popular saying. Needless to say, as a child I was lost half the time they spoke to me, never sure of what they meant to say when they were talking about camels going through the eye of a needle. All of this is to say that I kept going back to ask them to tell me more and explain what was it that they meant. In retrospect, I think that was all part of their plan: before I knew it I was in love with idioms. This manuscript wouldn't exist without them.

This manuscript would also not exist without the support and encouragement of my advisers: Grant Armstrong who over the long, long, long, long years somehow did not give up on me despite my numerous shortcomings and Rajiv Rao who has played the role of mentor since we became friends many years before I even attempted to start this doctorate business. I would be remiss if I did not acknowledge all the professors from both the Spanish and Portuguese department and the Language Science department: Fernando Tejedo, Cathy Stafford and Eric Raimy; their doors have always been open for me.

Finally, without my support system I would have not been able to complete this project. Scott Lipke, *you're simply the best, better than all the rest, better than everyone, anyone I've ever met.* The Chicago family who adopted me when I moved to the American Midwest: my second mom Daisy Ferguson, my uncle Alan Ferguson, and my cousin-brother Will Ferguson. The folks at *The Carpentries* who taught me how to turn raw data into insights, Sarah Stevens and Tobin Magle, you are the best data science instructors I know. The folks at SSCC, Doug Hemken and Jason Struck, thank you for your patience and generosity helping me understand what the numbers were saying. To the wonderful Gail Tanner and Morgan Zimmer who breathe life and warmth to the Department of Spanish and Portuguese. Lastly, to my proofreaders - Ben Gerber, Mateo Burner,

[1] a *refrán* is a witty saying, a maxim of common use

Emmanuel Herzog, Dan Schmidt, Daniel Manzano, and Joe Alvear - who had to suffer through so many half-baked manuscripts and my proclivity to generate typos beyond what is humanly possible, thank you all.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Defining an idiom

At its most general, an idiom is a peculiar fragment of language where what you see or hear is not what you get, semantically speaking. Instead, speakers of a language have agreed to a somewhat arbitrary and conventionalized meaning that may not be related to the individual bits that make up the idiom. I use the non-technical term "fragment" because speakers sense idiomaticity in a wide variety of constructions.

In fact, idiomatic fragments can be found across most if not all categories in natural languages. By way of an example, we can see in (1) idioms in every traditional grammatical category such as nouns (1a), adjectives (1b), verbs (1c), participles (1d), pronouns (1e), and adverbs (1f). With proverbs, we even see idioms that are clausal in nature (1g). Adjacent to idioms, we also see collocations (1h), where what is idiomaticized is the prescribed order in which the nouns must appear, c.f. "*the white and black cow."

(1) a. sopa boba
soup dumb
"food given to poor people at convents"

b. de brocha gorda
of brush fat
"of bad or poor quality"

c. estirar la pata
stretch the leg
"die"

d. hecho una Magdalena
made a Magdalene
"cry inconsolably"

e. cada quisque
each each one
"each one"

f. vuelto mierda
turn.PTCP shit
"very quickly"

g. cada ladrón juzga por su condición
each thief judge by their condition
"it takes one to know one"

h. blanco y negro
white and black
"black and white"

In the literature, we see that linguists have provided numerous definitions of what an idiom is.

- "I shall regard an idiom as a constituent or series of constituents for which the semantic interpretation is not a compositional function of the formatives of which it is composed" (Fraser 1970:22).
- "Roughly, it is a phrase (or sentence) which is conventionally used with a meaning different from its constructed literal meaning (if it has one)" (Davies 1982:68).
- "Idioms have a meaning that is not a simple function of the literal (i.e., non-figurative) meaning of their parts and that they manifest a high degree of conventionality in the choice of component lexical items" (O'Grady 1998:279).
- "We all know about idioms — phrasal combinations of morphemes with unpredictable non-compositional meanings —" (Marantz 1996:5).
- "Idioms are conventionalized linguistic expressions which can be decomposed into potentially meaningful components and exhibit co-occurrence restrictions that cannot be explained in terms of rule-governed morphosyntactic or semantic restrictions" (Martin 2010:81).

As can be seen in the list above, *conventionality* and *compositionality* are two properties shared by all the definitions. Nunberg et al. (1994)[2] identifies *transparency* as an

[2]It should be noted that Nunberg et al. (1994) uses the term *compositionality* not in the usual Fregean

additional property present in all idiomatic expressions. These three properties exist in a continuum that ranges from low to high. Despite Nunberg et al's unfortunate terminological switcheroo, this tripartite set of properties has a lot of descriptive power in that, when taken all together, these three properties are a good indication of whether a fragment is idiomatic and how easily a native speaker can identify its meaning.

Assuming an informative context is provided along with the idiom, a capable speaker would be able to deduce what a novel idiomatic expression means if it has low conventionality, high compositionality, and high transparency. For example, the expression *center divider* is highly compositional and highly transparent, but it also exhibits lower conventionality as one can also hear *center divide*, *middle divider*, *middle divide*. Compare *center divider* to *couch potato* where a reasonable variation like *sofa potato* does not seem to exist. *Couch potato* differs from *center divider* in that it is a highly conventionalized expression.

All of this is to say that despite being present in everyday language, the foundational question of what makes an idiom is hard to answer plainly and with a comprehensive set of criteria. It looks prudent instead to approach idiomaticity as an aggregate property resulting from these three factors coming together[3].

Considering how linguistically diverse idioms are, I chose to concentrate on just one of the categories listed in (1). I narrowed my focus to address just verbal idioms exclusively. Moreover, I further narrowed the investigation to verbal idioms of the type verb-object (VO idioms) only.

When I first became involved with this research project, I was interested in investigating the aspectual differences that result from interpreting a sentence containing an idiom literally and idiomatically. It is evident that the two interpretations of a sentence con-

sense of word, but rather to mean that some idioms distribute their idiomatic meaning among the constituents of the expression.

[3]It would be remiss not to include these additional paralinguistic tendencies that tend to co-occur with idioms, as Nunberg et al. (1994) also notes. Idioms tend to be used in informal register. Idioms tend to be used to capture situations of social importance.Finally, idioms tend to rely on metaphorical and figurative meaning for their motivation.

taining an idiom result in two different events. Whether those two events share the same aspectual class has not received a clear answer in the literature. I thought my research could get us closer to an answer by shedding some empirical light on this issue.

The reasons I chose to restrict the verbs I study herein to only transitive verbs was multifold. Firstly, I wanted a homogeneous research subject; having all sentences be of the same syntactic structure makes comparison across individual sentences straightforward. Secondly, one can find transitive VPs that are part of every aspectual class (Vendler 1967; Dowty 1979), but the same cannot be said of intransitive VPs. Thirdly, transitive VPs are the most common type of VPs (Fraser 2014:215), and so it follows that verbal idioms belonging to every aspectual class would be easier to find within the transitive category. As noted by McGinnis, "any aspectual class of nonidiomatic VPs also contains idiomatic VPs" (McGinnis 2002:667).

Lastly, I wanted to control for the object. All the expressions included in the corpus consisted of an object preceded by a definite article. It is well known that the object of a transitive verb can have significant effects on the aspectual structure of the event denoted by a sentence (Krifka 1998; MacDonald 2008). We can see the aspectual class of the event resulting from the same predicate changes from activities (2a) to accomplishments (2b) by virtue of a MASS to COUNT change in the nature of the object.

(2) a. The dude drank beer for an hour/*in an hour.
Intended reading: an activity.

b. The dude drank a jug of beer *for an hour/in an hour.
Intended reading: an accomplisment.

With the VO idiom restriction in place, I moved on to build my corpus. I relied heavily on newspapers, social media, and the *Dictionary of the Spanish Language* (Real Academia Española 1996) to gather the initial candidates for the corpus. However, given the popularity of some ditransitive VO idioms, I decided to include a few of them in the

final corpus for a grand total of 13 transitive and 7 ditransitive verbs. It should be noted that all the analysis carried out herein was limited to the transitive expressions only. I hope to include intransitive and ditrasitive verbs in future research.

1.2 Chapter Organization and Findings

If I were to summarize the problem my dissertation sought to address, it would be centered around the tension that exists between the two members of the literal-idiomatic pair. For the vast majority of verbal idioms, the idiomatic interpretation coexists with the literal interpretation. Each interpretation results in a different meaning, but is meaning the only place where we see a difference? Is there an aspectual relation between literal and idiomatic interpretations? Do speakers have systematic ways of distinguishing between the two? Does one interpretation condition the other? These were the questions that guided my research.

With those concerns in mind, the study morphed into an examination of differences. No longer was I planning to look at differences in aspectual structure alone, but at potential differences in the realization the two sentences could have. Once I decided to look at differences in realization, I felt methodologically obligated to also look at perception as a way to validate whatever findings the production experiment yielded. As a result, I organized my research around three experiments: a production, a perception, and a grammatical judgement experiment. Each experiment seeking to look into differences in production, perception, and aspectual class respectively.

Experiment 1, the production experiment detailed within chapter 2, looked methodically at 20 sentences that 13 native speakers produced when I asked them to report what they knew about a pair of fictional events both of which could be summarized using the same sentence. One of the events provided a context where the literal reading of the phrase was the salient interpretation, whereas the context of the second event was con-

ducive to an idiomatic interpretation.

The resulting pair of literal-idiomatic sentences should be, according to the received wisdom, identical as it is widely assumed that the literal-idiomatic interpretations are indistinguishable (Katz and Postal 1963; Fraser 1970) from one another and context is required for speakers to properly resolve the idiomatic ambiguity. As far as I know, this is the first attempt in the literature to validate experimentally this assumption that they must be identical.

I show in chapter 2 that despite this standing assumption in the literature, speakers involved in *Experiment 1* produced idiomatic sentences that were shorter in duration with narrower F0 ranges than their literal counterparts. This study did not probe too deep into the size of these effects however. As a result, I am hesitant to posit that speakers rely widely on this intonational strategy to signal the effect of saying something, but with the intent of meaning something else.

Nevertheless, these findings align themselves with recent research done on intonation in Spanish (Face 2001, 2002; Hualde 2003; Nibert 1999; Rao 2013a, 2013b) and contribute to the field by virtue of being the first study done in Spanish that focuses on the phonetic correlates of idiomatic usage. These combined results suggest that we are beginning to find converging evidence that speakers of Spanish (and perhaps intonational languages at large) manipulate various phonetic variables to achieve a larger number of pragmatic effects which go beyond simply distinguishing between declarative statements and interrogatives.

Remarkably, these results for Spanish matched the findings from Van Lancker and Canter (1981) for F0 range and duration for English idiomatic sentences. Despite the fact that we employed different elicitation methods, our combined findings show that both Spanish and English speakers manipulate duration and F0 range for idiomatic purposes.

Not all the linguistic variables examined — F0, F0 range, intensity, duration, and

speech rate — were found to play a role according to the data produced in *Experiment 1*. Moreover, in the transition from *Experiment 1* to *Experiment 2*, I found a mismatch between the variables that were significant in the production phase and those that were significant in the perception phase.

In chapter 3, I describe the results of *Experiment 2* which attempted to discern if there were cues that would help speakers in determining whether a phrase had a literal or idiomatic interpretation in the absence of context. To accomplish that, I isolated the literal idiomatic pairs of sentences produced by the speakers in *Experiment 1* and used those sentences as the input for participants in *Experiment 2* to judge idiomatic intent.

In other words, if it is true that speakers manipulate F0 range and duration to mark idiomatic intent, then it follows that the participants in *Experiment 2* would have been able to pick up on those cues to tell apart literal sentences from the idiomatic ones. Since the significant variables in production were F0 range and duration, the expectation would be that both F0 range and duration would be found to be significant predictors in perception.

This expectation was not met. The findings from *Experiment 2* showed that for the participants enrolled in the study, differences in F0 range were the only statistically significant predictor. It appeared that despite the fact that speakers were shortening duration to signal idiomaticity, participants did not pay attention to this cue and relied exclusively on how different F0 range was between the two realizations.

Experiment 2 did bring additional evidence to bear on the larger claim that the two realizations of an idiomatically ambiguous sentence must be identical. I show that they are not. The combined findings from both experiments indicate that there are measurable differences between the two realizations as some participants were able to reliably judge intent.

Experiment 2 produced another surprising finding: the differences found were located

at the start of the sentences, specifically, the subject and verb constituents. I was expecting the differences to be located, at least partially, in the verb and object constituent; in part for the reasons I mentioned earlier that led me to choose VO idioms exclusively, but also because the verb-object combination is a very productive idiom construction in English with the majority of idioms being part of this category (Bruening 2010:532). This is in contrast with the paucity of idioms formed by the verb and the subject. I know of no studies that have reported frequencies of different kinds of idioms in Spanish, but I suspect the proportion is similar for both languages.

In chapter 4, I describe the results of an experiment that probed the aspectual structure of the literal-idiomatic pair and sought to discern (a) if participants understood the aspectual class of the two members of the pair to be the same, and (b) if a systematic relationship between the two interpretations could be established. To carry out *Experiment 3*, I chose a selected subset of the phrases used in *Experiment 1*. For each one of those phrases, I crafted four contextual situations, two literal and two idiomatic. For the literal and idiomatic contexts, I further crafted a situational context that attempted to target the start of the event and another to target the end of the event.

Targeting the start and the end of the event in this way allowed me to implement two of the aspectual tests commonly used in the literature (Vendler 1967; Dowty 1979) to determine the aspectual class of a VP. By doing so, I had a clear way to compare what aspectual class the participants had in mind for the literal and idiomatic events.

That the aspectual class of the two realizations must be identical has remained an open issue in the field. It is assumed by some linguists (Marantz 1997; McGinnis 2002) that the aspectual class of the idiomatic event matches that of the literal event. However, other linguists (Glasbey 2003; Glasbey 2007) have reported cases where there appear to be aspectual mismatches between the two interpretations.

My results provide empirical evidence that supports Glasbey's position because they

show that, while it may be possible for some idiomatic events to maintain the aspectual class of the literal event, the results drawn from the participants' grammatical judgments indicate that is not a necessary condition.

The second, more ambitious goal was to find evidence supporting the idea that by knowing the aspectual class of the literal event, one can say something meaningful about the aspectual class of the idiomatic event. Although the majority of the results provide evidence supporting the hypothesis that the idiomatic event stands in a *defective relation* to the literal event, aspectually speaking, the picture these results paint is fuzzy.

At this time, the most charitable interpretation of my results would be to conclude that there is a tendency for idiomatic events to be as aspectually complex as the literal events or less complex, but critically not more complex, where aspectual complexity is defined as the number of aspectual features — dynamicity, telicity, duration — present in the event in question.

The generalization that I am after, plainly put, was that idiomaticity cannot introduce aspectual complexity. *Experiment 3* showed that for the majority of the participants this was the case: participants both agreed with expected aspectual results and rejected those sentences where aspectual complexity was artificially increased by manipulating the aspectual tests participants were asked to judge.

Following all these results, the conclusion I put forward here is that the assumption of identity between the two members of the literal-idiomatic pair should be re-examined. I show that from a production perspective, the evidence indicates that the two realizations were different, with idiomatic realizations exhibiting shorter duration with narrower F0 ranges. Moreover, the results show that a subset of the participants in *Experiment 2* were capable of judging intent beyond what can be reasonably explained by chance.

Finally, aspectually speaking, my results show that Spanish-speaking participants, when evaluating situational contexts, found differences between the event described by the

literal interpretation and the event described by the idiomatic interpretation. Furthermore, my results also provide evidence in support of a systematic relationship between the two interpretations where the idiomatic event can be said to stand in *defective relationship* with the literal event because it is likely to have as many aspectual features or fewer.

1.3 Limitations of this study

The experiments detailed within these pages were not free of difficulties however. There are a few experimental shortcomings that may hinder the validity of the conclusions presented.

For *Experiment 1*, the order in which speakers interacted with the background stories could have been randomized. There is a concern that the differences in duration could be attributed to the order in which the sentences were produced.

In the experimental design, I chose to have the speakers produce their literal sentences first and their idiomatic sentences second. I feel that the way I elicited their responses by asking them to insert their summary statement into a carrier sentence caused the speakers to fall back into their normal speech patterns. However, short of re-recording the speakers, I have no way to realistically address this issue.

For *Experiment 2*, an exit poll asking participants to rate their degree of familiarity with the phrases they interacted with would have provided an additional control variable. In other words, I have no way to tell if the participants who were very good at judging intent did so because they were very familiar with the idiom they were judging.

The training material for *Experiment 3* could have been more developed in two specific ways. First, the examples modeled for the participants only included one of the tasks they were asked to perform; both tasks should have been modeled for the participants. Second, a way to check that participants understood the task would have lent robustness to the experimental design. This could have been achieved by the addition of filler sen-

tences similar in structure to the ones used in the tasks described in *Experiment 3* but with unrelated scenarios that have clear interpretations which would, in turn, be likely to produce uncontroversial responses.

1.4 Future research

The avenues for future research are numerous. First, the interaction between *Experiment 1* and *Experiment 2* should be further studied. This work looked at production and perception individually. A future iteration of this work should include considering simultaneously how *speaker* and *participant* variables affect the estimation of intent.

As mentioned, effect size has not been studied for any of the predictors mentioned. Future research should explore how strong the effect of F0 range in determining intent is.

Where the differences live in the sentence is also a worthwhile subject of future research. This falls neatly in the domain of interfaces between phonology, syntax, and semantics. My findings show that the differences were located at the start of the sentence within the subject and the verb constituents. This contrasts with the expectation that the object constituent of these sentences be involved somehow.

Lastly, while the findings for *Experiment 3* show that there is a tendency towards maintaining or losing aspectual complexity, but not gaining it, future research should examine how robust this effect is. Furthermore, I examined only idioms with a verb-object structure. Future research should expand the types of verbs considered to include intransitive and ditransitive verbal idioms.

2 Chapter 2: Experiment 1 — Production

2.1 Introduction

Humans can produce a wide range of noises with their larynx. One could grossly characterize these sounds as sounds we associate with speech and sounds we do not. Independent of whether a sound is associated with speech, it may contribute to the overall meaning of an utterance.

Of the sounds that are associated with speech, we observe that (a) no natural language employs the entirety of the sound inventory and (b) that not all sound types are used with the same frequency. The World Atlas of Language Structure Online (Dryer and Haspelmath 2013) reports that some human-produced sounds are significantly less common than others. Among the unusual sounds (<10%), these are commonly included: clicks (the Niger-Congo family), whistles (El silbo from La Gomera, Spain), trills, breathy vocalizations, pharyngeals, nasalizations, and coarticulations, among others.

Individual natural languages do not employ the entire set of sounds available for human production; instead, each natural language relies on a fraction of the larger sound inventory. This subset of noises, or segments, are phonemic, or relevant to speech; that is to say, within a language, speakers distinguish a subset of phonemic segments that have lexico-semantic reflexes in speech production.

Languages like Spanish and English rely exclusively on consonantal and vocalic segments to make these phonemic distinctions and are known as *intonational languages*. Other languages employ variations in pitch or tone as an additional strategy to signal phonemic significance. Some Asian languages, such as Chinese and Thai, are well-known for these uses of tone and are aptly named *tonal languages*. I will refer to pitch or F0 interchangeably throughout this chapter although, technically, F0 is associated more with production and pitch with perception.

Compare the Spanish lexical items in (3a) and (3b), where exchanging one consonantal segment in the onset has the effect of forcing a different interpretation.

(3) a. pan
"bread"
b. tan
"so" (intensifier)

Vowels are also phonemic, as shown in the contrast in (4).

(4) a. pan
"bread"
b. pon
"you put" (imperative)

Standard Mandarin is a language that relies on pitch to distinguish lexical or grammatical meaning. For instance, take (5), the well-known example of the syllable 'ma' (Grice and Baumann 2007:7), where changes in tone correspond to various lexical distinctions.

(5) a. mā
"mother"
b. má
"hemp"
c. mǎ
"horse"
d. mà
"scold"

This use of tone to disambiguate lexical items parallels what consonantal and vocalic segments did in (3) and (4). Variations in pitch appear to play a double role; in tonal languages, they act as a linguistic mechanism capable of signaling phonemic distinctions, but also, for all languages, tonal and intonational alike, they serve as a way for speakers to give rise to a variety of meanings.

These meanings can be classified as *informational* when they are related to the message or *affective* when they are related to the speaker producing the utterance. Gussen-

hoven (2002) explains a variety of intonation/form vs. meaning/function pairings through the use of *biological codes*. Of the codes, the *Effort Code* seems to be particularly relevant to the ideas that I investigate here. According to Gussenhoven (2002:4), greater expenditures of effort in speech production lead to "wider excursion of the pitch movement," with the concomitant effect that narrower ranges "may be used to signal negation or withdrawal of information."

However, conveyance of emotions isn't the only use speakers have found for pitch in non-tonal languages. Ladd (1996, 2008) claims that intonation in languages like English and Spanish – although his examples rely mostly on French – exhibit phonological structure. Ladd also claims that there is a division of labor between the linguistic and paralinguistic functions of intonation, with the linguistic uses of intonation revolving around "pragmatic meanings in a *linguistically structured* way" Ladd (1996:4).

Following the work of Pierrehumbert (1980), as well as subsequent influential work on the topic, such as Ladd (1996, 2008) and Gussenhoven (2004), intonation has been studied for English using the framework known as the Autosegmental-Metrical model of intonational phonology (AM). Since the early 1990s, AM has been extended to a number of natural languages, including Spanish. In fact, this phonological model has resulted in language-specific transcription systems conducted with the Tones and Break Indices Framework (ToBI). To trace the development of ToBI in Spanish, see Beckman et al. (2002), Face (2002), Face and Prieto (2007), Hualde (2002), Hualde2003, Hualde and Prieto (2015), Hualde and Prieto (2016), Prieto and Roseano (2010), and Estebas Vilaplana and Prieto (2008).

AM provides a way to represent two main types of phonological targets associated with pitch/tone: those bound by and anchored to the stressed syllable domain (i.e., pitch accents), and also tonal events associated with boundaries of prosodic elements, as determined by a mix of syntactic and information structure within the utterance, known

as boundary tones (Hualde 2002:2). Beckman et al. (2002), for Spanish, and Hirschberg (2006), for standard American English, report a set of different pitch accents that are meaning-contrasting. Both authors indicate that changes in intonation can be employed to distinguish declarative sentences from interrogatives, exclamatives, and imperative statements.

An utterance within AM is commonly represented by hierarchichally organized *tiers*. Figure 1 (adapted from Rao and Sessarego (2018:182)) displays how various elements of a sentence are represented within this framework.

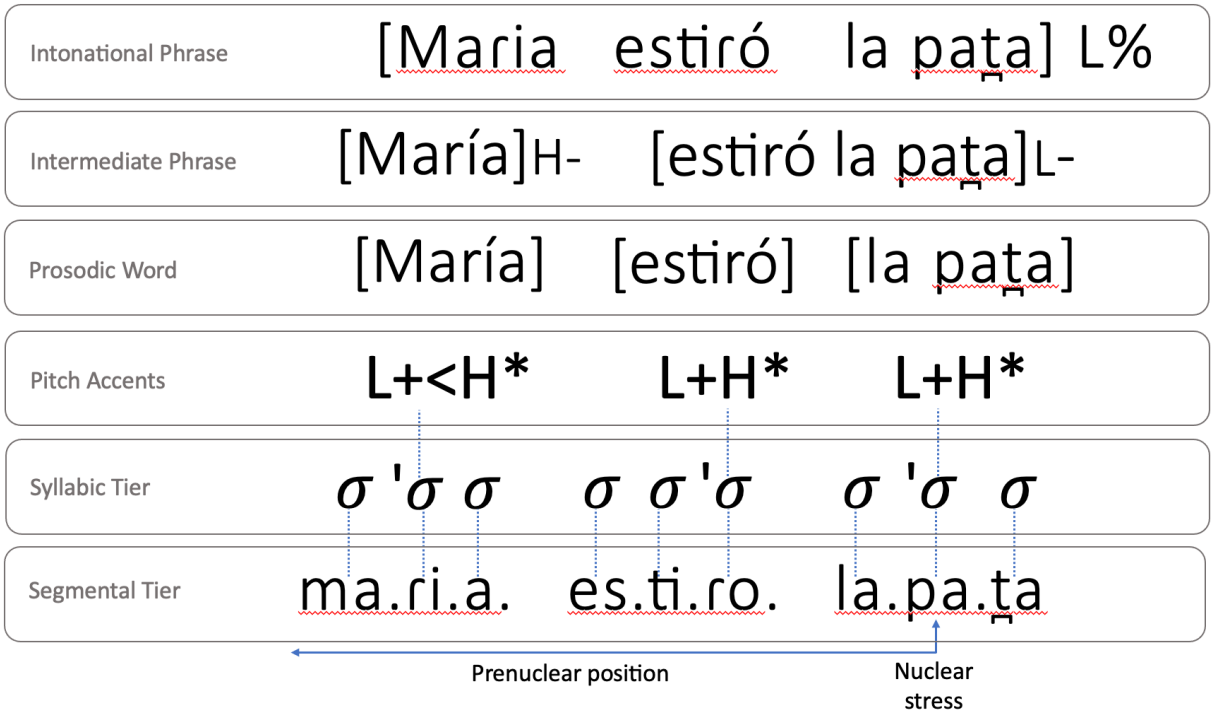


Figure 1: AM Diagram

Within Figure 1, the following information is presented for the utterance in question, *María estiró la pata* "María stretched her leg." Starting from the bottom up, the first tier displays the individual segments that make up the sentence. Above the segmental tier, the syllabic tier displays how the individual segments are put together into syllables[4] and also

[4]For the sentence shown here, I made the choice of collapsing the timing and stress tier into one *syllabic* tier. For simplicity's sake, I chose to disregard a commonly occurring phonological rule where an

indicates which are the stress-bearing syllables.

In order to paint a more detailed picture, I included, above the first three tiers, a Prosodic Word (PW) category that lists the three prosodic words included in the utterance. Prosodic words are most often content words bearing lexical stress. In Spanish, the correlates for stress are F0, duration, and intensity (Ortega-Llebaria 2006). I also included an Intermediate Phrase (ip) category listing the two intermediate phrases. Lastly, I included the intonational phrase (IP) which lists all the relevant prosodic elements, as they are shared by all the sentences contained in the corpus.

In addition to containing all the elements listed under it, IP also contains the boundary tone for the utterance, marked with L% in the diagram. Both ip and IP indicate parts of speech that "carry meaning and are terminal junctures in discourse" Rao (2016:55) as marked by phonetic cues such as perceptible pauses and final lengthening of segments.

A final note, Figure 1 also makes explicit which words appear in prenuclear or non-final position: *María* and *estiró* in this case. By contrast, *la pata* is the word occupying the nuclear or final position, and as such, *pata* is the most salient word in the highest level of constituency in this hierarchy, as Spanish is considered to obey the Nuclear Stress Rule, where final position of a prosodic domain is considered the most prominent (Chomsky and Halle 1968).

Within AM, *accent* is an ontological concept that refers to the relative values that can be ascribed to F0 within an utterance. Accent is particularly tied to F0 movement (ascending or descending) in or near a stressed syllable. These configurations of F0 excursions form monotonal or bitonal *pitch accents*, which in turn make up the inventory of pitch accents available for a given language variety. Pitch accents are anchored to content words bearing lexical stress (i.e., PWs). For Figure 1, those are [L+<H*], [L+H*], [L+H*].

unstressed /a/ following by an unstressed vowel is elided (Morgan 2010:102) in order to maintain a 1:1 correspondence between segments and syllables.

For Spanish, a recent inventory of pitch accents is shown in Figure 2 (adapted from Rao 2016:55). All this information together allows for a fuller description of speech phenomena, where it is easy to trace that content words bearing stress become anchors for the various excursions F0 is capable of achieving at the phrase-level for a specific language. The excursions correspond to the labels found under the *Pitch Accent* column. The schematics are an abstraction of the actual F0 movement observed in speech. Each schematic is split into three thirds where the vertical lines represent pre-tonic, stressed, and post-tonic syllables.

Pitch Accent	F0 Schematics	Description of F0 movement in/near stressed syllable
L+<H*		An F0 valley is anchored to the stressed syllable's onset and is followed by a rise through the entire stressed syllable that ends in a post-tonic peak.
L*+H		An F0 valley extends through the stressed syllable and is followed by an F0 rise at the stressed syllable's offset and continuing in the post-tonic syllable.
H*		An F0 plateau in the upper regions of the F0 range without a valley before it
L+(¡)H*		An F0 valley is anchored to the stressed syllable's onset and is followed by a rise to a peak that occurs within the stressed syllable, near its offset. The upstepped variant contains a peak higher than its preceding peak.
L*		F0 is flat and suppressed to the lowest portion of the F0 range.
L+H*		F0 is high at the stressed syllable onset and is followed by an F0 rise starting at the stressed syllable onset and continuing in the post-tonic syllable.
H+L*		F0 is high at the stressed syllable onset and descends throughout the stressed syllable, reaching a low point near its offset

Figure 2: Spanish Pitch Accents, Schematics and Descriptions

Spanish behaves like English in that it can use part of its phonological grammar, with different targets used to communicate different meanings (e.g., different emotions, attitudes, types of focus, etc.), in a pragmatically contrastive manner. Example (6), taken from Hualde (2002:3), suggests that Spanish employs differences in peak alignment (of a **pitch accent** associated with a stressed syllable) to signal pragmatic differences between a neutral declarative (6a) and a narrow focus declarative (6b). According to this strategy, an early rise of pitch in (6b) causes the peak to align with the tonic syllable and is

used to signal one reading, the narrow focus declarative, over another reading, the neutral declarative.

(6) a. emu**la**ban a mi hermano
they emulate.IPFV DOM POSS.1S brother
'they emulated my brother'
b. emu**la**ban a mi hermano
they emulate.IPFV DOM POSS.1S brother
'they emulated my brother' (and not mocked him)

Face (2002) expands on these observations, suggesting that pitch accents exhibit internal structure, and that speakers rely on them to signal differences in focus. Face argues that, in addition to alignment with the stressed syllable, modeling the pitch accents as L*+H for broad focus and L+H* for contrastive focus allows for a more accurate representation of data. According to Face, F0 peaks within a word's stressed syllable cue contrastive focus, whereas for broad focus, the peak of F0 occurs in a post-tonic syllable.

Nibert (1999) shows that intonation is also used to tease apart ambiguous syntactic structures. Her study looks at NPs modified by an AP, as in (7) *lilas y lirios amarillos* 'yellow lilacs and irises.' Nibert shows that speakers rely on F0 activity in the final syllable before a boundary (e.g., lengthening, high rise in F0, short pause, etc.) to disambiguate their intended meaning.

(7) a. lilas y lirios amarillos
lilacs and irises yellow
'yellow lilacs and irises'
b. [[lilas]H- [lirios amarillos]H-]L%
c. [[lilas y lirios]H- [amarillos]L-]L%

Nibert (1999) points out that (7a) is disambiguated by speakers using prosodic cues; (7b) has the intended meaning of 'lilacs and yellow irises,' whereas (7c) is interpreted as 'yellow lilacs and yellow irises.' This is expressed by means of intermediate prosodic phrases that envelope the constituent in a way that mimics the intended interpretation; that is, the H- target in (7c) is the intermediate prosodic phrase that envelops both 'lilas y

lirios,' signaling that the adjective is intended to take scope over them. This is in opposition to (7b), where the H- target marks the intermediate prosodic phrase containing 'lirios amarillos,' and the adjective's narrow scope over just the prosodic word 'lirios.'

Rao (2013a) explores intonational features of complaints as a speech act in four Mexican nationals living in the US. Rao finds evidence that speakers reliably produce higher mean F0 values when engaging in complaints. In addition to a high F0 mean, Rao also finds that speakers produced a circumflex accent rather than the final lowering of F0 that is common to many varieties of Spanish in broad focus contexts (Hualde 2002). The situations in which the circumflex accent was produced responded to several factors, which included the nature of the complaint, the gender of the complainer, and the social distance between the complainer and their interlocutor.

In his analysis, Rao (2013a) also addresses why complaint acts between friends are correlated with broader F0 ranges by relying on Gussenhoven's Effort, Frequency, and Production Codes. As Rao explains, the increased excursions have to do with "heightened sense of emotion and emphasis due to the less distance between interlocutors" (2013a:158). In other words, social proximity between people involved in this particular kind of speech act favors interactions where the speaker can be more blunt, more emphatic, and less socially guarded.

Rao (2013b) also shows that a variety of acoustic cues are used by speakers of Mexico City to signal sincerity vs. sarcasm in their speech. Pitch was found to be a significant factor, although it was not the only relevant one; gender, duration and intensity also appeared to play a role in conveying speakers' attitudes. Rao (2013b) found that, within the utterance, reductions of changes in F0 range specific to target words constituted a reliable indicator of sarcasm.

From Rao (2013a, 2013b) I borrowed the idea of priming my participants (Hartsuiker et al. 2016; Linzen et al. 2013; Schütze and Sprouse 2014) through the use of struc-

tured elicitation tasks. Similar to Rao (2013a), I crafted stories that rooted the participants' background knowledge of the event they were about to report. I also followed Rao's (2013b) lead by creating a template detailing the paralinguistic information shared by all the background stories I used.

As the studies mentioned above show, intonation has been looked at as a possible factor playing a role in speakers' attitudes toward the messages they are trying to get across. There is a type of speech that has been conspicuously absent from this inquiry, however: idioms.

As far as I know, the only assertion in the generative linguistics literature regarding the realization of idioms comes from Katz and Postal (1963) and Fraser (1970). Katz and Postal base their claim on principles of economy; according to these authors, systematic variation in stress patterns applying to phrases that have the potential of being interpreted ambiguously would result in "enormous complications of the phonological component" (Katz and Postal 1963:276–77).

Fraser (1970:26) also claims that the two potential realizations of a sentence containing an idiom are indistinguishable from one another. Like the previous authors, Fraser also relies on principles of economy, alleging that the complexity of phonological rules would increase precipitously if phrases that are identical at the surface level were to result in different realizations.

This chapter sets out to provide empirical evidence for the validity of the claim that the ambiguity of a sentence containing an idiom does not exhibit a reflex in its articulatory realization; in other words, the goal is to corroborate experimentally that the literal interpretation and the idiomatic interpretation result in identical utterances.

In the field of generative linguistics, idioms have been used as evidence to support a variety of claims. Larson (1988:340), for instance, uses discontinuous idioms, such as *"Lasorda sent his starting pitcher to the showers,"* to lend support to his double object

construction proposal. Larson points out that, under his analysis, the surface PP and the verb form a single thematic complex, which subsequently assigns a role to the surface complement "his starting pitcher."

Larson (1988:341) further notices that the locus for idiomaticity does not seem to lie with the constituent verb + object, but rather the object alone. Verbs do what they do unencumbered, only resulting in idiomatic meanings when they "interact compositionally with a semantically opaque NP."

Marantz (1997:207) revisits the observation that idiomaticity lives within the object alone by pointing out that, at least in the case of light verbs like *take*, the actual meaning of the entire VP seems to arise exclusively from the selected NPs; consider *take a leap*, *a leak*, *a piss*, *a break*, etc. Marantz also points out the scarcity of idioms where the external argument, rather than the internal argument, is the one idiomatized. It follows from Larson's double object construction that it would be impossible to isolate a constituent containing the verb and the external argument while excluding the internal one.

Jackendoff (1997) considers how prevalent "fixed expressions" are in everyday language use as he compiles his *Wheel of Fortune* corpus. Jackendoff explores a variety of scenarios where his Representational Modularity system provides a good account of set phrases, collocations, and idiomatic phrases in general. At no point, however, does Jackendoff suggest or hint that the phonological component of the interface needs to treat idioms differently from compositional phrases.

All of this is to say that there is one assumption that has been commonly held across the board: idioms are assumed to be produced the same way as their literal counterparts. This remains an uncontested – yet experimentally unverified – claim.

There is, however, experimental evidence from speech pathology and psycholinguistics that this assumption might need to be revisited. In a series of experiments, Van Lancker and Canter (1981) found out that participants were capable of recognizing the idiomatic

member of what they call *ditropically*[5] *ambiguous sentences* with a high degree of accuracy, which suggests that interlocutors may be able to pick up on cues to determine what the intended meaning should be.

The present work furthers the study of intonation in non-tonal languages in general, and in Spanish in particular, by evaluating the claim that speakers make no distinction in the production of literal versus idiomatic interpretation of sentences. In addition, this study will probe the use of F0, F0 range, intensity, duration, and speech rate as acoustic correlates and their role as disambiguation mechanisms to signal an idiomatic interpretation in sentences that are otherwise identical.

In sum, intonation in Spanish may constitute a meta-strategy that speakers employ to signal their intended interpretation when idioms are part of a speech act, not unlike what has already been done in the field regarding other pragmatic uses of intonation.

2.2 Methodology

Based on Katz and Postal (1963) and Fraser's (1970) claims for English, I began working under the assumption that there is no distinction between the realization of a sentence interpreted literally and the same sentence interpreted idiomatically. Furthermore, I also assume that the same is true for Spanish. This premise establishes the foundation for the null-hypothesis: there is no difference between idiomatic and literal intents for the same sentence. Sufficient evidence to reject H0 would come from an examination of F0, F0 range, duration, intensity, and speech rate, as Spanish speakers are found to rely on them, for example, to make distinctions related to stress (Ortega-Llebaria 2006), and to manipulate them in order to have global effects on the interpretation of utterances (Face 2001, 2002; Hualde 2003; Nibert 2005; Rao 2013b).

The overall experimental design involves three major, interconnected parts and their

[5]By ditropic, Van Lancker and Canter (1981) meant sentences that can be interpreted compositionally or non-compositionally

associated experiments: production (*Experiment 1*), perception (*Experiment 2*), and grammaticality judgments (*Experiment 3*). This chapter concerns itself with the first part: the production and execution of *Experiment 1*. In the process of designing *Experiment 1*, I took inspiration from existing work on Spanish intonation involving the use of a modified *Discourse Completion Task* (DCT).

DCT is a relatively new technique in Romance Linguistics studies associated with prosody and pragmatics. Unlike corpus or experimental techniques, DCT offers the advantage of facilitating the collection of speech that is semi-spontaneous in nature; that is, it facilitates the elicitation of comparable spontaneous data across speakers and varieties, and allows for the control of context and target sentences (Vanrell et al. 2018:214).

In particular, I was interested in having my participants produce a potentially ambiguous target sentence. For each of these target sentences, I crafted a context that favored one interpretation over the other: idiomatic or literal. I structured the way the participants interacted with the materials loosely following the intonational guided questionnaire of Prieto and Roseano (2010) and the elicitation tasks in Rao (2013a, 2013b) in that I created a set of hypothetical situations in which the speakers could anchor their background knowledge of the events containing the ambiguous sentences. Unlike Rao and Prieto & Roseano's tasks, I had my participants interact with the hypothetical situations more than once before they were asked to react to the situations and generate their speech samples.

To that effect, speakers read the stories aloud. After the participants were done reading each story, I asked them to think back about what they knew of the event and summarize what had occurred, using as much as possible "what happened was that" followed by the target sentence in question. This was done to normalize the prosodic environments in which the target sentences appeared by recreating the idea of what a carrier phrase does for single nouns (Butera 2018; Prieto and Roseano 2010; Rao 2015).

During the course of the recording sessions, which lasted 28 minutes on average, the participants encountered the same statement twice. These paired-up statements were equally distanced from one another by approximately 15 minutes. The end goal was to produce a minimally different pair of statements whose key difference was the speaker's intent.

The statements used in all three parts of the experiment come from a corpus of idiomatic expressions composed by a verb and its object that I collected from various sources; chiefly among them is the online dictionary authored and maintained by the Royal Academy of the Spanish Language (Real Academia Española 2020). All the expressions are listed in Table 1. The criterion for inclusion in the corpus was mostly syntactic: all members of the corpus consist of idioms formed by a verb, a definite article, and the verb's object. A subset of the corpus also included a second object, which will be later used to build on the findings for transitive statements.

Table 1: Idiomatic Corpus

VO Statement	Literal Interpretation	Idiomatic Interpretation
aplastar la oreja	press one's ear flat	sleep
beber los vientos por alguien	drink the winds for someone	fall in love with someone
cantarle las cuatro verdades a alguien	sing the four truths to someone	be brutally honest with someone
comerse la olla	eat the cooking pot up	worry too much
cruzar la mirada	cross the eyesight	make eye contact briefly
darle la lata a alguien	give the tin can	annoy someone
estirar la pata	stretch one's leg	die
recoger las velas	gather the sails	leave
matar el gusanillo	kill the little worm	drink alcohol on an empty stomach
romper el hielo	break the ice (with a hammer)	start a friendly conversation with a stranger
sacarle las castañas del fuego a alguien	take someone's chestnuts out of the fire	help someone in a risky situation
sacarle la pata del barro a alguien	take someone's leg out of the mud	help someone
sacarle la piedra a alguien	remove a stone for someone	annoy someone
soltar el trapo	drop the rag	laugh
tender las velas	unfurl the sails	leave
tirarle los tejos a alguien	throw the pebbles at someone	flirt
tocar la vihuela	play the vihuela guitar	do nothing
quemarse las pestañas	burn one's eyelashes	study a lot
tirar la toalla	throw the towel	give up
apretarse el cinturón	cinch one's belt	save resources due to scarcity

2.3 The Setup

I met each participant for a short pre-recording conversation in a private room. I had coffee and pastries available to ease the participants into the tasks I prepared for them, and to avoid as best I could the *Observer's Paradox* (Labov 1973, 2008). My intention was to make participants feel they were helping complete my study rather than being a subject in the experiment. During the pre-recording conversation, I explained that the general purpose of my research project was to gain a better understanding of idiomatic expressions. In particular, we wanted to study how/if there is a systematic relationship between a sentence intended to be interpreted literally and the same sentence when interpreted idiomatically.

During this time, I also went over the consent process, as outlined by the IRB protocol. I provided each participant with a copy of the consent form. After reading it together, they agreed to sign it. A copy of the consent form can be found in Appendix A.

For the next step, the participants completed a pre-interview questionnaire to collect demographic information such as age, gender, education level[6], and Spanish variety. Although this study does not aim to pursue a sociolinguistic inquiry into idioms, I decided to include these factors, as they constitute the cornerstone upon which many language variation studies are based. In addition, in future variations of this experiment, I may include these variables the analysis for a fuller picture of the phenomenon. A copy of the demographic questionnaire can be found in Appendix B.

Prieto and Roseano (2010) and Hualde and Prieto (2015) report significant prosodic variation across the Spanish-speaking world; for instance, Prieto and Roseano (2010) report that the nuclear pitch accent of broad focus for the Castilian, Cantabrian, Canarian, Mexican, Ecuadorean Andean, Argentinian and Mexican varieties is best described as L*. In contrast, Venezuelan Andean and Chilean varieties exhibit a nuclear pitch accent best

[6]Although education level was collected for each participant, it will not be considered in the current analysis.

described by L+H*. Similarly, Hualde and Prieto (2015) detail how pitch accents differ across Spanish varieties for a fairly extensive number of contexts, which include: broad focus, narrow focus, epistemically biased statements, yes/no questions, wh-questions, imperatives, and requests.

Labovian studies have found age to be a factor related to language variation, as younger speakers tend to be more innovative and prone to use non-standard forms, while older speakers tend to be more conservative in their language choices (Medina-Rivera 2011:36). In addition to its sociolinguistics relevance, age also has a physiological effect; that is, certain aspects of language production are affected by the aging process (Burke and Shafto 2004), resulting in phonological and semantic effects, such as slips of the tongue and retrieval difficulties (Ouyang et al. 2020).

Gender could have an effect on the results of this study as well. Similar to age, gender is seen as a sociolinguistic variable heavily involved in language change. Although there are different accounts regarding language differences between men and womem, see (Labov 1990) and (Rissel 1989). The general consensus is that gender plays a role in it. Physiologically, it is well known that, at least for F0, men and women consistently exhibit marked differences in their F0 mean productions (Traunmüller and Eriksson 1995).

The following step consisted of completing the Bilingual Language Profile (BLP) (Birdsong et al. 2012), which is a tool designed to gauge language dominance for a number of language pairings, with English-Spanish and Spanish-English being just being one of the available ones. Together with education level, Spanish variety and BLP scores will not be included in the analyses contained in this study, but it is not unreasonable to believe such factors could lead to stratified explanations of the phenomenon under study, as I plan to continue investigating the phenomenon in future research.

After completing the BLP questionnaire, I modeled for the participants the actual task they would be required to perform. I shared with them an example of the background

stories they would be working with and the type of response they would be required to produce. Any questions they had about the process were answered. Afterwards, we moved on to the recording booth.

The recording process took place inside a sound-proofed recording booth. We spent approximately 45 minutes to complete each interview. I continued to engage the participants as they became familiar with the workflow by interspacing the tasks with conversation unrelated to the experiment. All the background stories were organized using a binder and divided into literal and idiomatic halves. I recorded the entirety of the interview and guided each participant through each story. I took the time to answer questions regarding the stories and/or vocabulary items as the participants moved through the binder.

Speech samples were recorded using a *Zoom* H6 microphone. The audio captured was stored directly to a *SanDisk* memory card as .wav files. The audio files were batch -preprocessed using *Audacity* to isolate individual utterances. The audio files were subsequently processed on a *MacBook Pro* using *Praat* (Boersma and Weenink 2020) to extract F0, duration and intensity measurements.

After the recordings were completed, I asked each participant to complete an exit questionnaire, where I collected information regarding how familiar they were with the phrases they encountered during the course of the experiment. A copy of this form can be found in the Appendix D.

2.3.1 Participants

For this part of the study, Spanish instructors were contacted using the email list maintained by the Department of Spanish and Portuguese at a large university in the Midwestern US. In the email sent to them, instructors were asked to participate in the study and also to forward this invitation to students enrolled in their classes. The criteria for inclusion in the study were: being 18 years of age or older, holding and/or working

on a college degree[7], and being a native speaker of Spanish who immigrated to the US as an adult. The target number of participants for Phase 1 was 15 to 20 people, although only 13 participants volunteered their time. A copy of the invitation email can be found in Appendix E.

In addition to age, the following demographic information was collected from the participants: gender, language variety, number of years living in the US. The population sample (n=13) is split almost evenly between males (7) and females (6). The mean age in the sample is 31.31 years old. All participants self-reported Spanish dominance, with a mean Spanish BLP score of 191.8 and a mean English BLP score of 114.62. The number of years living in the US ranges significantly, from 0 to 17 years, with an average of 7.5 years. All participants are college graduate students pursuing a masters or a Ph.D. Table 2 displays all the demographic data.

Table 2: Demographic Information

Participant	Age	Gender	Variety Spoken	Spanish Score	English Score	Years in the US	BLP Score	Education Level
p01	31	male	Northern Peninsular	191.88	108.24	7	83.64	Masters
p02	29	male	Andean-Pacific	206.95	89.26	0	117.69	Masters
p03	30	female	Andean-Pacific	194.96	108.33	0	86.63	Masters
p04	31	female	Mexican	209.40	99.88	11	109.52	Masters
p05	32	male	Mexican	184.79	137.66	12	47.13	Masters
p06	28	male	Northern Peninsular	174.71	107.33	4	67.38	Masters
p07	36	male	Andean-Pacific	190.69	138.75	12	51.94	Masters
p08	33	female	Andean-Pacific	186.51	139.48	8	47.03	Masters
p09	34	female	Canarian	206.41	99.16	8	107.25	Masters
p10	35	male	Canarian	205.23	96.70	5	108.53	Masters
p11	31	male	Chilean	183.97	131.85	17	52.12	Masters
p12	24	female	Northern Peninsular	169.90	103.61	3	66.29	Masters
p13	33	female	Chilean	188.06	129.85	11	58.21	Masters

2.3.2 Stimuli and sentence production

Once in the recording booth, participants were asked to take a seat at a preset distance from the microphone. The microphone stand was adjusted to account for individual height. I explained to each participant that, as much as possible, it would be best if they

[7]The homogeneity of the sample in terms of education was an additional factor in excluding Education Level as an explanatory variable.

did not move around too much in order to minimize variation for the intensity variable. I started the recording by asking each participant what brought them to the US in order to give them the opportunity to get comfortable with the equipment setup, and to give me a chance to adjust the microphone settings for individual height.

Recall that the goal of this phase of the experiment is to record each participant as they produce a target sentence. Prior to eliciting the target sentence, each speaker read a background story for each of the target sentence pairs. The paired-up background stories were not sequenced one after the other; speakers went through the literal half first, and after a break, they completed the idiomatic half. In practice, this means that each element of each pair was distanced from its counterpart by approximately 15 minutes.

The background stories are intended to prime (Hartsuiker et al. 2016; Linzen et al. 2013; Schütze and Sprouse 2014) the participants toward a literal reading or an idiomatic reading by supplying a context that is conducive to only one of those interpretations. The background stories were written in such a way that they provide the participants with the same amount of information, thus reducing the amount of white (semantic or pragmatic) noise.

Each background story was structured similarly and contained the following information:

1. A sentence introducing the character or characters of the story using a copula verb.
2. A sentence that places the character in a context that is conducive to the desired outcome.
3. A finite/tensed sentence containing the VO target sentence.

By way of an example, (8) lists the background story conducive to the literal interpretation for one of the tested idioms, *sacar las castañas del fuego*. This expression interpreted literally means 'pull the chestnuts from the fire.' For the idiomatic interpretation, (9) lists its corresponding background story. The Real Academia Española (2020) lists this

expression with the idiomatic meaning of *ejecutar en beneficio de alguien algo de lo que puede resultar daño o disgusto para sí* 'carry out an action that may result in damage or displeasure for the benefit of someone else.' English has the equivalent phrase "pull someone's bacon/fat/chestnuts out of the fire." A list of all the background stories can be found in Appendix B.

(8) a. Juan es un chef de pastelería muy famoso, su amigo Luis es su asistente de pastelería.
"Juan is a famous pastry chef. His friend Luis is his assistant chef"
b. Anoche, después de cerrar la tienda, Juan y su amigo Luis comenzaron a preparar una receta nueva con castañas en lugar de nueces.
"Last night, after closing the store, Juan and his friend Luis started a new recipe that uses chestnuts instead of walnuts"
c. Mientras Juan preparaba la masa, Luis le sacó las castañas del fuego a Juan.
"While Juan was making the dough, Luis pulled the chestnuts from the fire for Juan"

(9) a. Juan es un chico privilegiado.Luis, su padre, es uno de los jueces del tribunal supremo de justicia.
"Juan is a privileged young man. Luis, his dad, is a Supreme Court Justice"
b. Anoche después de unos tragos en el bar, la policía detuvo a Juan por conducir bajo los efectos del alcohol.
"After a few drinks at the bar, the police stopped Juan because he was driving under the influence"
c. Como siempre, su padre intervino y le sacó las castañas del fuego a Juan.
"As always, his dad intervened and helped him out of this situation"

After finishing reading each background story, I asked each participant *¿Qué pasó?* 'What happened?' Participants answered by embedding their response within a structure of the type 'What happened was that' followed by the target sentence. As mentioned earlier, this was done to normalize the prosodic context in which the summarizing statement was produced by using a concept analogous to what a carrier phrase does during nonce or nonsense word elicitation. Elicitation of words spoken in isolation can create complications related to speech rate, as well as semantic variations, as speakers may try to imagine different contexts for the isolated sentences that, in turn, may affect the way they produce them (Gibbon et al. 1997:100).

Effectively, each participant produced two minimally different sentences that look like (10) and (11), where arguably, the only distinction is the intended meaning the speaker had in mind when they produced them.

(10) Luis le sacó las castañas del fuego a Juan.(literal)
Luis CLIT.3S take.PRET the chestnuts from the fire DOM Juan
'Luis took the chestnut from the fire for Juan'

(11) Luis le sacó las castañas del fuego a Juan.(idiomatic)
Luis CLIT.3S take.PRET the chestnuts from the fire DOM Juan
'Luis helped Juan out'

2.3.3 The Data Set

Existing literature in the field of prosody indicates that changes in syllable duration, intensity, and variations in F0 have been found to play a role in disambiguating pragmatic meanings across languages [Adachi (1996); Cheang and Pell (2008); Cheang and Pell (2009); Culpeper (2005); Fonagy (1971); Haiman (1998)). For Spanish in particular, Ortega-Llebaria (2006) finds that speakers of Spanish rely on F0, duration, and intensity as the key acoustic correlates of stress distinctions. Nibert (1999), Face (2001, 2002), and Rao (2013b) have found that different permutations of these three variables give rise to a variety of pragmatic meanings. As a result, the data collected was structured around these three variables. This section describes how the data was extracted from sentences like (10) and (11).

Each statement was looked at from two different perspectives: phrase level and constituent level (i.e., subject, verb, object). This was done to mirror the AM's hierarchical structure presented in Figure 1 where the *intonational phrase* corresponds to the entire utterance, and each of the three participants of the event correspond to the *prosodic words* making up each of the sentences in the corpus. See Figure 3, containing *María estiró la pata* 'María stretched her leg.' All measurements were carried out using Praat.

At the phrasal level, I looked at the entire statement and recorded maximum, mini-

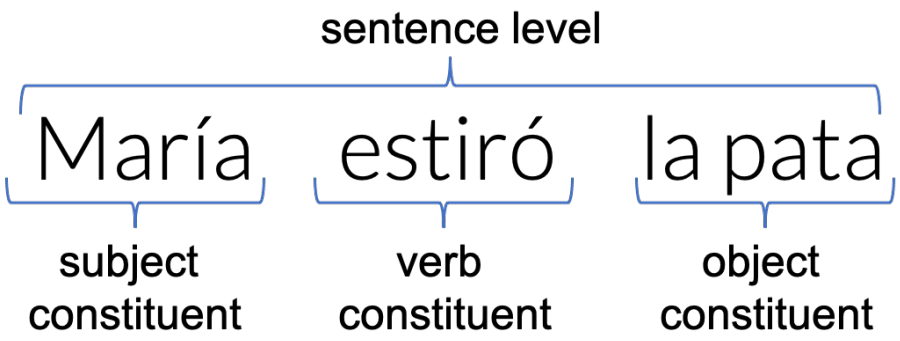


Figure 3: Phrasal and Constituent Level

mum, and mean values of F0. Duration for the entire sentence was also recorded. Finally, for intensity, I recorded maximum, minimum, and mean values.

I subsequently looked at the sentence at the constituent level. I recorded maximum, minimum, and mean values of F0 for the subject, verb, object ,and second object constituent (when present). I repeated the process for the intensity of each constituent. For duration, I recorded both the duration of the stressed syllable including onset, stressed vowel and offset, as well as the duration of the entire constituent.

To summarize, each literal-idiomatic pair of statements produced 72 data points for transitive statements, $(9_{phrase} + 9_{constituent} \times 3) \times 2_{intent} = 72$, as detailed below in Table 3.

Table 3: Recorded variables per statement

Phrasal Level	Constituent Level
Maximum F0 in Hz	Maximum F0 in Hz
Minimum F0 in Hz	Minimum F0 in Hz
Mean F0 in Hz	Mean F0 in Hz
Duration in ms	Duration in stressed syllable in ms
Maximum intensity in dB	Duration of the constituent in ms
Minimum intensity in dB	Maximum intensity in dB
Mean intensity in dB	Minimum intensity in dB
F0 range	Mean intensity in dB
Speech rate in words per ms	F0 range in dB

In anticipation of the data analysis needed for this chapter, as well as for *Chapter 3* and *Chapter 4*, I also controlled for the following variables:

1. the gender of each participant, recorded as MALE or FEMALE.
2. how familiar each speaker was with each idiomatic phrase, recorded on a 1-5 Likert scale.
3. whether the two sentences were in fact identical or if some lexical accommodation was needed, recorded as a logical variable, TRUE if they are identical or FALSE otherwise.
4. if aspectual loss is expected, as predicted by independent tests, recorded as a logical variable, TRUE or FALSE.
5. ditransitivity, recorded as a logical variable, TRUE or FALSE.
6. if the idiomatic mean is distributed across the constituents of the sentence, in the sense of Nunberg et al. (1994). This was recorded as a logical variable,TRUE or FALSE.
7. the location of the peak of F0 activity, recorded as SUBJECT, VERB, OBJECT, or OBJECT 2.

2.3.4 A first look at the data

The entire dataset is composed of 20 literal-idiomatic paired up sentences produced by 13 speakers. In this section, I will describe what the data look like both from the point of view of how each statement was produced by the participants, as well as from the perspective of how each participant produced their individual statements. I believe this will give us a clearer understanding of the data I am working with. F0 is expressed in hertz (Hz), duration in milliseconds (ms), and intensity in decibels (dB).

Table 4 shows the collected data for statement 01, *aplastar la oreja*, with the literal meaning of 'to press one's ear flat' and the idiomatic meaning of 'sleep.' I show values of sentence-level F0 only. Not shown for space reasons are the rest of the collected F0 values for the subject, verb and object constituents. Phrasal and constituent recorded values for duration and intensity are also not shown.

Table 5 shows the recorded F0 values for the verb constituent produced by participant 08 across all the sentences this individual produced. For space reasons, only the recorded F0 values for the verb constituent are shown. The data collected for every constituent are

Table 4: Statement 01: aplastar la oreja . Recorded values at phrase level

Participant	Intent	Max F0	Mean F0	Min F0	Duration	Max Intensity	Mean Intensity	Min Intensity
01	literal	217.41	132.55	113.61	1095.69	75.99	68.36	52.18
01	idiomatic	152.52	128.99	105.47	1293.04	73.52	67.57	46.66
02	literal	140.43	123.99	104.39	1274.58	78.85	74.04	52.92
02	idiomatic	132.69	112.19	101.09	1229.06	80.11	75.47	51.51
03	literal	315.69	212.15	175.29	1141.91	71.68	63.87	45.54
03	idiomatic	288.47	204.04	166.92	1216.12	74.93	66.54	40.12
04	literal	275.95	202.70	167.87	1024.03	77.94	71.92	51.36
04	idiomatic	291.15	174.74	127.76	1219.91	77.12	69.00	46.96
05	literal	154.63	136.51	120.80	1064.55	72.27	66.85	49.87
05	idiomatic	158.45	130.25	110.60	1041.59	68.16	62.38	46.88
06	literal	137.83	51.37	92.39	1771.01	75.67	71.56	47.56
06	idiomatic	127.60	111.90	82.83	898.32	78.19	70.52	57.65
07	literal	134.79	107.32	88.46	1694.23	72.90	68.93	46.34
07	idiomatic	121.32	99.43	81.79	1518.21	73.37	67.82	49.60
08	literal	282.09	198.18	151.92	1319.48	77.36	67.98	43.85
08	idiomatic	343.74	217.84	109.60	1325.68	75.98	68.55	49.07
09	literal	256.86	170.10	202.83	1524.22	71.68	66.85	39.87
09	idiomatic	289.42	220.69	160.72	1368.91	73.04	68.93	40.99
10	literal	185.26	157.09	118.35	1609.76	75.75	69.55	49.76
10	idiomatic	187.46	147.21	114.07	1039.68	68.50	64.87	50.31
11	literal	174.84	126.97	102.51	1274.04	77.33	70.73	49.38
11	idiomatic	169.41	139.42	121.20	899.80	76.19	66.61	41.82
12	literal	238.76	196.63	179.29	1066.15	67.75	63.22	49.34
12	idiomatic	249.15	194.24	169.42	986.62	71.20	66.00	46.55
13	literal	318.79	256.32	216.60	1500.07	82.36	72.85	42.69
13	idiomatic	320.44	217.84	162.15	1264.62	79.09	71.39	50.04

identical. The data collected from the individual constituents differ from the data collected at the phrase level in that phrasal duration only has a single recorded value, while a constituent's duration has two: across the entire constituent and for just the stressed syllable.

In organizing the data, I followed what is known as *tidy data* (Grolemund and Wickham 2017), which is a standard way of mapping the meaning of a dataset to its structure. A dataset is messy or tidy depending on how rows, columns and tables are matched up with observations, variables and types. Tidy data follows three basic principles: (1) each variable forms a column, (2) each observation forms a row, and (3) each type of observational unit forms a table.

Finally, the data also could be looked at from the perspective of gender, given that

Table 5: Participant 08: Recorded values for F0 verb constituent

Intent	Max F0	Mean F0	Min F0	Duration	Max Intensity	Mean Intensity	Min Intensity
literal	193.41	172.11	154.53	502.10	72.01	65.24	46.53
idiomatic	249.00	207.87	164.13	548.21	74.67	68.57	49.07
literal	218.24	186.90	175.28	268.16	69.40	67.94	62.64
idiomatic	217.66	195.78	181.00	375.33	72.37	69.67	56.51
literal	206.79	184.02	157.13	356.53	73.11	67.39	47.42
idiomatic	224.16	190.77	164.46	398.40	73.45	67.51	43.78
literal	191.51	158.42	146.16	317.60	70.33	64.80	50.01
idiomatic	192.62	172.88	159.07	362.09	69.04	65.20	31.19
literal	239.74	192.97	172.58	573.94	72.41	66.07	47.38
idiomatic	247.25	206.57	186.48	591.51	72.61	65.98	37.53
literal	181.95	180.47	177.75	184.93	68.43	66.33	63.85
idiomatic	192.07	182.79	171.80	181.59	72.14	69.93	66.79
literal	228.94	186.53	153.34	407.42	68.70	64.66	42.96
idiomatic	291.81	229.83	178.81	534.35	74.47	68.66	48.74
literal	242.79	202.70	154.59	501.47	70.35	65.80	54.18
idiomatic	224.20	195.38	154.37	506.90	69.99	64.03	40.30
literal	269.90	190.91	162.92	441.73	73.65	68.54	44.87
idiomatic	224.20	196.39	180.84	434.55	73.06	69.57	54.58
literal	234.05	183.61	162.45	502.19	71.80	65.97	43.91
idiomatic	241.48	186.27	165.05	408.43	74.14	68.99	54.16
literal	219.17	191.29	177.98	308.85	70.09	64.70	52.45
idiomatic	261.23	199.95	177.10	294.85	71.93	65.42	35.48
literal	227.17	197.93	181.03	287.21	69.23	64.40	49.90
idiomatic	216.83	199.82	176.75	333.50	68.87	64.11	36.38
literal	220.82	199.32	188.24	308.35	77.47	70.95	62.04
idiomatic	229.39	191.70	176.15	343.93	69.61	64.49	45.57
literal	229.99	160.20	147.59	483.75	68.85	63.89	53.88
idiomatic	250.08	196.61	177.06	567.66	74.61	68.11	47.63
literal	227.68	194.10	166.75	397.23	73.07	69.51	51.81
idiomatic	247.97	191.42	154.82	405.34	71.70	67.65	55.00
literal	205.09	190.96	175.55	272.89	72.15	68.42	45.05
idiomatic	217.74	179.52	160.18	297.79	71.40	66.30	49.57
literal	242.77	181.77	148.88	338.62	70.58	64.74	43.44
idiomatic	236.65	204.88	189.47	372.54	71.90	66.48	36.48
literal	253.41	200.09	162.48	320.32	73.47	69.31	38.99
idiomatic	256.22	197.88	176.43	297.58	74.64	68.77	55.72
literal	255.34	208.09	152.77	227.18	75.03	72.24	57.35
idiomatic	228.44	179.27	154.56	287.71	68.49	64.04	42.84
literal	256.70	181.55	155.54	532.70	74.71	67.91	44.42
idiomatic	266.01	178.00	134.42	593.68	76.24	67.47	39.56

the typical value of F0 varies significantly based on this variable. It is a well-known fact that human males' vocal cords and larynxes are larger than those of females. Both of these anatomical differences result in male adults having "lower pitch of voice and lower vowel resonances in comparison with the female adult" (Ohala 1983:12), with mean values of F0 of 120 Hz for men and 210 Hz for women (Traunmüller and Eriksson 1995). The subsequent analysis could be split by gender. All the steps of the analysis would be identical. Table 6 shows the mean values of F0 for male and female participants across all statements for the object constituent.

Table 6: Average F0 maximum, mean, and min values for the object constituent

Intent	Gender	Max F0 Object	Mean F0 Object	Min F0 Object
literal	male	134.18	116.99	103.83
literal	female	236.12	200.51	173.73
idiomatic	male	134.00	117.42	105.77
idiomatic	female	235.94	201.49	176.82

Table 6 confirms that for my participant population there is a noticeable difference in the average maximum, mean, and minimum F0 values between the female and male participants in the sample for the literal and idiomatic intents.

The disadvantage of analyzing the data this way is that the amount of data points to draw conclusions from, as well as to use to look for the difference proposed in H1, would be effectively halved. In order to get around that problem and be able to compare populations with such different means, the data can be normalized by z-scoring (Winter 2020:87). Z-scoring is a common statistical operation that allows comparing samples from different populations by linearly transforming both populations to a common standard.

Z-scoring involves two steps: *centering* the data by subtracting the distribution's mean from all individual values, and *scaling* the data by dividing all individual values by that distribution's standard deviation. Winter (2020:89) states that "standardizing may

help in making variables comparable." Examples of the use of this procedure are Song et al. (2002), who use z-scoring to normalize phonetic environments, and Blanco (2016), who normalizes accentedness ratings in order to account for listeners using a rating scale differently.

Therefore, z-scoring allows us to compare the male and female participants by putting them on equal footing. Each individual score now represents how far and in which direction that score deviates from its distribution's mean, expressed in units of the distribution's standard deviation. Z-scoring has the additional advantage of preserving the relationships that exist within the raw scores because it is a linear transformation (Schütze and Sprouse 2014:18).

Looking at the aggregated data in tabular form allows for the quick inspection of what values might be off compared to the rest of the recorded values. But ultimately, a tabular view is not a great indicator of the quality of the data recorded. To examine whether the data are normally distributed, and also to spot possible outliers, the data were initially visualized thusly: Figure 4 shows the variables maximum, mean, and minimum F0 at the sentence level per participant. On the vertical axis, the respective F0 values for each of the recorded sentences were plotted. Each participant's production is split into their literal response at the top and their idiomatic response at the bottom.

As expected, Figure 4 shows that each participant produced their maximum, mean, and minimum values clustered around the values' respective means. The plots show that there is -within each participant – variation across all 20 sentences they produced. In addition, the plot shows that some participants score greater F0 values for all three indicators than others. Cross-referencing that with the demographic data confirms that the higher F0 values were produced by participants 03, 04, 08, 09, 12, and 13, all female participants.

Figure 5 shows the same data z-scored by each individual's production. Generally, observations of real world phenomena tend to cluster around mean values and exhibit vari-

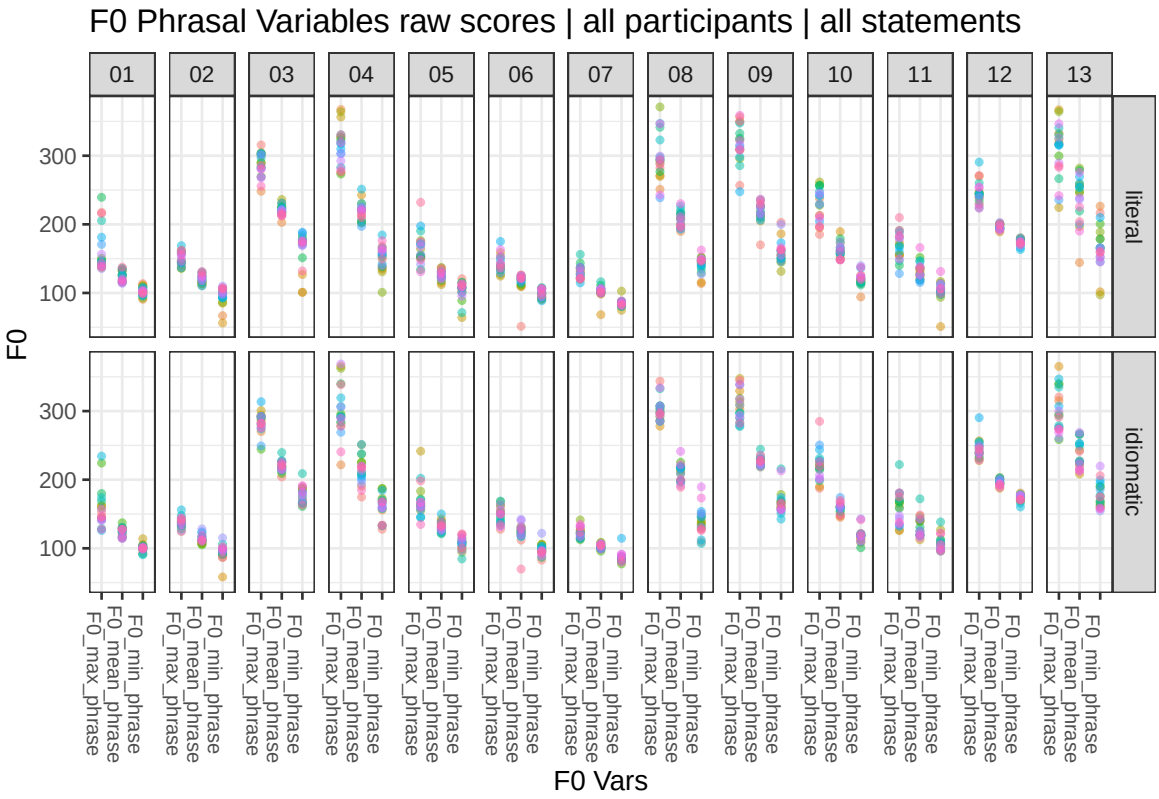


Figure 4: F0 phrasal variables, raw scores

ation along a Gaussian curve; that is 95% of the data is expected to fall within 2 standard deviations of the mean. Once z-scored, we see that the male-female variation is gone; this has the methodological advantage of allowing for comparison among genders. We further see that variation within each variable has also been normalized. As a consequence, for each of the statements the participants produced and each of the variables recorded, the observed values oscillate around a mean of zero. Furthermore, having z-scored the values allows us to easily spot outliers.

We see that for some participants, their productions are tightly clustered, while for others, there is more of a spread. How a person manipulates their F0 range seems to be a factor of "language, type of text, type of discourse and emotional state of the speaker" (Traunmüller and Eriksson 1995). Given that the experiment controlled for these variables, I chalk up this variation to individual production.

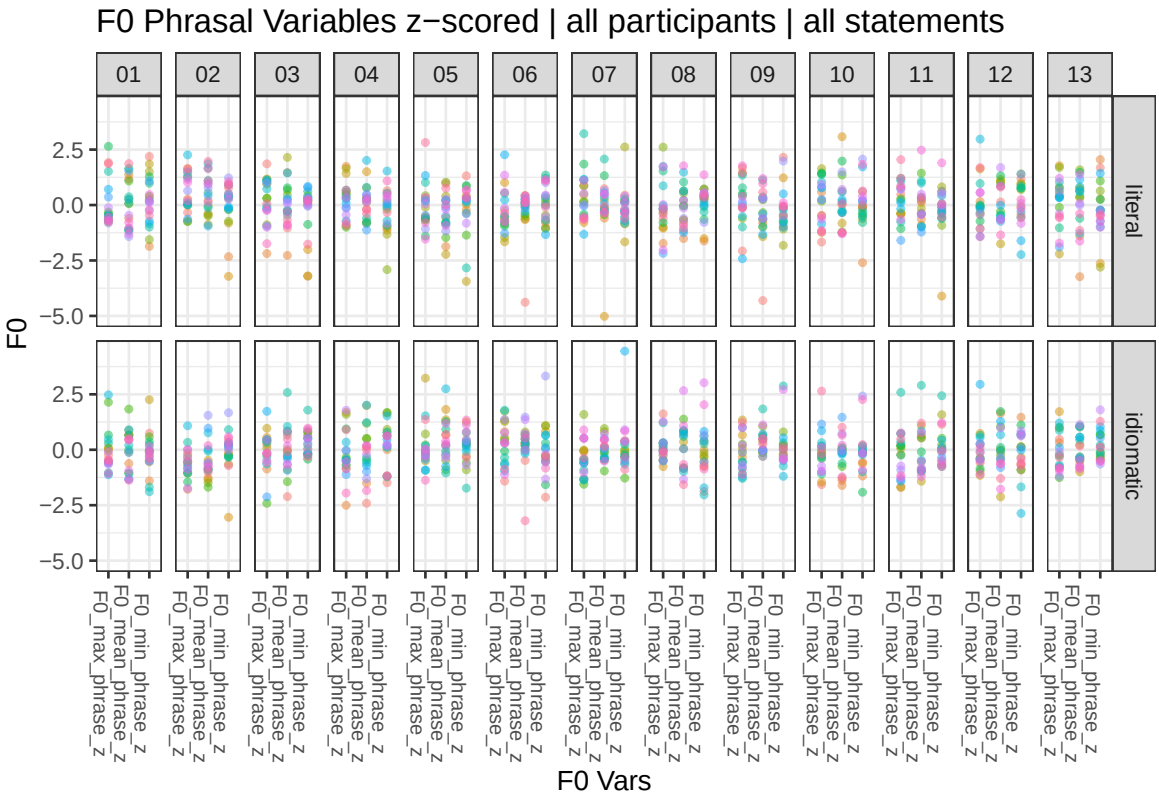


Figure 5: F0 phrasal variables, z-scores values

While the plots in Figure 4 and Figure 5 do an effective job of showing that every one of the participants exhibit a degree of variation in their production across their recorded variables and statements, it is not easy to spot how normally distributed their data are. Data distribution is an important factor to look at because it gives us an idea of how reliably the values were recorded. If the data consists mostly of outliers, it would impossible (or very unreliable) to draw conclusions from them.

2.3.5 Quality of the data

As hinted at earlier, observations taken from naturally occurring events tend to find themselves spread along well-established patterns or distributions. Recorded values of height, weight, reaction time to a stimulus, number of items a person can recall, typical duration of a syllable in a language, and a large number of other natural phenomena tend to cluster around the population's mean for that parameter and then thin out as they drift

away from said mean in what is known as a *normal distribution* (Aron et al. 2013).

A normal distribution, also often called a Gaussian or bell-shaped curve, exhibits a variety of properties that makes data conforming to this distribution easy to analyze and study. Chiefly among them is the clustering of observations around the mean of the population, as well as the predictable frequency of how those values will appear as they move farther away from the mean. This expected behavior is captured by the *Central Limit Theorem*, which states that "if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population with replacement, then the distribution of the sample means will be approximately normally distributed. This will hold true regardless of whether the source population is normal or skewed, provided the sample size is sufficiently large" (usually $n > 30$) (Sullivan 2016:69).

There are two main ways in which I checked for the normality of the collected data: (a) by visually inspecting the data and (b) by applying statistical tests for normalcy.

Figure 6 shows Quantile-Quantile (QQ) plots for F0 maximum values at the sentential level for all 13 participants. The 45º line is a hypothetical normal distribution and the dots are individual observations along that line; the closer the dots are to the line, the closer to an idealized normal distribution the observations are. As the plots show, a large fraction of the observations are tightly clustered around the line and contained within the interval between -1 standard deviation and 1 standard deviation. All of this is to say that the plots suggest the majority of data conforms closely to a normal distribution.

Figure 7 shows the distribution of the data as a density plot for the recorded maximum values of the verb constituent. A density plot displays the actual observed values and compares it to a theoretical normal distribution with the same mean and standard deviation as the samples. Since the raw samples were z-scored, the data are being compared to a normal distribution with a mean of zero and a standard deviation of one. As Figure 7 shows, the recorded values for all participants – except perhaps for participant 01 and

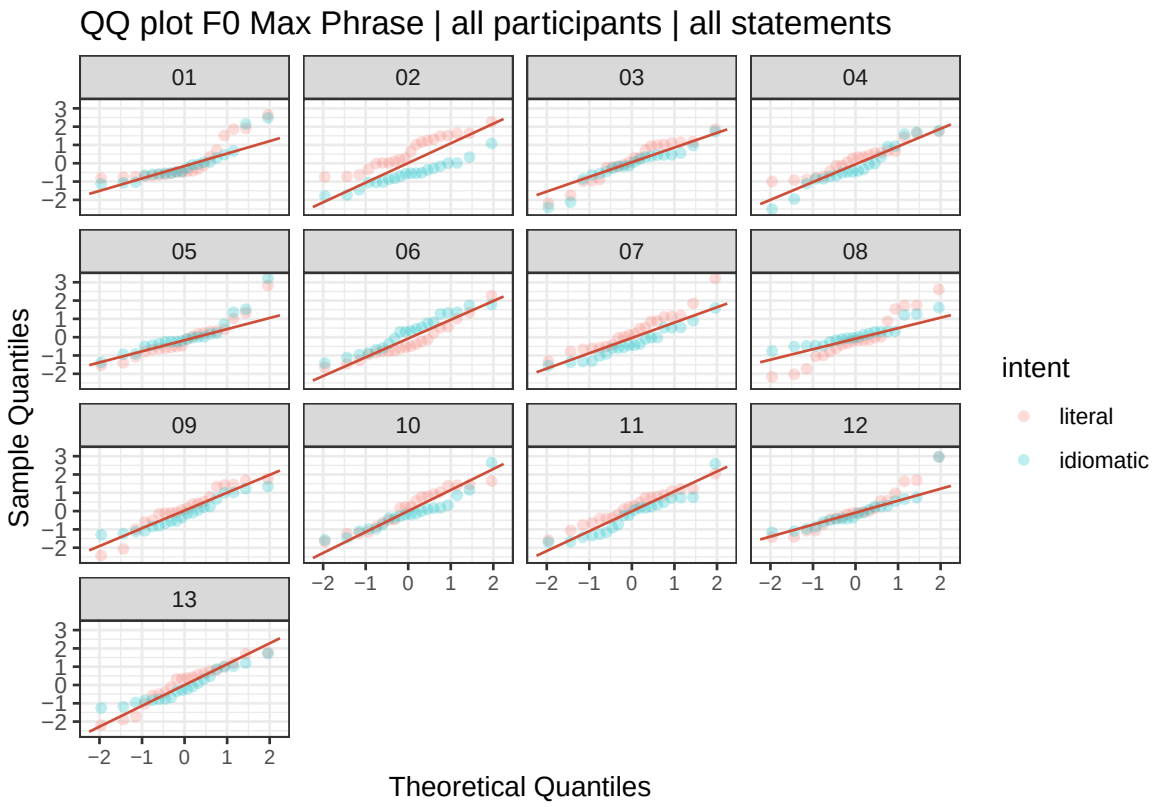


Figure 6: QQ plots F0 Maximum phrase | All participants | All statements

participant 11 - are fairly normally distributed.

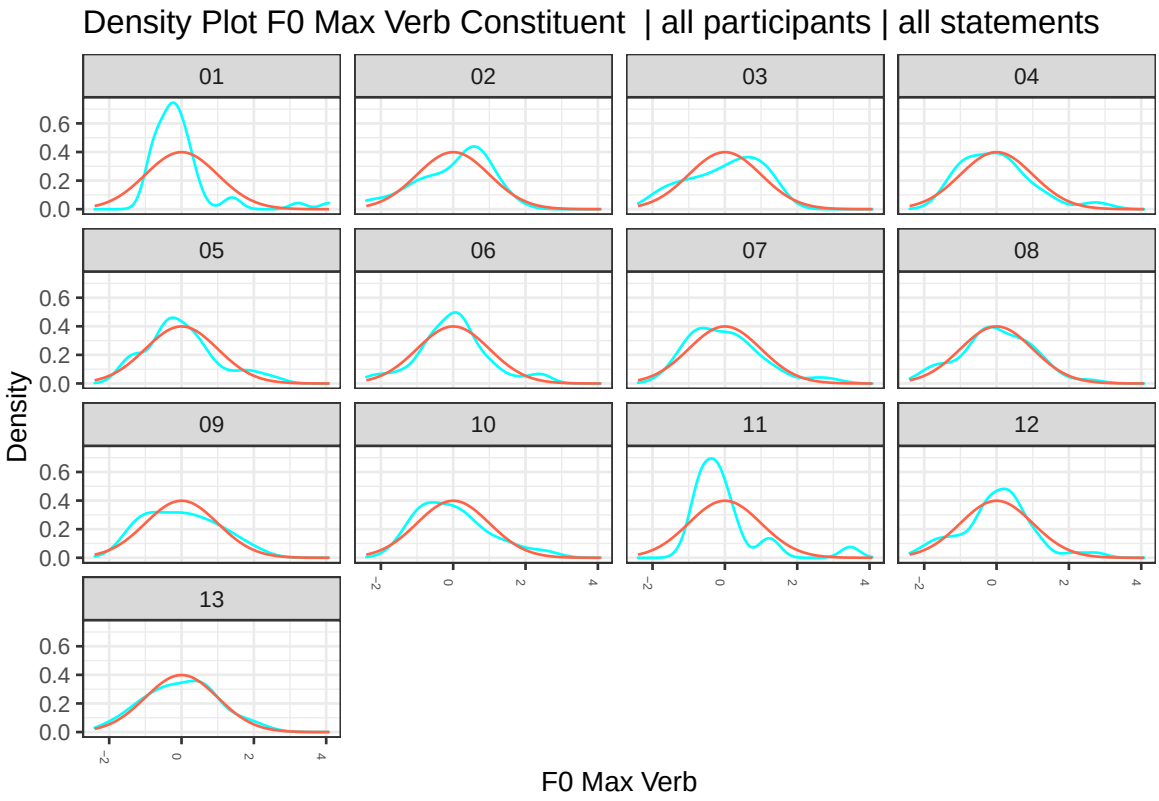


Figure 7: Density plots F0 Max Verb Constituent | All participants | All statements

Figure 8 provides a bird's eye view of all the F0 data for the maximum values of F0 at the phrase level by using boxplots. Boxplots are highly effective way of seeing the recorded data distribution. The boxplots in Figure 8 tell us information about central tendency and dispersion of the collected data. Critically, Figure 8 shows that:

1. The majority of the data for all speakers are contained within the first quantile and the third quantile. The dots represent potential *outliers*.
2. Per participant, the center of the data isn't the same for the literal and idiomatic interpretations.
3. The maximum values and minimum values do not appear to be consistent across participants.
4. The spread does not appear to be consistent across intents; that is, for some speakers, the F0 excursion is greater in the idiomatic intent, and for others, it is greater in the literal one, regardless of gender.
5. The median for the idiomatic interpretation is lower for 11 out of the 13 participants. This means that the upper 50% of the idiomatic data is consistently below the upper 50% of the literal data. This, in turn, suggests that there might be a difference

between the two realizations.

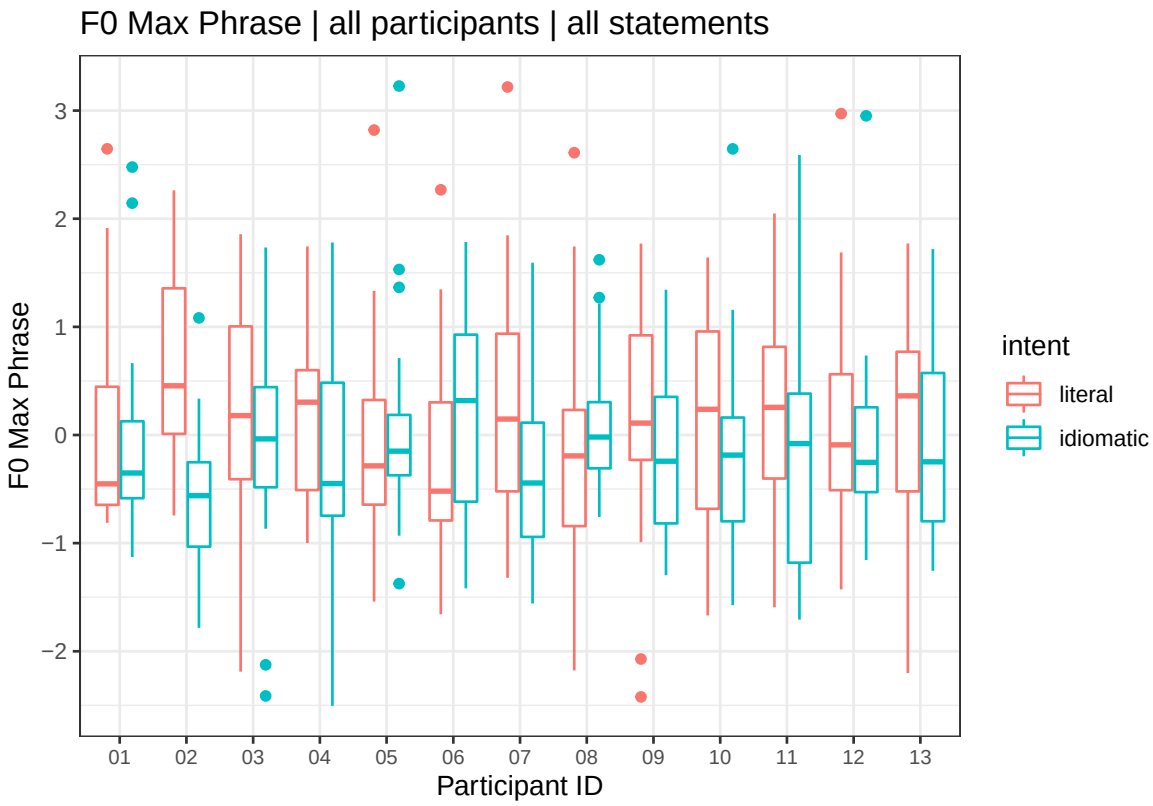


Figure 8: Density plots F0 Max Phrase Level | All participants | All statements

Lastly, I also checked for normality using a Wilcoxon test. A Wilcoxon test compares the recorded observations against an idealized normal distribution (Bedford Wassink and Bender 2012). When I applied these tests to the recorded variables, they indicated that the data were normally distributed. As a way of example, Table 7 shows the results of the Wilcoxon test for the mean values of F0 at the phrase level. The p-value > 0.05 indicates that the data is normally distributed. Given the results of the visual inspections and the statistical tests, I will move forward with the assumption that the data is indeed normally distributed. Normally distributed data is an assumption that needs to be met for the use of the statistical tests that I employ in the results section.

Table 7: Test of Normalcy for F0 Mean Phrase

Variable	Test	p-value
F0 Mean Phrase	Wilcoxon rank sum test	0.93

2.4 Are there differences between the two interpretations?

Thus far, I have shown that the data collected for *Experiment 1* consist of 20 pairs of minimally-different statements produced by 13 speakers, and that these data are normally distributed. The goal of this section is to evaluate the existing claim in the literature that there is no distinction in the way speakers produce idiomatic vs. literal sentences. In other words, *ceteris paribus*, two (at the surface) identical sentences would have identical realizations despite the fact that speakers intend for each member of the pair to have a different interpretation.

To motivate my particular research, I started with the claim that there isn't any difference in the realization of idiomatic and literal sentences. Recent research in the area of prosody has opened the door to question that claim; for instance, "changes in prosody and melodic contour are assigned a certain grammatical and/or pragmatic meaning according to the context in which it is used, such as an emphatic statement or a question expressing surprise" (O'Rourke 2012:173).

We know that languages in general use changes in F0 in two different ways, depending on whether they are intonational or tonal. As an intonational language, Spanish can rely on intonation to mark clause-level meaning, such as declaratives vs. interrogatives. In addition to this basic usage, we have seen recent research focusing on other pragmatic meanings, such as marking contrastive focus, syntactic disambiguation, and signaling sarcasm. Recall that the goal of this chapter was to probe if these uses could be extended to include idiomatic disambiguation.

To accomplish this goal, the first step is to ascertain if there are acoustic differences between the two realizations, and if so, where such differences are located. To that end, I started by looking at one of the recorded variables: *F0_max_phrase*. F0_max_phrase is the maximum value of F0 across the entire sentence, as measured using Praat. The plot in Figure 9 is similar to the plot shown in Figure 8 except that the horizontal line within each box represents the mean of an individual production of F0_max_phrase (as opposed to the median).

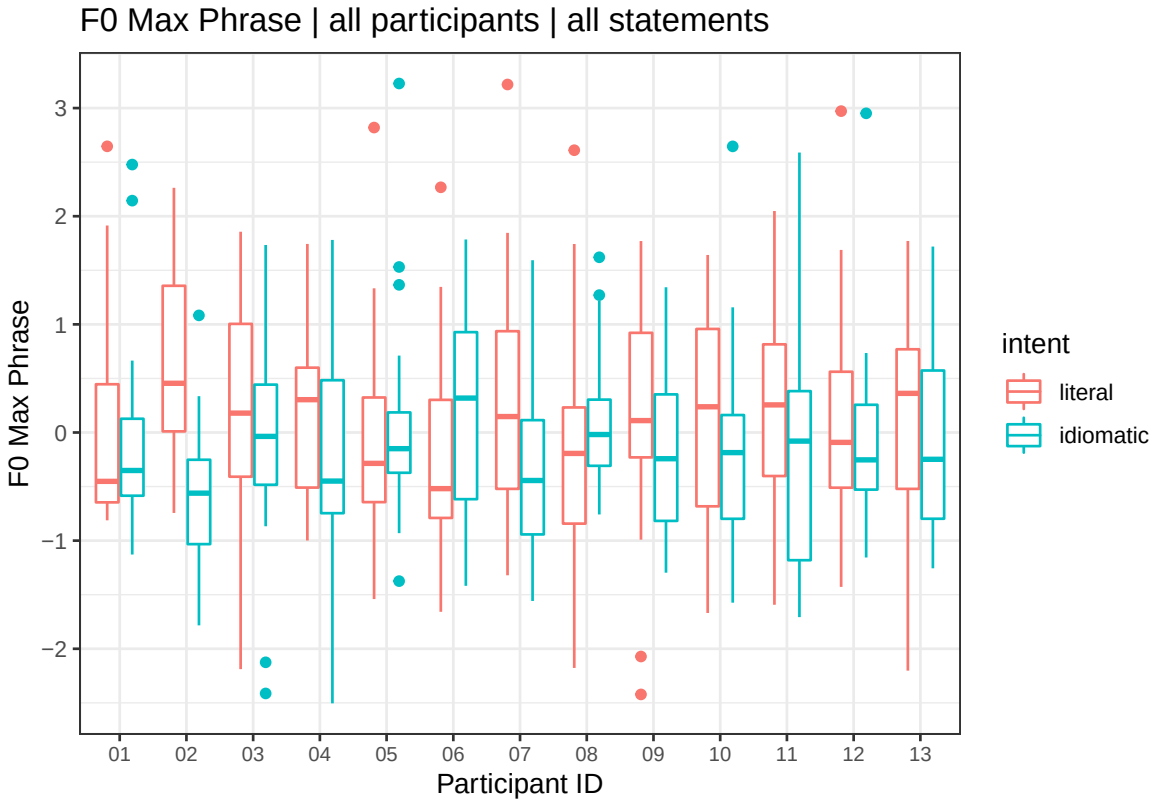


Figure 9: Box plots F0 Max Phrase Level | All participants | All statements

As the plot in Figure 9 shows, every participant produced the two sentences in the pair differently; that is, the mean of the idiomatic sentence is different from that of the literal sentence. The same analysis that was applied to F0 max for the phrase was repeated for the rest of the variables listed in Table 3 in order to be as thorough as possible.

Although this is a good first step, the situation isn't as simple as spotting the means

and concluding that they are different for most of the participants. The expected normal variation within each individual's production – as represented by the total height of a box – may very well account for these observed differences; in other words, these differences can be the result of individual variation rather than idiomatic intent.

In the following section, I apply a statistical test to find out if the observed differences are statistically significant, which would help rule out that they are due to chance or a product of individual variation.

2.5 Results

2.5.1 F0 Results

In order to determine if the observed differences are statistically significant and not due to individual variation, I carried out a paired t-test between observations. A paired t-test compares two means that are from the same individual (Winter 2020:176). It is conducted to determine whether there is statistical evidence that the mean difference between paired observations on a particular outcome is significantly different from a specified value, which in this case is zero, meaning both intents have equal means.

The statistical results will either support H_0 ($p > .05$) or give us grounds to reject H_0 ($p < .05$), in which case we consider the alternative that there are actual differences in the way speakers produce their idiomatic sentences. All results are presented in raw scores, but all calculations were done with z-scored values to eliminate interference due to gender differences in F0. As mentioned earlier, z-scoring is a linear transformation that preserves relationships existing in the data. Raw scores are human friendlier units than z-scores for the purposes of interpretation. See Table 8 for a summary of the results.

Table 8: t-test results for the F0 variable

	Maximum F0	Mean F0	Minimum F0
Sentence Level			
literal	223.89	167.76	128.24
idiomatic	219.66	168.28	131.2
p-value	< 0.05 *	> 0.05	> 0.05
Subject Constituent			
literal	212.93	186.17	164.15
idiomatic	210.49	187.87	167.58
p-value	> 0.05	> 0.05	$>$ 0.05
Verb Constituent			
literal	188.73	160.61	142.6
idiomatic	187.21	162.1	144.52
p-value	> 0.05	> 0.05	< 0.001 ***
Object Constituent			
literal	181.59	155.84	136.34
idiomatic	181.39	156.5	138.8
p-value	> 0.05	> 0.05	> 0.05

Table 8 shows the raw mean values for the variables F0 max, F0 mean, and F0 min for sentence level and individual constituents. It also shows that there are significant differences found for the maximum value of F0 at the sentence level for all speakers, as well as the minimum value of F0 for the verb constituent. When I initially fitted the model, I was concerned that most of the variation in the model was due to the differences in F0 production between female and male participants, as well as individual differences in production.

I addressed this concern by z-scoring the F0 values, which normalized the differences caused by gender. I ran an additional linear regression controlling for participants and gender, which yielded virtually the same results. A linear regression is another way to model the relationship between the dependent variable, or F0 values in this particular case, and the explanatory variable, or the literal and idiomatic intent of the speaker.

An advantage of a linear model is that it is programmatically very simple to include additional explanatory variables. The result of including gender in the model is that it

does not seem to have an effect in explaining the variation captured by the data, and for that reason, I decided to rely on t-tests rather than linear models. Recall that a paired t-test allows us to compare paired observations, which in this case results in comparing the literal production vs. the idiomatic production for each participant across all the sentences they produced. The tests indicate if there is a difference, and if so, that the difference is not likely due to chance.

The same basic procedure was followed to examine the rest of the constituents. Similar to the sentence level results, the constituent results show the raw values for all three observed variables. For the verb constituent, F0 min exhibits a significant difference between the idiomatic and the literal realization at the .001 level. For the object constituent, none of the observed variables yielded significant differences between the literal and idiomatic intent.

2.5.2 Intensity Results

The workflow followed in the previous section was repeated for the intensity predictors; the data were collected and organized exactly the same way as the F0 data were. A visual inspection of the intensity data reveals that the bulk of the data are normally distributed, as shown in Figure 10.

Figure 11 shows that participants exhibit a noticeable amount of variation in intensity. In order to be able to compare across all participants, all recorded intensity variables were z-scored by participant. Recall that z-scoring normalizes the raw scores by each individual's production, which puts them on equal footing and allows for comparison.

I subsequently analyzed the intensity data by looking at the predictors at the sentence and constituent level. The values presented in Table 9 were obtained by splitting the data into literal and idiomatic intent and calculating the mean of each predictor. Values are presented in raw dB, but all calculations were carried out using z-scored values. The

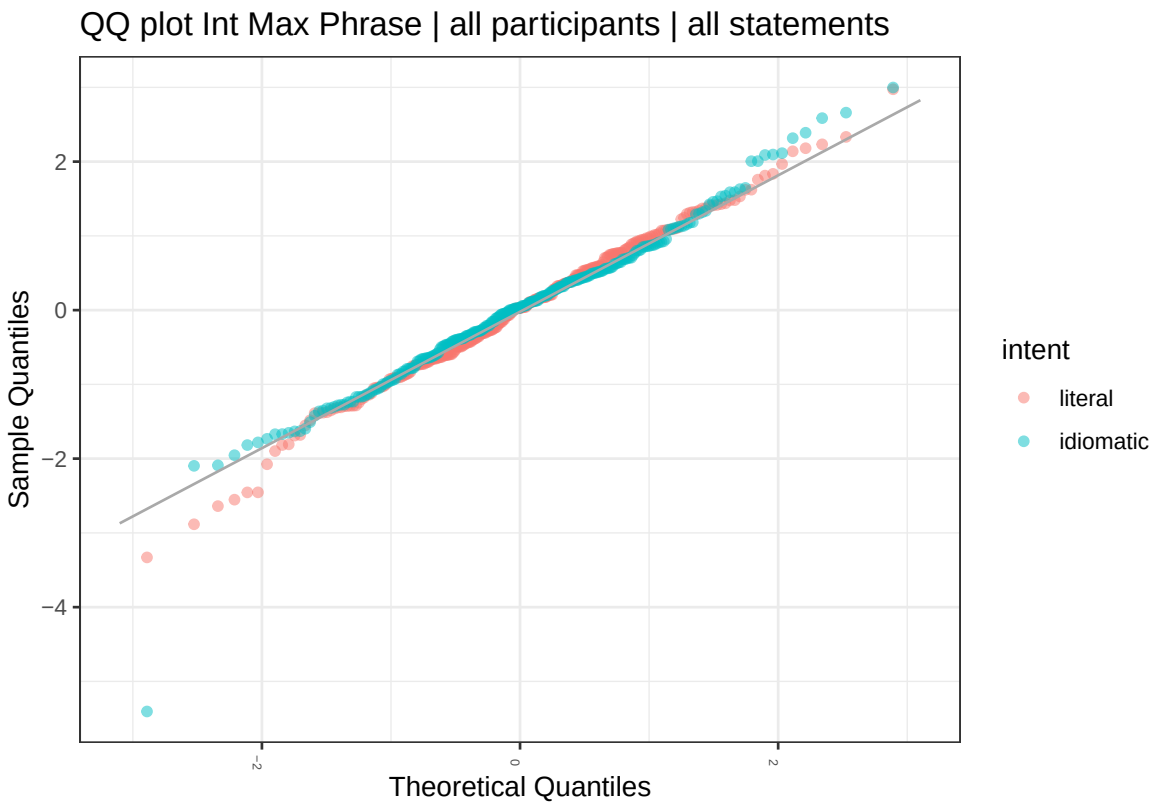


Figure 10: QQ Plot for Mean Intensity at the sentence level

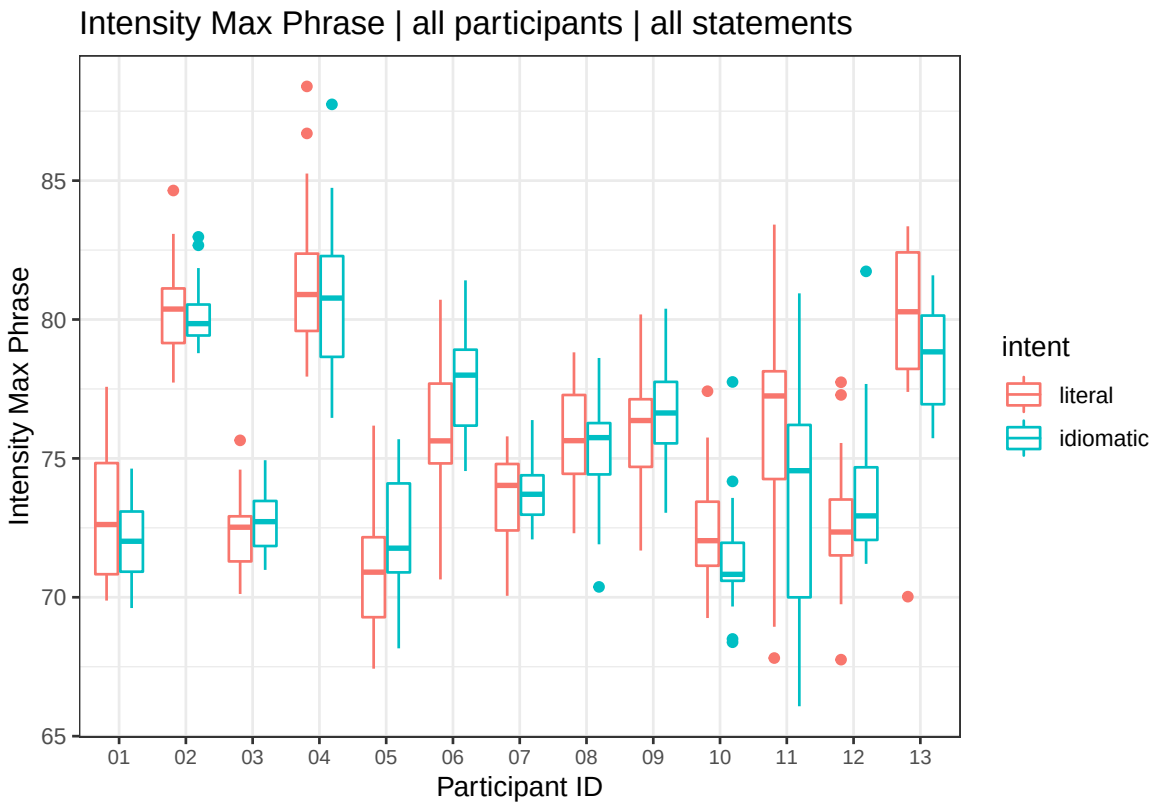


Figure 11: Boxplots plots Intensity Max Phrase Level | All participants | All statements

p-values were calculated by carrying out paired t-tests for each pair of literal and idiomatic observations per sentence across all participants.

The values shown in Table 9 indicate that there are no significant differences for the maximum, mean, and minimum values of intensity at the sentence or constituent level; in other words, from the point of view of intensity, the participants in the sample exhibited no difference in the way they produced their idiomatic and literal sentences.

Table 9: t-test results for the Intensity variable

	Intensity Max	Intensity Mean	Intensity Min
Sentence Level			
literal	75.45	69.12	46.01
idiomatic	75.24	69.08	46.47
p-value	> 0.05	> 0.05	> 0.05
Subject Constituent			
literal	74.87	71.22	56.17
idiomatic	74.74	71.35	56.75
p-value	> 0.05	> 0.05	> 0.05
Verb Constituent			
literal	72.86	68.8	51.35
idiomatic	72.97	68.93	52.02
p-value	> 0.05	> 0.05	> 0.05
Object Constituent			
literal	70.6	66.02	49.88
idiomatic	70.6	66.18	50.83
p-value	> 0.05	> 0.05	> 0.05

2.5.3 Duration Results

Duration values were recorded differently from F0 and intensity values in that, in addition to sentence and constituent values, the duration of the stress syllable (DSS) was also recorded. Unlike the F0 parameter, there is no indication that participants exhibit systematic duration differences by gender; in other words, the mean duration of female and male participants on average was expected to be the same for all constituents which was indeed the case, as shown in Table 10.

Table 10: Mean values for the duration variable

	Duration Phrase	Duration Subject	Duration Verb	Duration Obj
male	1.50	0.37	0.34	0.40
female	1.57	0.37	0.35	0.43

However, in retrospect, there was a flaw in the experimental design when I chose to use different names for the subject constituents; some of them were polysyllabic, such as 'Teresa,' while others were monosyllabic, like 'Juan.' This decision had implications for duration in that some of the differences among sentences would be the result of 'Teresa' being, in principle, three times as long as 'Juan.' As a result, in order for duration to be comparable across the 20 statements in the corpus, its values were z-scored.

In the process of visualizing the results, the QQ plot resulted in the best evidence that speakers shortened their duration when their intent was idiomatic, as shown in Figure 12. As indicated earlier, a QQ plot displays the data against a theoretical normal distribution. For the duration data, Figure 12 shows a fairly normal distribution. In addition, this plot also shows that for the vast majority of the data collected, the participants were systematically producing idiomatic sentences differently than their literal counterparts.

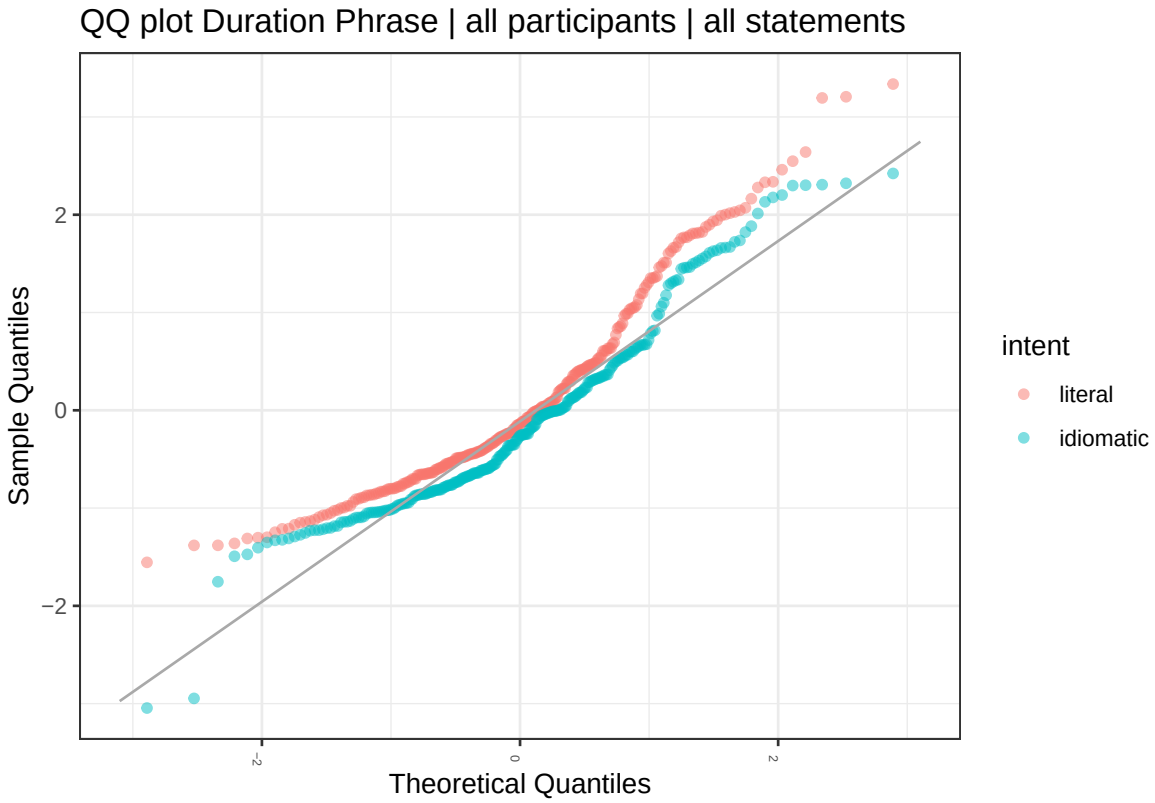


Figure 12: QQ Plot for Mean Duration at the sentence level

These duration results provide the most robust evidence in support of rejecting H0 (see Table 11 for mean values across constituents). At the sentence level, the difference between literal and idiomatic intent is 118.55 ms on average, with the idiomatic statements being shorter than the literal ones for all participants and all statements.

Table 11: Mean Duration per intent (ms)

	Phrase	Subject	Verb	Object
literal	1.59	0.39	0.36	0.43
idiomatic	1.48	0.35	0.33	0.40
difference	118.55	33.86	28.85	23.37

The next logical step in corroborating what the QQ plot in Figure 12 showed consists of applying a paired t-test comparing mean values of duration per intent. The t-tests carried out showed that there are highly significant differences in mean duration for the

sentence level, and the subject and verb constituents, as reported in Table 12. No significant differences were found for the object.

Table 12: t-test results for the Duration variable

Score	Duration Phrase	Duration Subject	DSS Subject	Duration Verb	DSS Verb	Duration Object	DSS Object
literal	1.59	0.39	0.23	0.36	0.17	0.43	0.17
idiomatic	1.48	0.35	0.21	0.33	0.15	0.4	0.17
p-value	< 0.001 ***	< 0.001 ***	< 0.001 ***	< 0.001 ***	< 0.001 ***	> 0.05	> 0.05

2.5.4 F0 Range Results

F0 range is defined as the difference between the maximum and minimum F0 value within a constituent, as shown in Figure 13. F0 range as a predictor variable has been included in studies previously mentioned (Face 2001; Rao 2013a, 2013b), and has been found to be susceptible to various factors, such as the language spoken, the type of discourse, and the speaker's emotional state (Traunmüller and Eriksson 1995:1).

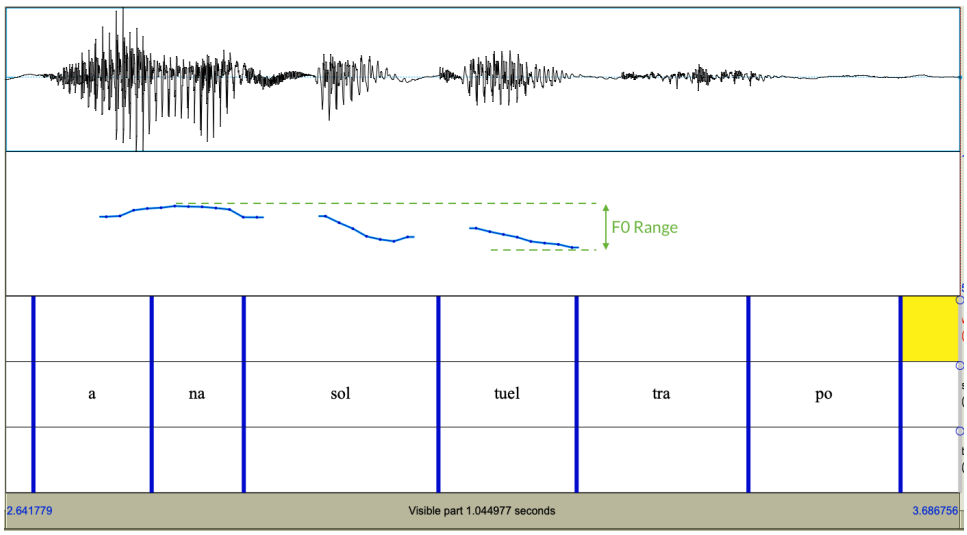


Figure 13: F0 Range Calculation

In order to avoid variation in the sample due to gender-induced differences, all recorded values were z-scored by individual participant. Table 13 shows mean values in Hz for the phrase, subject, verb and object constituent split by intent. A visual inspection shows that there is a difference across the board between the literal and idiomatic intent.

Furthermore, we see that F0 range decreases as we move from the beginning of the sentence to sentence-final position; this is to be expected, as neutral declaratives in Spanish are known to exhibit downstepping (Prieto 1998).

After calculating F0 range for each participant, each statement, and each intent, I carried out paired t-tests for the literal against the idiomatic values and tested if the difference was significantly different from zero. A significant result would indicate that there is a difference in the F0 range between the two interpretations, and that this difference is unlikely to be due to chance. P-values for F0 range are reported in the third row of Table 13. I found significant differences in F0 range at the sentence level, as well as in the verb constituent. No significant differences were found for the subject or object constituent.

Table 13: Mean F0 range per constituent - all participants

	F0 Range Phrase	F0 Range Subject	F0 Range Verb	F0 Range Object
literal	95.65	48.78	46.13	45.25
idiomatic	88.47	42.91	42.69	42.59
p-value	< 0.05 *	> 0.05	< 0.001 ***	> 0.05

2.5.5 Speech Rate

The last variable I looked into was *speech rate*. Speech rate is defined as the number of syllables divided by the entire duration of the sentence in syllables per second. Using this approach, Rao (2013b) found that slower speech rates correlated with sarcastic speech. The tendency to increase the duration of utterances when the intent is sarcastic has been noted in other languages as well (Adachi 1996 for Japanese; Haiman 1998 for English, Greek, Russian, and Tagalog; Culpeper 2005 for British English; Cheang and Pell 2008, 2009 for Canadian English and Cantonese).

The speech rate data are normally distributed, as Figure 14 shows. As I found no evidence that there are speech rate differences between genders, there was no need to z-score the speech rate observations.

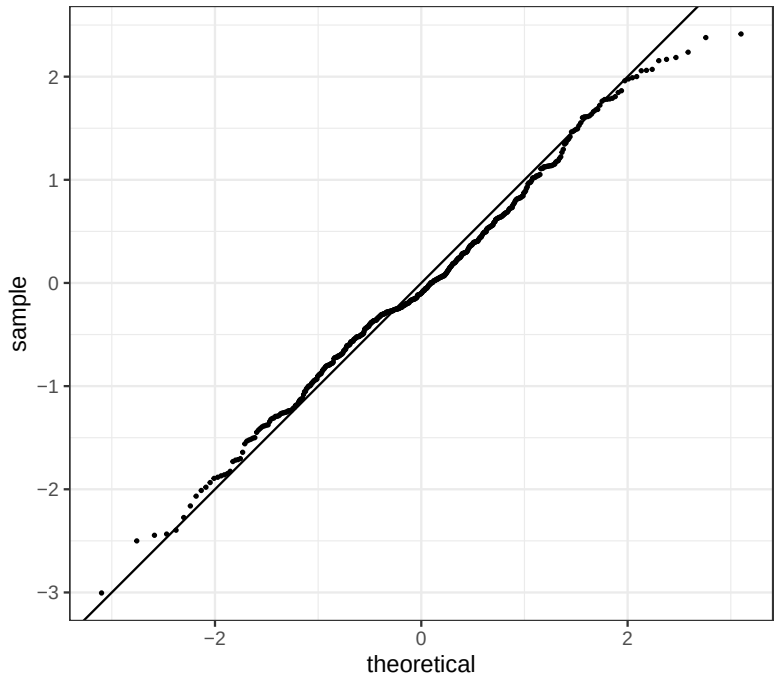


Figure 14: QQ Plot Speech Rate

In order to determine if there are significant differences between the speech rate of literal statements and idiomatic statements, I carried out a paired t-test comparing the production of each participant across sentences. The t-test revealed that there is a statistically significant difference between the speech rate of literal sentences and that of the idiomatic sentences for the participants in this study. Table 14 shows speech rate mean values across speakers and statements. The third row displays the p-value for the t-test.

Table 14: Speech rate t-test results (ms)

	Mean Speech Rate
literal	6.08
idiomatic	6.93
p-value	< 0.01 **

2.6 Discussion and Conclusions

The results in the previous section point to one important finding: the standing assumption that no difference exists between a sentence interpreted literally and the same

sentence interpreted idiomatically may very well need to be revisited. The results suggest that speakers can manipulate F0 and duration in their speech to mark if each of their context-dependent production is to be interpreted literally or idiomatically.

That F0 and duration play a role in this phenomenon isn't that surprising. Independent evidence already pointed to these variables having an important role in distinguishing lexical stress, and consequently, meaning indirectly. Ortega-Llebaria (2006) explicitly identified F0, duration, and intensity as the cues that speakers of Spanish pay attention to when discerning lexical stress.[8] Similarly, Beckman et al. (2002) and Hualde and Prieto (2015) identified multiple meaning contrasting pitch accents from most varieties of Spanish.

My results agree with the findings of Van Lancker et al. (1981:24) for English. The authors found that literal-idiomatic ambiguous sentences differ in "sentence duration, pausing, pitch contours, and word duration" in such a way that these factors "mark utterances as either literal or idiomatic." It appears that Spanish, like English, relies on changes in duration to help interlocutors disambiguate these sentences.

The end goal of this study was to accurately describe where and how these differences are manifested in speech as speakers use these cues to signal their intended meaning. To that end, Table 15 summarizes the findings of the previous sections. For F0, the first of the variables observed, we see that the maximum value of F0 at the phrase level is lower for the idiomatic sentences. We also see that the minimum value of F0 in the verb constituent is greater for the idiomatic interpretation compared to the literal one. This result was found to be significant at the .05 threshold.

[8]As an example, consider, for instance, the triplet *número/numero/numeró*, where, by virtue of switching the syllable position where stress falls, we see a noun (número), a sentence in the present (I number), or a sentence in the past (s(he) numbered).

Table 15: Summary of Significant Results for Experiment 1

Variable	p-value
F0	
Phrase F0 Max	< 0.05 *
Verb F0 Min	< 0.001 ***
Duration	
Phrase Duration	< 0.001 ***
Subject Duration	< 0.001 ***
Subject Stressed Syllable Duration	< 0.001 ***
Verb Duration	< 0.001 ***
Verb Stressed Syllable Duration	< 0.001 ***
F0 Range	
Phrase F0 Range	< 0.05 *
Verb F0 Range	< 0.001 ***
Speech Rate	
Speech Rate	< 0.01 **

The duration variables showed highly significant differences (p-value $< .001$) for the subject and the verb constituents. For all three predictors, a decrease in duration for the idiomatic interpretation was observed. No differences were found in the object constituent between the two interpretations. Duration was also measured for the stressed syllable of each constituent, where we saw that the idiomatic sentences had shorter stressed syllables in the subject and the verb constituent. No evidence was found that the stressed syllable in the object was different between intents.

F0 range yielded similar results, where significant differences (p-value $<.05$) were found in F0 range at the phrase level and very significant differences were found within the verb constituent. For all three constituents, the idiomatic interpretation exhibited decreased F0 range when compared to the literal interpretation.

Taking all three acoustic correlates together - F0, duration, and intensity - seems to paint a coherent picture: idiomatic interpretations translated in “compressed” sentences as far as these predictors are concerned. From the point of view of F0, at the phrase level,

a lower value of the ceiling of F0 and an unchanging floor would result in smaller excursions. In a similar fashion, for the verb constituent, a higher value of the floor of F0 meant that the F0 activity would have a smaller excursion in the verb constituent, given that maximum values of F0 remained unchanged.

From the point of view of duration, idiomatic sentences were also compressed, as they were shorter than their literal counterparts by 118 ms on average. F0 range painted the same compressed picture F0 did, with idiomatic statements being realized with narrower F0 ranges compared to literal statements. Similar to F0, F0 range exhibited significant differences located at the phrase level and the verb constituent.

At this point, I want to compare these results with those of Rao (2013b). This study, like Rao's, worked on data elicited by utilizing DCT scenarios where the aim was to produce semi-spontaneous speech. In his study, Rao reported that Mexican male speakers also exhibited a narrower F0 range for their sarcastic sentences. This result is notable, not because I want to suggest that sarcasm and idiomaticity are the same thing, but rather because both sarcasm and idiomaticity create a situation where the speakers are engaging in a mode of communication where "I'm saying this and I don't mean it" (Rao 2013b:49). In other words, what both the present results and Rao's results have in common is that both involve instances of primed speech production where speakers are meant to not be taken literally.

A second observation has to do with the lack of predictors in the object constituent. At first blush, this would look like an unremarkable observation, except that we know that some idioms exhibit a loss of notional objects. Take the expression *kick the bucket*, where, when understood idiomatically (i.e. "die"), there isn't a notional bucket to kick. The idiomatic interpretation corresponds to an event where there is only one participant involved. If this mechanism exists and speakers exploit these correlates to signal what interpretations they are trying to convey, it may not be unreasonable to assume those correlates will

live in the constituents that are guaranteed to be involved in meaning formation and not in a constituent that at times may be present and at times may not.

Recall that earlier, I indicated that there was some evidence that narrower ranges could be used to "signal negation or withdrawal of information" Gussenhoven (2002:4). In future research, I plan to look into the relationship between the narrowing of F0 ranges as a function of the loss of aspectual features.

The next chapter is a perception experiment where participants will be asked to validate the results found here. Recall that the outcome of this chapter's experiment was a pair of sentences that differed minimally in the speaker's intent, see (6) and (7). In the next chapter, participants will be played one of the two statements and be asked to judge if the speaker's intention was literal or idiomatic by means of a forced choice task.

The participants' estimation of each sentence's intent in the form of "literal" or "idiomatic" will be tallied against the intent that *Experiment 1*'s speakers had in mind when they produced those sentence. The resulting proportions of "correct" estimates against "incorrect" estimates will give a good indication of whether or not Spanish speakers are able to discern idiomatic intent in situations where context is unavailable to them.

3 Chapter 3: Experiment 2 — Perception

3.1 Introduction

In the previous chapter, I endeavored to make the case that current research in Spanish prosody has led to a deeper understanding of the manner in which Spanish speakers use intonation. Prior to Nibert's (1999) *lilas y lirios amarillos*, the consensus was that Spanish speakers' use of intonational strategies was narrowly confined to signaling a switch from a declarative to an interrogative.

Indeed, recent research in the field has caused that horizon to expand. Among these new ideas, we find that intonation is thought to perform a variety of roles that include syntactic disambiguation by means of intermediate prosodic phrases; focus resolution by manipulating the alignment of F0 peaks to their anchoring stressed syllables; and signaling sarcasm by manipulating F0 and duration. In *Chapter 2*, I concluded that the uses Spanish speakers have for intonation may need to be expanded to include idiomatic disambiguation as well.

This chapter is about finding empirical validity for those conclusions. As a result, the aim of this chapter is to investigate whether interlocutors can pick up on the intent a speaker has when producing a sentence that contains an idiom. In particular, in *Experiment 1*, I found that the correlates duration and F0 range were statistically significant in how the pool of 13 speakers contrasted their literal and idiomatic sentences. With that end in mind, the focus of *Experiment 2* is to ascertain the effect those two predictors have in the ability of interlocutors to correctly estimate the intent of sentences exhibiting the potential to be idiomatically ambiguous.

By virtue of its design and implementation, *Experiment 2* also gathered a significant amount of demographic linguistic data associated with the participants. In the forthcoming analysis, I will explore whether these factors have any bearing on the ability of partici-

pants to estimate idiomatic intent.

An answer in the positive can give us insights into the way different groups use prosody to negotiate meanings that are additive to the literal interpretation of a sentence. An answer in the negative, assuming that other factors prove to be significant, may be taken as evidence that prosody is used in Spanish in general as a disambiguating mechanism to identify idiomatic intent.

This chapter is organized as follows: section 3.2 covers relevant work done in the field that has impacted the experimental design choices I made. I also discuss work that guided me in choosing what linguistic and non-linguistic variables I included in the analysis. Section 3.3 covers how the experimental design choices came together, the participants, the independent variables, and the general workflow of the experiment and its implementation. In section 3.5, I explain how I used the information I gathered from the participants in order to determine if they could successfully estimate intent from a context-free sentence. Section 3.6 details the findings and their implications.

3.2 Literature Review

3.2.1 Linguistic Variables

In this section, I review a few studies that were influential in the design choices I made for *Experiment 2*. The first study had to do with an experiment focused on idioms. The other studies were not focused on idioms *per se*, but were tangentially influential in helping motivate what factors to include in the analysis.

The first study, and the most relevant experiment to my research, was carried out by Van Lancker and Canter (1981). In the spirit of full disclosure, I was not aware of this research until mid 2019. It was encouraging from a methodological perspective, since we made many similar design choices. It was also encouraging that many of their findings were compatible with my own as I show in the *Analysis* section later in this chapter.

That being said, there was a design choice where my experiment and Van Lancker and Carter's (1981) diverged. I chose to elicit the input in a sequential manner rather than in a contrastive manner. To contextualize, Van Lancker and Canter (1981:66) had their participants read idiomatically ambiguous sentences in pairs, one member after the other. They asked their participants to "attempt to convey, as unambiguously as possible, a literal meaning in one instance of saying the sentence, and an idiomatic meaning in the other instance."

In my design, I chose to have my participants learn the background story behind each interpretation and then relate what they knew of the event back to the interviewer. It is my belief that such a way of eliciting the sentences would result in something closer to spontaneous speech. Remarkably, both methods coincide in finding shorter idiomatic sentences and -arguably- narrower F0 ranges. I say arguably, because Van Lancker and Canter (1981:68) did not measure F0 range, but they did find "fewer and weaker pitch-accents contributing to a relatively smooth intonation contour" which, given the nature of speech signals, could only mean narrower F0 ranges.

I took an additional element from Van Lancker and Canter (1981): differentiate how to present the input to the listener during the perception tasks. The idea behind this strategy was to allow the participants to listen to the manner in which a speaker distinguishes their production of literal and idiomatic sentences before asking them to render judgment. I wanted to know if prior exposure to this contrast was a factor in identifying idiomatic intent. As a result, my participants were randomly assigned to stimuli that offered this contrast or proceeded directly to ask them to estimate intent.

Lisker and Abramson (1964) carried out measurements of Voice Onset Time (VOT) of word-initial stop consonants across a number of languages. VOT is defined as the time that lapses between the release of a stop and the start of phonation. In their view, VOT offers a single measurement that successfully manages to convey the differences between

stop consonantal segments.

Abramson and Lisker (1973) showed that, by manipulating VOT, they could map where participants differentiate voiced vs. voiceless stops in English and Spanish speakers. They found that this single adjustment caused participants to reliably tell the segments apart. They also found that where the burst is located in reference to start of phonation is language specific.

From Abramson and Lisker, I learned to pay attention to demographic variables that could potentially be used to group together varieties of Spanish. This initial study sparked the idea that changes contained in the speech signal can trigger perceptual judgments in the speakers of the language. I realize that I might be reaching by switching from discrimination of segments to discrimination of propositions, but an empirical approach to the situation may provide an answer to validate this assumption.

Ortega-Llebaria (2006) and Ortega-Llebaria and Prieto (2010) have been relevant to my study in that they informed what phonetic cues to pay attention to when developing the data that became the input for *Experiment 2*. According to Ortega-Llebaria and Prieto's research, speakers of Spanish are sensitive to these cues when determining lexical stress: F0, intensity and duration.

Secondly, Ortega-Llebaria (2006:105) presented the idea that information from the speech signal can be seen from a statistical perspective as a two-step process, where the initial step is phonetic, related to the acoustic signal and ultimately best approached by "bottom-up statistics which analyze all the variation displayed by the relevant cues as a source of useful information in finding distinct modes for each phonetic category."

The second step implies a re-analysis of the acoustic cues that give rise to meaningful phonological categories and is, critically, language specific. This secondary categorization process explains why the same acoustic cues can be seen as voiced in one language and voiceless in another. It is the interaction and interpretation of these phonetic cues that

result in meaningful categories in the language. Thirdly, Ortega-Llebaria's conclusion that it is possible to disentangle phonetic cues in such a way that we see them work in a two-tiered system - bound in the segment domain to determine stress, and at a suprasegmental level to indicate sentence level accentedness - is noteworthy.

The ability of speakers to repurpose phonetic cues led me to consider the idea that the same process could be applied to the phonetic cues contained in a speech sample to help participants decide how to interpret a sentence that is, at its core, ambiguous. As shown by the studies discussed in *Chapter 2*, it is possible to modulate speech in order to make elements more salient. This saliency, in turn, can be used to signal what choice is the one the speaker wants their listeners to adopt.

3.2.2 Extra-linguistic Variables

In addition to linguistic variables, I was also interested in the effect that non-linguistic variables could have in the ability of participants to discern intent. The motivation behind the collection of these variables can be subsumed by the idea of *stratification*. Stratification as a concept is used in most sociolinguistic studies and one that is closely related to the (socio-)*linguistic variable* (Labov 1966; Tagliamonte 2006).

A linguistic variable represents choices in language whereas stratification corresponds to the various factors, both linguistic and extra-linguistic, that help explain why some choices cluster around speaker groups. Labov (1990), for instance, identifies gender and social class as driving factors behind language variation with most men using non-standard forms and women driving linguistic change. Additionally, studies such as Paboudjian (2011) discuss sociocultural factors and their effects on prosodic variation.

In *Chapter 2*, for example, gender was a relevant factor related to F0 production and necessitated methodological accommodations (z-scoring) in order to be able to compare F0 values across gender. Unlike the gender and social reasoning in the paragraph above,

women and men cluster around typical F0 values for anatomical reasons (Traunmüller and Eriksson 1995). Specific to gender and intonation, Henton(1995:43) indicates that female speakers exhibit "larger variations at the 'lower' acoustic levels in the frequencies, bandwidths and amplitudes of vowel formants F1-F4." For additional discussion about gender and variation specific to Spanish, see studies such as Holmquist (2011).

Similar to gender, age has been found to be a relevant factor in language variation. Mitchell and Kingston (2014), for instance, report that aging adults perform poorly on pitch and amplitude discrimination. Furthermore, Seddoh et al. (2020) find that adults exhibit diminished ability to decode prosodic information. For additional discussion on age along these lines, see Ben-David et al. (2019). For Spanish-specific studies linked to age and aging, see Cameron (2011).

In addition to the aforementioned factors known to be involved in language variation, I was also interested in nativeness. There are known differences between L1 and L2 speakers across a variety of linguistic phenomena. Zárate-Sández (2015), for instance, reports differences for intonational targets among Spanish monolinguals, adult English-speaking learners of Spanish, and early Spanish-English bilinguals.

Specific to idiomaticity and nativeness is the idea that idiomatic use is closely related to L2 proficiency. Jackendoff (1992, 1997) points out that native speakers employ a large number of idiomatic and fixed expressions in their everyday interactions. Indeed, a hallmark of fluency is the ability of speakers, L1 and L2 alike, to choose from among the existing options the ones that that are not only grammatical but also "natural-sounding." Pawley and Syder (1983:191) use the terms "nativelike selection" and "nativelike fluency" in reference to these choices speakers routinely make. From that, it follows that the speech of speakers not making the correct choices is marked. For further discussion about the role of idioms as a measure of proficiency, see Vanderniet (2015).

There are reasons to collect geographical data as well. Differences between Spanish

varieties are well-documented in the literature. See Lipski (1996, 2008) for a general survey detailing variety-specific features in the lexicon, phonology, syntax and social dimensions. Hualde and Prieto (2015) also covers issues of lexical, morphological, and syntactic differences across Spanish varieties. These authors, along with those such as Prieto and Roseano (2010), also go into great detail about intonation and prosodic features and their variation observed in European and American varieties of Spanish.

I want to close this section by acknowledging that the interplay between speech and idiomatic intent that I am looking to study is but a small subset of research on interfaces. It is known that markedness and acentedness are used to make explicit intermediate intonational phrases that signal the correct parsing of a syntactic structure, which in turn, has reflexes in the interpretation of a clause (Nibert 1999, 2000). The list of phenomena associated with what Jackendoff (1972) calls "association with focus," and its concomitant sensitivity to phonological prominence, is extensive. Among them, we see quantificational adverbs, determiner quantifiers, counterfactuals, generics, modals, comparatives, superlatives, and negation (Howell 2012:21).

Furthermore, for the scope of this study, I confined the approach to an exclusively phonetic one with the understanding that as progress is made, the research can be expanded into a hybridized model where phonological categories can be brought into play[9]. Indeed, research in the field commonly approaches phenomena with observable phonetic explanatory variables first and then further refines them by introducing phonological categories accounting for meaningful stratification of the data. Among the options that I could pursue to refine the analysis started here, the inclusion of pitch accents and boundary tones within the AM Theory and SP_ToBI seems like a clear choice (Pierrehumbert 1980; Ladd 1996, 2008; Beckman et al. 2002; Face and Prieto 2007; Hualde and Prieto 2015).

However, what exactly constitutes these individual bits of acoustic information is not

[9]For further discussion about how to model speech perception and phonology, see Hume and Johnson (2001)

that clear to begin with; there is a lot of noise and variation both within individuals and across linguistics groups. Beddor (2017:1) states that, "Empirical studies that address this question provide clear evidence that perception is a malleable, dynamic, and active process. Findings show that listeners perceptually factor out, or compensate for, the variation due to context yet also use that same variation in deciding what a speaker has said."

All of this is to say that what is presented here is just a first attempt to find what is the same and what is different between idioms and the literal phrases that host them. As I tune in to the articulatory mechanics that are employed to signal idiomaticity, the description can be expanded to include phonological categories and how these interface with structure and meaning.

3.3 Experimental Design

This chapter is about validating the findings from *Chapter 2*. Recall that in *Chapter 2*, I conducted a production experiment *(Experiment 1)* with the help of 13 adult, Spanish native speaker participants. They all read a background story that was conducive to a literal or idiomatic interpretation of a sentence composed by a verb and its object. After reading and discussing the background story, I elicited a summarizing statement from the speakers by asking *what happened?* The speakers responded by encapsulating the summarizing statement in a carrier sentence of the type *what happened was ...* followed by the sentence in question.

The end result of *Experiment 1* was 20 pairs of sentences where each member of the pair differed from the other only in the intent the speaker had in mind at the time of production. Statistical analyses run on these pairs revealed that the speakers in the sample produced the idiomatic member of the pair shorter by approximately ~130 ms on average with a p-value=.001. In addition to duration, the F0 range of the idiomatic phrases was narrower by 27 Hz on average as well, with a p-value=.05.

This set the stage for the present experiment. The literal-idiomatic pairs produced in *Experiment 1* constituted the input for the participants in the perception experiment *(Experiment 2)*. The participants in *Experiment 2* were asked to listen to the members of these pairs and then estimate what the speakers' intent was at the time of production.

Following work done in English (Van Lancker and Canter 1981), the input was presented to the participants according to two different formats. Figure 15 shows the input for *Set A*, in which participants had the opportunity to listen to the complete set first. Participants were then asked to judge the intent of the individual members of the pair, one sentence at a time, in random order.

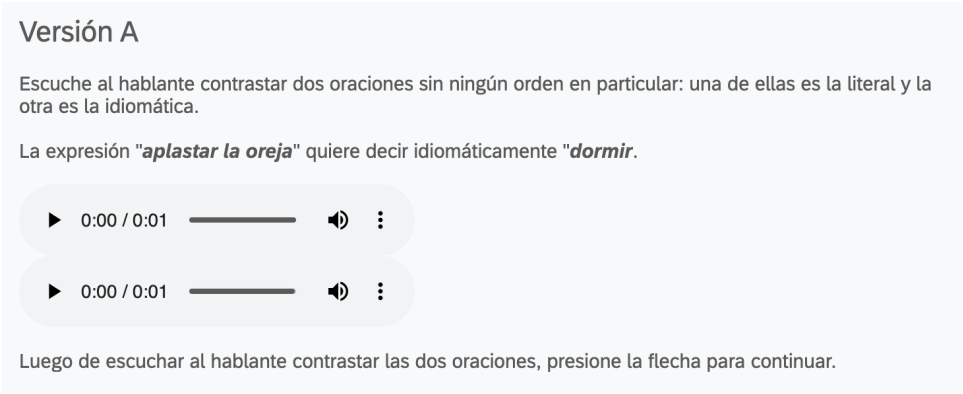


Figure 15: Set A input: complete set presented first

Alternatively, participants may have been presented with *Set B* input, as in Figure 16. *Set B* presented the participants with an individual member of the pair chosen at random. Then they were asked to estimate the speaker's intent without the benefit of having listened to the contrasting pair first. What set a participant was paired with was randomized as well.

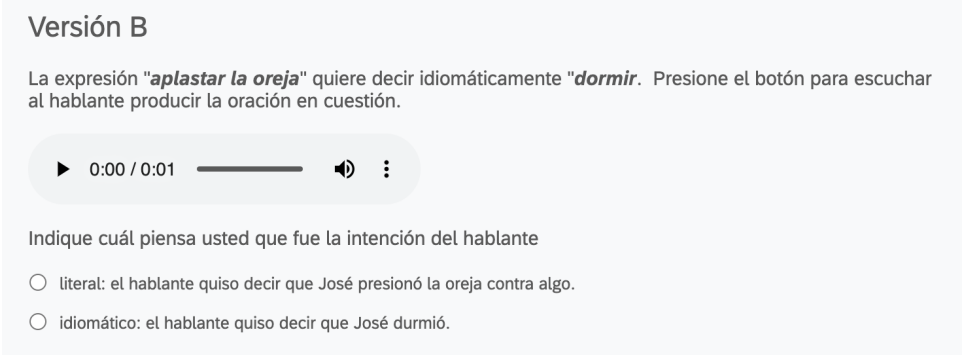


Figure 16: Set B input: without members of the pair being contrasted.

Of the 20 pairs of sentences produced in *Experiment 1*, a subset of 4 pairs were chosen to be the stimuli for the participants in *Experiment 2*. The inclusion criteria for the stimuli was twofold: (a) the literal and idiomatic sentences were absolutely identical and (b) they were the most commonly known expressions among the speakers of *Experiment 1*. In addition, 1 of the 13 speakers from *Experiment 1* was not included as part of the stimuli as his sentences contained a large number of pauses and hesitations.

In sum, the stimuli for the perception experiment consisted of 4 pairs of sentences produced by 12 speakers. The members of each pair were minimally different; the intent of the speaker at the time of production was the one difference. All fours pairs had a similar structure and consisted of a subject, a verb, and its object. There were a total of 96 sentences, stripped of context, and with a duration of approximately 1 second long each.

$$Number\ of\ stimuli = Speakers \times Sentences \times Intents \times Sets \Rightarrow 12 \times 4 \times 2 \times 2 \Rightarrow 192$$

In actuality, however, each participant interacted with only a subset of 16 sentences. These 16 sentences came from 2 speakers producing all 4 pairs. All pairings made between speakers and participants were randomized. The way in which a participant saw the stimuli was also randomized choosing between *Set A* and *Set B*.

The resulting output from each participant interacting with the stimulus was a judgment indicating whether the participant believed the statement was intended to be interpreted literally or idiomatically. In practical terms, each participant produced a response that was either correct or incorrect. The participants were, of course, blind to the intent of the speaker.

A correct response resulted from a participant indicating that a sentence was to be interpreted literally when the speaker meant it literally or idiomatically when the speaker meant it idiomatically. In other words, when the response and the intent of the speaker in the recording matched. Incorrect results come from a mismatch. Either the participant responded literal when the speaker's intent was idiomatic or the other way around, an idiomatic response when the intent was, in fact, literal.

3.3.1 Implementing the Perception Experiment

Experiment 2 was instantiated on *Qualtrics* (Qualtrics, Provo, UT). Qualtrics is the survey platform officially supported by UW-Madison. Potential participants get access to the study by means of an anonymous link as was required by the IRB protocol. Concerns of security and anonymity were managed by the Qualtrics platform and security protocols. In addition, all survey submissions were acquired via an anonymous link.

3.3.2 Structure of the Study

The first page of the study presented the potential participant with a general description of the research which included ways to contact the researchers and a statement about the goals of the research. This page also included a link to a more detailed explanation of the study and their rights as participants. The last relevant piece of this page required the potential participant to give their consent to participate in the study. Only after doing so were participants presented with the various tasks of the study.

The study went live on November 24 2020. On November 25, the website hosting the

study experienced the largest number of visits, 325 hits. The day of the study release, an invitation to participate was posted on the Department of Spanish and Portuguese of UW-Madison's website as well as the Latin American, Caribbean, and Iberian Studies' website. The email explained that the study was looking for native and near-native Spanish speakers to be participants in a linguistic experiment about idioms and encouraged them to participate. The email also asked the email recipients to forward the link to people who would be interested in becoming a participant both within and outside their academic circles.

Figure 17 gives an idea of the interest the study raised. The plot displays website hits throughout the month. The plot shows a downward trend that peaks on the release day and tapers off as the days went by. The study was open for a month, from November 24 to December 24, 2020. Each of the subsequent peaks observed in the plot corresponds to an outreach effort done through my social networks (personal emails, Facebook and Twitter). After the second peak, I decided to tap into my social media connections in an effort to capture participants outside of academic circles in order to diversify the sample.

The study itself was composed of three distinct parts:

1. a perception experiment.
2. a grammaticality judgment experiment.
3. a production experiment.

All three parts are independent, with idioms being the common thread throughout all the tasks. This chapter, as mentioned earlier, deals with the first part: the perception experiment. Before engaging in the perception tasks described earlier, participants were also asked to provide answers to these demographic questions:

1. Age of the participant, in years
2. Gender with which the participant identifies
3. Nativeness, whether a participant considers themself a native speaker

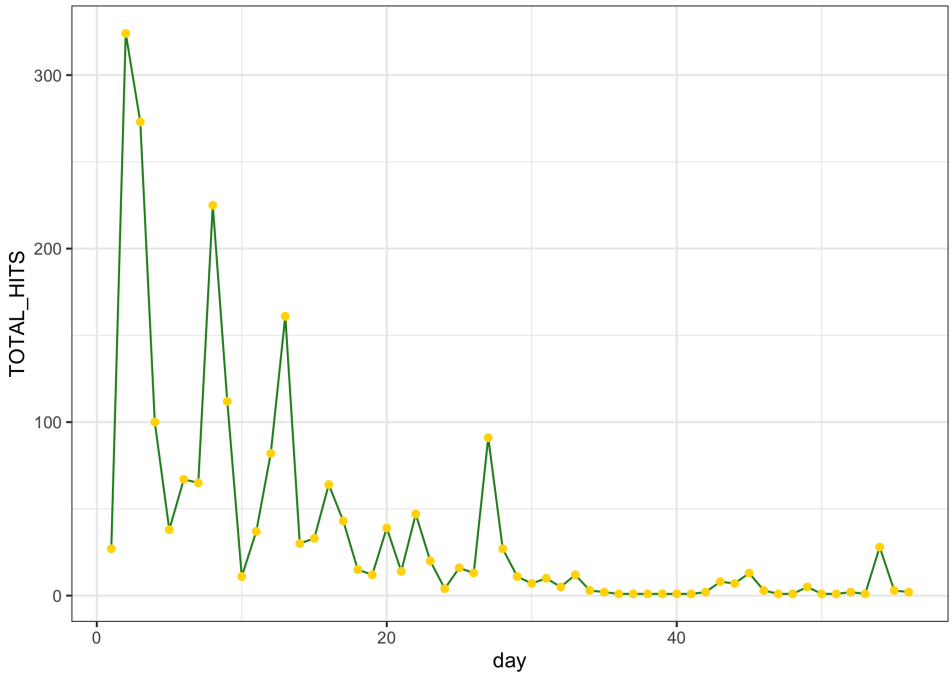


Figure 17: Website hits in a month

4. Country of childhood
5. Country of residence
6. Variety of Spanish spoken
7. Self-assessed speaking skills using a five-point Likert scale
8. Self-assessed listening listening skills, using a five-point Likert scale
9. Self-assessed writing skills, using a five-point Likert scale
10. Education level, from incomplete high school to a doctorate

Sociolinguistics has shown us that the answers to some linguistic phenomena may come packaged in an stratified manner that provides a better sense of how people use language. Some of these strata may be external to language, like country of birth or socioeconomic status, some may be linked to physiological reasons like pitch range. It may be the case that the perception of idiomatic intent constitutes a linguistic feature of a particular variety, or perhaps age (and its associated cognitive growth and decline) correlates with the ability to discern the intent of a speaker when the speaker engages in a speech act in

which they do not wish to be understood literally. As Mack (2019:315) puts it, the inclusion of these demographics is motivated by "the premise that social variables are essential to understanding how language works."

3.3.3 Perception Experiment

In this section I give a detailed description of what a participant was asked to do in order to complete the perception tasks. I also explain how the input was presented to the participants and what a participant response to a stimulus would look like.

Once a participant agreed to be part of the study and completed the demographic questions, they proceeded with the perception tasks. Behind the scenes, Qualtrics selected 2 speakers from the pool of 12 speakers in a random manner. These two speakers provided the input for the perception tasks for that particular participant.

The input was identical irrespective of the speaker and consisted of four pairs of sentences. Each member of the pair was identical to the other, except that at the time of production, the speaker was summarizing an event that was conducive to a literal interpretation in one case, and to an idiomatic interpretation in the other. The summarizing statement was the same in either case. The four pairs of sentences had the same structure: a subject, a tensed verb and its object. These four pairs of sentences are listed in (12) to (15).

(12) a. José aplastó la oreja
José flatten.PRET the ear
'José pressed his ear flat (against something)'

b. José aplastó la oreja.
José flatten.PRET the ear
'José slept'

(13) a. Juan rompió el hielo.
Juan break.PRET the ice

'Juan broke a piece of ice (with an ice-pick)'

b. Juan rompió el hielo.
Juan break.PRET the ice
'Juan started a conversation in a friendly way'

(14) a. Pedro tiró la toalla.
Pedro toss.PRET the towel
'Pedro tossed the towel'

b. Pedro tiró la toalla.
Pedro toss.PRET the towel
'Pedro gave up'

(15) a. Juan se apretó el cinturón.
Juan REFL tighten.PRET the belt
'Juan tightened his belt'

b. Juan se apretó el cinturón.
Juan REFL tighten.PRET the belt
'Juan saved (money or resources)'

Figure 18 outlines the entire perception experiment. It consists of five relevant steps that a participant needed to follow in order to complete all the tasks. Demographic questions aside, all pairings of speakers and participants, the choice of Set A or B, as well as the sentence order for each participant were randomized.

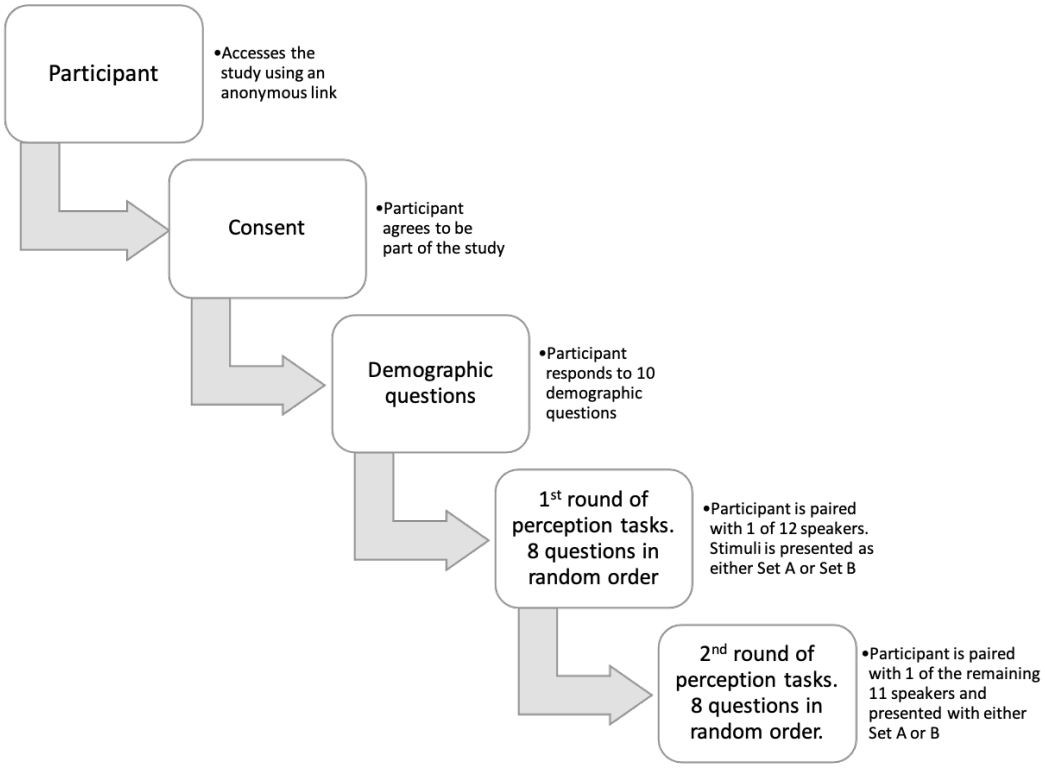


Figure 18: Perception Test Experimental Design

3.3.4 Meeting the participants

By the cutoff date, December 24, 2020, 584 people had participated in the study. Of those, n=485 completed all the perception tasks presented to them. The present analysis will be conducted using the completed data. The participant sample is significant for a variety of reasons:

1. It is a large sample compared to what we see in most perception studies
2. It is very diverse data. It is diverse by several simultaneous criteria which I describe below.

A critical element of a good sample is the requirement that the sample has to be representative of the population under study. Inaccurate results and conclusions may be reached when the sample is a distorted representation of the population. Sample biases can be caused by over, or under, representation within the categories and variables used

to model the phenomenon under study, by failing to stratify the phenomenon along sound groups or by small sample sizes, among other factors. See Visser et al. (2000) for a discussion of sampling and biases in Social Sciences research. See also Bakker (2010) for a discussion specific to language sampling.

Andringa and Godfroid (2020:135) point out that these issues are particularly critical for sociolinguistics and SLA studies as the conclusions drawn may not be generalizable to the population at large but rather be idiosyncratic to the group dominating the sample. In their meta-analysis of studies published in the SLA field, they found that 88% of all samples consisted of university students. This alone locks in on a number of constraints in the sample: age, socioeconomic status, and language variety spoken, among others.

For this study I attempted to find participants beyond the university environment. As mentioned earlier, I tapped into my social media network in order to ask every contact to complete the study and to forward it to anyone meeting the inclusion criterion of native or near nativeness. This resulted in the diverse sample whose specifics I detail below.

3.3.4.1 Gender Table 16 shows the breakdown of the sample by gender. The majority of the participants identified as women in the sample. The sample is composed, gender-wise, thusly n_{female} 57.61%, n_{male} = 41.15%, and n_{other} = 1.23%

Table 16: Gender Distribution

gender	n
male	200
female	280
non-binary	4
no answer	2

3.3.4.2 Age The sample contains a wide distribution, varying from participants as young as 18 to participants as old as 81. Table 17 shows the number of participants per

age bracket. The majority of the participants are contained in the 30-50 age bracket. For IRB reasons, the responses from participants younger than 18 were discarded.

Table 17: Age distribution

Age Brackets	n
18-29	92
30-39	147
40-49	152
50-59	63
60-69	19
70-79	10
80-89	3

3.3.4.3 Language variety Thanks to the experiment being conducted fully online, the sample population is composed of participants from several Spanish varieties. Participants provided information regarding where they grew up in order to ascertain their childhood variety. There are 28 self-identified different childhood varieties in the sample. Similarly, participants were also asked to identify the variety of Spanish they speak. There are 22 self-identified different adult varieties in the sample. Table 18 shows the 10 most frequent self-identified Spanish varieties as children and as adults.

Table 18: Childhood and Adult Variety.

standard_child_variety	n	standard_adult_variety	n
USA	132	Spain	110
Venezuela	100	Mexico	108
Spain	75	Venezuela	100
Mexico	66	Argentina	25
Chile	17	Chile	19
Peru	14	USA	19

Figure 19 gives us an overview of the most common varieties chosen as a child and as an adult. An additional interesting point that we can glean from the data is that there

are fewer adult varieties than childhood varieties. This chapter won't be expanding on that tangent, but that means that 6 people in the sample identify their Spanish as having changed from the Spanish they spoke as a child to the variety they speak as adults.

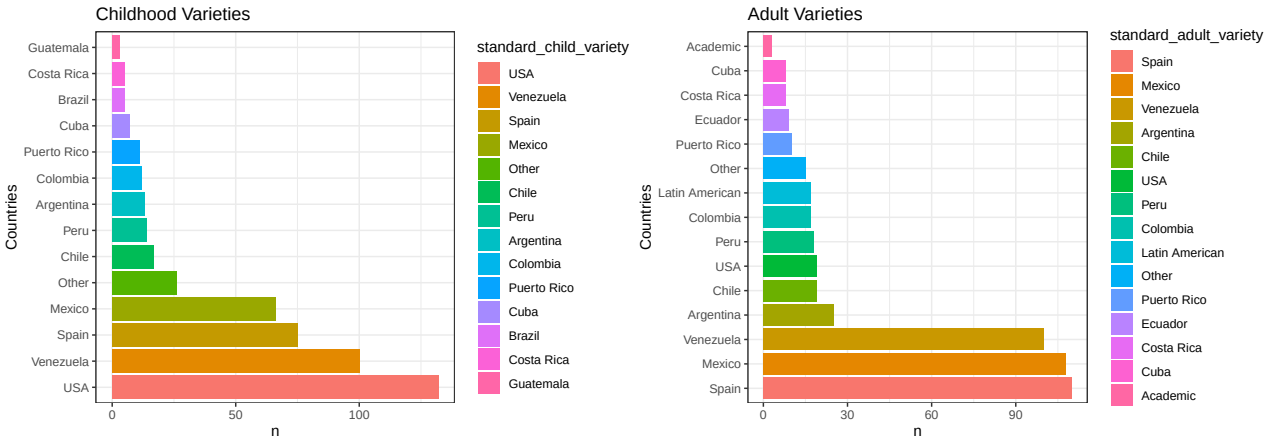


Figure 19: Child and Adult Varieties

3.3.4.4 Country of residence Participants reported 36 countries of residence in the sample. The majority of the participants live in the USA, making up 60.08% of the total sample. The second most frequent place of residence is Spain, at 10.91%, followed by Mexico and Venezuela, at 5.76% and 4.12%, respectively. To get an idea of the geographical spread of the sample, Figure 20 shows where the participants of *Experiment 2* live. Countries in teal indicate that a participant lives in that country.

3.3.4.5 Communicative Skills Participants were asked to self-assess how proficient they were at these communicative skills. Both natives and non-native participants rated their skills very highly with a substantial majority placing themselves at the "Excellent" level for all three categories. Figure 21 below shows the frequency distribution of the speaking skill within the sample.

The overall distribution of both natives and non-natives was the same; more people rated themselves as excellent speakers/listeners/writers than good. More people rated themselves "Good" than they did "Average." Within speaking skills, a notable difference

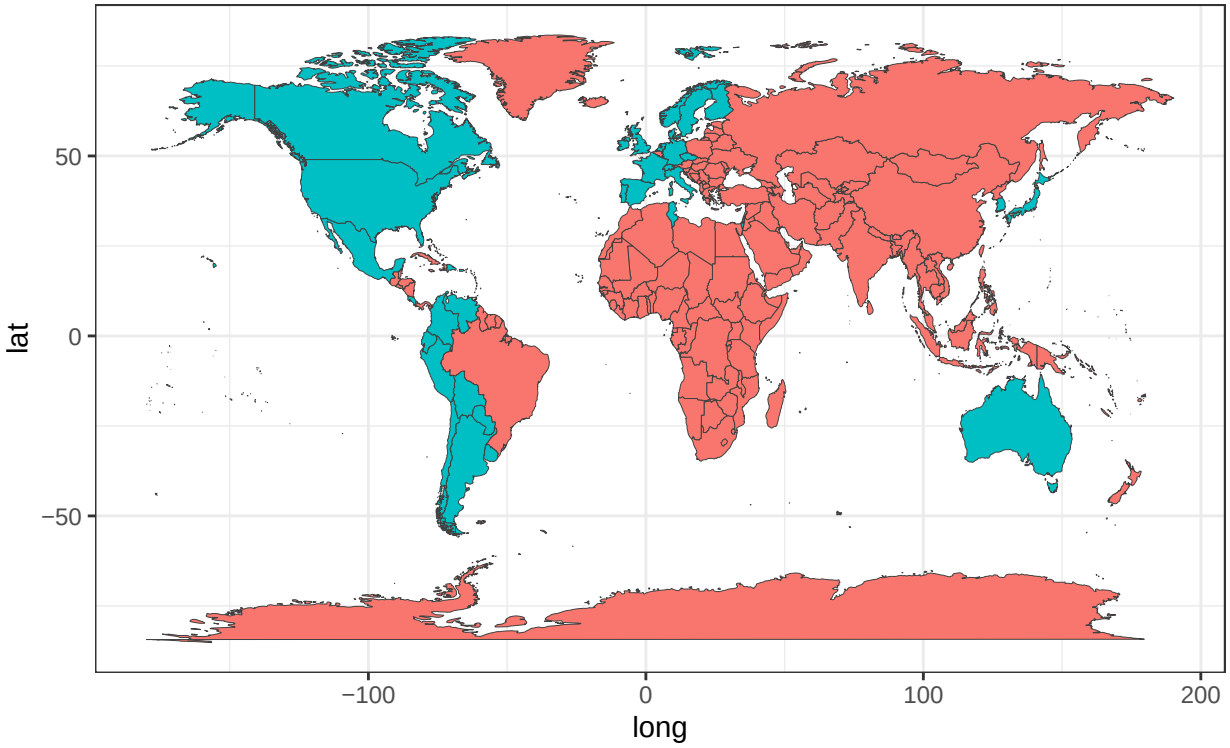


Figure 20: Experiment 2 - Participants' Country of Residence

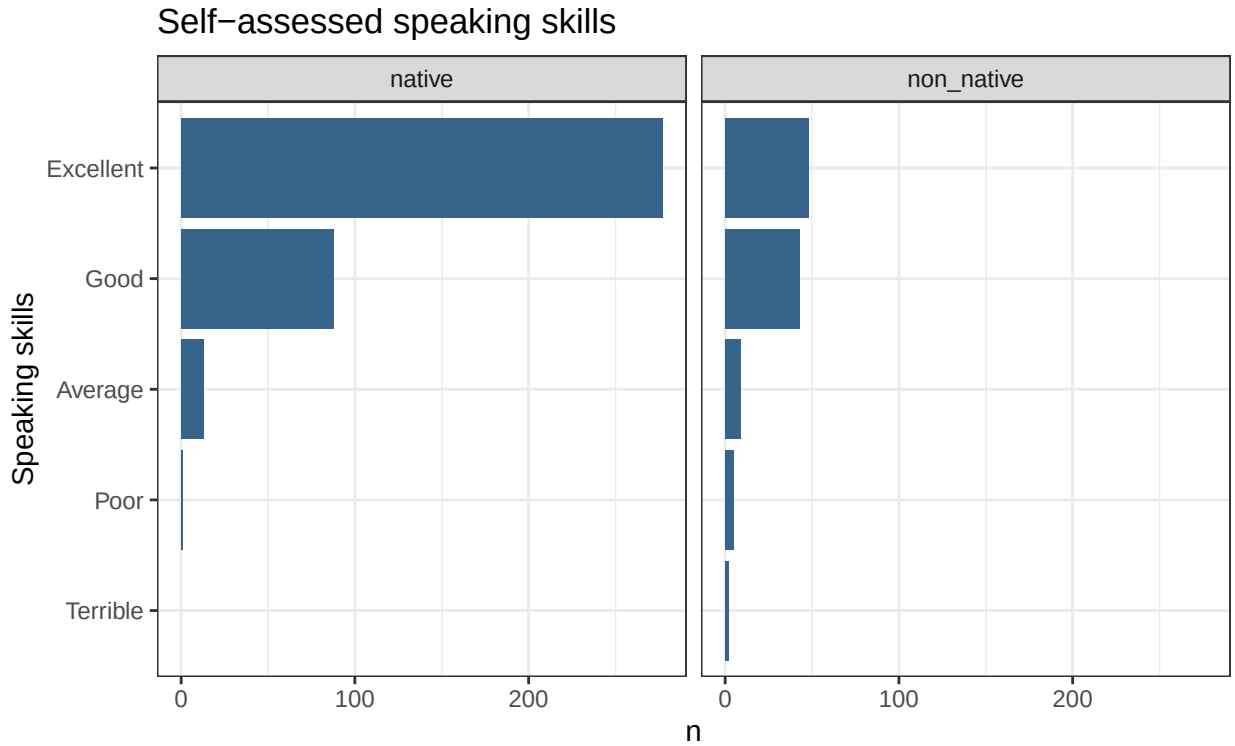


Figure 21: Speaking Skills

between natives and non-natives participants is that ratio of excellent to good for the native population is more than two to one for natives whereas for non-natives, they are virtually the same.

For listening skill, the overall tendency is the same, as Figure 22 shows. For native speakers, the ratio of excellent-to-good is much higher, about 3 to 1. For non-natives, the ratio remains the same; about the same number of non-native speakers rate themselves as excellent as they do good.

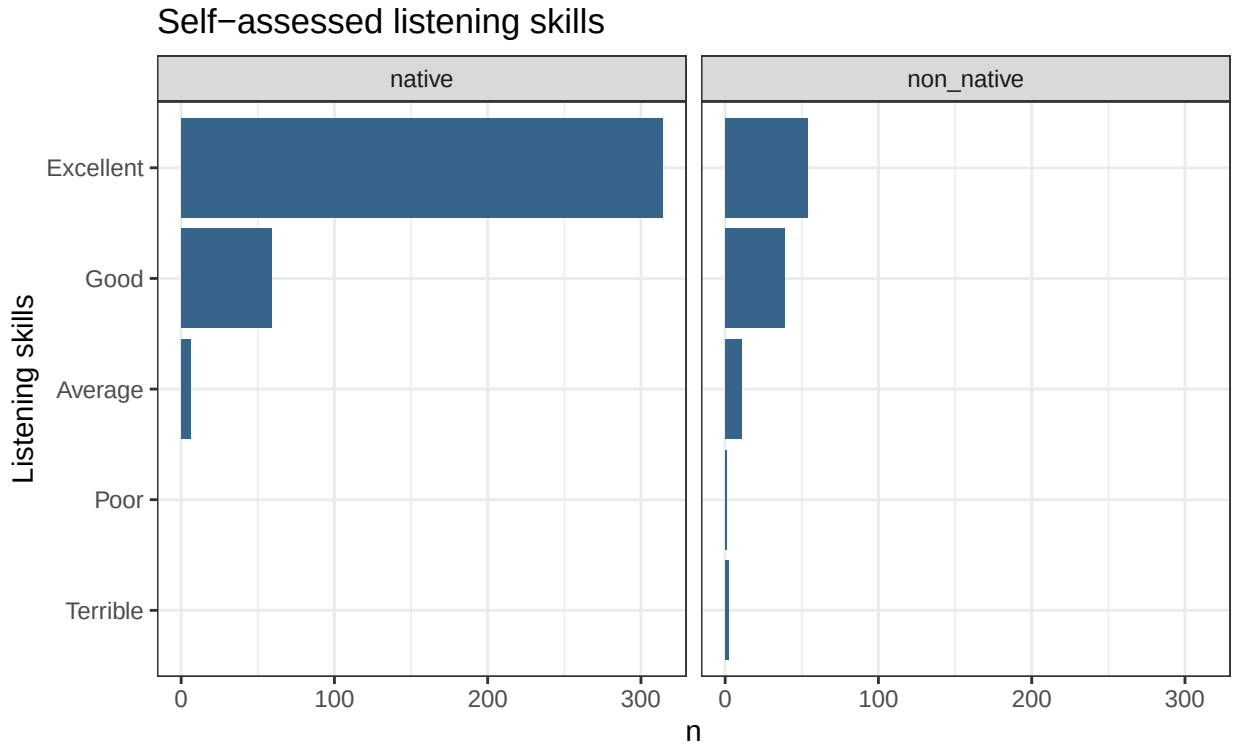


Figure 22: Listening Skills

The last skill surveyed was writing. Figure 23 shows the count distribution. Overall, participants appear more critical of their writing skills. Both for native and non-native speakers, the ratio of excellent to good is decreased. In addition, more people rated themselves as average for writing than they did for speaking and writing.

To summarize, the speaking, listening and writing data collected shows a similar distribution, both within and across nativeness. That is to say, both native and non-native

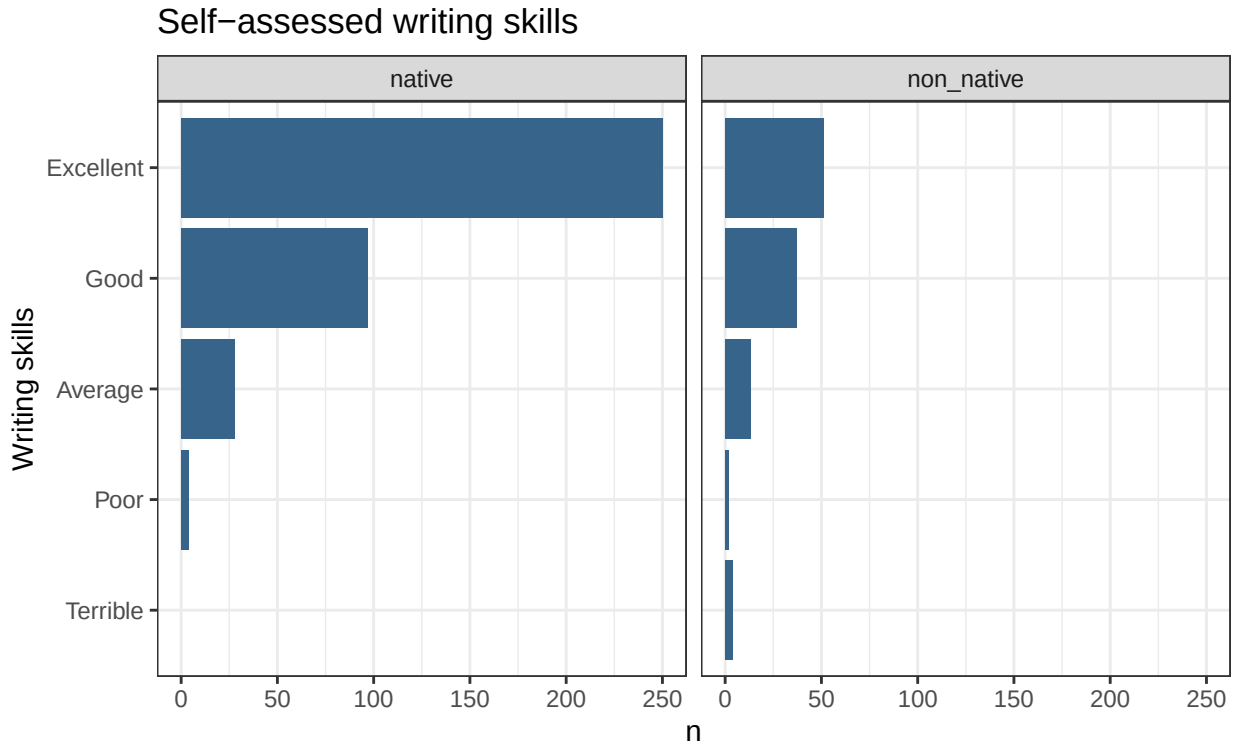


Figure 23: Writing Skills

speakers rated their skills consistently toward the highest rating of the scale. For non-native participants, however, the ratio of excellent to good is smaller than for natives. It appears that non-native speakers are more critical of their proficiency in their L2 performance, which isn't necessarily a surprising finding.

3.3.4.6 Education The very last piece of demographic information collected was education level. This was also self-reported. Given the way the study was set up and the distribution channels used, it is not surprising that a significant percentage (90.12%) of the participants in the sample reported college level education or higher. Figure 24 shows the breakdown of the sample population. A notable difference between the native and non-native participants is that most non-native participants have either a masters or a doctorate. For the native participant, college is the most common answer.

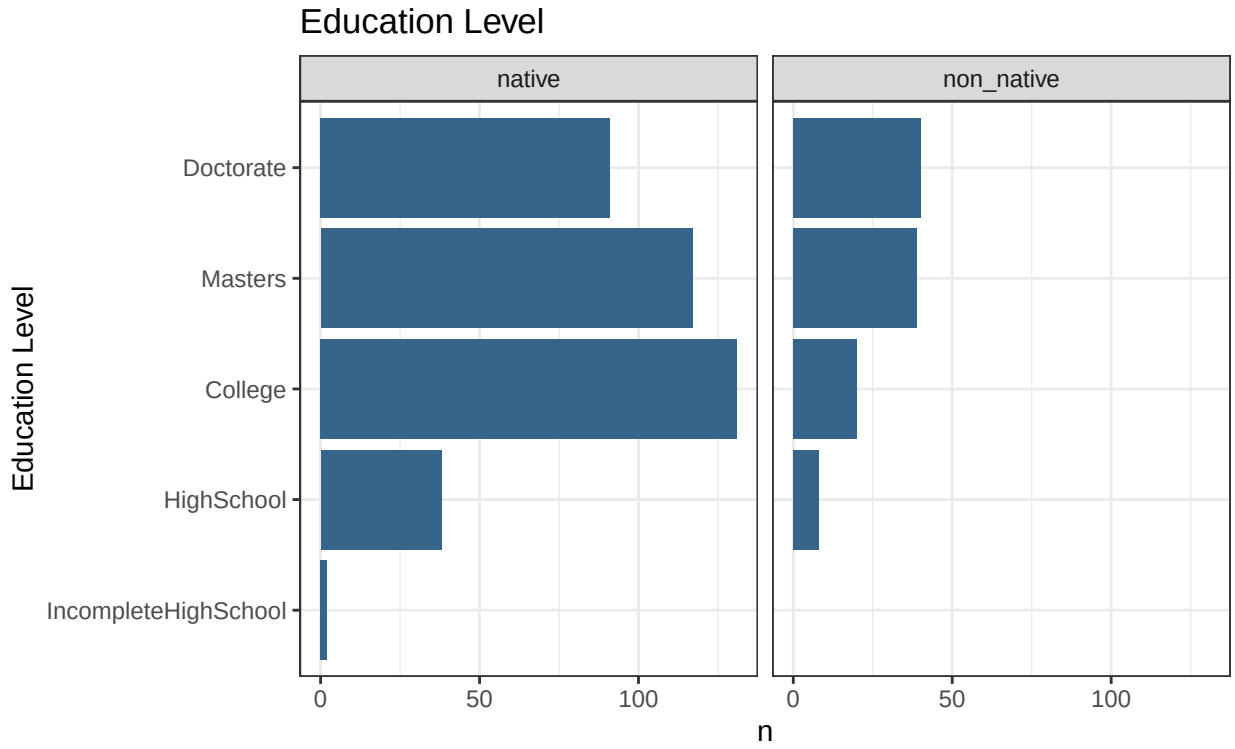


Figure 24: Education Level

3.4 From Participants to Observations

In section 3.3.4, I provided a description of the population sample that participated in the perception experiment. The main takeaways were the size of the population ($n_{population}$ = 486), the gender distribution (n_{men} = 41.15% and n_{women} 57.61%); and the age distribution, where the majority of the participants (97.33%) are between the ages of 20 and 69.

Looking at the data set from the point of view of participants isn't the only way to process the results; in fact, it is probably the least conducive way to organize the data for analyses that can reveal existing patterns within the data. The reason behind this assertion is that each participant contributed more than one observation. In other words, the data needs to be reshaped from a participant-based *wide-format* to an observation-based *long format*, where participants become one of the explanatory variables following what is known as *tidy data* (Wickham 2014).

The data set thusly transformed consists of 7772 observations and 19 explanatory variables, or 147668 data points. In addition to the explanatory variables listed in Section 3.3, the entire data set shaped into a long format contains:

1. progress: a percentage indicating how much of the entire experiment each participant completed. For the perception experiment, all participants have a 100% completion rate.
2. duration in seconds: this indicates the time lapsed from the start of the study to the end.
3. participant id: a factor with 486 levels. The id number was sequentially assigned to the participants.
4. standard country of residence: a factor indicating the country of residence.
5. standard childhood variety: a factor indicating the variety spoken as a child.
6. standard adult variety: a factor indicating the variety spoken as an adult.
7. speaker: a factor indicating the speaker from *Experiment 1* who produced the input sentence.
8. statement: a factor indicating the statement being evaluated.
9. intent: a factor indicating the intended interpretation the speaker had for each statement.
10. set: a factor indicating if the input was presented following *Set A* or *Set B*.
11. response: a factor indicating what each participant believes the intent of the speaker was.

3.4.1 Quality of the data

There are two potential issues with the duration data: too short and too long. Too short (people taking less than 6 minutes is just 2% of the sample), so it might not be a problem. Upon initial inspection of the data, too long did seem like a problem. Table 19 below shows summary statistics for duration in minutes. Table 19 shows that the data is

very much skewed towards people taking way longer than what the average person took to complete the study.

Table 19: Summary Statistics for Duration in Minutes

duration
Min. : 4.00
1st Qu.: 11.97
Median : 17.82
Mean : 101.75
3rd Qu.: 29.42
Max. :5550.90

I suspect the issue occurred each time a participant failed to submit their response. After a week, Qualtrics automatically closes that participant's submission and auto-submits it, which of course distorts the duration that a typical participant takes to complete the study.

Figure 25 shows the duration distribution scaled to those values shorter than 3 times the median. It looks pretty normal, but not symmetric, although there is a likely reason as to why it cannot be symmetric: there are no participants who took zero seconds in completing the survey. An unscaled plot would show a very long tail for the reasons already discussed.

Other than those two potential issues, the data does not suffer from the concerns seen in the production experiment given that most of the data produced are categorical. The data that are not categorical, are unlikely to have any bearing on the responses provided by the participants.

3.5 Analysis

Initially, I wanted to approach *Experiment 2* by looking at the results using a wide lens angle in order to tell how well participants estimated the speaker's intent in each of

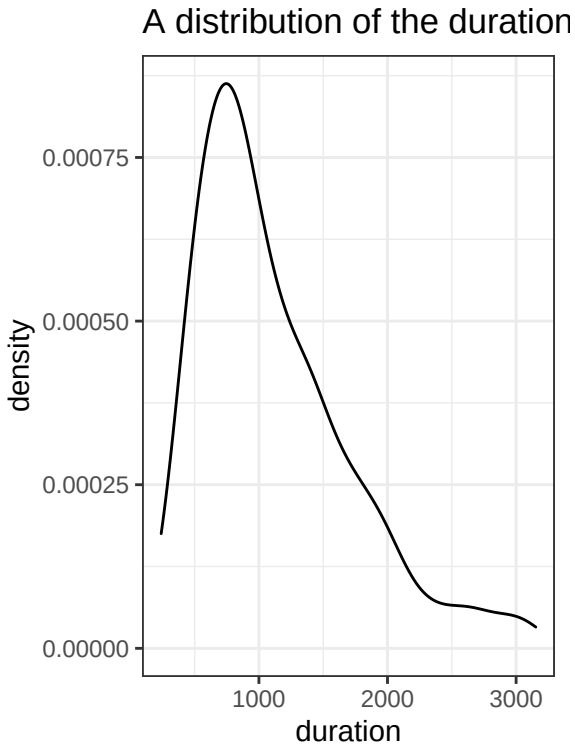


Figure 25: Duration to survey completion in minutes

the context-less sentences. The set up is that we have two variations of the same sentence. The null hypothesis is that participants are unable to tell apart the two variants (literal and idiomatic) of the sentences. In other words, people can hear no difference in the way sentences are produced when the only distinction between those sentences is the intent the speaker had in mind when they produced the sentences.

The alternative hypothesis is that they do. Based on *Experiment 1*, we have some evidence that, in fact, speakers make a distinction in how they produce their sentences based on their intent; idiomatic sentences are shorter in duration and they have a narrower F0 range compared to their literal counterparts. The alternative hypothesis is that participants *are able* to pick up on these differences and correctly perceive what the intent of the speaker was based solely on the acoustic differences present in the speech signal.

To validate this hypothesis, *Experiment 2* was carried out. In it, n = 486 participants tried to determine the intent of 12 speakers. Each speaker produced 4 sets of identical

sentences that were ambiguous in their interpretation. The input for each participant consisted of the isolated sentences, stripped of context, and randomized in the order they were presented. The results can be seen in Table 20.

Table 20: Results for Entire Population Perception Experiment

intent	idm	lit
idm	2014	1872
lit	1930	1956

Table 20 should be read this way: the rows display the *intention* the speaker had in mind at the time of production; the columns display what *intention* the participants indicated for each sentence they heard. Table 20 shows that *Experiment 2* participants' perceptual judgments are overall correct more often than not. The margin seems to be razor thin, however.

To get an objective sense of how successful participants are at perceiving the speakers' intents, I applied a McNemar test on these tabular results. The McNemar test is a non-parametric test for paired nominal data. Nominal data, also known as categorical data, corresponds to data that can be represented by categories. Specific to this experiment, the intent of the speakers and the response of the participants were both recorded as *literal* or *idiomatic*.

A McNemar test is commonly used to determine if there are differences to be found on a dichotomous dependent variable between two related groups. A McNemar test is similar to the t-tests carried out in the previous chapter, except that the variable is dichotomous in nature and not continuous. In the context of *Experiment 2*, the dependable variable was intent. The dichotomous outcomes are *literal* and *idiomatic*. The two related groups correspond to the speakers from *Experiment 1* who produced the statements and the participants in *Experiment 2* who estimated intent.

Assumptions for a McNemar test:

- There is one nominal variable with two categories - this assumption is met. The nominal variable being *intent* and the categories being *literal* and *idiomatic*.
- The two groups in the observations must be mutually exclusive. We meet this criterion, as sentences had to be judged as literal or idiomatic through the forced-choice experimental design.
- The sample must be a random sample. This assumption is met as participants were asked but not forced to participate in the study.

Table 21 shows that when the entire sample is considered, participants are not great at discerning the speakers' intent. In other words, with a p-value = 0.36, there isn't sufficient evidence to reject the null-hypothesis and entertain the idea that participants are able to tell apart the intent of the speaker.

Table 21: McNemar Results on the Raw Scores

Statistics	p_value	Method
0.8546	> 0.05	McNemar's Chi-squared test with continuity correction

3.5.1 Literal heads or idiomatic tails or how this linguistic experiment isn't different from a coin toss

I designed *Experiment 2* around one interaction: have a participant listen to a sentence stripped of context, and then ask the participant, through a forced-choice question, to decide what kind of event the speaker was picturing when they produced that sentence.

Recall from *Chapter 2* that the consensus tells us that each identical pair of sentences in the stimuli is, well, identical. Nothing in the literature, with the notable exception of Van Lancker and Canter (1981), indicates that once context is removed people would not be able to tell apart which sentence corresponds to which intent. Assuming that people

are exclusively guessing, there are four possible outcomes, two of which would constitute a successful guess, as illustrated below in Figure 26.

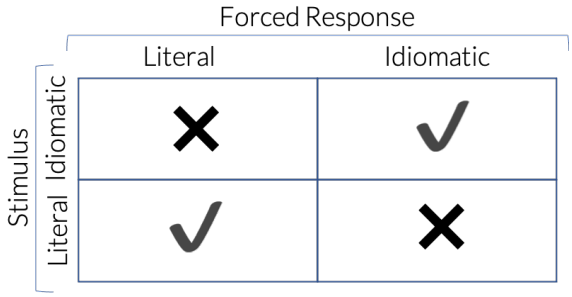


Figure 26: Experiment 2: Possible Outcomes

Looked at from that perspective, the job I asked the participants to carry out isn't that different from asking them to guess if a coin will land heads or tails, or in this situation, 16 coins, to be exact. By all accounts, barring the presence of something in the sentence itself, the participants should be estimating the intent following a 50-50 chance, just like a coin toss.

In order to build a baseline for comparison, I carried out what 16 fair coin tosses would look like in order to calculate what the probability for success along an incremental line from zero correct guesses all the way to 16 correct guesses would be. Such theoretical values come from the binomial distribution calculating the probability from 0 to 16 successes, assuming 16 trials, and a probability of success of 50%. Each of those probabilities multiplied by the total number of participants (485) would result in the expected correct guesses *if* the participants were guessing, as the literature suggests.

Rounding down to the integers, the expected results would look like what we observe in Table 22 below. The numbers on the top row represent the possible number of correct guesses, the numbers on the bottom row are the expected numbers to find within *Experiment 2's* parameters.

Table 22: Expected Number of Successes in 16 trials

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	2	14	66	216	517	948	1355	1524	1355	948	517	216	66	14	2	0

What Table 22 shows us is that most people would be guessing 8 correct sentences by chance. Scaled to our sample population, out of 7760 observations, we should expect to see 1524 observations to correspond to 8 correct estimations. The numbers decrease symmetrically, with 1355 observations correctly estimating intent 7 or 9 times, and so on towards the extremes of the scale, with zero observations making all correct estimations or estimating incorrectly all 16 times. This is represented below in Figure 27, a symmetrical bar plot centered on 8 correct guesses and tapering down toward all correct responses or all incorrect responses.

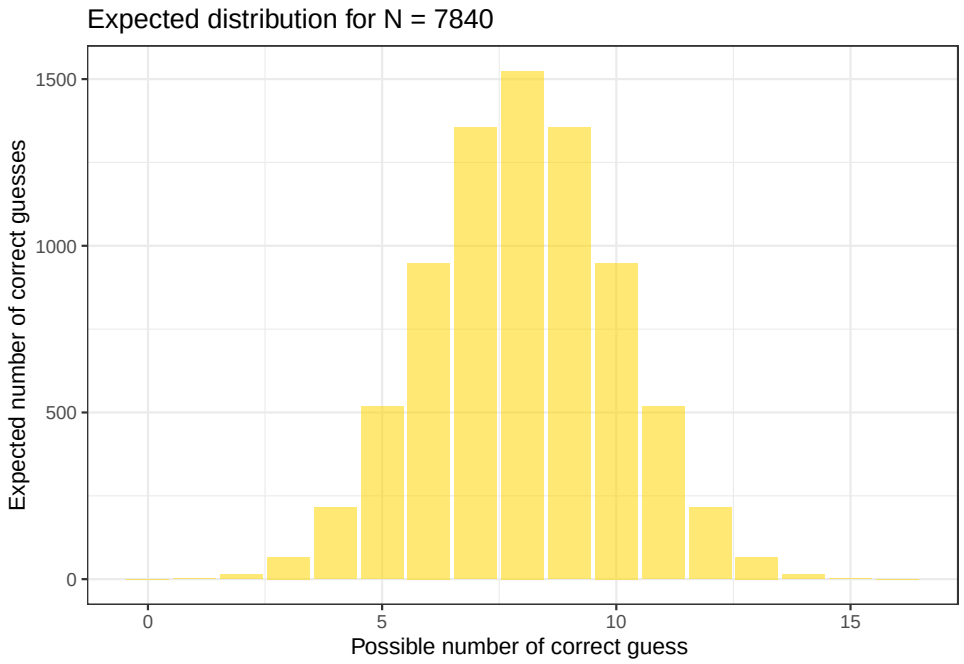


Figure 27: Theoretical Binomial distribution

The intuition behind this is easy to follow, even though it is very unlikely to be able to guess all 16 coin tosses correctly; the larger the number of times that we attempt to

make all those guesses, the more likely it is that someone would guess them all correctly.

Moving from the theoretical coin tosses to the experimental judgments made by the participants, we see in Table 23 that the numbers don't quite match. The frequency distribution in Figure 28 reveals that fewer observations are located around the center of the distribution. That means that more observations are correct than expected by chance, but also that there are more incorrect observations than what chance would predict.

Table 23: Observed Number of Correct Estimations in 16 sentences

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
16	0	112	80	432	512	816	976	1408	1008	992	720	416	128	80	32	32

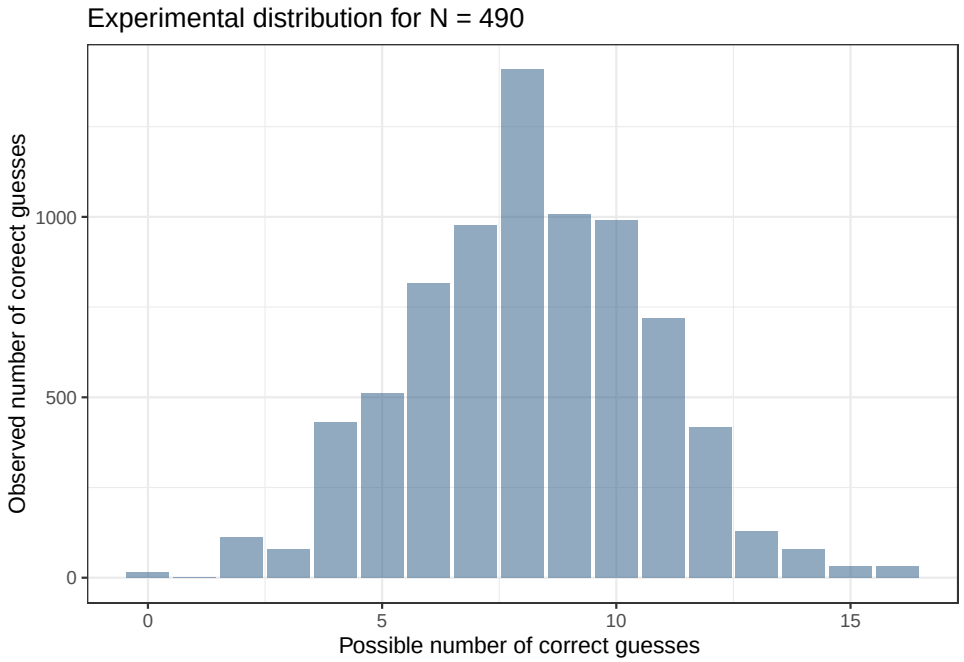


Figure 28: Experimental distribution

The next step is to figure out if the differences can be attributed to chance. A rough initial comparison between the two distributions done through a chi-squared test indicates that the two distributions are *not* the same (with a p-value of 2.2e-16). In other words, it is astronomically unlikely that the observed values are what they are by chance. Looking

at the two distributions superimposed on each other (see Figure 29), it becomes clearer that the observed distribution (in blue) is skewed to the right. In other words, participants are guessing better than a 50-50 chance would allow. The next step is to fine tune what is causing people to guess better or worse than chance would dictate.

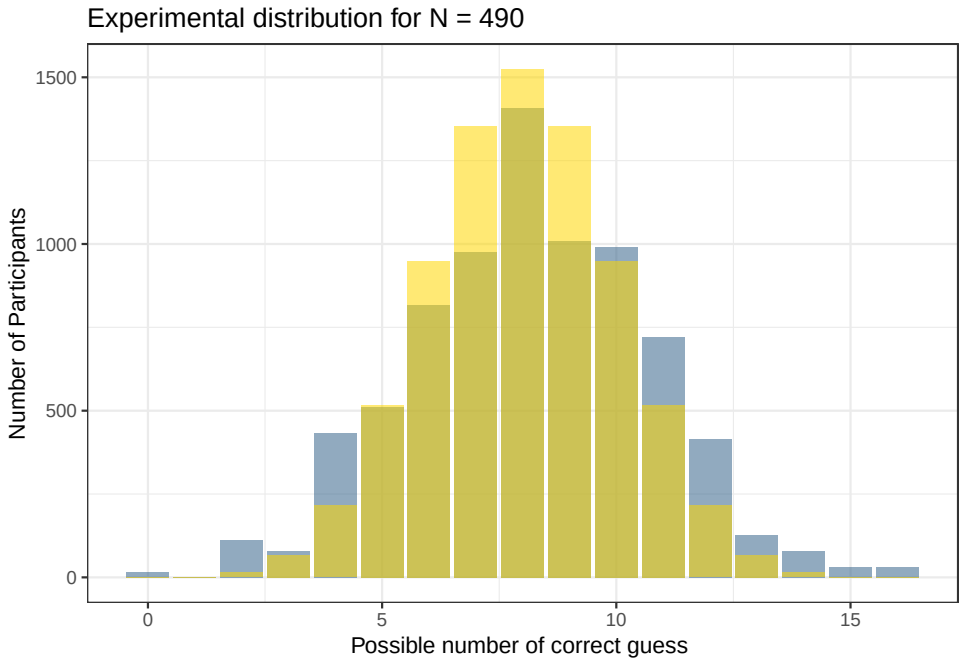


Figure 29: Experimental distribution

Comparing the number of expected results to observed results per successful guess category, I found 344 worse-than-expected-observations and 637 better-than-expected-observations, which means that, overall, people are judging correctly the intent of speakers better than chance would otherwise indicate.

3.5.2 Significant factors in estimating (idiomatic) intent

Recall that when looking at the bigger picture, the results of *Experiment 2* were less than encouraging. In general, people do not appear to be able to estimate speakers' intent when the sentence was isolated from its context. Recall also that when looking at it from the perspective of participants guessing the intent of speakers across the number of

possible correct guesses, I found that a number of participants were more successful than chance at estimating the intent of the speakers they were paired with.

In future iterations of this study, the next logical step would be to zoom in on the subset of participants who estimate intent with a higher success rate than chance and cross-reference those with the speakers who are better at getting their intent across. However, moving forward, I would not work with the subset of participants who are better-than-chance guessers. Instead, I would continue to work with the entire population in order to identify below what factors appear to influence how successful participants are at estimating idiomatic intent.

3.5.3 Explanatory Variables and their effect on estimating speaker intent

In this section, I examine the effect that each of the recorded explanatory variables had on predicting the intent of the speaker. In order to quantify the impact each variable had, a model was fit using logistic regression, which is a form of regression analysis used when the independent variable is dichotomous in nature. In order to satisfy this requirement, I created an indicator, *correct response*.

Correct response is a logical indicator that is set to TRUE when the the participant's response matches the intent of the speaker for each observation, and FALSE otherwise; in other words, when the participant correctly guessed what the speaker's intent was. In the process of fitting the model, I probed the relationship between *correct response* and the various explanatory variables. Those results are described below.

3.5.3.1 Age As was mentioned earlier, the population sample was very diverse when looked at from the perspective of age; however, age does not seem to be a significant factor in explaining the number of correct responses each participant estimated. Figure 30 shows a mostly constant line hovering around 8 correct guesses across the age variable.

The result of fitting the data to a logistic regression model `glm(correct_response`

`~ age,data, family = binomial)` corroborates that age is not a significant factor, as shown in Table 24. The information displayed in Table 24 is the standard output for a regression model. Below, I describe the information that these values convey. This information applies to all the regression output tables in this chapter.

1. 2 rows identified as *intercept* and the variable in question. For Table 24, the variable is the age of the participants.

2. Estimate: In many instances the value of the intercept isn't interpretable, but it is needed for the model to work. In the simplest sense possible, the intercept here represents how close to a successful guess a person of age zero would be. For a unit increment of age, the odds of successfully estimating intent increases by 0.0032.

3. Standard error: The standard error of a regression indicates the average difference of the observed values from the regression line. It indicates how wrong the regression model is on average using the units of the response variable.

4. z-value: It is the particular statistic score computed by the model. It corresponds to the Estimate divided by the Standard Error. This z statistic is used to compute the p-value.

5. p-value: Arguably the first value most people look at when deciding how good the model is. A low p-value is preferable and it represents how likely it is for the data to be due to chance. A smaller p-value corresponds to a smaller possibility of encountering the data distributed the way it is exclusively due to chance.

Table 24: Age logistic regression results

	Estimate	Std. Error	z-value	p-value
Intercept	-0.0674	0.0765	-0.8811	> 0.05
Age	0.0028	0.0019	1.5143	> 0.05

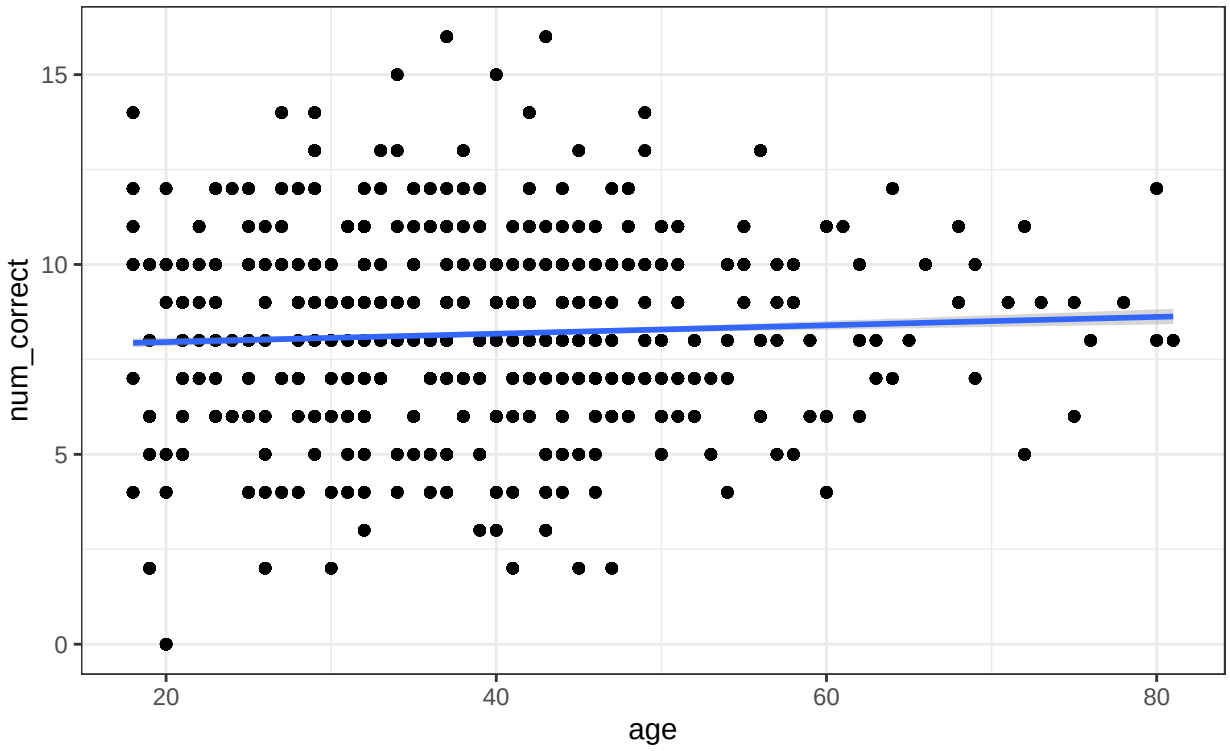


Figure 30: Age regression plot

3.5.3.2 Gender Recall the sample population was divided into four categories: female, male, non-binary, and no response. The distribution by gender is shown in Table 25. Given how small the percentages of participants outside the traditional male-female categories are, the analysis below will exclude the non-binary and no-response categories.

Table 25: Age Distribution in the Sample

gender	Freq
male	41.19
female	57.73
non-binary	0.82
no response	0.41

Using those metrics, after fitting the data to a logistic regression model `glm(correct_response ~ gender,data, family = binomial)`, the results in Table 26 indicate that gender is a significant factor in estimating speaker intent with women

being better at estimating intent than men are.

Table 26: Logistic regression results for Gender

	Estimate	Std. Error	z-value	p-value
Intercept	-0.0263	0.0354	-0.7429	> 0.05
Gender[female]	0.1138	0.0463	2.4572	< 0.01 **

An interesting point regarding age is that despite age not being a good predictor individually, when considered in an additive model using age and gender, the model tells us that older participants are better at predicting intent than younger participants. In addition, female participants become better at estimating intent, but its significance remains within the < 0.01 threshold. The results are shown in Table 27.

Table 27: Logistic regression results for Age and Gender

	Estimate	Std. Error	z-value	p-value
Intercept	-0.1879	0.0856	-2.1936	< 0.05 *
Gender[female]	0.1259	0.0467	2.695	< 0.01 **
Age	0.0039	0.0019	2.0719	< 0.05 *

3.5.3.3 Nativeness Table 28 shows that for the participants in *Experiment 2*, being a native or a near-native speaker of Spanish did not make a difference in their ability to estimate idiomatic intent.

Table 28: Logistic regression results for Nativeness

	Estimate	Std. Error	z-value	p-value
Intercept	0.0574	0.0257	2.2349	< 0.05 *
Nativeness[non-native]	-0.0645	0.0547	-1.1773	> 0.05

3.5.3.4 Communicative Skills and Education As mentioned in earlier sections, the population sample was pretty homogeneous as far as how participants rated themselves in

their communicative skills. Regarding their education, most participants indicated they held a college degree or higher. None of the education levels or communicative scores appear to be good predictors of idiomatic intent.

Table 29: Logistic regression results for Communicative Skills and Education

	Estimate	Std. Error	z-value	p-value
Intercept	-0.2468	0.1507	-1.6374	> 0.05
Listening	0.0152	0.0696	0.2188	> 0.05
Speaking	0.0415	0.076	0.5453	> 0.05
Writing	0.009	0.0565	0.1593	> 0.05
Education Level	0.0194	0.0247	0.7873	> 0.05

3.5.3.5 Geographical and dialectal factors None of the geographical or dialectal predictors appear to have any bearings on the participants' ability to estimate speaker intent. Although I am aiming for parsimony in the evaluation of all my predictors, the results of carrying out the analysis for all the levels were very long tables that were, ultimately, uninformative. For those reasons, I chose not to include them here.

3.5.3.6 Stimuli Input Format Discriminating how the input was presented to participants was motivated by Van Lancker and Canter (1981), who found that it made no difference for their participants whether they had the opportunity to listen to their ambiguous sentences being contrasted before they had to judge them or simply proceeded to judge one member of the pair. I replicated this distinction in the input of *Experiment 2*.

Recall that participants saw their sentences in two different ways, which I named *Set A* and *Set B*. In *Set A*, participants had the opportunity to listen to the speaker they were paired with contrast the two members of the pair before they were asked to judge each member in random order. *Set B*, on the other hand, asked the participants to directly judge a member of the pair without the benefit of listening to the two contrasting sentences first.

After a visual inspection, it appears that participants who interacted with *Set A* did better at estimating the intent of the speaker they heard, as depicted in Figure 31. *Set A* participants did better for 10, 12, 13, and 14 guesses.

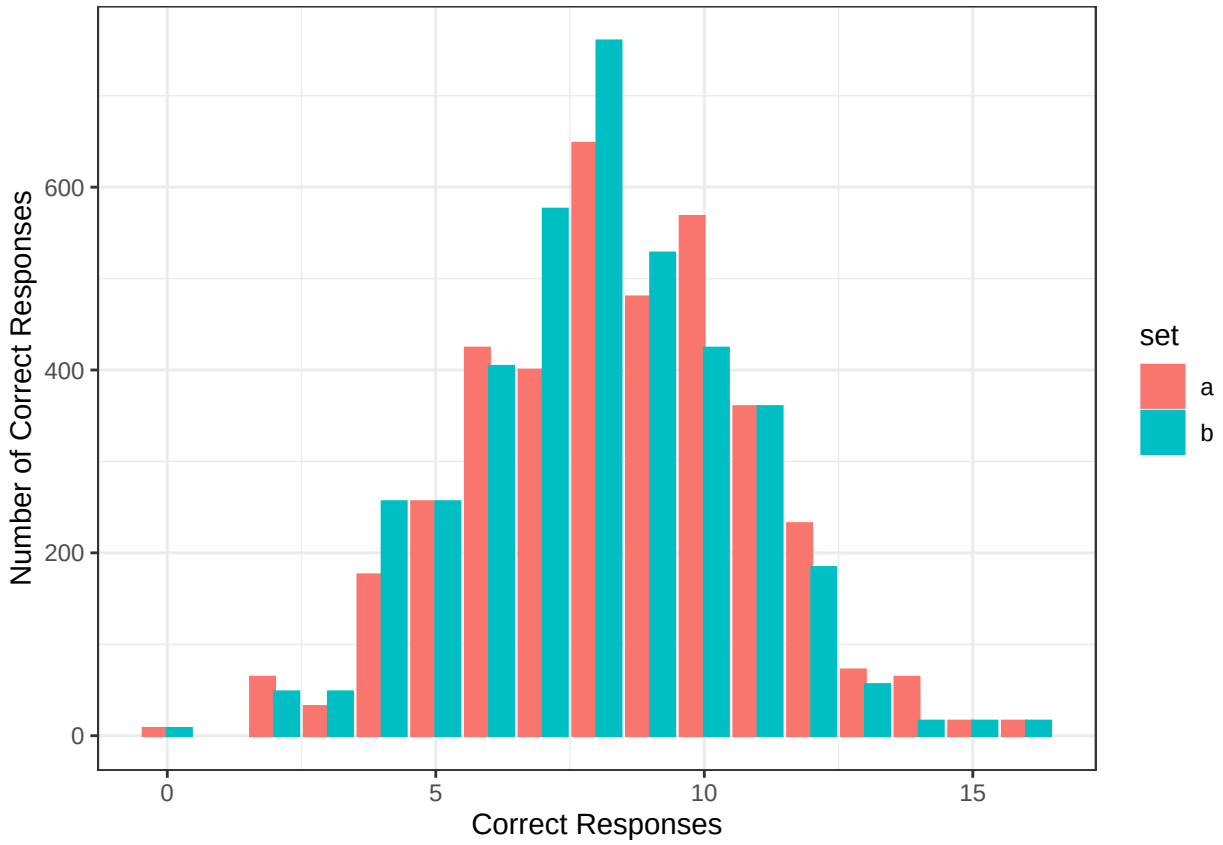


Figure 31: Regression Plot for Set

However, after fitting a logistic regression model `glm(correct_response ~ set, data,family = binomial)`, participants who interacted with *Set B* did not perform significantly different from *Set A* participants, as shown in Table 30.

Table 30: Logistic regression results for input set

	Estimate	Std. Error	z-value	p-value
Intercept	0.0755	0.0324	2.3305	< 0.01 **
Set B	-0.0634	0.0454	-1.396	> 0.05

3.5.3.7 Duration and F0 range differences The last set of variables I used to fit the model was the differences in F0 range and duration between the literal and the idiomatic sentences. Recall from *Experiment 1* in *Chapter 2* that both duration and F0 range were found to be significantly different between the two intents. In fact, it was this difference that drove the perceptual experiment; it was just one half of the picture to find out that speakers produced systematically shorter idiomatic sentences with narrower F0 ranges. The perception experiment took a stab at determining if native and near-native speakers were capable of picking up on those differences to determine intent.

Van Lancker and Canter (1981:68) concluded that idiomatic phrases were characterized by "shorter durations" and "fewer or weaker pitch-accents (contributing to a relative smooth intonation contour.)" I take their findings as compatible with my findings from *Experiment 1*, despite the fact that Van Lancker and Canter did not measure F0 range directly.

Ideally, the participants would be able to perceive these differences in duration and F0 range and use them to estimate speaker intent. The model tested if this was indeed the case. Fitting the model with these individual predictors reveals that only F0 range is statistically significant for the entire sample.

This model considering both predictors `glm(correct_response ~ dur_diff + range_diff, data,family = binomial)` is shown in Table 31.

Table 31: Logistic regression results for Duration and F0 range differences

	Estimate	Std. Error	z-value	p-value
Intercept	0.0472	0.0252	1.8748	> 0.05
Duration Difference	-0.0212	0.1064	-0.1995	> 0.05
F0 range Difference	-0.0019	6e-04	-3.1904	< 0.001 ***

3.6 Conclusions

Although the present experiment does belong under the umbrella of *perception* studies[10], it differs significantly from the majority of Spanish perception studies in that the phenomenon I am trying to investigate does not live in a segment or correlate with many sociolinguistic traits. The phenomenon in question is, of course, the estimation of idiomatic intent. For the purposes of this experiment, this estimation was narrowly defined as a cue in the speech signal that alerts the interlocutor that the speaker does not wish that their sentences be interpreted literally.

Recent literature on Spanish intonation challenges the idea that literal and idiomatic interpretation are acoustically identical. Idiomaticity is understood to be resolved in the semantics/pragmatics phase of interpretation when the concomitant context is considered. It follows that if context is removed, interlocutors would have no additional resources to disambiguate the speaker's intent - or - as it has been suggested (Van Lancker and Canter 1981:68), they will have to default to the idiomatic interpretation given the high frequency co-occurrence of these collocations given that phrases of this type are more frequently used with idiomatic interpretations in mind.

Neither of these two premises is borne out by the data collected in this experiment; that is, more participants predict intent better than chance would allow for and there is no evidence that participants default to an idiomatic interpretation when their estimation is incorrect.

The number of observations defaulting to idiomatic when the interpretation was literal is 1958, whereas the number of observations defaulting to literal when the interpretation was idiomatic is 1890. The difference between those two weighted by the total number of observations is approximately 1.5%, hardly an indication that when wrong, people are defaulting to one interpretation over the other.

[10]See Chappell (2019) for a review of current issues related to perceptual studies in Spanish

As alluded to at the start of this section, only one of the individual sociolinguistic variables yielded statistically significant results. This is surprising given that other studies focusing on intonation have found differences along some of these variables; for example, both Seddoh et al. (2020) and Ben-David et al. (2019) found age to be a relevant factor. Considered in tandem, however, when I modeled age and gender in an additive effects model, it did result in statistically significant differences, as Table 27 showed.

The one variable that did yield statistically significant results was gender, with women being better than men at estimating idiomatic intent. Although the sample did include some participants outside the traditional male-female roles, I chose to exclude them from the analysis on the grounds that there were too few of them and they were *very* good at estimating idiomatic intent. As a result, their presence in the sample distorted the model considerably. Current research on intonation and gender revolves heavily around the idea of gender roles, their correlation with speech patterns, and perceptual judgments associated with one's intonational profile. See Hancock et al. (2015) and Henton (1995) for additional discussion of the topic along those lines.

This finding about gender begs for additional research in order to gauge how strong this effect is. Earlier in the literature review, I mentioned that the received wisdom from sociolinguistics is that women are the driving force behind linguistic change (Labov 1990). It is also known that women manipulate their F0 into broader excursions that men do for biological reasons (Traunmüller and Eriksson 1995), but also for contextual reasons McConnel-Ginet (1978). Hansen Edwards (2008:254) explains that when looked at from a purely biological standpoint, gender "does not seem to be a significant factor in L2 pronunciation accuracy," but social circumstances associated with gender may be conducive to opportunities that foster better instruction, for example. In other words, the effect may not be directly associated with gender but rather with favorable situations that people of that gender commonly experience. I wonder if the same can be said of women being adept at manipulating pitch over wider ranges, as Henton (1995:42) reports. It may be the case

that female speakers' manipulation of pitch has the concomitant effect of making them better at perceiving it.

I expected the variable I called *nativeness* to have some effect in explaining the variation in the correct estimates. In the literature, a milestone of near-native competency is the appropriate use of idiomatic phrases, as argued for in Vanderniet (2015). Alas, nativeness does not seem to play a role in the ability of participants to correctly estimate the intent of the speaker they were paired with. Given that the criteria for inclusion in the study was to be a native or a near-native speaker of Spanish (each participant decided if they qualify or not), it is not entirely surprising that nativeness ended up not being significant, as all participants were highly proficient in their use of Spanish. This insight does leave a door open for future research where non-native speakers can be asked to perform a similar task in order to gauge their accuracy in predicting intent. If increased proficiency correlates with increments in accuracy predicting idiomatic intent, then idiomaticity may be used as an indicator of proficiency.

Whether or not participants had the opportunity to hear a speaker contrast the two sentences before they were asked to estimate the intent did not result in significant differences either. These results match what Van Lancker and Canter (1981:67) found in their experiments: "the pairing of ditropic sentences is not a necessary condition for correct identification." My results corroborate that the cues present in the speech signal seem to be sufficient for a subset of the participants to estimate intent.

In the *Analysis* section, the entire population sample was split into average guessers, worse-than-chance guessers, and better-than-chance guessers. The rationale behind this decision revolved around the way the experiment was modeled. A contingency matrix (Figure 26) revealed that asking participants to estimate intent from a sentence stripped of context is not different from asking them to guess the outcome of a coin toss; both phenomena follow a binominal distribution with a 50-50 chance of guessing correctly. To ascer-

tain the effect of duration and F0 range in predicting intent, a multiple logistic regression model was fit using both differences in duration and differences in F0 range to account for the binary outcome of the *correct response* variable.

Correct response was a derived variable set to TRUE when the intent of the speaker and the estimation of the participant matched, FALSE otherwise. The binary nature of *correct response* made a logistic regression modeling a natural choice to examine the data. Choosing F0 range differences and duration differences was motivated by the results of *Experiment 1* because both variables were found to be statistically significant when literal sentences were compared to their idiomatic counterparts.

The logistic regression model showed that only F0 range differences were statistically significant for the entire population sample. This was a surprising result given that we knew from *Experiment 1* that speakers exhibited statistically significant differences both for duration and for F0 range between their literal and idiomatic realizations.

The general results, although promising, are not without limitations; for one, even though there are more better-than-chance participants at estimating intent, the experiment revealed that there are also worse-than-chance participants. At this moment, I don't have a plausible rationale to explain why some people are worse at guessing. I did show, however, that this distribution is extremely unlikely to be caused by chance. The chi-squared test conducted on the raw counts indicated that the odds of observing people guessing the way they did due to chance are infinitesimally small.

Another issue with the analysis is that F0 range as an explanatory variable, despite being significant, does not manage to be *very good* at explaining the success rate of the better-than-chance participants. What that means is that, at the end of the day, prosodic cues are not a deterministic factor in conveying idiomatic meaning. In other words, context continues to be required and prosodic cues may be just there to nudge the interlocutor in the right direction.

The last issue with the study as it was conducted is that I failed to control for a variable that in hindsight could have a significant impact on how successful someone is at estimating idiomatic intent: familiarity with the idiom. I should have asked each participant how familiar they were with the idioms in the study prior to the their participation. It seems reasonable to assume that one would be more adept at identifying intent with phrases one has used or heard before.

The main conclusion of this chapter, considering all the relevant factors, is that a subset of the participants was successful at estimating the intent of the speakers they were paired with. With this evidence in hand, the natural next step is to challenge the received wisdom that literal and idiomatic realizations of otherwise identical utterances are phonetically the same. Future research leading to the identification of the factors on which the successful participants rely on to tell idiomatic intent can, in turn, lead to discard this premise at least for a subset of the Spanish-speaking population.

In the next chapter, I explore the differences in the aspectual structure of literal sentences and idiomatic sentences by asking participants to render grammatical judgments on expressions of the familiar subject-verb-object type that have been the focus of the previous two chapters.

4 Chapter 4: Experiment 3 — Acceptability judgments

4.1 Introduction

The previous two chapters and their accompanying experiments were focused on establishing if there were acoustic differences between sentences interpreted compositionally and the same sentences interpreted idiomatically. In *Chapter 2*, based on the speech samples produced by a pool of 13 speakers and a corpus of 20 sentences, I concluded that the speakers produced idiomatic sentences that were significantly shorter and had significantly narrower F0 ranges.

In *Chapter 3*, I sought to verify experimentally if native and near native participants could pick up on the cues present in the speech signal to estimate the intent of speakers. Using the sentences created in *Chapter 2*, I stripped them of their associated context in order to use them as stimuli for the participants of *Experiment 2*. I proceeded to ask the participants to estimate the intent of the speaker producing the sentences they were paired with in order to gauge how successful they were at estimating intent. I concluded that a subset of the 480 participants were able to estimate the intent of the speaker from the sentences they heard well beyond what chance would allow for.

No demographic variables, with the exception of gender, played a role in accounting for successfully estimating intent. Women in the sample were statistically better at estimating intent than men were. Of the two linguistic variables considered, only F0 range was a statistically significant predictor of intent, although the overall effect did not seem to be strong enough to conclude that speakers could rely exclusively on F0 range to tell apart idiomatic intent in a reliable manner. Identifying specifics about the mechanics of F0 in idiom recognition, as well as details about what makes a speaker more successful at getting their intent across, are clear avenues for future research.

This chapter marks a departure from the experimental line of inquiry pursued in

the previous two chapters by shifting the focus to the differences present in the aspectual structures of the events denoted by the two interpretations. Recall that throughout the previous chapters, the subject of the research has been a complete sentence composed by a verb and its object with an ambiguous interpretation oscillating between literal and idiomatic. *Chapter 4* seeks to probe if there is a predictable relationship between the aspectual structure of the literal event and that of the idiomatic event.

The way these two events stand in relationship to one another has been a source of continuous debate. The opinions range from linguists (Katz and Postal 1963; Marantz 1996:5; O'Grady 1998) who think that idiomatic meaning — and its concomitant aspectual structure — is completely idiosyncratic, to some researchers who argue that the aspectual structure of the idiomatic event is completely dependent on the aspectual structure of the literal event (McGinnis 2002, 2005). Other researchers argue for a happy medium where the aspectual structure of the idiomatic event is partially determined by the aspectual structure of the literal one (Glasbey 2003; Glasbey 2007).

In the course of creating the corpus and choosing which idiomatic phrases to use for the various tasks I prepared for my participants, I had to test what aspectual class each of the literal and idiomatic sentences belong to. These aspectual classes, as well as the tests devised to determine aspectual membership to a class, are largely based on the work of Vendler (1967) and Dowty (1972). In the briefest way possible, aspectual class refers to a classification of verbal predicates based on homogeneous behavior within the grammar of the language. These aspectual classes presuppose the presence of aspectual features within the structure of the event. The tests themselves consist of situational contexts where the resulting readings are taken as evidence of the presence of these aspectual features at some level of representation.

According to the most accepted taxonomy, although there are critics (Walková 2012) and revisions (Smith 1991), there are four aspectual classes: states, activities, achieve-

ments and accomplishments. The order in which I list the aspectual classes will become relevant in this chapter. To the best of my knowledge, aspectual classes have not been looked at from a feature-based analysis the way I approach features here.

If we assume that differences between classes can be captured by features, then it seems natural that the accumulation of aspectual features allows for a hierarchy of sorts among the aspectual classes. States form the simplest predicates with no aspectual features to speak of. Successive addition of aspectual features leads to activities by adding dynamicity, achievements by adding telicity, and accomplishments by the addition of duration.

It should be noted that the "addition" of aspectual features needs to happen in that specified order (duration > telicity > dynamicity). It does not seem possible for a verbal predicate to denote a telic event if that event is not dynamic also. The same holds true for an event exhibiting duration; it also has to be telic. The inverse relationship does not hold; a telic event does not have to be durative nor does a dynamic event need to be telic. The accruing of aspectual features follows the hierarchy presented above.

A novel observation came from these early classification attempts at creating my idiomatic corpus; for the 20 idiomatic phrases that made up my corpus, I noticed that for every idiom, it was the case that the idiomatic event belonged to the same aspectual class as the literal event **or** lower in the hierarchy presented. I did not find for any of them an instance where the aspectual class of the idiomatic event ranked higher up in this hierarchy.

Critically, these observations provided empirical evidence that while it is possible for the idiomatic event to be in the same aspectual class as the literal event, it is also possible for the idiomatic event to be part of a different aspectual class. Based on my observations, when there was a difference between the two events, the difference was always in the direction indicated in the hierarchy I described earlier. In other words, if an aspectual

component is lost, the loss follows this hierarchy: duration > telicity > dynamicity.

Therefore, the goal of this chapter is to corroborate, and if proved valid, formalize the intuition that there is a dependency between the idiomatic event and the literal event and that the dependency can be described in terms of the aspectual class of the literal event. In other words, the literal event's aspectual structure may be used to predict what aspectual structure the idiomatic event can have.

The rest of the chapter is organized according to the following outline: section 4.2 provides the grammatical tests used throughout the chapter to show which aspectual class a VP belongs to. In this section, I also show how each aspectual class behaves for each of the contexts the aspectual tests create. Section 4.3 details the experimental design, shares the participants' demographic information, introduces the idiomatic corpus with which the participants interacted, and states what alternative hypotheses are being considered.

Results are spread out in two sections, section 4.4 focuses on whether participants judged the aspectual structures of the literal and the idiomatic events as the same or different. Section 4.4.6 explores if the found differences lend support to the idea that a *defective relationship* exists between the idiomatic and the literal event.

The first set of results suggests that it is ill-advised to maintain the position that the two events must share the same aspectual class, as proposed in McGinnis (2002, 2005) given that the empirical evidence provided by the participants contradict such a position.

The second set of results take on a more nuanced view of the participants' responses where I compared the expected results put forth by the tests to the participants' judgments of contexts which necessitated that the idiomatic event be part of an aspectual class higher up in the hierarchy proposed above. The expectation is that the participants will reject these interpretations as ungrammatical and by doing so shed light on the direction in which the idiomatic event stands in relationship with the literal one.

4.2 Aspect and Idioms

4.2.1 Lexical aspectual classes

Lexical or inner aspect[11] refers to the linguistic phenomenon which deals with how speakers conceptualize the way events change as time lapses. I start from the premise that events in the world are different from how we describe them linguistically. We only represent a handful of real-world properties of events in our linguistic encoding of them. Rappaport Havov et al (2010:2) indicate this fact about languages and the elements in the world they referred to: "research focused on the linguistic representation of events have revealed that only a subset of these properties are linguistically significant."

Whereas actual events are structured in ways that are by no means obvious, events in the linguistic sense arise from an intersection of various elements within the sentence which include: lexico–semantic properties of the verb, referential properties of the verb's arguments, and the temporal/spatial properties of VP modifiers. One can determine both how significant these elements are, and which way they contribute to sentence meaning through the use of grammatical/aspectual tests that will be introduced in the following section.

As the grouping of these linguistically significant properties abstracts away from actual events, it helps to define a "template for the linguistic representation of events" (Rappaport Hovav et al. 2010:2). These kinds of representations of events are known collectively as *event structure*. Event structure as a linguistic concept carries several layers of information, as I'll describe below.

The study of these internal temporal properties of events has been documented in the literature as a typology of aspectual classes commonly referred as to AKTIONSART. Based on the work of Vendler (1967) and Dowty (1979), there are four[12] proposed aspec-

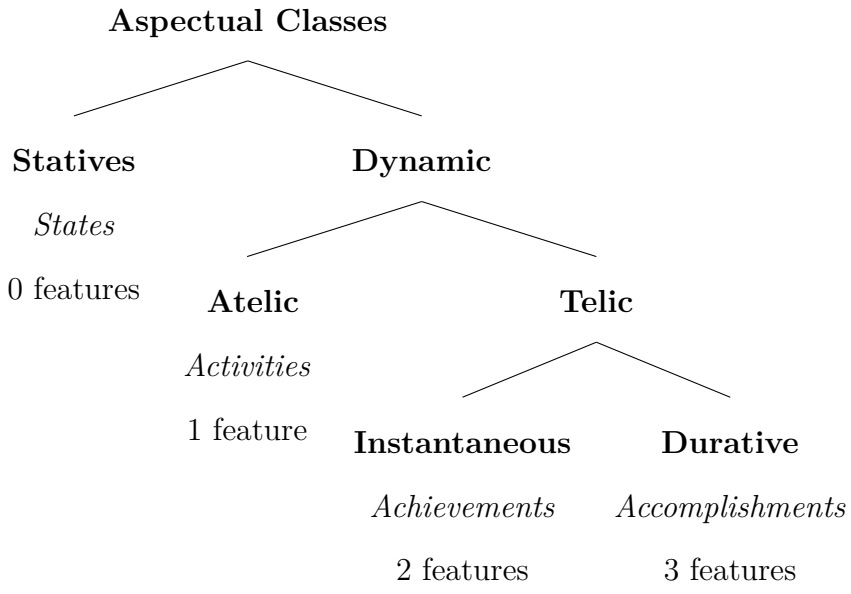
[11]This is in opposition to grammatical or outer aspect which anchors an event in time and provides a temporal perspective to reference time

[12]I'll assume that semelfactives are part of the aspectual class of achievements as they only appear to

tual classes. Predicates belonging to a class exhibit homogeneous behavior when made to interact with aspectual tests. A group of predicates that respond to these tests in a coherent, predictable way are then thought to belong to the same aspectual class.

I suggest below that aspectual classes can be minimally distinguished using three features. These three features are dynamicity, telicity, and durativity. The order in which I listed these features is significant: predicates in natural languages appear to accumulate these features in order. In other words, if a predicate is telic, it is also true that the predicate is dynamic. The reverse does not seem to hold; a predicate can be dynamic, but not telic. As a result, these features can be understood to be stacked on each other in such a way that, as features are progressively added, a predicate can be sorted from one aspectual class into another in a hierarchical manner, as shown in (16). Following this idea of hierarchical ordering, verbal predicates can be grouped into STATES, ACTIVITIES, ACHIEVEMENTS, and ACCOMPLISHMENTS.

(16)



The first of these features, dynamicity, splits all possible events, which I will refer to as eventualities from now on, into two sets, states and events proper. Events are predi-

differ from achievements in the transient nature of the object they take and the presence of a resulting state.

cates that are said to occur or happen where an action involves some kind of change. By contrast, predicates lacking dynamicity are static; they are states. States are said to hold for an unspecified amount of time. States exhibit internal homogeneity throughout.

As a result of this homogeneity, there is no evidence or suggestion of progression toward a culminating point. States do not appear to have natural boundaries or limits. Examples of states are shown in (17).

(17) a. saber la respuesta
know det answer
'know the answer'

b. ser alto
be tall
'be tall'

c. tener hambre
have hunger
'be hungry'

Telicity refers to the presence or absence of a natural endpoint for an action. Verbs denoting actions with such endpoints are said to be *telic*. By contrast, actions lacking these endpoints are said be *atelic*. Traditionally, the presence of a singular endpoint was considered sufficient to classify a verbal predicate as telic. In a more recent analysis, MacDonald (2008) suggests that for a verb to be telic, a startpoint or an endpoint are needed, and most likely both. Kearns (2011) discusses *boundedness* as an umbrella term for both inceptive and a terminative telos points of an event.

Kearns (2011:163) indicates that there is considerable consensus in current research regarding the subinterval property of the event denoted by a predicate. "It is not as much a test as a way of understanding telicity that can be used to verify the classification of a predicate into a category" (Kearns 2011:163). The general intuition behind the subinterval property has to do with looking inside the event to see if the action is homogeneously

carried out throughout the event completion.

For telic events, let I_a be such that the entire event is contained within I_a. One can also imagine a shorter interval, I_b such that $I_b \subset I_a$. This subinterval I_b does not contain the entire telic event by virtue of being a proper subset of I_a. So, for instance, say *running a mile* occurs over I_a, any segmentation of the event over I_b would not be *running a mile*. We say in this case that the subinterval property is absent.

For atelic events, the subinterval property yields an interesting insight. Consider this time, *running*. If we take running over an interval I_a, we can find smaller subintervals I_b as before. Contrary to what happened with telic events, any segmentation of the event done over I_b is still *running*. The result is that atelic events exhibit the subinterval property because they are homogeneously spread over the duration in which they happen; we can find shorter subintervals where it would be the case that the event occurs. For telic events, on the other hand, any interval that does not contain the duration of the entire event can not be said to be the event proper.

Concordantly, telicity splits verbal predicates into those with no inherent limits or endpoints and those which do exhibit these limits as part of the predicate's meaning. As a result, activities are atelic predicates, whereas achievements and accomplishments are telic. Examples of telic predicates are shown in (18), whereas examples of atelic predicates are listed in (19).

(18) a. cruzar la calle
cross the street
'cross the street'

b. leer un libro
read a book
'read a book'

c. florecer
bloom
'bloom (said of a flower)'

d. avistar un pájaro
spot a bird
'spot a bird'

(19) a. correr
run
'run'

b. respirar
breathe
'breathe'

c. lavar
wash
'wash'

d. crecer
grow
'grow'

The third feature, duration, refers to the time that must lapse as an action progresses toward its culmination or endpoint. Actions that are instantaneous are said to have no duration; they culminate as soon as they start. Verbs of cognition are commonly considered to have no duration. Consider for instance, *'Mary heard the news at 10:00 pm'*. One instant before 10:00 pm, Mary was not aware of the news, and a second after, she was. Kearns (2011:157) posits that one way to think of these kinds of verbs is to conceptualize them as "front edges" of states; that is, they are "idealized to the momentary transition between not being aware" of something and then being aware of it. Examples of events with no duration are listed in (20).

(20) a. darse cuenta
give.CL count
'realize'

b. notar
notice
'notice'

c. llegar
arrive
'arrive'

d. parpadear
blink
'blink'

By contrast, some events are not instantaneous; they denote events where time has to lapse in order for an event to be said to have occurred. These events are dynamic; an action progresses towards its endpoint. When the endpoint is reached, these events are said to have culminated. Examples of durative events are listed in (21).

(21) a. estirar el brazo
stretch the arm
'strech one's arm'

b. llenar un envase
fill a container
'fill a container'

c. correr medio maratón
run half a marathon
'run half a marathon'

d. escribir un poema
write a poem
'write a poem'

Having introduced dynamicity, telicity, and duration as the key properties that let us group verbal predicates into aspectual classes, I will proceed to introduce the actual grammatical/aspectual tests I rely on to ascertain whether a particular VP exhibits such properties.

4.2.1.1 Using acceptability judgment tests to establish aspectual membership: dynamicity and the progressive In terms of aspectual tests, the *progressive*, the *time-span* construction, and the *durative phrase* play a critical role in that these constructions

make evident an aspectual feature is present in the sentence and its concomitant event. The first of these acceptability judgment tests that I employ is the *progressive*.

Consider the Spanish sentences in (22).

(22) a. Graciela tiene un orzuelo.
Graciela have.PRS a sty
'Graciela has a sty'

b. Marina carga un libro.
Marina carry.PRS a book
'Marina carries a book'

c. Moura avista un pájaro.
Moura spot.PRS a bird
'Moura spots a bird'

d. Carmen Luisa bebe una copa de vino.
Carmen Luisa drink.PRS a glass of wine
'Carmen Luisa drinks a glass of wine'

All sentences in (22) denote an event with two participants: the internal argument and the external argument. All the sentences share the same basic syntactic structure. Placing those sentences in the progressive results in diverging interpretations that depend on their lexical aspectual properties. This is shown in (23) below.

(23) a. *Graciela está teniendo un orzuelo.
Graciela be have.PTCP a sty
'Graciela is having a sty'

b. Marina está cargando un libro.
Marina be carry.PTCP a book
'Marina is carrying a book'

c. Moura está avistando un pájaro.
Moura be spot.PTCP a bird
'Moura is spotting a bird'

d. Carmen Luisa está bebiendo una copa de vino.
Carmen Luisa be drink.PTCP a glass of wine

'Carmen Luisa is drinking a glass of wine'

(23a) is hard to parse, if not ungrammatical. (23b) means Marina is somewhere in the middle of the book-carrying event. (23c) means that Moura already has spotted a bird a few times and probably will do so a few more times. For (23d), Carmen Luisa is roughly halfway in her wine-drinking event. Only (23b) and (23d) give rise to the same general interpretation.

The fact that (23a) behaves dissimilarly from (23b), (23c) and (23d) suggests that even though all sentences in (23) have the same structure, the verbs involved are not the *same* kind of verbs. Whatever is different between them is sensitive to the progressive and appears to group (23b), (23c) and (23d) together. De Miguel (1999:3012) indicates that predicates like (23a) do not denote the *occurrence* of an event, but rather that said event *holds* homogeneously in time. De Miguel explains that it is precisely this constancy in time that conflicts with the progressive and causes the unacceptable readings.

In addition to refusing to play nice with the progressive, states also reject being part of imperative statements, as in (24b). Finally, when a statement expressing a state is uttered in present tense, that statement does not get interpreted with a habitual reading, as in (24c), but instead they get interpreted as holding at the time of speech. Compare (24c) to *Sinforoso corre* 'Sinforoso runs,' where the available interpretation is that Sinforoso habitually runs.

(24) a. *Serafina está sabiendo la respuesta
Serafina be.PRS know.PRS.PTCP the answer
'Serafina is knowing the answer'

b. *Rosendo, ¡sé alto!
Rosendo be.IMP tall
'Rosendo, be tall!'

c. #Sinforoso tiene hambre
Sinforoso have.PRS hunger

'Intended: Sinforoso is habitually hungry vs He is hungry right now '

In sum, predicates that are not felicitous in the progressive, are states. Conversely, a predicate that is compatible with the progressive is, therefore, taken to be dynamic.

4.2.1.2 Using acceptability judgment tests to establish aspectual membership: telicity, the durative phrase and the time-span construction To ascertain telicity, I followed a strategy that mirrors what I did with the progressive. I used two adverbial phrases not selected by the verb head that trigger similar clustering behaviors. García Fernández (1999:3135) indicates that, together with grammatical tense, grammatical aspect, and lexical aspect, adverbial phrases form a patchwork of constructions that work together to give rise to expressions of time in the language.

The two adverbial phrases that I use in this chapter to probe for telicity are *aspectual adverbial phrases of quantification* headed by *en* 'in' and *durante* 'during/for,' which I will call the *time span construction* and the *durative phrase*, respectively, to match MacDonald's (2008) terminology. Consider how our four sentences interact with the *time span* construction in (25) below.

(25) a. *Graciela tuvo un orzuelo en 10 minutos.
Graciela have.PRT a sty in 10 minutes
'Graciela had a sty in 10 minutes'

b. *Marina cargó un libro en 10 minutos.
Marina carry.PRT the book in 10 minutes
'Marina carried the book in 10 minutes'

c. Moura avistó un pájaro en 10 minutos.
Moura spot.PRT a bird in 10 minutes
'Moura spotted a bird in 10 minutes'

d. Carmen Luisa tomó una copa de vino en 10 minutos.
Carmen Luisa drink.PRT a glass of wine in 10 minutes
'Carmen Luisa drank a glass of wine in 10 minutes'

When made to interact with the *time span* construction, the four sentences cluster again, but differently from the ones in (23). Namely, the *time span* splits the four sentences into two even groups: (25a) and (25b) yield unacceptable results, whereas (25c) and (25d) are acceptable when interacting with the *time span* construction. It is worth noting that even though (25c) and (25d) result in grammatical sentences, they do not yield the same interpretation. That is to say, the salient interpretation of (25c) is an inceptive one by virtue of being a punctual event (García Fernández 1999:3134). By contrast, (25d) yields a terminative reading, i.e., the event requires 10 minutes to lapse in order to reach its culmination.

García Fernández (1999) explicates that these adverbial phrases provide information about the temporal development of the event. The *time span* adverbial indicates how much time lapses as the event progresses from its beginning to its end *including* the event's telos points. By comparison, the *durative* phrase indicates lapsing of time by the indicated amount *excluding* telos points. The disconnect between the *time span's* inclusion of telos with the lack of telos in (25a) and (25b) causes the unacceptability.

Predicates with clear telos, like (25c) and (25d) are incompatible with the *durative* phrase because their telos, and their culmination, are excluded from the event and consequently, they are not technically finished actions. Critically, this lack of overlap places both adverbial phrases in complementary distribution, as (26) shows below.

(26) a. Graciela tuvo un orzuelo durante dos semanas.
Graciela have.PRT a sty for 10 minutes
'Graciela had a sty for 10 minutes'

b. Marina cargó el libro durante 10 minutos.
Marina carry.PRT the book for 10 minutes
'Marina carried the book for 10 minutes'

c. *Moura avistó un pájaro durante 10 minutos.
Moura spot.PRT a bird for 10 minutes
'Moura spotted a bird for 10 minutes'

d. *Carmen Luisa tomó una copa de vino durante 10 minutos.
Carmen Luisa drink.PRT a glass of wine for 10 minutes
'Carmen Luisa drank a glass of wine for 10 minutes'

As a result, compatibility with the *time span* is the classical test for telicity, whereas compatibility with the *durative* phrase is taken as an indication that a predicate is atelic. In the following section, I will look into what are the assumed consequences that a phrase is compatible or incompatible with these constructions.

4.2.1.3 Using acceptability judgment tests to establish aspectual membership: duration, the almost and the takes-x-time adverbials The last of the three features required to fully establish the aspectual class of a VP is duration. The simplest way to test for the durativity of a predicate is to check how it interacts with the present progressive. The salient interpretation of a predicate with duration in the present progressive is an *in-progress* interpretation. By contrast, an instantaneous predicate yields an *iterative* interpretation. This difference is illustrated in (27a) and (27c) below where the interpretations are that Frank is in the process of running and Mirlen is in the process of crossing the street. By contrast, in (27b), the interpretation is that Zurlay has already snapped her fingers and probably will continue to do so a few more times.

(27) a. Frank Ramón está corriendo
Frank Ramón be run.PTCP
'Frank Ramón is running'

b. Zurlay está tronando los dedos
Zurlay be.PST snap.PRS.PTCP the fingers
'Zurlay is snapping her fingers'

c. Mirlen está cruzando la calle
Mirlen be cross.PTCP the street successively
'Mirlen is crossing the street repeatedly'

We can account for these differences in interpretation by distinguishing *whether* duration is external to the VP or internal to the VP. For an internal duration reading, an

activity (27a) and an accomplishment (27c) share the *in-progress* interpretation. This corresponds to the activity proper or to the activity-phase of the accomplishment.

However, if duration is seen as external to the VP, in other words, we think of duration as taking scope over the entire event, we see that only the telic predicates can successfully interact with the progressive. Consider now the same activities, achievements, and accomplishments as they interact with the adverbial *sucesivamente* 'repeatedly' in (28).

(28) a. #Frank Ramón está corriendo sucesivamente
Frank Ramón be run.PTCP successively
'Intended reading: Frank Ramón is running repeatedly'

b. Zurlay está tronando los dedos sucesivamente
Zurlay be snapping.PTCP the fingsers successively
'Intended reading: Zurlay is snapping her fingers repeatedly'

c. Mirlen está cruzando la calle sucesivamente
Mirlen be cross.PTCP the street successively
'Intended reading: Mirlen is crossing the street repeatedly'

We see that telic predicates yield a *series-of-similar-events* interpretation whereas the atelic predicate yields a *single-event* interpretation (terminology borrowed from MacDonald (2008)).The felicitousness of (28b) and (28c) hinges on the fact that both the street-crossing event and the blinking event consist of a series of individual street-crossing and blinking subevents strung together by the adverbial *sucesivamente*. To string one after the other, we need one subevent to end before the next one can start. By contrast, in (28a), if we look at the way we conceptualize *correr* 'run,' i.e., without an intrinsic terminative telos point, a subevent can't end before the next one can begin.

For these reasons, and perhaps in a case of poorly chosen terminology, despite the fact that all dynamic predicates **are** compatible with the progressive, only telic events are considered to have duration. This has been noted already for states in Smith (1991:487), "unlike eventives, states compatibility with the durative phrase should not be taken as

evidence of duration." I will consider duration to be narrowly defined as the time that needs to lapse between the point in time when an event starts and the point in time when it culminates.

Readings to tease apart duration can be hard to elicit exclusively by the progressive. For that reason, I will employ two additional tests: the *takes-x-time* and the *almost* adverbials. Both tests are understood to be sensitive to the presence of telos points in such a way that when speakers use them, they can give rise to different readings based on which telos point is targeted or available.

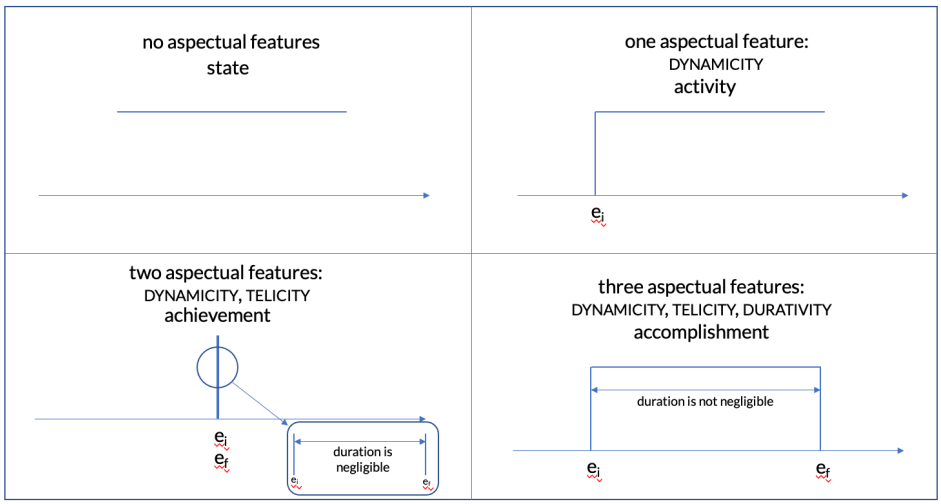


Figure 32: Number of Features and Aspectual Classes

Keeping in mind Figure 32, shown above, consider (29) below, where we can see how the *takes-x-time* test works.

(29) a. Frank Ramón tomó 10 minutos en correr
Frank Ramón take.PRET 10 minutes in runINF
'Frank Ramón took 10 minutos to run'

b. Zurlay tomó 10 minutos en tronar los dedos
Zurlar take.PRET 10 minutes in snap the fingersINF
'Zurlay took 10 minutes to snap her fingers'

c. Mirlen tomó 10 minutos en cruzar la calle
Mirlen take.PRET in cross.INF the street
'Mirlen took 10 minutos to cross the street'

All three statements in (29) are acceptable in an inceptive interpretation; that is, after 10 minutes have lapsed, the three actions took place, or 10 minutes lapsed and Frank Ramón started to run, Zurlay started to snap her fingers, and Mirlen started to cross the street. (29c) has an additional reading not available to the other two statements. The available reading is a terminative one, where Mirlen started to cross the street, and after 10 minutes, she reaches the other side of the street.

One way to look at this discrepancy is to imagine that *takes-x-time* is able to target either the inceptive telos point e_i or the terminative telos point e_f. Under this assumption, a predicate like *correr* in (29a) can only give rise to one reading, the inceptive one, since it only has the one inceptive telos e_i. A predicate like *snap one's fingers* (29b), despite having both e_i and e_f teloi, also results in a single reading, as both e_i and e_f effectively happen simultaneously given that the event is thought to be instantaneous, as shown in Figure 32. (29c) is the one that gives rise to both readings, as duration is, critically, the additional aspectual element that allows for the targeting of both e_i and e_f.

Almost is the other test I employ to probe durativity. It mimics the use of the *takes-x-time* test in that it also targets the inceptive or terminative telos points of an event, but the resulting interpretations are different. *Almost* yields a counterfactual reading for all three statements in (30) below, in the sense that none of them came to happen.

(30) a. Frank Ramón casi corrió
Frank Ramón almost run.PRET
'Frank Ramón almost ran'

b. Zurlay casi tronó los dedos
Zurlay almost snap.PRET the fingers
'Zurlay almost snap her fingers'

c. Mirlen casi cruzó la calle
Mirlen almost cross.PRET the street
'Mirlen almost crossed the street'

As is the case with the *takes-x-time* examples, in addition to the counterfactual reading, (30c) also gives rise to a second reading: an incompletive interpretation. Unlike (30a), where once Frank Ramón started running, it is the case that he has run; or (30b), where Zurlay, once she started snaping her fingers, she was done snapping them just as quick. By contrast, it is the case that Mirlen can start crossing the street, but halfway in the process of crossing the street, she can change her mind and turn around, effectively never finishing the street-crossing event.

Up to this point, I have discussed the three aspectual properties that a predicate may exhibit. I have shown that we can rely on aspectual tests to determine if that property is present in a sentence. Moreover, I have shown that the presence of dynamicity, telicity, and durativity account for differences in a predicate's aspectual structure, and consequently, in its aspectual class. Figure 32 gives an abstract representation of how aspectual classes differ from one another, as seen exclusively from the point of view of aspectual features.

In addition to the internal temporal information described above, event structure may also reflect external temporal information; that is, how an event is examined or viewed from various temporal perspectives. For an observer, independent of how an event is minimally constructed, it can be accessed as it develops, but it can also be observed from an outside perspective, where it is presented as a completed state of affairs. This basic difference between the two perspectives gives rise to the well-known *pretérito/imperfecto* distinction in Spanish and the more general perfective/imperfective viewpoint aspect.

I have described, thusly, the process I will follow to determine the aspectual membership of a predicate. To summarize, these steps are taken to be an indication of aspectual

features present in the event. The assumption is that the grammar of a language is sensitive to these aspectual features and this sensitivity is expressed in the form of salient or available readings. The three aspectual features are dynamicity, telicity, and durativity.

To test for dynamicity, we check to see if the predicate in question is acceptable in the progressive. To test for telicity, I use the adverbial phrases of duration *in-x-time* and *for-x-time*. Compatibility with the former, in a single-event reading, is taken as an indication that the event is telic. On the other hand, compatibility with the latter, also in a single-event reading, is taken as an indication that the event is atelic. Finally, in order to test for durativity, I will use *takes-x-time* and *almost* adverbials. Both adverbials give rise to **two** readings when the predicate denotes a durative event and only **one** reading otherwise, for a non-durative event.

4.2.2 Modeling aspectual classes with features

In the previous section, I described how to discern if any one of the three aspectual properties — dynamicty, telicity, and duration — are present in a sentence by means of applying a subset of aspectual tests. To make explicit the number of aspectual features present in an event, I will rely on simple cardinality. In other words, given a complete sentence, by applying the tests described earlier, I can determine if that sentence is dynamic; if so, that counts as one feature present. I can also determine telicity, by applying the appropriate aspectual tests. If the sentence is telic, that counts as another feature added. Finally, using the *take-x-time* and *almost* tests, I can determine if a sentence is durative. When it is, that counts as another feature.

As a result, by applying the set of tests described earlier, I have provided an algorithm that classifies predicates into aspectual classes by simply keeping track of the features present. This process, in turn, allows for a comparison of events by comparing the number of features present. This process can be directly applied to the pair of events resulting from interpreting a sentence literally and idiomatically. Assuming one knows the

number of features of a literal event, can we say anything meaningful about its concomitant idiomatic event? In other words, can we circle back to the opposing views expressed by McGinnis and Glasbey regarding how predetermined an idiomatic event is in relationship to its literal event?

In addition to using aspectual features to classify predicates into aspectual classes, I also want to explore if the accrual of aspectual features can also function as an indicator of a different property, which could be used to account for the observed hierarchy of events. Let us think of this emerging property as *aspectual complexity*, in the sense that, as features are gathered and we move from *states* to *accomplishments*, the denoted event becomes more complex as the number of aspectual features increases.

This idea of complexity has been explored in analyses such as Dowty's (1979) decomposition of predicates, where he indicates that aspectual classes can be understood by means of a "bisentential analysis" of the semantic primitives CAUSE, DO, and BECOME, as shown in(31) (examples taken from (Dowty 1979:123)).

(31) a. Statives: $\pi_n(\alpha_1, ..., \alpha_n)$.(John knows the answer)

b. Activities: $\textsc{do}(\alpha_1, [\pi_n(\alpha_1, ..., \alpha_n)])$.(John is walking)

c. Achievements: $\textsc{become}[\pi_n(\alpha_1, ..., \alpha_n)]$.(John discovered the solution)

d. Accomplishments: $[[\textsc{do}(\alpha_1, [\pi_n(\alpha_1, ..., \alpha n)])]\textsc{cause}[\textsc{become}[\rho_m(\beta_1, ..., \beta_m)]]]$

Dowty's analysis for each of the aspectual classes becomes notationally more complex in order to account for an increasing number of subevents and participants required to model events.

Other researchers have explored lexical aspectual classes using experimental methods to probe for complexity. For instance, McKoon and Macfarlane (2002) probe the behavior of various verbal predicates using four different experiments that resulted in longer processing times for sentences with externally caused change of states. They also found

that lexical decision response times were slower when they contrasted externally caused change-of-state verbs, like *shake* with internally caused ones, like *shudder*, showing that less complex events can be accessed more rapidly than more complex ones.

McKoon and Macfarlane (2002) relied on simpler (than Dowty's) event templates to separate the various predicates they worked with. Their event templates also make clear that, as we move along the hierarchical order, more complexity is introduced. This is because more participants and more subevents are required to model an event as a whole, as seen in Table 32 (examples taken from (McKoon and Macfarlane 2002:3)). Although McKoon and Macfarlane (2002) do not specifically consider aspectual classes in their experiment, their findings support the idea that there are differences in processing time between states, on the one hand, and events, on the other.

Event template name	Template	Example
State	x(IN STATE)	x(LOVE), x(EXIST)
Activity	x(ACT)	x(RACE), x(HIT)
Change of State	x(BECOME IN STATE)	x(BECOME ARRIVED), x(BECOME BLOOMED)
Externally caused change of state	α CAUSE(x BECOME IN STATE)	α CAUSE(x(BECOME BROKEN)), α CAUSE(x(BECOME FADED))

Table 32: Event templates

In a similar fashion, Gennari and Poeppel (2003) conducted two experiments seeking to corroborate that eventives (activities, achievements, and accomplishments) induce longer processing times than statives (states). Gennari and Poeppel (2003) found that simpler semantic complexity, exemplified by sentences containing activities, results in shorter processing times than those sentences containing activities, achievements, and accomplishments. As Gennari and Poeppel (2003:29) put it, "the more complex the structure, the longer the time to access, or activate, those structures."

In addition to determining the aspectual class of a predicate, I want to explore if

the cardinality of the aspectual features present in an event can act as a proxy for how complex an event is. There is a sense that the hierarchy depicted in (16) has an empirical basis that goes beyond simple intuition and it is supported by two basic points:

(1) As mentioned earlier, in section 4.2.1, aspectual properties accumulate in sequential order; in order to be durative, a predicate must also be telic, and in order to be telic, a predicate must also be dynamic. The reverse does not seem to hold.

(2) Despite the fact that psycholinguistic studies haven't always found clear-cut results that delve into the aspectual granularity that I pursue here, they have consistently found that events proper are more complex than states.

With this goal in mind, the remainder of the chapter is set to explore the idea that we can, in fact, say something meaningful about the way an idiomatic event stands in relationship to its literal event.

4.2.3 Idioms and Aspect

Everything done up to now has been for the purpose of finding a reliable way to identify the aspectual class of a predicate. I now intend to return to the general theme that has been threaded throughout the previous two chapters: idioms. Idiomatic phrases form the other half explored in this chapter. Idioms are fascinating objects of linguistic inquiry because they fly in the face of a foundational principle of how human languages work: idioms violate Frege's Principle of Compositionality. Jacobson (2014:42) indicates "The meaning of the whole is computed in some predictable way from the meaning of the parts." Again Jacobson (2014:42) states: "Surely at least most of language has to be compositional, or there would be no way to systematically combine units into larger meaningful expressions which has never been uttered or heard before."

Despite the appeal of considering language to be compositional in nature, other researchers (Sinclair 1991; Jackendoff 1992, 1997) have argued in favor of idiomaticity as

a prevalent factor in day-to-day speech production. This position is buttressed predominantly by the high frequency of words and phrases in syntactic positions that should be open to more variation than what is observed. The Idiom Principle (Sinclair 1991:110) addresses this: "the principle of idiom is that a language user has available to him a large number of semi-constructed phrases that constitute single choices, even though they might appear to be analysable in segments."

In this regard, idiom usage can be seen as a way to gauge language proficiency in that choices speakers make in day-to-day interactions result in language sounding more or less marked. Pawley and Syder (1983:191) call the process of making these natural-sounding choices "nativelike selection" and "nativelike fluency." Thus, choosing the right idiomatic option may have a wide range of SLA repercussions, as not operating in the "nativelike" zone will necessarily result in "marked" speech. This section will remain focused on aspect and idioms of the form VO exclusively, but the ground seems fertile for further research addressing idiomatic usage and the different choices made by L1 and L2 speakers. See Titone et al (2015) for differences in monolingual and bilingual idiom processing and Vanderniet (2015) for a case study of idioms used as a measure of proficiency.

For the remainder of this chapter, the focus will remain on differences in aspectual structure that arise when a potentially ambiguous sentence is interpreted idiomatically, as compared to the aspectual structure that results from a literal interpretation. This issue remains unresolved in the literature with some linguists arguing that the aspectual structure of the idiomatic event is compositional and not subject to the idiosyncrasy one expects of idioms. Chiefly among these opinions is McGinnis's (2002, 2005), although her position is also echoed in Marantz (1997), O'Grady (1998), Fraser (1970), and Katz and Postal (1963), among others.

McGinnis argues that an idiomatic VP is fully compositional and its aspectual structure arises from the contributions of its literal building blocks. In her own words: "for any

idiomatic VP with a verb V_n, there will be a nonidiomatic VP with V_n that has the same aspectual properties" (2002:669). If my understanding of her statement is correct, McGinnis argues that the aspectual structure of the idiomatic event is the same as the aspectual structure of the literal event. Even in the case of idioms where there are no literal counterpart, McGinnis maintains that by looking at a VP with *similar* syntactic constituents, one would be able to ascertain the aspectual structure of the idiomatic event. Again, McGinnis (2002:668) "even if a VP has a noncompositional idiosyncratic meaning, it will have a compositional structural meaning. Specifically, it will have the same aspectual properties as any VP with the same syntactic properties."

There are a few cases where McGinnis (2002:670) does present examples of an aspectual discrepancy between the two interpretations, see (32) taken from McGinnis (2002). McGinnis proposes that this aspectual differences is explained by ambiguity of *get* as an agentive and non-agentive verb and consequently, she argues, the two interpretations arise because of "differences in clause structure."

(32) a. The mouse stopped getting the cheese (accomplishment)

b. #The car stopped getting Harry's tongue (achievement)

c. #The teacher stopped getting Harry's homework (achievement)

Counter to this position is that of Glasbey's (2003; 2007) who indicates that while it is true that idiomatic VPs matching their literal VPs exist, it appears to also be true that for some idiomatic VPs, there are aspectual mismatches. For the claim that both events must have the same aspectual structure, Glasbey provides as a counter example the idiomatic phrase "paint the town red" which is a literal accomplishment, but tests as an idiomatic activity when it is interpreted as "to go out drinking, dancing, etc." Glasbey concludes that while it is possible that the two events can coincide aspectually, it is not necessary.

Furthermore, Glasbey tentatively identifies differences in thematic relations, as being

a factor involved in aspectual compositionality. According to Glasbey, the process of building up the event from constituents that have undergone a change in thematic relations results in events that are different, or as she puts it "the difference here is that, because the thematic relations may be different in the idiomatic interpretation than in the literal interpretation, the result of the aspectual composition may also be different" (2003:48).

Along the lines of Glasbey (2003; 2007), other linguists have found correlates that account for an idiom's idiosyncratic behavior. For instance, Nunberg et al. (1994) report that when the idiomatic components of an idiom are segmented into the idiom's constituents then those idioms can be subjected to a variety of operations, like passivization, while retaining their idiomatic meaning. Nunberg et al. (1994) conclude that idiomatic phrases should be subdivided into *idiomatically combining expressions* and *idiomatic phrases* proper, with the former being compositional but the later fully idiosyncratic.

Mateu and Espinal (2007:55), similarly, show that if a language belongs to a verb-framed language, in the sense of Talmy (1985), one will not find in its inventory of viable sentences constructions such as "complex, telic path-of-motion constructions like *The boat floated out of the cave*" or "complex resultative constructions like *John cooked the pot black*." Moreover, these constraints on what structures are possible extends to the available idioms in the language.

Given the differing ideas about the relation between aspectual interpretations of literal and idiomatic VPs, I designed an experiment to test which hypothesis may better account for any differences in interpretation.

4.3 Experimental Design

Chapter 4 is about finding empirical evidence for the intuition that the aspectual relationship between compositional and idiomatic interpretations of the same sentence results in what I will call *a defective relationship*, where by *defective* I mean that an id-

iomatic event may use the aspectual features present in its literal event or discard them, but critically, it may not add aspectual features. The evidence I gathered to support this position comes from the aspectual tests discussed earlier and the grammatical intuitions of participants.

Speaker intuitions form the basis of grammatical judgments used in modern linguistics. Chomsky (1957, 1986) makes clear the importance of speakers' introspection regarding the well-formedness of sentences and the grammaticality (read:validity) of syntactic operations. As Cowart (1997:1) remarks, this constituted a significant turn from previous linguistic approaches: "Chomsky instead embraced introspective judgments as an essential tool for the exploration of mental resources underlying language use."

This reliance on speakers' intuition is not free of criticism or difficulties, however. Schütze (1986) voices the concern that without a proper methodology of grammaticality judgments, the criticism directed at them cannot be properly addressed. The two major criticisms directed at grammaticality judgments are the idea that theoretical linguists build *theories of intuition*, as opposed to the actual competence that allows for everyday language production and use, as well as the bias and subjectivity colored by a linguist's own views.

Experiment 3 leverages speakers' intuition by asking them if it is accurate to use a given sentence to describe a certain state of affairs. All participants responses consisted of a *forced choice* type of answer in the form of "yes, the situation can be described using this sentence" or "no, the situation cannot be described using this sentence." For each tested interpretation, a context was crafted.

These contexts were crafted while keeping an eye on "controlling context" in the sense of Cowart (1997). It is known that a variety of extrasyntactic factors affects the acceptability of statements (Cowart 1997:47); among these factors, we see the frequency and familiarity of a lexical item, the length of an item, the semantic properties of an item, and

an item's interaction with the other components of the sentence being tested. Lastly, a participant's own personal linguistic history can affect their response.

In order to present a stimulus as homogeneously as possible, all scenarios were maximally similar; they all had the same structure and informational content. Each scenario was composed of three sentences, as described below.

The grammatical judgment corpus was made up of four idiomatic phrases presented earlier and repeated here as (33), where each participant was asked to judge only two of those phrases. Here are the specifics about the judgment tasks:

(33) a. estirar la pata
stretch the leg
lit: "stretch one's leg"; idm: "die"

b. soltar el trapo
release the rag
lit: "drop the rag"; idm: "burst out laughing"

c. tender las velas
unfurl the sails
lit: "unfurl the sails"; idm: "take advantage of a favorable situation"

d. quemarse las pestañas
burn.CL the eyelashes
lit: "burn one's eyelashes"; idm: "study hard"

- Each sentence had two possible interpretations: a literal interpretation and an idiomatic interpretation.
- For each interpretation, two contexts were crafted: a context that targets the start of the event and a context that targets the end of the event. The previously mentioned *almost* and *takes-x-time* adverbials were used to guide the process.
- In order to present a stimulus as homogeneously as possible, all scenarios had the same structure and the same informational content. Each scenario consisted of a question asking if the target sentence accurately described the context containing the

almost or *takes-x-time* adverbial.

- Each participant was, therefore, asked to judge 8 pairings of sentences and contexts: 2 *interpretations* × 2 *targets* × 2 *sentences to judge*.
- The two adverbials that can target the start or end of an event were split evenly between the four statements. That is to say, two of the four statements used the *almost* adverbial and the remaining two used the *takes-x-time* adverbial.

4.3.1 Participants

Experiment 3 participants are a subset of the participants involved in *Experiment 2*. The inclusion criteria was completion of all the grammatical judgments tasks. I will not replicate the detailed demographic breakdown of the sample population as I did for *Experiment 2*, but I will provide a general recount of the demographic makeup.

- n = 282
- n_{men} = 118, or 41.84%
- n_{women} = 160, or 56.74%
- $n_{non\text{-}binary}$ = 4, or 1.42%
- n_{native} = 226, or 80.14%
- $n_{non\text{-}native}$ = 56, or 19.86%

Age distribution of the participants is shown in Table 33 below.

Table 33: Age distribution

Age Brackets	Count
[10-18)	0
[18-20)	8
[20-30)	60
[30-40)	79
[40-50)	90
[50-60)	28
[60-70)	12
[70-80)	4
[80-90)	1

4.3.2 Corpus makeup

The corpus used for this experiment is composed of four statements. All four statements share the same syntactic structure; they are composed of an external argument, a verb, and its internal argument. The four sentences in question are listed in (34).

(34) a. estirar la pata
stretch the leg
lit: "stretch one's leg"; idm: "die"

b. soltar el trapo
release the rag
lit: "drop the rag"; idm: "burst out laughing"

c. tender las velas
unfurl the sails
lit: "unfurl the sails"; idm: "take advantage of a favorable situation"

d. quemarse las pestañas
burn.CL the eyelashes
lit: "burn one's eyelashes"; idm: "study hard"

The four idiomatic phrases in (34) form a subset of the larger idiomatic corpus employed in *Experiment 1*. Recall that *Experiment 1* was conducted using a corpus formed

by twenty sentence pairs. The current experiment was run on four of those twenty sentences. The selection criteria for these four sentences was three-fold:

a. These four expressions were among the better-known expressions by the pool of 13 native-speakers involved in *Experiment 1*.
b. The four expressions had the same syntactic structure, i.e., subject + verb + object (this filtered out 7 pairs that had a second object).
c. The four expressions were produced by the participants involved in *Experiment 1* were absolutely identical (this filtered out statements where there were minute differences between the members of the pair).

4.3.3 Membership to aspectual classes

Each member of the four pairs was tested utilizing the aspectual tests described earlier, in section 4.2. Figure 33 below details each step of the process used to establish membership to an aspectual class.

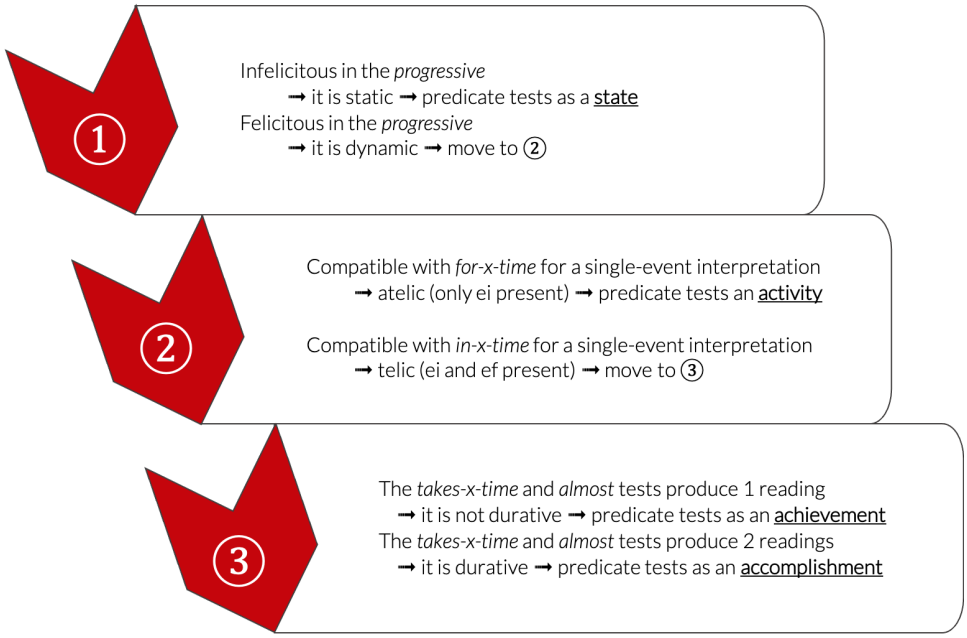


Figure 33: Establishing Aspectual Membership

Following the process outlined in Figure 33, the first step determined if the four state-

ments are dynamic; that is, if they interact in a felicitous manner with the progressive. Those results are seen in (35) below:

(35) a. Firuláis está estirando la pata
Firuláis be stretch.PTCP the leg
Lit: "Firuláis is stretching his leg"; idm: "Firuláis is dying"

b. Pedro está soltando el trapo
Pedro be drop.PTCP the rag
lit: "Pedro is dropping the rag"; idm: "Pedro is bursting out laughing"

c. Pedro está tendiendo las velas
Pedro be unfurl.PTCP the sails
lit: "Pedro is unfurling the sails"; idm: "Pedro is taking advantage of a favorable situation"

d. Pedro está quemándose las pestañas
Pedro be burn.PTCP the eyelashes
lit: "Pedro is burning his eyelashes"; idm: "Pedro is studying hard"

The felicitousness of (35) suggests that all four statements are dynamic in both their literal interpretation and their idiomatic interpretation. In this case, as far as dynamicity is concerned, the literal event and the idiomatic events are the same.

The next step consists of determining telicity for both readings of each pair. To determine whether a predicate denotes a telic event, I'll rely on compatibility with the time-span construction on a single-event reading. Atelicity is confirmed by compatibility with a durative phrase. The results can be seen below in (36) for atelic and (37) for telic events.

(36) a. *Firuláis estiró la pata durante 10 minutos
Firuláis stretch.PRET the leg for 10 minutes
Lit: "*Firuláis stretched his leg for 10 minutes"; idm: "*Firuláis died for 10 minutes"

b. *Pedro soltó el trapo durante 10 minutos
Pedro drop.PRET the rag for 10 minutes

lit: "*Pedro dropped the rag for 10 minutes"; idm: "*Pedro bursted out laughing for 10 minutes"

c. *Pedro tendió las velas durante 10 minutos
Pedro unfurl.PRET the sails for 10 minutes
lit: "*Pedro unfurled the sails for 10 minutes"; idm: "?Pedro took advantage of a favorable situation for 10 minutes"

d. *Pedro se quemó las pestañas durante 10 minutos
Pedro .cl burn.PRET the eyelashes for 10 minutes
lit: "*Pedro burned his eyelashes for 10 minutes"; idm: "?Pedro studied hard for 10 minutes"

The literal members of each of the four pairs are not compatible with the durative phrase, which suggests they are telic events. For the idiomatic members, (36a) and (36b) test as incompatible with the durative (according to my intuition). Sentences (36c) and (36d) do not yield as clear of a reading; I will set that difference aside for the moment to see what the time-span construction reveals about them.

(37) a. Firuláis estiró la pata en 10 minutos
Firuláis stretch.PRET the leg in 10 minutes
Lit: "Firuláis stretched his leg in 10 minutes"; idm: "Firuláis died in 10 minutes"

b. Pedro soltó el trapo en 10 minutos
Pedro drop.PRET the rag in 10 minutes
lit: "Pedro dropped the rag in 10 minutes"; idm: "Pedro bursted out laughing in 10 minutes"

c. Pedro tendió las velas en 10 minutos
Pedro unfurl.PRET the sails in 10 minutes
lit: "Pedro unfurled the sails in 10 minutes"; idm: "Pedro took advantage of a favorable situation in 10 minutes"

d. Pedro se quemó las pestañas en 10 minutos
Pedro .CL burn.PRET the eyelashes in 10 minutes

lit: "Pedro burned his eyelashes in 10 minutes"; idm: "*Pedro studied hard in 10 minutes"

The readings for (37) are a little clearer to get. When interpreted literally, all four sentences in (37) are compatible with the time-span construction. Given that all four literal statements show incompatibility with the durative phrase and compatibility with the time-span construction, I conclude that in their literal interpretation, the examples in (37) are all telic predicates.

When interpreted idiomatically, (37a), (37b), and (37c) are compatible with the time-span construction. Taking both results together, I conclude that (34a), (34b), and (34c) are telic predicates. (34b) shows an aspectual discrepancy between the literal and the idiomatic member of the pair. As things stand, (34d) when interpreted idiomatically, tests as an aspectual activity.

The last step in establishing the aspectual membership of each statement consists of establishing duration. To do that, I employ the *takes-x-time* and *almost* adverbials, as described in section 4.2.1.3). Both adverbial phrases work in the same basic way. Assuming the event is telic, and thus exhibits as part of its meaning both an inceptive telos point and a terminative telos point, the adverbials work by attempting to target each telos point. If the targeting results in a singular reading, the predicate is instantaneous. By contrast, if two interpretations arise, the predicate is durative. Instantaneous events are achievements, whereas durative events are accomplishments.

This last step is critical for two reasons: For most of the sentences in(34), duration is the only difference between the literal and the idiomatic interpretations of the sentences; additionally, this is the key piece of information provided by participants in *Experiment 3*.

4.3.4 Establishing duration

In a nutshell, Experiment 3 consists of asking participants to judge whether each reading in each pair is instantaneous or durative by targeting the event's inceptive and terminative telos points.

In terms of the way the experiment was set up, the four sentences composing the grammatical judgment corpus (34) were paired up with either the *almost* adverbial or the *takes-x-time* adverbial. Expressions (34a) and (34d) were tested with the *almost* adverbial, whereas (34b) and (34c) were paired up with the *takes-x-time* adverbial. Participants were asked to judge just two out of the four possible expressions. The main motivation for this decision was to keep the survey short and reduce cognitive effort and fatigue.

To illustrate the process each participant completed, I will walk through the steps of the survey. Recall that the first part of the survey formed the basis for *Experiment 2*. After participants were finished with that part of the survey, they were presented with a page that introduced the grammatical judgment section. A copy of that page is displayed in Figure 34. A translated version of this page can be found in Appendix H.

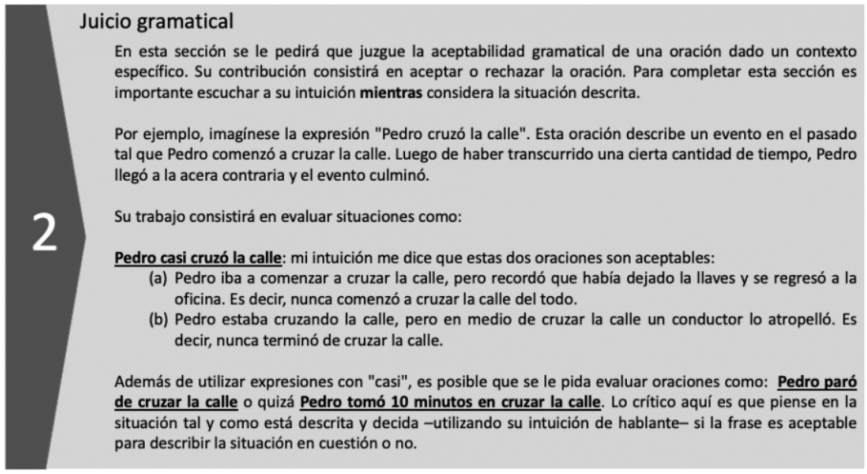


Figure 34: Introduction to Grammatical Judgments

First, the introduction section sought to provide a transition from the perception tasks of *Experiment 2*, so that participants were cognizant that they would switch to a

completely different task. In addition, this page sought to train the participants for the task they were about to perform. It did that by giving a specific example of what the task involved and what kind of responses they were expected to produce.

Behind the scenes, *Qualtrics* paired up which two of the four statements in (34) each participant would judge in a random manner. Once a participant was paired with one of the four sentences, they were asked to judge all four scenarios; this resulted in each participant judging a total of eight sentences.

4.3.4.1 Walkthrough with *estirar la pata* and the *almost* adverbial Consider the expression (34a) *estirar la pata*, as it is made to interact with the *almost* adverbial below in (38).

(38) a. Firuláis casi estiró la pata
Firuláis almost stretch.PRET the leg
Lit: "Firuláis almost stretched his leg"

b. Firuláis casi estiró la pata
Firuláis almost stretch.PRET the leg
idm: "Firuláis almost died"

The literal interpretation of *estirar la pata* and the *almost* adverbial (38a) result in two distinct readings: the counterfactual reading and the incompletive reading. As a result, I conclude that *estirar la pata* is an aspectual accomplishment. In comparison, (38b) yields just one reading: the counterfactual reading. *Estirar la pata*, when interpreted idiomatically, is an achievement.

Below I detail the steps that a participant had to follow when paired up with *estirar la pata* and the *almost* adverbial. The target language text appears first and the free translation is enclosed in parentheses.

1. I briefly introduced what the expression means when used literally and when used idiomatically. For the training material, I decided to used the word *figurative* to refer

to idiomatic readings because I had the sense people would be more familiar with it.

- Literalmente, la expresión **estirar la pata** ocurre cuando un animal extiende una de sus extremidades. En situaciones informales o en broma, la pata se puede utilizar para referirse a la pierna de una persona. (Literally, the expression **stretch one's leg** occurs when an animal stretches one of their limbs. In informal situation or in jest, the word "pata" can be used to refer to the leg of a person.)
- Figurativamente, estirar la pata es una expresión que significa "morir." (Figuratively, **stretch one's leg** is an expression that means to "die").

2. The four crafted scenarios are listed below, although participants interacted with them in random order. The response collected from the participant was the same regardless of the expression or the interpretation. It consisted of a statement of agreement or disagreement as to whether the expression can be used to accurately describe each situation.

    1. **Intended use: literal interpretation & counterfactual reading**
        - Hablando de su mascota Firuláis, ¿puede decir Pedro que "Firuláis casi estiró la pata" para indicar una situación literal donde Firuláis tenía que estirar la pata (para que le revisarán una herida), pero el dolor era tan intenso que no pudo ni siquiera comenzar a estirarla?
        - Speaking of his pet Firuláis, can Pedro say that "Firuláis almost stretched the leg" to indicate a literal situation where Firuláis had to stretch the leg (to have a wound checked), but the pain was so intense that he couldn't even start stretching it?
    2. **Intended use: literal interpretation & incompletive reading**
        - Hablando de su mascota Firuláis, ¿puede decir Pedro que "Firuláis casi estiró la pata" para indicar una situación literal donde Firuláis comenzó a estirar la pata, pero no logró terminar de hacerlo? Es decir, comenzó a

estirarla, pero el dolor le impedía estirarla completamente.

- Speaking of his pet Firuláis, can Pedro say that "Firuláis almost stretched the leg" to indicate a literal situation where Firuláis began to stretch the leg, but could not finish doing it? That is, he began to stretch it, but the pain prevented him from stretching it fully.

3. **Intended use: idiomatic interpretation & counterfactual reading**
    - Hablando de su mascota Firuláis, ¿puede decir Pedro que "Firuláis casi estiró la pata" para indicar una situación figurada donde pasó algo que podría haber causado la muerte de Firuláis, pero terminó no ocurriendo, así que Firuláis sigue vivo? Es decir, no llegó a morir.
    - Speaking of his pet Firuláis , can Pedro say that "Firuláis almost stretched the leg" to indicate a figurative situation where something happened that could have caused Firuláis's death, but ended up not happening, so Firuláis is still alive? That is, he did not die.
4. **Intended use: idiomatic interpretation & incompletive reading**
    - Hablando de su mascota Firuláis, ¿puede decir Pedro que "Firuláis casi estiró la pata" para indicar una situación figurada donde comenzó a morir, pero no logró terminar de morir?
    - Speaking of his pet Firuláis, can Pedro say that "Firuláis almost stretched the leg" to indicate a figurative situation where he began to die, but could not finish dying?

The process described in this section was repeated for the expressions *quemarse las pestañas* which also interacted with the *almost* adverbial. For the adverbial *takes-x-time*, the two expressions tested were *tender las velas* y *soltar el trapo*. All three thorough descriptions can be found in Appendix G.

The following two sections detail the experimental results. I decided to present the results in a staggered manner in order to address two different but related issues. The

first issue has to do with the differences in aspectual structure between the literal and idiomatic events. Section 4.4 sheds light on which of the two opposing views is supported by the empirical evidence gathered in this experiment. In section 4.4.6, I explore how the idiomatic event stands in relationship to the literal event by focusing on the number of aspectual features present in each event.

4.4 Results

In this section, I seek to find experimental validation for the ideas presented in sections 4.2.2 and 4.3.3. Recall that in section 4.2.2, I indicated that one possible way to look at aspectual classes is by grouping them according to the number of aspectual features present in an event. Earlier in section 4.3.3, I followed individual steps to determine which aspectual class each of the four statements tested belongs to. These results are put together in Table 34 below.

Table 34: Expected Results

Statement	Interpretation	Dynamicity	Telicity	Durativity	Features	Class
Estirar la pata						
	literal	yes	yes	yes	3	accomplishment
	idiomatic	yes	yes	no	2	achievement
Quemarse las pestañas						
	literal	yes	yes	yes	3	accomplishment
	idiomatic	yes	no	no	1	activity
Soltar el trapo						
	literal	yes	yes	no	2	achievement
	idiomatic	yes	yes	no	2	achievement
Tender las velas						
	literal	yes	yes	yes	3	accomplishment
	idiomatic	yes	yes	no	2	achievement

According to the results in the column labeled "Features," three out of the four statements tested exhibit a change in aspectual membership; that is to say, the literal interpretation and the idiomatic interpretation of those three sentences denote different events, aspectually speaking. Only one of the statements, *soltar el trapo*, retains its aspectual

class for both interpretations.

As has been mentioned, there are differences of opinion regarding the way the literal event stands in relationship with the idiomatic event from an aspectual point of view. McGinnis (2002, 2005) expresses the view that the literal event specifies the aspectual structure of the idiomatic event and, as a result, they have to be the same[13]. Contrary to this view, Glasbey (2003; 2007) argues that the idiomatic event can, in fact, be different.

Experiment 3 provides a way to assess these two positions by asking participants to judge the two events that result from interpreting a sentence literally and idiomatically using the adverbials *almost* and *it takes-x-time*. The distribution of participants' responses can then be compared. If McGinnis is correct, the distribution of responses should match. If, on the other hand, the distributions do not match, that would support Glasbey's position.

To make explicit what *Experiment 4* sets out to find, let us start by establishing the null hypotheses H_{01}, H_{02}, and the alternative hypothesis, H_1 and H_2.

The null hypothesis H_{01} corresponds to a state of affairs where the idiomatic interpretation exhibits a number of aspectual features equal to the aspectual features of the literal event, as argued in McGinnis (2002, 2005). To put it differently, the aspectual class can be the same, H_{01}, or different, H_1.

The alternative hypothesis H_1 represents Glasbey's (2003; 2007) position. It entertains the idea that the aspectual features of the idiomatic event do not have to match those of the literal event.

[13]unless additional factors are considered, such as differences in agentivity, which in her view lead to differences in clause structure

4.4.1 Observed results for *estirar la pata*

In the experiment, participants were asked to judge duration by judging if the sentences could be used to accurately describe the contexts crafted in section 4.3.4.1. Figure 35 shows the participants' results for *estirar la pata*.

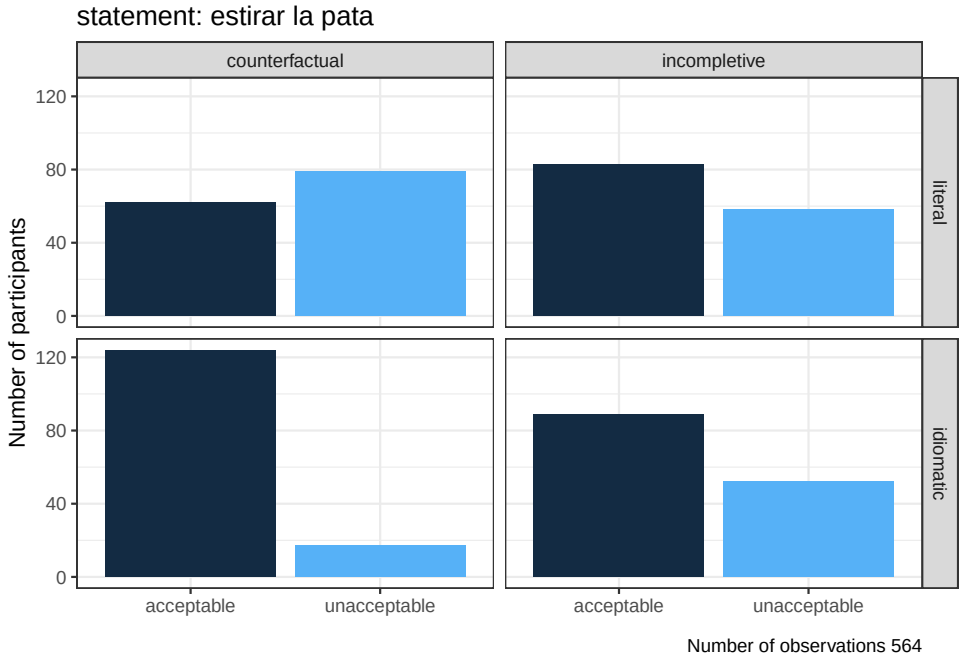


Figure 35: Observed results for estirar la pata

A visual inspection of the observed results shows how participants judged the two interpretations. When interpreted literally, *estirar la pata* is judged to be acceptable only for one scenario, with most participants favoring the incompletive scenario while disfavoring the counterfactual scenario. According to these results, participants conceptualize literal *estirar la pata* as an aspectual achievement.

The results for the idiomatic interpretations show that participants judged as acceptable both the counterfactual and the incompletive scenario. Accordingly, idiomatic *estirar la pata* is judged as an accomplishment.

4.4.2 Observed results for *quemarse las pestañas*

The second statement tested with the *almost* adverbial was (34d) *quemarse las pestañas* "burn one's eyebrows/study hard." In the experiment, participants were asked to judge the acceptability of each event when used to describe two scenarios. The scenarios to judge were described above, in section 4.3.4.1. Figure 36 shows the participants' results for *quemarse las pestañas*.

The expectation is that participants will judge as grammatical both crafted scenarios for the literal interpretation, but will find that, for the idiomatic interpretation, only one of the crafted scenarios is acceptable, the counterfactual.

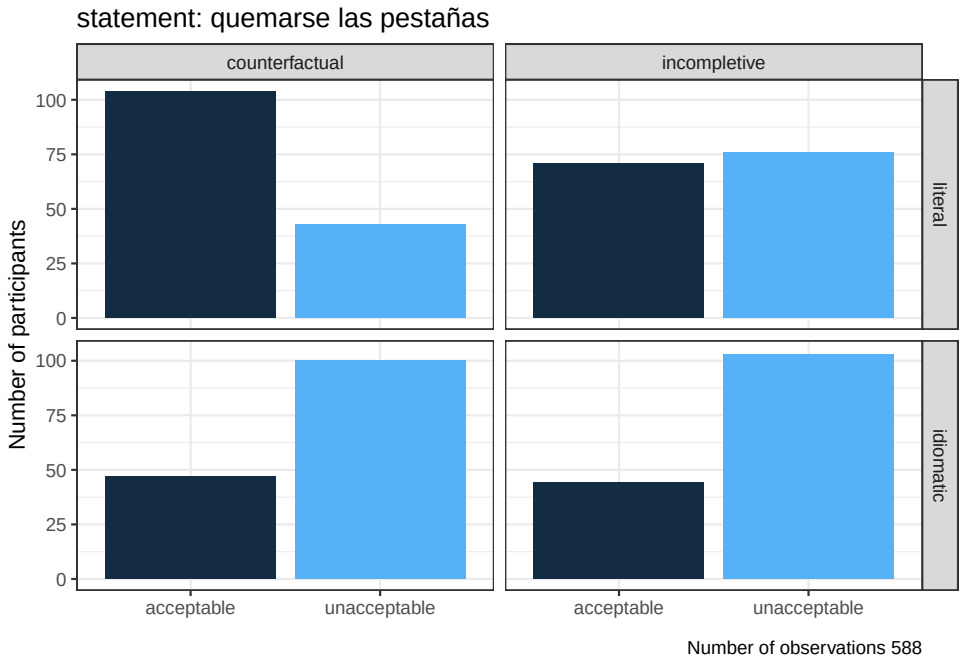


Figure 36: Observed results for quemarse las pestañas

Upon inspection of the observed results, we see that, when interpreted literally, *quemarse las pestañas* is judged to be acceptable only for one scenario. Most participants favored the counterfactual scenario and disfavored the incompletive scenario. According to these results, participants conceptualize literal *quemarse las pestañas* as an aspectual achievement.

The results for the idiomatic interpretations show that participants judged as unacceptable both the counterfactual and the incompletive scenario. However, given the independent results from (37), which confirm its atelicity, I will move forward assuming that *quemarse las pestañas* is, in fact, an activity.

4.4.3 Observed results for *soltar el trapo*

The expression (34b) *soltar el trapo* "drop the rag/burst out laughing" is one of the two statements tested with the *take-x-time* adverbial. The task participants were asked to complete was similar to the previous tasks already described, in that for each interpretation, participants were asked to judge two scenarios, one targeting the start of the event and the other targeting the end of the event. The scenarios to judge were described Appendix F.

Figure 37 below shows the participants results for *soltar el trapo*.

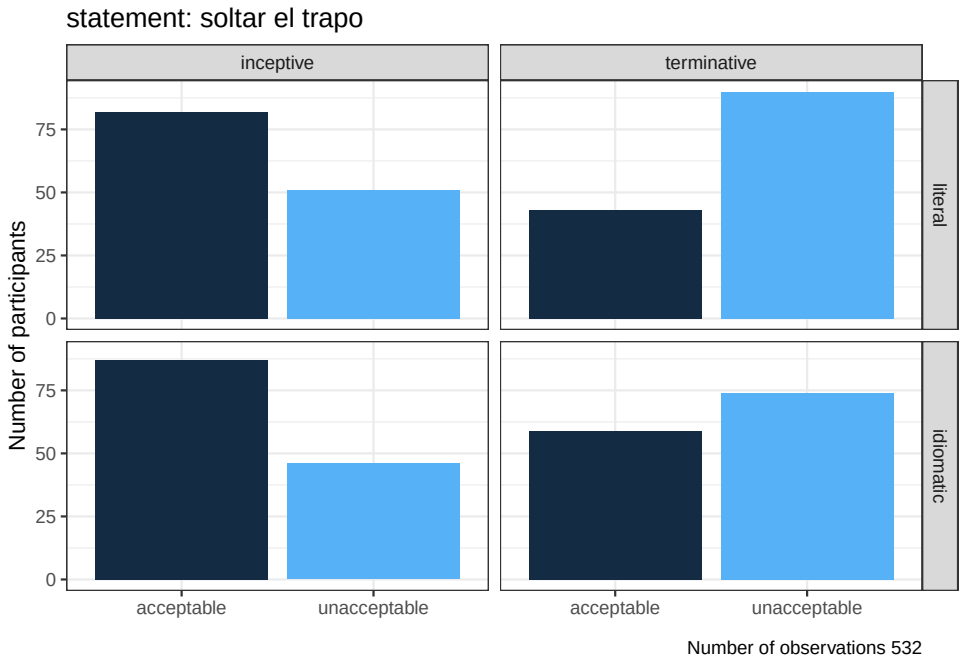


Figure 37: Observed results for soltar el trapo

The results for *soltar el trapo* show that, for both interpretations, participants do not

judge the scenarios as equally acceptable. A majority of the participants favored the inceptive reading while disfavoring the terminative reading. In aspectual terms, participants judged both interpretations to be achievements.

4.4.4 Observed results for *tender las velas*

(34c) *tender las velas* "unfurl the sails/take advantage of a situation" is the second statement tested with the *takes-x-time* adverbial. The task participants were asked to complete was similar to the previous tasks already described in that it was composed of four scenarios, each of which attempted to target the start and the end of the events.

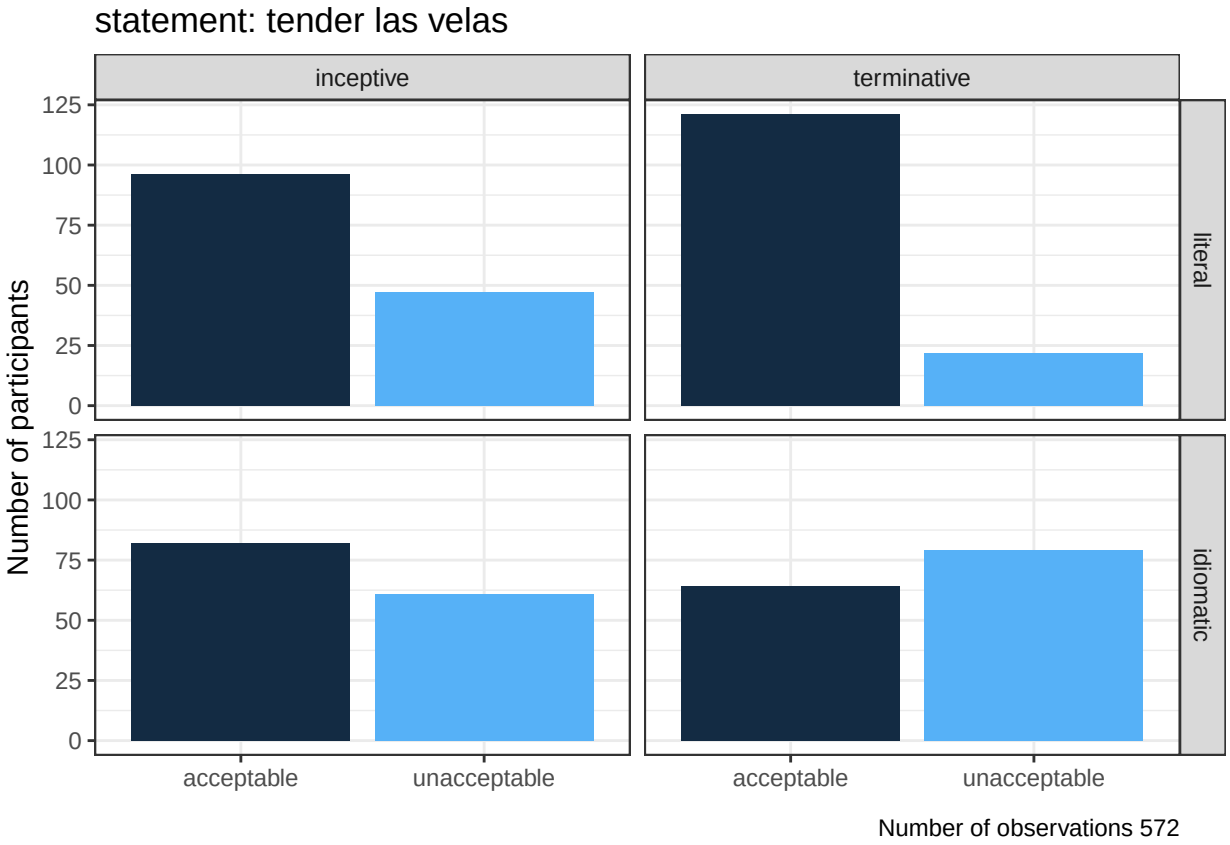


Figure 38: Observed results for tender las velas

Figure 38 shows that the participants judged the two interpretations differently. For the literal interpretation, participants judged both scenarios as acceptable, and as a result, literal *tender las velas* is an accomplishment. For the idiomatic interpretation, participants

favored the inceptive reading while disfavoring the terminative reading. Idiomatic *tender las velas* is found to be an achievement.

4.4.5 Discussion of Results for H_1

The overall objective of Section 4.4 was to provide experimental validity for the conflicting claims about the aspectual structures that result from interpreting an idiom literally or idiomatically which are addessed by H_{01} and H_1

The experimental results are summarized as follows:

- For *estirar la pata*
  - literal interpretation: only the incompletive scenario is judged as acceptable → literal *estirar la pata* is an **achievement**.
  - idiomatic interpretation: both the counterfactual and the incompletive scenario are judged as acceptable → idiomatic *estirar la pata* is an **accomplishment**.
- For *quemarse las pestañas*
  - literal interpretation: only the counterfactual scenario is judged as acceptable → literal *quemarse las pestañas* is an **achievement**.
  - idiomatic interpretation: both the counterfactual and the incompletive scenario are judged as unacceptable → idiomatic *quemarse las pestañas* is an **activity**.
- For *soltar el trapo*
  - literal interpretation: only the inceptive scenario is judged as acceptable → literal *soltar el trapo* is an **achievement**.
  - idiomatic interpretation: only the inceptive scenario is judged as acceptable → idiomatic *soltar el trapo* is an **achievement**.
- For *tender las velas*
  - literal interpretation: both the inceptive and the terminative scenario are judged as acceptable → literal *tender las velas* is an **accomplishment**.
  - idiomatic interpretation: only the inceptive scenario is judged as acceptable →

idiomatic *tender las velas* is an **achievement**.

All these results taken together constitute counter evidence to the claim that the aspectual structure of the two events must be same. As seen above, the events can have the same aspectual structure, as is the case for *soltar el trapo*, but they do not need to match, as was demonstrated for *estirar la pata*, *quemarse las pestañas*, and *tender las velas*.

In sum, these results provide evidence to reject H_{01} and consider the alternative H_1; that is, the literal and idiomatic events do not need to belong to the same aspectual class.

4.4.6 Results for H_2

Now that we have settled on the fact that the aspectual class of literal and idiomatic events can be different, this section explores *the ways in which* they can be different. In the course of working with the sentences that formed my corpus of idiomatic expressions, I observed that this difference seems to be directional. Namely, for all the idiomatic statements I have come across, I have noticed that the two members of the pair stand in a *defective relationship*, where the idiomatic member can have as many aspectual features as the literal member, but critically, no more.

This observation is novel and challenges the common view that there is ungoverned arbitrariness in what an idiomatic phrase can mean regardless of its literal counterpart, as exemplified in this quote "We all know about idioms — phrasal combinations of morphemes with unpredictable non-compositional meanings —" (Marantz 1996:5). This part of the chapter attempts to find support for the view that aspectually, at least, we find a constraint that limits what idioms can mean.

To put it in a different way, if idioms can freely modify the interpretation of an expression, we should expect to see sentences that are literal states (STT), but become idiomatic activities (ACT), or achievements (ACH), or accomplishments (ACC). In fact, we

should be able to see idiomatic re-interpretations go in all possible directions and cover all possible combinations.

Figure 39 below, provides a visual representation of what we should expect to see naturally occurring if idiomaticity is completely (aspectually) unconstrained. The horizontal axis shows the four possible aspectual classes a literal event can be, while the vertical axis shows the four possible aspectual classes of the idiomatic event. Assuming complete arbitrariness, we should be able to find idioms manifesting all possible combinations, which I represent here with a green shade.

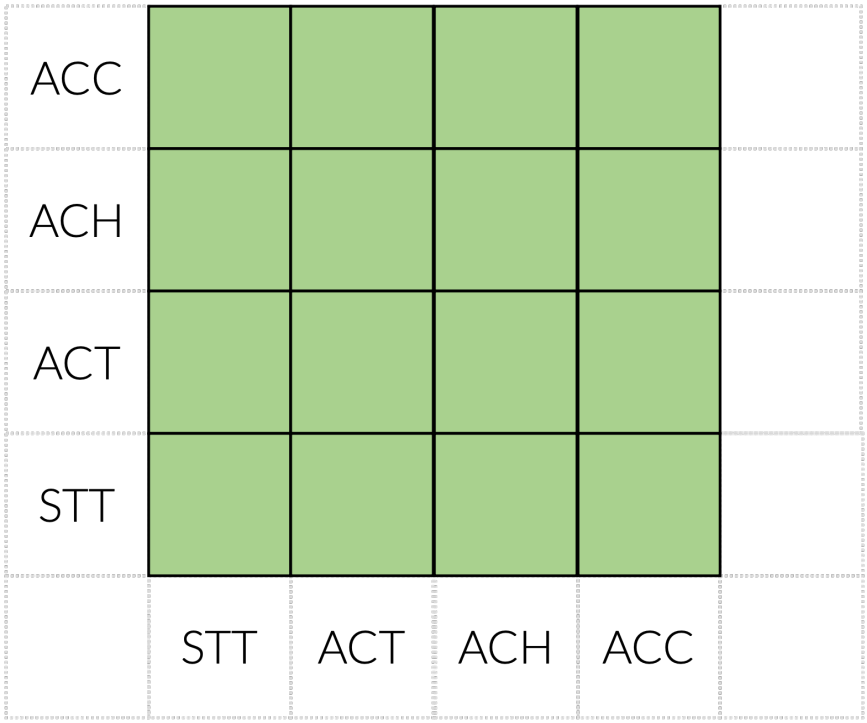


Figure 39: Expected Literal-Idiomatic

What I found instead, is that the literal aspectual class does appear to limit the aspectual class of the idiom. For the idiomatic expressions in my corpus, I did not find instances where the idiomatic interpretation was part of an aspectual class with a greater number of aspectual features, or higher up in the y the axis than the literal event. To put it plainly, it seems as if the literal interpretation limits the idiomatic interpretation. This

state of affairs in represented below in Figure 40 where the idiomatic interpretation can go as high as the literal interpretation or lower.

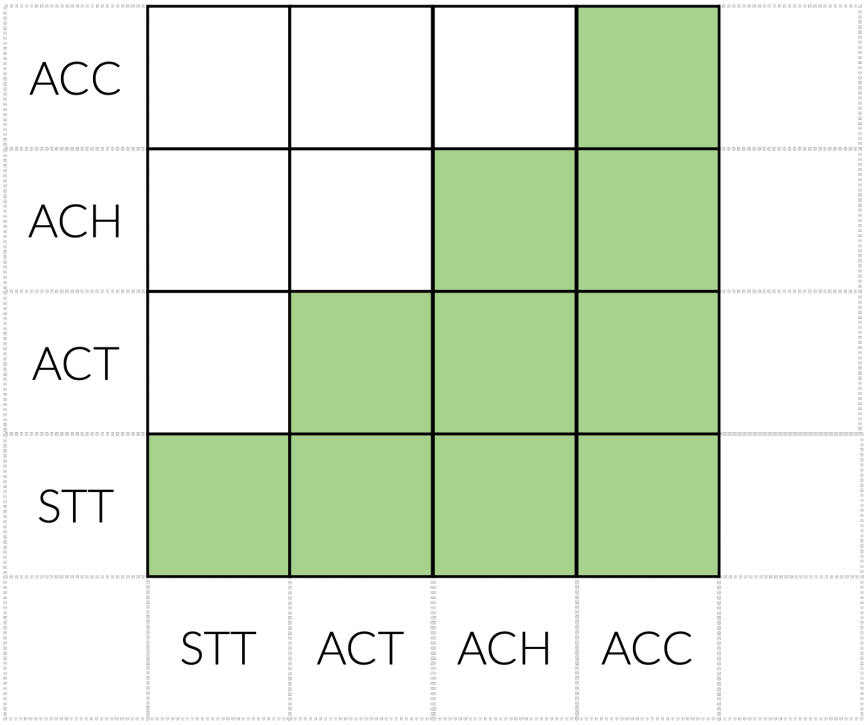


Figure 40: Expected Literal-Idiomatic

To find evidence that this may be the case, I rely on inductive reasoning, the intuitions of the participants, and some statistical way of comparing these intuitions with the distributions of raw counts from the participants' answers.

The aspectual tests described earlier give us a baseline for the aspectual membership of the statements. I call these *expected results*. The responses from the participants can also be used to experimentally determine the aspectual membership of the idiomatic phrases. These are called *observed results*.

These results form a count distribution that can be compared to get an objective sense of how similar/dissimilar the expected and observed results are across interpretations (literal vs idiomatic) and across scenarios (start-of-event vs end-of-event).

- *soltar el trapo* exhibits no loss of aspectual features. This means that the distribu-

tion of responses from the participants should be about the same for both the literal and the idiomatic interpretations.

- *estirar la pata* and *tender las velas* exhibit a loss of one aspectual feature, whereas *quemarse las pestañas* exhibits a loss of two aspectual features. For these three sentences, participants' responses are expected to be distributed differently.

The alternative hypothesis H_2, therefore, corresponds to a state of affairs where the idiomatic interpretation stands in a *defective relationship* to the literal interpretation. In other words, the H_2 describes situations where the aspectual class can either be maintained or not, but there is a systematic relationship between the two – if different, the number of features of the idiomatic interpretation must be fewer than those in the literal one.

By contrast, the opposite view, held by H_{02}, is that the number of aspectual features of the idiomatic event can, in fact, exceed the number of aspectual features of the literal event.

In order to validate these two views, the experiment went into the world and asked participants to judge by proxy these relationships. It sought to accomplish this by asking participants to judge the duration of both literal and idiomatic events, assuming that both events are in fact durative, and therefore, supporting H_{02}.

It is important to note that based on the *expected results*, there are two kinds of evidence being collected. For *soltar el trapo*, the evidence I hope to gather is weaker, as I predict that participants will identify the aspectual class of *soltar el trapo* as an achievement and proceed to collect the responses to find out if they do.

The test for the remaining three expressions makes a stronger prediction; it predicts that crafting scenarios that are consistent with H_{02} will result in unacceptable grammatical judgments by the participants. The state of affairs embodied by H_{02} is, of course,

contrary to the expected results detailed in Table 34.

In terms of how this will be tested, for the statement *soltar el trapo* 'drop the rag,' participants' responses should be equally distributed, as shown in Table 34, where the aspectual class of the literal and the idiomatic interpretations is, in both cases, an achievement.

For the remaining three statements, participants are asked to judge a scenario that is not expected to exist. This carries the expectation that participants should systematically reject those three sentences. If the results are consistent, I will have grounds to disregard H_{02} and consider the validity of the alternative hypothesis. As Hales (2019) puts it, "The very nature of an inductive argument is to make a conclusion probable, but not certain, given the truth of the premises."

4.4.7 Quantifying observed results

The descriptions I provided in sections 4.4.1 - 4.4.4 constitute a bird's-eye view of how participants reacted to literal and idiomatic interpretations as they judged how accurate those sentences were at describing the crafted scenarios that were designed to highlight the beginning or the end of an event. The assumption is that a durative event would result in both scenarios being judged as acceptable.

In this section, I review those same results with a more quantitative eye in order to gather evidence to reject or uphold H_{02}. To accomplish that, I performed a chi-squared test on the same count results presented earlier. A chi-squared test is a hypothesis testing method that is used to determine whether there are statistically significant differences between expected frequencies and observed frequencies. These observed values must correspond to mutually exclusive categories. The experimental design meets this assumption, as participants provided their judgments through a forced choice task; that is to say, participants could only indicate that they judge the statement to be acceptable to describe a

given scenario or not.

Through my experimental design, I compared the participants' literal and idiomatic judgments within each of the scenarios. By using the aspectual tests described in section 4.3.3, I independently determined whether the distribution of literal and idiomatic judgments is supposed to be the same or different. With this baseline established, I can then get a sense of how close the observed results are to each other across scenarios and/or across interpretations.

The interpretation of the chi-squared results is fairly straightforward; a statistically significant result indicates that the distributions are **not** the same, and consequently, that there is sufficient evidence to reject H_{02}, i.e., we should entertain the idea of accepting H_2. On the other hand, p-values above 0.05 indicate that the distributions are the same, and therefore, the evidence is insufficient to reject H_{02}.

To recap, Table 34 presented a summary of what the aspectual tests tell us to expect about the four statements tested. According to those results,

- *estirar la pata* is expected to be a literal accomplishment and an idiomatic achievement. The responses for the counterfactual scenarios should be the **same** for both the literal and idiomatic interpretations. However, for the incompletive scenario, only the literal interpretation should be acceptable; as a result, the distribution of responses should be **different**.

- *quemarse las pestañas* is expected to be a literal accomplishment and an idiomatic activity. The responses should pattern the same way as *estirar la pata*; i.e., the distribution of responses should be the **same** for the inceptive scenario, but **different** for the terminative scenario.

- *soltar el trapo* is expected to be a literal achievement and an idiomatic achievement. The distribution of responses should be the **same** for both the inceptive scenario and

the terminative scenario across both interpretations.

- *tender las velas* tested as a literal accomplishment and an idiomatic achievement. The distribution of responses should be the **same** for the inceptive scenario, but **different** for the terminative scenario.

With those expectations clearly stated, the results for the two statements that were tested with the *almost* adverbial are described in the following subsections.

4.4.7.1 *Estirar la pata* Given that *estirar la pata* tested as a literal accomplishment and an idiomatic achievement, I expected the two interpretations to differ only in duration. Participants were expected to **accept both literal interpretations**, but **accept only one** idiomatic interpretation.

Participants' responses for the counterfactual contexts can be seen in Table 35. Participants found the counterfactual scenario unacceptable for the literal interpretation. By comparison, participants found acceptable the counterfactual scenario for the idiomatic interpretation.

Table 35: Estirar la pata - counterfactual readings

	Literal	Idiomatic
Acceptable	62	124
Unacceptable	79	17

Table 36 summarizes the results for the incompletive scenario. Participants found the incompletive scenario to be acceptable for both the literal and the idiomatic interpretations.

Table 36: Estirar la pata - incompletive readings

	Literal	Idiomatic
Acceptable	83	89
Unacceptable	58	52

The expected results are not borne out by the data. For the literal interpretation, participants were expected to accept both the counterfactual and the incompletive readings. In reality, more participants found the counterfactual scenario unacceptable, but found the incompletive reading acceptable.

Table 37 shows the results of a chi-squared test run on the literal responses. It confirms that the counterfactual vs incompletive distributions are not the same, contrary to expectations.

Table 37: chi-squared results for literal 'estirar la pata'

statistic	p-value	method
5.68	0.02	Pearson's Chi-squared test with Yates' continuity correction

For the idiomatic interpretation, the expectation is that participants find acceptable only one of the scenarios. The reality, however, is that the majority of participants found both interpretations to be acceptable. However, both scenarios are not accepted equally. Table 38 shows that the distributions of responses are significantly different. We also know from Tables 35 and 36 that participants favored the counterfactual scenario over the incompletive.

Table 38: chi-squared results for idiomatic 'estirar la pata'

statistic	p-value	method
22.18	0	Pearson's Chi-squared test with Yates' continuity correction

The results for the idiomatic interpretation are problematic for H_2. Having participants accept both readings suggests that *estirar la pata* behaves as an accomplishment, and therefore, exhibits one more aspectual feature than its literal counterpart.

4.4.7.2 ***Quemarse las pestañas*** The expression *quemarse las pestañas* was also tested with the adverbial *almost*. The two scenarios crafted to test the expression were a counterfactual and an incompletive. *Quemarse las pestañas* is expected to behave as a literal accomplishment and an idiomatic activity.

I expected the two interpretations to differ not only in duration but also in telicity. Even though participants were formally tested only to judge duration, recall that aspectual features are hierarchically stacked in such a way that achievements and activities both lack durativity. Therefore, a scenario that presupposes duration *should* be judged ungrammatical also.

Table 39 shows the observed responses for the counterfactual scenarios of *quemarse las pestañas*. For the literal interpretation, participants found the counterfactual scenario to be acceptable. For the idiomatic interpretation, participants found the counterfactual scenario unacceptable.

Table 39: Quemarse las pestañas - counterfactual readings

	Literal	Idiomatic
Acceptable	104	47
Unacceptable	43	100

Table 40 presents the participants' responses for the incompletive scenario. For the literal interpretation, participants found the incompletive scenario unacceptable. For the idiomatic interpretation, participants also found the incompletive scenario unacceptable.

Table 40: Quemarse las pestañas - incompletive readings

	Literal	Idiomatic
Acceptable	71	44
Unacceptable	76	103

We can probe the distributions for the literal responses across both scenarios. Partici-

pants are expected to accept both interpretations, given that in *quemarse las pestañas*, the application of the aspectual tests indicated a literal accomplishment. The results shown in Table 41, however, indicate that this is not the case; participants only accepted the counterfactual scenario, not the incompletive. For these participants, literal *quemarse las pestañas* behaves as an aspectual achievement.

Table 41: chi-squared results for literal 'quemarse las pestañas'

statistic	p-value	method
14.46	0	Pearson's Chi-squared test with Yates' continuity correction

For the idiomatic interpretation of *quemarse las pestañas* "to study hard," the expectation was that participants would reject the incompletive scenario but accept the counterfactual, as activities do not exhibit duration but do have an inceptive telos. This result is not borne out by the responses. Probing participants' responses further, we see that they judged as ungrammatical both scenarios about the same, as the results of the chi-squared tests in Table 42 show.

Table 42: chi-squared results for idiomatic 'quemarse las pestañas'

statistic	p-value	method
0.06	0.8	Pearson's Chi-squared test with Yates' continuity correction

4.4.7.3 ***Soltar el trapo***

Soltar el trapo "drop the rag/burst out laughing" is one of the two expressions tested with the adverbial *takes-x-time*. This adverbial operates similarly to the way the *almost* adverbial does. It can give rise to two different readings according to what telos point it targets, assuming that the predicate is also durative. The two crafted scenarios used in the experiment sought to be compatible with the inceptive telos point for one scenario and the terminative telos point for the other. As before, accomplishments should be acceptable for both scenarios, whereas achievements should be

acceptable in only one of the scenarios.

Soltar el trapo was expected to be an achievement for both the literal and the idiomatic interpretation. The expectation was that participants' responses should be acceptable for only one of the scenarios. Table 43 shows the distribution of responses for the inceptive scenario. Table 44 shows the distribution for the terminative scenario.

Table 43 shows that, for the literal interpretation, participants found the inceptive scenario to be acceptable. For the idiomatic interpretation, participants found the inceptive scenario to be acceptable as well.

Table 43: Soltar el trapo - inceptive readings

	Literal	Idiomatic
Acceptable	82	87
Unacceptable	51	46

Table 44 shows the results for the terminative scenarios. For the literal interpretation, participants found the terminative scenario to be unacceptable. For the idiomatic interpretation, participants also found the terminative scenario to be unacceptable.

Table 44: Soltar el trapo - terminative readings

	Literal	Idiomatic
Acceptable	43	59
Unacceptable	90	74

Recall from Figure 37 that the distribution was remarkably similar both across interpretations and scenarios, as was expected from the application of aspectual tests. Only one of the two scenarios is acceptable for both the literal and idiomatic interpretations. Furthermore, participants accepted the inceptive reading over the terminative reading for both interpretations.

The application of a chi-squared test to the literal results indicates that the two distributions of responses are indeed different, as shown in Table 45. The distribution of responses for the idiomatic is also different, as shown in Table 46. In both cases, the results confirm that participants are judging only one of the scenarios as acceptable.

Table 45: chi-squared results for literal 'soltar el trapo'

statistic	p-value	method
21.79	0	Pearson's Chi-squared test with Yates' continuity correction

Table 46: chi-squared results for idiomatic 'soltar el trapo'

statistic	p-value	method
11.07	0	Pearson's Chi-squared test with Yates' continuity correction

We can also get a sense that the rate of acceptability with which participants judged both scenarios is the same across interpretations. From Figure 37, it is evident that participants found the inceptive scenario acceptable and judged the incompletive scenario as unacceptable. Judging by the visuals, it looks as if the rate of acceptability is the same across interpretations. Tables 47 and Table 48 confirm this; both scenarios have p-values $>$ 0.05, which means they have the same distribution.

Table 47: chi-squared results for inceptive 'soltar el trapo'

statistic	p-value	method
0.259623	0.6103793	Pearson's Chi-squared test with Yates' continuity correction

Table 48: chi-squared results for terminative 'soltar el trapo'

statistic	p-value	method
3.577834	0.0585555	Pearson's Chi-squared test with Yates' continuity correction

4.4.7.4 ***Tender las velas*** The expression *tender las velas* "unfurl the sails/take advantage of a situation" was the second expression tested with the *takes-x-time* adverbial. The crafted scenarios described in Appendix G were used to target the inceptive telos and the terminative telos. Table 34 shows that, according to the aspectual tests, *tender las velas* is expected to be a literal accomplishment but an idiomatic achievement. Therefore, the expectation was that participants would accept both scenarios for the literal interpretation, but only one of the scenarios for the idiomatic interpretation.

Table 49 shows the participants' responses. For the literal interpretation, participants judged the inceptive scenario as acceptable. For the idiomatic interpretation, participants also found the inceptive scenario acceptable.

Table 49: Tender las velas - inceptive readings

	Literal	Idiomatic
Acceptable	96	82
Unacceptable	47	61

Table 50 presents the terminative scenario results. For the literal interpretation, participants found the terminative scenario acceptable. For the idiomatic interpretation, however, participants found the terminative scenario to be unacceptable.

Table 50: Tender las velas - terminative readings

	Literal	Idiomatic
Acceptable	121	64
Unacceptable	22	79

The observed results for *tender las velas* matched exactly what was expected, i.e., both scenarios are acceptable in the literal interpretation, but only one scenario is found to be acceptable for the idiomatic interpretation. Table 51 shows that a chi-squared test

conducted on the literal responses reveals that participants did not judge both scenarios equally. Participants favored the terminative scenario over the inceptive one.

Table 51: chi-squared results for literal 'tender las velas'

statistic	p-value	method
10.95	0	Pearson's Chi-squared test with Yates' continuity correction

For the idiomatic interpretation, participants' responses also patterned as expected, judging as acceptable only one of the scenarios. Participants judged the inceptive scenario as acceptable, but indicated that the terminative scenario was unacceptable to describe the idiomatic event. The chi-squared test in Table 52 shows that the distribution seen in these responses is statistically significant at the 0.05 threshold of significance, and therefore, different.

Table 52: chi-squared results for idiomatic 'soltar el trapo'

statistic	p-value	method
4.04	0.04	Pearson's Chi-squared test with Yates' continuity correction

The way the responses patterned for each scenario showed that participants judged as acceptable both interpretations about the same for the inceptive scenario, as shown in Table 53, where a p-value > 0.05 indicates that the distributions are the same.

Table 53: chi-squared results for inceptive 'tender las velas'

statistic	p-value	method
2.514253	0.1128211	Pearson's Chi-squared test with Yates' continuity correction

For the terminative scenario, however, participants' responses diverged noticeably. Participants judged as acceptable the literal interpretation of the event while rejecting the

idiomatic interpretation. Table 54 shows that the distribution for the observed responses is as different as they can be, with a p-value of virtually zero.

Table 54: chi-squared results for inceptive 'tender las velas'

statistic	p-value	method
48.00086	0	Pearson's Chi-squared test with Yates' continuity correction

4.5 Discussion about Observed vs Expected Results

Table 55 summarizes the observed results from the participants' grammatical judgments and contrasts them with the expected results obtained from the application of aspectual tests described in section 4.3.4. Participants' responses patterned as expected when they judged sentences containing the *takes-x-time* adverbial. Participants judging sentences containing the *almost* adverbial did not provide a clear picture, perhaps even antagonistic to H_2. In the reminder of this section I will describe these results in detail.

Table 55: Expected vs Observed Results

Intent	Exp. # Feat	Exp. Class	Exp. Readings	Exp. Incept.	Exp. Term.	Obs. Readings	Obs. Incept.	Obs. Term.	Obs. Class
Estirar la pata									
literal	3	accomplishmen	2	yes	yes	1	no	yes	achievement
idiomatic	2	achievement	1	no	yes	2	yes	yes	accomplishment
Quemarse las pestañas									
literal	3	accomplishmen	2	yes	yes	1	yes	no	achievement
idiomatic	1	activity	1	yes	no	0	no	no	activity
Soltar el trapo									
literal	2	achievement	1	yes	no	1	yes	no	achievement
idiomatic	2	achievement	1	yes	no	1	yes	no	achieve-ment
Tender las velas									
literal	3	accomplishmen	2	yes	yes	2	yes	yes	accomplishment
idiomatic	2	achievement	1	yes	no	1	yes	no	achievement

Recall that the *almost* adverbial works in two different ways, depending on whether the start or the end of the event is being targeted. If the target is the inceptive telos point, it results in a *counterfactual* reading, characterized by a state of affairs that critically did not even start. If, by contrast, the target is the end of the event, the resulting reading is *incompletive*. In this case, the state of affairs denoted by the predicate started, but did not

culminate. The *almost* adverbial is used to differentiate telic predicates that have duration versus those that are not durative. In practical terms, I used the *almost* adverbial to tell apart accomplishment-type predicates from the other two types of dynamic predicates.

Similar to the *almost* adverbial, the *takes-x-time* adverbial also works by targeting both telos points. The resulting readings are, however, different. When the target is the inceptive telos, it results in an *inceptive* reading that denotes the time it takes for the event to start. By comparison, when the target is the terminative telos, the resulting reading is *terminative* and denotes the time that lapses from the moment the event starts to the time when it culminates.

With that as a backdrop, consider the specific results below.

4.5.1 *Almost* adverbial

4.5.1.1 *Estirar la pata* Both statements tested with the counterfactual adverbial produced results that do not match the expected behavior for neither literal interpretations nor idiomatic interpretations.

For the statement *estirar la pata*, the literal interpretations for both the counterfactual and the incompletive readings were expected to be acceptable, but only the incompletive reading was observed to be acceptable by the participants. This is not necessarily a huge problem, as it is known that achievement-like predicates are sometimes understood to include an activity reading event that, although not technically part of the predicate, is difficult to dissociate from the event itself (Kearns 2003).

The results for the idiomatic interpretations are more problematic. The expectation was that only one of the two scenarios would be accepted by the participants, but the observed results indicate that participants accepted both the counterfactual and the incompletive readings. This, in turn, is evidence in favor of an accomplishment interpretation for *estirar la pata*; that is to say, aspectually speaking, *estirar la pata* is aspectually more

complex than literal *estirar la pata*, a clear violation of H_2.

4.5.1.2 *Quemarse las pestañas*

The observed results for literal *quemarse las pestañas* are also problematic in that only the counterfactual reading was found to be acceptable by the participants. This is contrary to the expectation that both the counterfactual and the incompletive would be acceptable. According to these results, participants judged *quemarse las pestañas* as a literal achievement, as opposed to the accomplishment-like predicate predicted by the application of the aspectual tests.

For the idiomatic interpretation, the results for *quemarse las pestañas* did not conform to the expectations in that neither the counterfactual nor the incompletive scenarios were judged as acceptable by the participants. This presents a problem in that at least one of the scenarios should have been judged acceptable. From these results alone, I cannot gather evidence to judge the duration of idiomatic *quemarse las pestañas*. However, the independent evidence from the progressive tests, (see (35d)) that indicates that idiomatic *quemarse las pestañas* is an activity.

4.5.2 *Takes-x-time* adverbial

In contrast to the results observed with the *almost* adverbial, the *take-x-time* adverbial produced results that matched exactly the expectations set by the tests from section 4.3.4.

4.5.2.1 *Soltar el trapo*

Soltar el trapo was expected to behave as an achievement both for the literal and the idiomatic interpretation. The distribution of participants' responses confirmed that both interpretations are accepted for just one of the scenarios. Furthermore, participants' responses agree that the inceptive scenario is judged acceptable, while the terminative scenario is judged as unacceptable for both interpretations.

4.5.2.2 ***Tender las velas***

Tender las velas was expected to be an aspectual accomplishment for the literal event. The results from the participants confirmed that they judged the literal event as an accomplishment by accepting both scenarios. For the idiomatic event, the expectation was that the participants would judge the event as an achievement. The observed results matched this expectation, as participants judged only the inceptive scenario as acceptable.

In addition to the results described above, participants' responses also provide negative evidence by matching the prediction that participants will judge as ungrammatical an idiomatic event that is assumed to have more aspectual features than its counterpart literal event. Indeed, participants found this scenario unacceptable, thus lending support to the claim that the idiomatic interpretation of an event cannot be aspectually more complex that its concomitant literal event.

4.5.3 Are "good guessers" a factor?

The experiment involving *estirar la pata* presents at least two obvious difficulties. For one, the responses from the literal event did not match the expected aspectual class set up by the application of the aspectual tests described in Figure 33. Contrary to the expectation of being accomplishments, participants judged the literal event as achievements, regardless of nativeness.

For another, the responses about the idiomatic event indicate that participants conceptualize idiomatic *estirar la pata* not as an instantaneous event, but rather one with duration, which is an obvious problem for the hypothesis I put forth. However, as mentioned earlier, achievement-like predicates can be sometimes hard to interpret, given that it is common for speakers to add an activity-like phase leading to the culmination of an event proper. Kearns (2003:634) describes how an independent event, such as *journeying*, can be seen as a supporting event that leads to an *arriving* event, which in turn, is interpreted as the natural culmination of *journeying*.

Given the way the dataset was collected, it would be impossible to further test the participants to determine if they are entertaining this type of bi-eventual structure for idiomatic *estirar la pata* with their responses.

However, seeing how the results suggest competing views of the event, I set out to explore what would happened if I isolated a subset of the overall results by considering the grammatical judgments from *only* those participants whose judgments for the literal event matched the expected results provided by the aspectual tests. The intuition behind this was to find out if the participants whose responses matched the answers predicted by the aspectual tests have anything else different to say about the idiomatic event.

The expectation is that participants whose answers match the results of the aspectual tests would also produce answers in line with the alternative hypothesis H_2. The total number of participants who interacted with the statement *estirar la pata* was 141. Of those, 62 had responses matching the aspectual tests for the counterfactual scenarios. For the incompletive scenarios, 83 participants had responses matching the tests .

Recall that literal *estirar la pata* "to stretch one's leg" should correspond to an accomplishment, whereas idiomatic *estirar la pata* "die" should be an achievement. Figure 41 shows the distribution of responses from the subset of participants whose answers for the literal event matched the results of the aspectual tests.

For these participants, the number of acceptable responses for the counterfactual scenario far surpasses the number of acceptable responses for the incompletive scenario, which would make idiomatic *estirar la pata* an achievement.

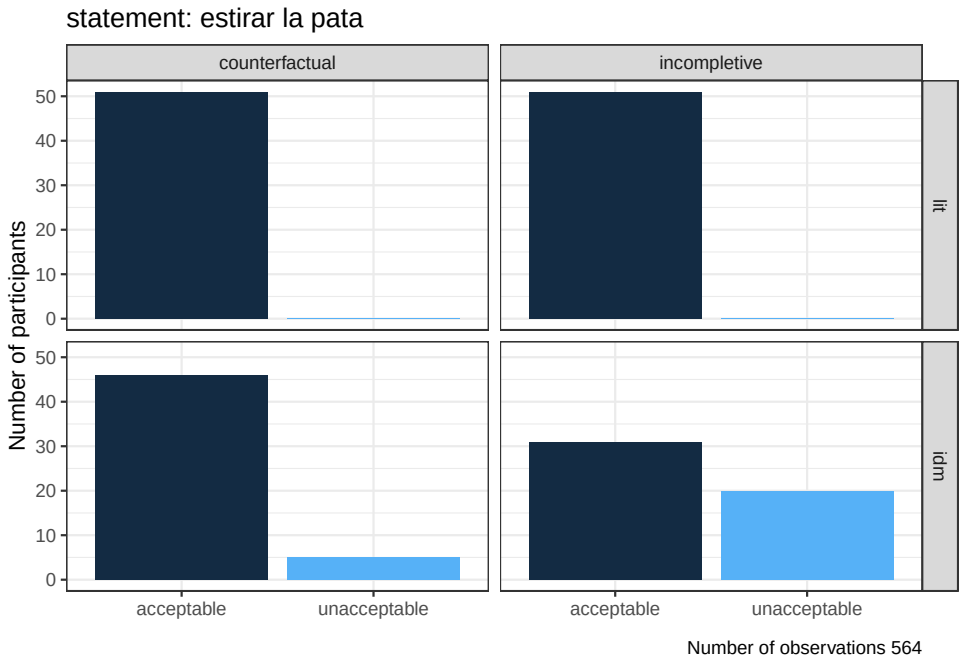


Figure 41: Responses for estirar la pata

To further investigate to this claim, I proceeded to compare the count distributions of the responses. The expectation is that participants would find equally acceptable the counterfactual scenarios for both interpretations. By contrast, participants should find acceptable the incompletive scenario for the literal event only.

In other words, the results of a chi-squared test applied to the counterfactual scenarios should not be significant because the distributions should be the same; that is, its p-value should be greater than 0.05. For the incompletive scenarios, the expectation is that the p-value would be significant, which would indicate that the distributions of responses are different, statistically speaking.

The two distributions to be tested are shown below in Table 56 for the counterfactual scenario and Table 57 for the incompletive scenario.

Table 56: Estirar la pata - counterfactual scenario

	Literal	Idiomatic
Acceptable	51	46
Unacceptable	0	5

Table 57: Estirar la pata - incompletive scenario

	Literal	Idiomatic
Acceptable	51	31
Unacceptable	0	20

Indeed, the results of chi-squared tests indicate that the two interpretations are the same for the counterfactual scenario, as shown in Table 58, and different for the incompletive scenario, as shown in Table 59.

Table 58: chi-squared results for counterfactual scenario

statistic	p-value	method
3.36	0.07	Pearson's Chi-squared test with Yates' continuity correction

Table 59: chi-squared results for incompletive scenario

statistic	p-value	method
22.45	0	Pearson's Chi-squared test with Yates' continuity correction

These results suggest that the participants who agree with the aspectual tests applied to the literal scenarios would also judge the idiomatic scenarios in a manner that is consistent with the alternative hypothesis H_2 under consideration.

4.6 Conclusions

This chapter set out to probe the aspectual structure of idiomatically ambiguous sentences by relying on the grammatical judgments of native and near-native participants. In order to accomplish this goal, I designed an experiment to leverage the intuition of 282 participants by asking them to decide if a sentence could be used to accurately describe a scenario. I collected their responses as *acceptable* or *unacceptable* through a forced-choice task.

I tested four ambiguous sentences, crafting two scenarios for each sentence, one literal and one idiomatic, for a total of eight possible scenarios. The key difference between the literal and idiomatic member of each pair was duration. In addition to controlling for sentences and scenarios, I relied also on two adverbial constructions, the *almost* adverbial and the *takes-x-time* adverbial, to gather duration data. Each participant was randomly paired up with a combination of adverbials, scenarios and sentences.

To glean what aspectual structures the participants had in mind for each of those combinations, I relied on the measure of *aspectual complexity*, which I defined as the number of aspectual features present in the event, to assign membership to one of the traditional Vendler-Dowty aspectual classes. The responses I collected took the form of both positive and negative evidence, as understood in the field of language acquisition (Marcus 1993).

I then proceeded to compare the aspectual class of the literal event to that of the idiomatic event for each of the four sentences in the corpus, with the intention of validating or falsifying two working hypotheses. The first hypothesis, H_1, was motivated in part by the opposing views entertained in McGinnis (2002, 2005) and Glasbey (2003; 2007). In broad strokes, McGinnis posits that the literal event subcategorizes the idiomatic event, while Glasbey takes the stance that those two events can be different, aspectually speaking.

This chapter sheds some light on this discussion. According to the results of the experiment, participants can and do conceptualize the two events arising from idiomatically ambiguous sentences differently, as shown in Table 55. While participants understood *soltar el trapo* as an aspectual achievement for both the literal interpretation of the event "drop the rag" and the idiomatic interpretation "burst out laughing," the results quite clearly show that they can also interpret the events as aspectually different, as they did for *estirar la pata*, *quemarse las pestañas* and *tender las velas*. This chapter presents empirical evidence that it is more likely that Glasbey is correct; an event resulting from interpreting a sentence idiomatically does not have to match the literal event.

The second hypothesis, H_2, examined in this chapter was motivated by the observation that in the instances when the two events differed in aspectual membership, that difference appeared to be directional. That is to say, the event denoted by the idiomatic member of the pair appears to be aspectually less complex; in other words, it belongs to an aspectual class with fewer aspectual features than the literal member. The alternative was not observed; it does appear to be the case that the idiomatic event can exhibit a larger number of aspectual features than the literal event.

At first blush, the results seem to partially support H_2 as seen in table 60. In particular, there is a stark difference between the results coming from the participants who interacted with the *takes-x-time* adverbial and those results coming from the participants who were paired with the *almost* adverbial.

Table 60: Expected vs Observed Results

Intent	Test	Exp. Class	Exp. Readings	Exp. Incept.	Exp. Term.	Obs. Readings	Obs. Incept.	Obs. Term.	Obs. Class
Estirar la pata									
literal	almost	accomplishmen	2	yes	yes	1	no	yes	achievement
idiomatic	almost	achievement	1	no	yes	2	yes	yes	accomplishment
Quemarse las pestañas									
literal	almost	accomplishmen	2	yes	yes	1	yes	no	achievement
idiomatic	almost	activity	1	yes	no	0	no	no	activity
Soltar el trapo									
literal	takes-x-time	achievement	1	yes	no	1	yes	no	achievement
idiomatic	takes-x-time	achievement	1	yes	no	1	yes	no	achievement
Tender las velas									
literal	takes-x-time	accomplishmen	2	yes	yes	2	yes	yes	accomplishment
idiomatic	takes-x-time	achievement	1	yes	no	1	yes	no	achievement

Participants who were paired with the *take-x-time* adverbial judged their sentences as expected. It is worth nothing that the evidence produced by the *soltar el trapo* statement resulted in a type of confirmation data in the sense that the aspectual tests gave us an expectation that was fulfilled by the participants' responses. The second statement, *tender las velas* resulted in a more stringent type of test because it resulted from a negative evidence type of prediction. Whereas the results from *soltar el trapo* simply needed to match the aspectual tests, the results from *tender las velas* predicted that participants would reject both inceptive and terminative readings and settle with only one.

If we were to stay with just the *take-x-time* results, this experiment would have produced incontrovertible evidence in favor of H_2. Alas, the results from the *almost* adverbial were not nearly as clear. Both statements tested with the *almost* adverbial produced messy results. The expected results from the aspectual tests classified *quemarse las pestañas* as an activity. Participants paired with this statement found both scenarios unacceptable, which made determining duration for *quemarse las pestañas* problematic. However, given that *quemarse las pestañas* is felicitous with the progressive, I concluded that it is an activity, so in Table 60, *quemarse las pestañas* is listed as such. Regardless of these results, it would not be a problem for H_2, as an activity exhibits fewer aspectual features

than an achievement.

The last tested statement, *estirar la pata*, did produce surprising results. Contrary to the expectation that the literal event be an accomplishment, participants judged literal *estirar la pata* as an achievement. Furthermore, participants' responses placed idiomatic *estirar la pata* in the accomplishment class. This is a clear counter example in violation of the generalization that this chapter set out to investigate. In Section 4.5.3, I explored a possible explanation as to what might be behind the unexpected results, as it is known in the literature that it is common for some speakers to introduce an activity-like phase for achievements whose culmination denotes a change of state.

Weighing in on the results produced by all four statements does not paint the crisp picture I was hoping for, but it does lend support for H_2; idiomatic events are likely to be members of an aspectual class that has as many or fewer aspectual features as the concomitant literal event does, but no more aspectual features.

5 References

Abramson, Arthur; and Leigh Lisker. 1973. Voice-Timing Perceptionin Spanish Word-Initial Stops. *83rd Meeting of the Acoustical Society of America*. 1–8. doi:https://doi.org/10.1016/S0095-4470(19)31372-5.

Adachi, Takanori. 1996. Sarcasm in Japanese. *Studies in Language* 20. 1–36.

Andriga, Sibel; and Aline Godfroid. 2020. Sampling Bias and the Problem of Generalizability in Applied Linguistics. *Annual Review of Applied Linguistics* 40. 134–142.

Aron, Arthur; Elliot J Coups; and Elaine N Aron. 2013. *Statistics for Psychology*. Pearson Education Inc.

Bakker, Dik. 2010. Language Sampling. *The Oxford Handbook of Linguistic Typology*.

Beckman, Mary E.; Manuel Díaz-Campos; Julia Tevis McGory; and Terrell A. Morgan. 2002. Intonation Across Spanish, in the Tones and Break Indices Framework. *Probus* 14. 9–36.

Beddor, Patrice Speeter. 2017. Speech Perception in Phonetics. doi:10.1093/acrefore/9780199384655.0 https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-62.

Bedford Wassink, Alicia; and Emily Bender. 2012. Computational Methods in Linguistics. Universidad de Washington, ms. http://courses.washington.edu/cmling/Introduction.pdf.

Ben-David, Boaz M.; Sarah Gal-Rosenblum; Pascal H. H. M. van Lieshout; and Vered Shakuf. 2019. Age-Related Differences in the Perception of Emotion in Spoken Languages: The Relative Roles of Prosody and Semantics. *Journal of Speech, Language, and Hearing Research* 62. 1188–1202.

Birdsong, D.; L. M. Gertken; and M. Amengual. 2012. Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism 2019. https://sites.la.utexas.edu/bilingual.

Blanco, Cynthia Patricia. 2016. Cross-Language Speech Perception in Context: Ad-

vantages for Recent Language Learners and Variation Across Language Specific Cues. Thesis.

Boersma, Paul; and David Weenink. 2020. Praat: Doing Phonetics by Computer. http://www.praat.org/.

Bruening, benjamin. 2010. Ditransitive Asymmetries and a Theory of Idiom Formation. *Linguistic Inquiry* 16. 519–562.

Burke, Deborah M.; and Meredith A. Shafto. 2004. Aging and Language Production. *Current Directions in Psychological Science* 13. 21–24. doi:10.1111/j.0963-7214.2004.01301006.x. https://pubmed.ncbi.nlm.nih.gov/18414600.

Butera, Briana J. 2018. A Lenition Continuum: Acoustic Variability of Spanish Stop Consonants. Thesis.

Cameron, Richard. 2011. Aging, Age, and Sociolinguistics. *The Handbook of Hispanic Sociolinguistics*, ed. by Manuel Díaz-Campos. Wiley-Blackwell.

Chappell, Whitney. 2019. *Recent Advances in the Study of Spanish Sociophonetic Perception*. Vol. 21. Issues in hispanic and lusophonic linguistics (IHLL). John Benjamins Publishing Company. doi:https://doi.org/10.1075/ihll.01cha.

Cheang, Henry; and Marc Pell. 2008. The Sound of Sarcasm. *Speech Communication* 50. 366–381.

Cheang, Henry; and Marc Pell. 2009. Acoustic Markers of Sarcasm in Cantonese and English. *Journal of the Acoustical Society of America* 126. 1394–1405.

Chomsky, Noah; and Morris Halle. 1968. *The Sound Pattern of English*. Harper & Row.

Chomsky, Noam. 1957. *Syntactic Structures*. The Netherlands: The Hague.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.

Cowart, Wayne. 1997. *Experimental Syntax. Applying Objective Methods to Sentence Judgments*. Sage Publiscations, Inc.

Culpeper, Jonathan. 2005. Impoliteness and Entertainment in the Television Quiz Show: The Weakest Link. *Journal of Politeness Research* 1. 35–72.

Davies, Martin. 1982. AIdiom and Metaphor. *Proceedings of the Aristotelian Society.* 67–85.

De Miguel, Elena. 1999. El Aspecto Léxico. *Gramática Descriptiva de La Lengua Española*, ed. by Ignacio Bosque and Violeta Demonte. Espasa.

Dowty, David R. 1972. Studies in the Logic of Verb Aspect and Time Reference in English. Thesis.

Dowty, David R. 1979. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ.* Dordrecht: Kluwer Academic Publishers.

Dryer, Matthew S.; and Martin Haspelmath. 2013. The World Atlas of Language Structures Online (WALS) 2019. https://wals.info/.

Estebas Vilaplana, Eva; and Pilar Prieto. 2008. La Notación Prosódica Del Español: Una Revisión de Sp_ToBI. *Estudios de Fonética Experimental* 17. 263–283.

Face, Timothy. 2001. Focus and Early Peak Alignment in Spanish Intonation. *Probus* 13. 113–246.

Face, Timothy. 2002. Spanish Evidence for Pitch-Accent Structure. *Linguistics* 40. 319–345.

Face, Timothy; and Pilar Prieto. 2007. Rising Accents in Castillian Spanish. *Journal of Portuguese Linguistics* 5.

Fonagy, Ivan. 1971. Synthèse de l'ironie [Overview of Irony]. *Phonetica* 23. 42–51.

Fraser, Bruce. 1970. Idioms Within a Transformation Grammar. *Natural Language & Linguistic Theory* 16. 29–312.

Fraser, Bruce. 2014. On the Relationship Between Type and Token Frequency. *Journal of Quantitative Linguistics* 21. 199–222.

García Fernández, Luis. 1999. Los Complementos Adverbiales Temporales. La Sub-

ordinación Temporal. *Gramática Descriptiva de La Lengua Española*, ed. by Ignacio Bosque and Violeta Demonte. Espasa.

Gennari, Silvia; and David Poeppel. 2003. Processing Correlate of Lexical Semantic Complexity. *Cognition* 89. B27–B41.

Gibbon, Dafydd; Roger Moore; and Richard Winski. 1997. *Spoken Language System and Corpus Design*. Berlin/Boston, GERMANY: De Gruyter, Inc. http://ebookcentral.proquest.com/lib/wisc/detail.action?docID=3044590.

Glasbey, Sheila. 2007. Aspectual Composition in Idioms. *Recent Advances in the Syntax and Semantics of Tense, Aspect and Modality. Edited by Louis de Saussure, Jacques Moeschler, Genoveva Puskás*. 71–87.

Glasbey, Sheila R. 2003. Let's Paint the Town Red for a Few Hours: Composition of Aspect in Idioms. *Proceedings of the ACL Workshop the Lexicon and Figurative Language*, ed. by A. M. Wallington. ACL 2003.

Grice, Martine; and Stefan Baumann. 2007. An Introduction to Intonation - Functions and Models. *Non-Native Prosody. Phonetic Description and Teaching Practice.*, ed. by Trouvain and Ulrike, 25–51. Trends in linguistics. Studies and monographs. De Gruyter.

Grolemund, Garrett; and Hadley Wickham. 2017. *R for Data Science*. O'Reilly.

Gussenhoven, Carlos. 2002. Intonation and Interpretation: Phonetics and Phonology. *Speech Prosody 2002*.

Gussenhoven, Carlos. 2004. *The Phonoolgy of Tone and Intonation*. Cambridge University Press.

Haiman, John. 1998. *Talk Is Cheap: Sarcasm, Alienation, and the Evolution of Language*. New York, NY: Oxford University Press.

Hales, S. D. 2019. Thinking Tools: You Can Prove a Negative. *Think*. 109–112.

Hancock, Adrienne B.; Holly Wilder Stutts; and Annie Bass. 2015. Perceptions of Gender and Femininity Based on Language: Implications for Transgender Commu-

nication Therapy. *Language and Speech* 58. 315–333. doi:10.1177/0023830914549084. https://journals.sagepub.com/doi/abs/10.1177/0023830914549084.

Hansen Edwards, Jette G. 2008. Social Factors and Variation in Production in L2 Phonology. *Phonology and Second Language Acquisition*, ed. by Jette G Hansen Edwards and Mary L. Zampini. Philadelphia: John Benjamins.

Hartsuiker, Robert J; Saskia Beerts; Maaike Loncke; Timothy Desmet; and Sarah Bernolet. 2016. Cross-Linguistic Structural Priming in Multilinguals: Further Evidence for Shared Syntax. *Journal of Memory and Language* 90. 13–30.

Henton, Caroline. 1995. Pitch Dynamism in Female and Male Speech. *Language and Communication* 15. 43–61.

Hirschberg, Julia. 2006. Pragmatics and Intonation. *The Handbook of Pragmatics*, ed. by Laurence Horn and Gregory Ward. Wiley.

Holmquist, Jonathan. 2011. Gender and Variation: Word-Final /s/ in Men's and Women's Speech in Puerto Rico's Western Highlands. *The Handbook of Hispanic Sociolinguistics*, ed. by Manuel Díaz-Campos. Wiley-Blackwell.

Howell, Jonathan. 2012. Meaning and Prosody: On the Web, in the Lab and from the Theorist's Armchair. Thesis.

Hualde, José Ignacio. 2002. Intonation in Romance Introduction to the Special Issue. *Probus* 14. 1–7.

Hualde, José Ignacio. 2003. El Modelo Métrico y Autosegmental. *Teorías de La Entonación*. Ariel Lingüística.

Hualde, José Ignacio; and Pilar Prieto. 2015. Intonational Variation in Spanish: European and American Varieties Intro. *Intonation in Romance*, ed. by Sonia Frota and Pilar Prieto. First edition. [Oxford] : Oxford University Press, [2015] 2015. https://search.library.wisc.edu/catalog/9911172562902121.

Hualde, José Ignacio; and Pilar Prieto. 2016. Toward an International Prosodic Alphabet. *Journal of the Association for Laboratory Phonology* 7. 1–25.

Hume, Elizabeth; and Kyle Johnson. 2001. A Model of the Interplay of Speech Perception and Phonology. *OSJ Working Papers in Linguistics* 55. 1–22.

Jackendoff, Ray. 1972. MIT Press.

Jackendoff, Ray. 1992. Babe Ruth Homered His Way into the Hearts of America. *Syntax and Semantics*, ed. by Tim Stowell and Eric Wehrli, 26: 1250 Sixth Avenue, San Diego, California 92101: Academic Press Inc.

Jackendoff, Ray. 1997. Idioms and Other Fixed Expressions. *The Architecture of the Language Faculty*, 153–177. Linguistic Inquiry Monographs.

Jacobson, Pauline. 2014. *Compositional Semantics: An Introduction to the Syntax/Semantics Interface*. Oxford University Press.

Katz, Jerrold Jacob; and Paul Postal. 1963. Semantic Interpretation of Idioms and Sentences Containing Them. *Quarterly Progress Report of the MIT Research Laboratory of Electronics*. 275–285.

Kearns, Kate. 2003. Durative Achievements and Individual-Level Predicates on Events. *Linguistics and Philosophy*. 595–635.

Kearns, Kate. 2011. *Semantics*. Palgrave MacMillan.

Krifka, Manfred. 1998. The Origins of Telicity. *Events and Grammar*, ed. by Susan Deborah Rothstein. Springer-Verlag New York.

Labov, William. 1966. *The Social Stratification of English in New York City*. Center for applied linguistics. Cambridge: Cambridge University Press.

Labov, William. 1973. *Sociolinguistic Patterns*. University of Pennsylvania Press, Incorporated. https://books.google.com/books?id=hD0PNMu8CfQC.

Labov, William. 1990. The Intersection of Sex and Social Class in the Course of Linguistic Change. *Language Variation and Change* 2. 205–254. doi:10.1017/S0954394500000338.

Labov, William. 2008. Quantitative Reasoning in Linguistics. University of Pennsylvania.

Ladd, Robert. 1996. Intonational Phonology. *Cambridge Studies in Linguistics* 79. Xv–

334.

Ladd, Robert. 2008. *Intonational Phonology*. Cambridge studies in linguistics. Cambridge, UK: Cambridge University Press.

Larson, Richard K. 1988. *On the Double Object Construction. The MIT Press*. Vol. 19. 3.

Linzen, Tal; Alec Marantz; and Liina Pylkkänen. 2013. Syntactic Context Effect in Visual Word Recognition. *The Mental Lexicon* 8. 117–139.

Lipski, John M. 1996. *El Espanol de America (Linguistica) (Linguistica / Linguistics) (Spanish Edition)*. Madrid: Cátedra.

Lipski, John M. 2008. *Varieties of Spanish in the United States*. Washington, D.C.: Georgetown University Press.

Lisker, Leigh; and Arthur S. Abramson. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical. *Word* 20. 384–422.

MacDonald, Jonathan. 2008. *The Syntactic Nature of Inner Aspect - a Minimalist Perspective*. John Benjamin Publishing Company.

Mack, Sara. 2019. Of Intersectionality, Replicability, and Holistic Perspectives. *Recent Advances in the Study of Spanish Sociophonetic Perception*, ed. by Whitney Chappell, 21:315–325. Issues in hispanic and lusophonic linguistics (IHLL). John Benjamins Publishing Company. doi:https://doi.org/10.1075/ihll.21.12mac.

Marantz, Alec. 1996. "Cat" as a Phrasal Idiom: Consequences of Late Insertion in Distributed Morphology. *MIT*.

Marantz, Alec. 1997. No Escape from Syntax. Don't Try Morphological Analysis in the Privacy of Our Own Lexicon. 4. Penn working papers in linguistics. 201–225.

Marcus, Gary F. 1993. Negative Evidence in Language Acquisition. *Cognition*. 53–85.

Martin, Everaert. 2010. The Lexical Enconding of Idioms. *Syntax, Lexical Semantics and Event Structure*, ed. by Malka Rappaport Hovav, Ivy Sichel, and Edit Doron. Oxford University Press.

Mateu, Jaume; and M. Teresa Espinal. 2007. Argument Structure and Compositionality in Idiomatic Constructions. *The Linguistic Review* 24. 33–59.

McConnel-Ginet, Sally. 1978. Intonation in a Man's World. *Signs* 3.

McGinnis, Martha. 2002. On the Systematic Aspect of Idioms. *Linguistic Inquiry* 33. 665–672.

McGinnis, Martha. 2005. Painting the Wall Red for a Few Hours: A Reply to Glasbey (2003). *Snippets*. http://www.ledonline/snippets/.

McKoon, Gail; and Talke Macfarlane. 2002. Event Templates in the Lexical Representation of Verbs. *Cognitive Psychology* 45. 1–44.

Medina-Rivera, Antonio. 2011. Variationist Approaches: External Factors Conditioning Variation in Spanish Phonology. *The Handbook of Spanish Sociolinguistics*, ed. by Manuel Diáz-Campos. Wiley-Blackwell.

Mitchell, Rachel L. C.; and Rachel A. Kingston. 2014. Age-Related Decline in Emotional Prosody Discrimination: Acoustic Correlates. *Experimental Psychology* 61. 215–223. doi:10.1027/1618-3169/a000241.

Nibert, Holly. 1999. A Perception Study of Intermediate Phrasing in Spanish Intonation. *Advances in Hispanic Linguistics; Papers from the 2nd Hispanic Linguistics Symposium*. 231–247.

Nibert, Holly. 2000. Phonetic and Phonological Evidence for Intermediate Phrasing in Spanish Intonation. Thesis.

Nibert, Holly. 2005. The Acquisition of the Phrase Accent by Intermediate and Advanced Adult Learners of Spanish as a Second Language. *Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*. 108–122.

Nunberg, Geoffrey; Ivan A. Sag; and Thomas Wasow. 1994. Idioms. *Language* 70. 491–538.

O'Grady, William. 1998. The Syntax of Idioms. *Natural Languages and Linguistic*

Theory 2. 279–312.

O'Rourke, Erin. 2012. Intonation in Spanish. *The Handbook of Hispanic Linguistics*, ed. by Hualde, O'Rourke, and Olarrea, 173–191.

Ohala, J. J. 1983. Cross-Language Use of Pitch: An Ethological View. *Phonetica* 40. 1–18. doi:10.1159/000261678. https://www.karger.com/DOI/10.1159/000261678.

Ortega-Llebaria, Marta. 2006. Phonetic Cues to Stress and Accent in Spanish. *Proceedings of the 2nd. Conference on Laboratory Approaches to Spanish Phonetics and Phonology*. 104–118.

Ortega-Llebaria, Marta; and Pilar Prieto. 2010. Acoustic Correlates of Stress in Central Catalan and Castilian Spanish. *Language and Speech* 54. 73–97.

Ouyang, Mingkun; Xiao Cai; and Qingfang Zhang. 2020. Aging Effects on Phonological and Semantic Priming in the Tip-of-the-Tongue: Evidence from a Two-Step Approach. *Frontiers in Psychology* 11. 338–338. doi:10.3389/fpsyg.2020.00338.

Pawley, Andrew; and Frances Hodgetts Syder. 1983. Two Puzzles for Linguistic Theory: Nativelike Selection and Native Like Fluency. *Language and Communication*, ed. by Jack Richards and Schmidt Richard. Longman.

Pierrehumbert, Janet. 1980. The Phonetics and Phonology of English Intonation. Thesis. http://dspace.mit.edu/handle/1721.1/16065.

Prieto, Pilar. 1998. The Scaling of the l Values in Spanish Downstepping Contours. *Journal of Phonetics* 26. 261–282.

Prieto, P; and P Roseano. 2010. *Intonation Across Spanish, in the Tones and Break Indices Framework*. München.

Rao, Rajiv. 2013a. Prosodic Consequences of Sarcasm Versus Sincerity in Mexican Spanish. *Concentric: Studies in Linguistics* 39. 33–59.

Rao, Rajiv. 2013b. Intonational Variation in Third Party Complaints in Spanish. *Journal of Speech Sciences* 3. 141–168.

Rao, Rajiv. 2015. On the Phonological Status of Spanish Compound Words. *Word*

Structure, 84–118. Edinburgh University Press.

Rao, Rajiv. 2016. On the Nuclear Intonational Phonology of Heritage Speakers of Spanish. *Advances in Spanish as a Heritage Language*, 51–81. John Benjamins Publishing Company.

Rao, Rajiv; and Sandro Sessarego. 2018. The Intonation of Chota Valley Spanish: Contact-Induced Phenomena at the Discourse-Phonology Interface. *Studies in Hispanic and Lusophone Linguistics* 11. 163–192.

Rappaport Hovav, Malka; Edit Doron; and Ivy Sichel. 2010. *Lexical Semantics, Syntax, and Event Structure*. Kluwer Academic Publishers.

Real Academia Española. 2020. Journal Article. *Diccionario de La Lengua Española, 23a Ed [Versión 23.4 En Línea]*. https://dle.rae.es.

Real Academia Española, Diccionario escolar de la. 1996. Website. Espasa.

Rissel, Dorothy A. 1989. Sex, Attitudes, and the Assibilation of /r/ Among Young People in San Luis Potosí, Mexico. *Language Variation and Change* 1. 269–283. doi:10.1017/S0954394500000181. https://www.cambridge.org/core/article/sex-attitudes-and-the-assibilation-of-r-among-young-people-in-san-luis-potosi-mexico/B8E2DF12373F1907F766DB13026B2C3A.

Schütze, Carson. 1986. *The Emprical Base of Linguistics: Grammaticality Judgments and Linguistics Methodology*. Berlin: Language Science Press. doi:10.17169/langsci.b89.100.

Schütze, Carson; and Jon Sprouse. 2014. Judgment Data. *Research Methods in Linguistics*, ed. by Robert Podesva and Devyani Sharma. Cabridge University Press. doi:10.1017/CBO9781139013734.

Seddoh, Amebu; Afua Blay; Richard Ferraro; and Wayne Swisher. 2020. Prosodic Perception in Aging Individuals: A Focus on Intonation. *Current Psychology* 39. 1221–1233.

Sinclair, John. 1991. Evaluating Instances. *Corpus, Concordance and Collocation*.

Smith, Carlota. 1991. *The Parameter of Aspect*. Studies in linguistics and philosophy. The Netherlands: Kluwer Academic Publishers.

Song, Hoeun; Jaein Kim; Kyongrok Lee; and Jinyoung Kim. 2002. Phonetic Normalization Using z-Score in Segmental Prosody Estimation for Corpus-Based TTS System. https://www.isca-speech.org/archive/archive_papers/icslp_2002/i02_2393.pdf.

Sullivan, Lisa. 2016. The Role of Probability 2020. https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_probability/index.html.

Tagliamonte, Sali A. 2006. *Analysing Sociolinguistic Variation (Key Topics in Sociolinguistics)*. 1st Ed. Cambridge: Cambridge University Press.

Talmy, Leonard. 1985. Language Typology and Syntactic Description. *Language Typology and Syntactic Description*, ed. by Timothy Shoppen. Cambridge University Press.

Titone, Debra; Georgie Columbus; Veronica Whitford; Julie Mercier; and Maya Libben. 2015. *Contrasting Bilingual and Monolingual Idiom Processing*. Cambridge University Press.

Traunmüller, Hartmut; and Anders Eriksson. 1995. The Frequency Range of the Voice Fundamental in the Speech of Male and Female Adults. *Unpublished Manuscript*. https://www2.ling.su.se/staff/hartmut.old/f0_m&f.pdf.

Van Lancker, Diana; and Gerald J Canter. 1981. Idiomatic Versus Literal Interpretation of Ditropically Ambiguous Sentences. *Journal of Speech and Hearing Research*.

Van Lancker, Diana; Gerald J Canter; and Dale Terbeek. 1981. Disambiguation of Ditropic Sentences: Acoustic and Phonetic Cues. *Journal of Speech and Hearing Research*.

Vanderniet, Kyle Hugh. 2015. Idioms as a Measure of Proficiency. Thesis.

Vanrell, Maria del Mar; Ingo Feldhausen; and Lluïsa Astruc. 2018. The Discourse Completion Task in Romance Prosody Research: Status Quo and Outlook. *Methods in Prosody: A Romance Language Perspective*, ed. by Ingo Feldhausen, Jan

Fliessbach, and Maria del Mar Vanrell, 191–227.

Vendler, Zeno. 1967. Linguistics and the a Priori. *Linguistics in Philosophy*. Cornell University Press.

Visser, Penny S.; Jon A. Krosnick; and Paul J. Lavrakas. 2000. Survey Research. *Handbook of Research Methods in Social and Personality Psychology*. 223–252.

Walková, Milada. 2012. Dowty's Aspectual Tests: Standing the Test of Time but Failing the Test of Aspect. *Poznań Studies in Contemporary Linguistics* 48. 495–518.

Wickham, Hadley. 2014. Tidy Data. *Journal of Statistical Software* 59. 23. doi:10.18637/jss.v059.i10. https://www.jstatsoft.org/v059/i10.

Winter, Bodo. 2020. *Statistics for Linguists. An Introduction Using r*. Routhledge.

Zárate-Sández, Germán A. 2015. Perception and Production of Intonation Among English-Spanish Bilingual Speakers at Different Proficiency Levels. Thesis.

Appendix

Appendix A

UNIVERSITY OF WISCONSIN-MADISON

Research Participant Information and Consent Form

Title of the Study: Idiomatic Expressions in Spanish

Principal Investigator: Rajiv Rao (phone: 608-262-2093) (email: rgrao@wisc.edu)

DESCRIPCION OF THE RESEARCH

You are invited to participate in a research study about the production, perception, and grammatical judgments test of idiomatic expressions in Spanish. You have been asked to participate because you are native speaker of Spanish.

The purpose of the research is to gain a better understanding of the relationship that holds between a sentence intended to be interpreted idiomatically and the same sentence

intended to be interpreted literally.

The study will be carried out with participants who are graduate and undergraduate students at UW-Madison. Data will be collected from participants in a quiet room in Van Hise Hall. Participants will provide their answers via a microphone which will be recorded directly into a laptop computer.

Audio files will be made of your participation. These files will be used in two occasions: the first time to determine the presence of prosodic elements. Some of the audio recordings will also be used as stimuli for a perception test during the second phase of the study.

All audio recordings will remain in the possession of members of the study team. These files will be password protected and encrypted. All speech samples will be de-identified so that your personal information will not be accessible. The principal investigator may choose one or two short speech samples from select participants to display in a conference, a paper, or a class presentation. The recorded files will be kept indefinitely for future research. The files will be kept on UW's Box service accessible only to the study team.

I agree to share audio clips with the study team and to allow the team to maintain possession of them, given that my name and/or other personal information will not be used when discussing them.

☐ Check this box if you agree.

WHAT WILL MY PARTICIPATION INVOLVE?

If you decide to participate in this research you will be asked to carry out three different tasks in Spanish. You will first be asked to read a short background story. The story ends with you producing a summarizing statement which will be recorded. Second, you will be asked to listen to a statement similar to the one you produced for the first task.

You will be asked if the person that produced the statement meant it to be interpreted idiomatically or literally. Third, you will be asked to perform a grammatical judgement test.

Your participation will last approximate 90 minutes and will require three sessions.

ARE THERE ANY RISKS TO ME?

We don't anticipate any legal, economic, or social status risks to the participants. There is a minimal risk for breach of confidentiality. Also, throughout the session you may feel fatigued. If you do, you are welcome to request a break before continuing. Finally, we anticipate that some of the tasks presented to you might be challenging. However, you will first be trained to complete them and provided with plenty of examples.

ARE THERE ANY BENEFITS TO ME?

We don't expect any direct benefits to you from participating/participation in this study.

WILL I BE COMPENSATED FROM MY PARTICIPATION?

If you complete all three tasks your name will be entered into a drawing for an Amazon gift card in the amount of $50.

HOW WILL MY CONFIDENTIALITY BE PROTECTED?

To minimize any breach of confidentiality, the sessions will be conducted in a private room in Van Hise Hall by one of the study team members. All digital data collected will be associated with a code key. All reference to the data will be conducted by that code key and not by your name. The names, code keys, and consent forms will in a locked file cabinet. The electronic data will be encrypted and stored in UW's Box service. No electronic form will be maintained containing both your personal information and your code key.

While there will probably be publications as a result of this study, your name will not be used. Only groups characteristics will be published.

WHOM SHOULD I CONTACT IF I HAVE QUESTIONS?

You may ask question about the research at any time. If you have questions about the research after you leave today, you should contact the Principal Investigator Rajiv Rao at 608-262-2093.

If you are not satisfied with the response of the research team, have more questions, or want to talk to someone about your rights as a research participant, you should contact the Education Research and Social & Behavioral IRB Office at (608) 263-2320.

Your participation is completely voluntary. If you begin and change your mind you may end your participation at any time without any penalty.

Your signature indicates that you have read this consent form, had an opportunity to ask any questions about your participation in this research and have voluntarily consented to participate. You will receive a copy of this form for your records.

Name of participant (please print): _____

Signature of Participant: _____

Date: _____

Appendix B

Demographic data

* Required

1. First Name *

2. Participant ID

3. Age *

4. Gender *

Mark only one oval.

- Female
- Male
- Prefer not to say

5. Education Level *

Mark only one oval.

- Elementary
- Highschool
- College
- Masters
- PhD

6. Spanish Variety *

7. BLP Score

Appendix C

Experiment 1 - Background Stories

Instructions:

You will be given a short background story for each of the expressions tested in this study. After reading the background story, we would like you to imagine that someone asks you 'what happened?'. With the context provided by the background story in mind and the corresponding expression, we would like you to answer the prompt starting with 'it happens that **x**'

For example,

Background story:

- Mary is an older lady.
- She was diagnosed with cancer a few months ago when she had her annual checkup.
- Last night things took a turn for the worst; Mary kicked the bucket last night.

Expression to be tested:

- **kick the bucket** (lit. to die)

Someone asks you:

- What happened?

You'll summarize the event using the expression underlined:

- It happened that **Mary kicked the bucket last night**

Background story:

- Antonio es un chico muy travieso. Luis es su hermano menor.
- Anoche Luis le contó muchas historias de terror a Luis. Luis no pudo dormir nada.
- Antonio le metió en la cabeza a Luis que El Coco existía.

Expression to be tested:

- **meter en la cabeza** (lit. to persuade someone of something)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Antonio le metió en la cabeza a Luis que El Coco existía.

Experiment 1 - Background Stories

01

Background story:

- José es un espía de la CIA.
- Anoche logró infiltrarse en la casa del embajador de Rusia.
- Oculto fuera de la habitación, José aplastó la oreja, cuidadosamente, contra la pared para tratar de escuchar la conversación que estaba ocurriendo.

Expression to be tested:

- **aplastar la oreja**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que José aplastó la oreja.

Experiment 1 - Background Stories

02

Background story:

- A Ana le encanta salir a tomar con sus amigos
- Anoche Ana y su amigo Eduardo fueron al bar juntos
- Eduardo está tomando antibióticos ahora, así que Ana bebió las cervezas por Eduardo

Expression to be tested:

- **beber las cervezas**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Ana bebió las cervezas por Eduardo.

Experiment 1 - Background Stories

03

Background story:

- Miguel es un cantante famoso. Su amigo Nicolás es el director del coro.
- Anoche durante el ensayo, Miguel finalmente completó su sección sin ningún error.
- Miguel le cantó las cuatro estrofas a Nicolás cuando comenzó el segundo acto.

Expression to be tested:

- **<u>cantar las cuatro estrofas</u>**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Miguel le cantó las cuatro estrofas a Nicolás.

Experiment 1 - Background Stories

04

Background story:

- Luisa es una crítica gastronómica famosa.
- Ayer visitó el restaurante que recién recibió una estrella Michelin por sus paellas.
- Luisa se comió la paella, ¡no podía parar!

Expression to be tested:

- **comerse la paella**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Luisa se comió la paella.

Experiment 1 - Background Stories

05

Background story:

- Juan y Ana son estudiantes de primer año de la universidad.
- Ayer Juan y Ana aparcaron el carro en un estacionamiento cercano.
- Ana y Juan cruzaron la calle para llegar a la residencia.

Expression to be tested:

- **cruzar la calle**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Juan y Ana cruzaron la calle.

Experiment 1 - Background Stories

06

Background story:

- Luisa es la hermana mayor de Juancito.
- Anoche Luisa estaba preparándose para ir a una cita, pero no lograba decidir qué conjunto llevar.
- Juancito le dio una falda a Luisa para que se viera genial.

Expression to be tested:

- **dar la falda**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Juancito le dio la falda a Luisa.

Experiment 1 - Background Stories

07

Background story:

- Teresa es una instructora de yoga.
- Ayer le sorprendió mucho descubrir que era el primer día de yoga de todos los participantes.
- Comenzó la clase muy pausadamente. Primero Teresa estiró la pierna. Comenzó con la izquierda, luego la derecha.

Expression to be tested:

- **estirar la pierna**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Teresa estiró la pierna.

Experiment 1 - Background Stories

08

Background story:

- Pedro es un marinero muy diestro.
- Anoche llegó al puerto de La Guaira luego de haber completado una vuelta al mundo.
- Pedro recogió las velas después de llegar al puerto.

Expression to be tested:

- **recoger las velas**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Pedro recogió las velas.

Experiment 1 - Background Stories

09

Background story:

- Rupertina es entomofóbica.
- Anoche mientras trataba de dormirse, Rupertina vio un gusanillo escurrirse sobre las mantas.
- Rupertina mató el gusanillo apenas lo vio.

Expression to be tested:

- **matar el gusanillo**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Rupertina mató el gusanillo.

Experiment 1 - Background Stories

10

Background story:

- Juan es un conductor de maquinaria pesada.
- Anoche hubo una tormenta de lluvia congelante sobre todo Wisconsin.
- Todo su camión estaba encapsulado en hielo. Para poder entrar a su camión, Juan rompió el hielo con un destornillador.

Expression to be tested:

- **romper el hielo**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Juan rompió el hielo.

Experiment 1 - Background Stories

11

Background story:

- Juan es un chef de pastelería muy famoso. Su amigo Luis es su asistente de pastelería.
- Anoche después de cerrar la tienda, Juan y Luis comenzaron a preparar una receta nueva con castañas rostizadas en lugar de nueces.
- Mientras Juan preparaba la masa, Luis le sacó las castañas del fuego a Juan.

Expression to be tested:

- **sacar las castañas del fuego**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Luis le sacó las castañas del fuego a Juan.

Experiment 1 - Background Stories

12

Background story:

- Pedro es un pastor de cabras. Petunia es una de sus cabras.
- Ayer mientras exploraba el río, Petunia se quedó atascada con una de las patas en el barro.
- Pedro le sacó la pata del barro a Petunia cuando la encontró al final del día.

Expression to be tested:

- **sacar la pata del barro**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Pedro le sacó la pata del barro a Petunia.

Experiment 1 - Background Stories

13

Background story:

- Luis es un padre muy cariñoso. Su hija se llama Teresa.
- Ayer mientras estaban en el parque, Luis notó que Teresa cojeaba cada vez que trataba de dar un paso.
- Luis descubrió que había una piedra dentro del zapato, así que Luis le sacó la piedra a Teresa del zapato.

Expression to be tested:

- **sacar la piedra**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Luis le sacó la piedra a Teresa.

Experiment 1 - Background Stories

14

Background story:

- Ana es una persona muy ordenada.
- Anoche Ana comenzó a limpiar la cocina pasando un trapo húmedo por encima de todas las superficies.
- De repente, notó algo moviéndose en el trapo, Ana soltó el trapo cuando vio una araña escurriéndose hacia su mano.

Expression to be tested:

- **soltar el trapo**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Ana soltó el trapo.

Experiment 1 - Background Stories

15

Background story:

- Luis es un navegante muy diestro.
- Anoche soplaba el viento en dirección al oeste. Los cielos estaban despejados.
- Luis tendió las velas y zarpó en dirección al próximo puerto.

Expression to be tested:

- **tender las velas**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Luis tendió las velas.

Experiment 1 - Background Stories

16

Background story:

- Ana y Pedro son personas muy abnegadas.
- Anoche los dos estaban construyendo casas para los indigentes de su ciudad.
- Mientras Ana estaba en el techo, Pedro le tiró las tejas para que las instalara en el techo.

Expression to be tested:

- **tirar las tejas**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Pedro le tiro las tejas a Ana.

Experiment 1 - Background Stories

17

Background story:

- Miguel es un músico muy talentoso.
- Anoche mientras estaba visitando a sus amigos de la universidad, le pidieron que tocara.
- Miguel tocó la vihuela toda la noche.

Expression to be tested:

- **tocar la vihuela**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Miguel tocó la vihuela.

Experiment 1 - Background Stories

18

Background story:

- Luis es un estudiante de postgrado.
- Ayer Luis pasó el día en la playa recogiendo especímenes, pero se olvidó de ponerse un sombrero.
- Al final del día, Luis se quemó las orejas.

Expression to be tested:

- **quemarse las orejas**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Luis se quemó las orejas.

Experiment 1 - Background Stories

19

Background story:

- Pedro es un chico muy desordenado.
- Ayer por la tarde fue a la piscina con sus amigos.
- Apenas llegó a casa, Pedro tiró la toalla en el medio de la cocina.

Expression to be tested:

- **tirar la toalla**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Pedro tiró la toalla.

Experiment 1 - Background Stories

20

Background story:

- Juan es un chico muy a la moda.
- Ayer Juan compró un pantalón de una marca muy lujosa.
- Luego de ponerse el pantalón nuevo, Juan se apretó el cinturón, se puso la camisa y salió al bar.

Expression to be tested:

- **<u>apretarse el cinturón</u>**

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Juan se apretó el cinturón.

Experiment 1 - Background Stories

21

Background story:

- José es un estudiante graduado.
- Este semestre, José está tomando 20 créditos. Cuando llega el viernes, está exhausto.
- Apenas llegó a casa, José aplastó la oreja para recuperar el sueño perdido.

Expression to be tested:

- **aplastar la oreja** (idm. *sleep*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que José **aplastó la oreja**.

Experiment 1 - Background Stories

22

Background story:

- Ana es una chica muy apasionada
- Anoche mientras estaba sentada al lado de una fogata, comenzó a conversar con Eduardo
- Casi inmediatamente, Ana bebió los vientos por Eduardo

Expression to be tested:

- **<u>beber los vientos por alguien</u>** (idm. *be in love with someone*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Ana bebió los vientos por Eduardo

Experiment 1 - Background Stories

23

Background story:

- Miguel es un político muy astuto. Nicolás es su opositor político más importante.
- Anoche durante el debate hubo un momento de intensa discusión sobre política internacional.
- Miguel le cantó las cuatro verdades a Nicolás cuando le dijo que era un político corrupto.

Expression to be tested:

- **<u>cantar las cuatro verdades</u>** (idm. *be brutally honest*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Miguel le cantó las cuatro verdades a Nicolás.

Experiment 1 - Background Stories

24

Background story:

- Luisa es una estudiante de derecho.
- Ayer pasó todo el día en la biblioteca Memorial preparándose para el examen, pero no estaba segura si había estudiado el material correcto.
- Luisa se comió la olla, ¡no podía parar!

Expression to be tested:

- **comerse la olla** (idm. *overthink or worry too much*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Luisa se comió la olla.

Experiment 1 - Background Stories

25

Background story:

- Ana y Juan son personas muy sociables.
- Ana decidió ir a la fiesta a las 6 pm. Juan también decidió ir a la fiesta a las 6 pm.
- Ana y Juan cruzaron la mirada apenas llegaron en la fiesta.

Expression to be tested:

- **cruzar la mirada** (idm. *make eye contacy briefly*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Juan y Ana cruzaron la mirada.

Experiment 1 - Background Stories

26

Background story:

- Luisa es la hermana mayor de Juancito.
- Anoche mientras Luisa trataba de terminar su ensayo, Juancito le hacía pregunta tras pregunta a Luisa.
- Juancito le dio la lata a Luisa para que continuara prestándole atención.

Expression to be tested:

- **dar la lata** (idm. *annoy excessively someone*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Juancito le dio la lata a Luisa.

Experiment 1 - Background Stories

27

Background story:

- Teresa es una señora mayor.
- El médico le diagnosticó cáncer pulmonar cuando fue a hacerse un chequeo anual hace unos meses.
- Las cosas empeoraron rápidamente. Anoche Teresa estiró la pata.

Expression to be tested:

- **estirar la pata** (idm. *die*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Teresa estiró la pata.

Experiment 1 - Background Stories

28

Background story:

- Pedro es un político corrupto.
- Ayer el senado votó a favor de impugnar a Pedro unánimemente.
- Pedro recogió las velas después de escuchar el resultado de la votación y renunció a su cargo.

Expression to be tested:

- **recoger las velas** (idm. *leave suddenly*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Pedro recogió las velas.

Experiment 1 - Background Stories

29

Background story:

- Rupertina es una ejecutiva de ventas con una cuenta muy importante ese día.
- Rupertina sabía que sólo con ginebra sería capaz de relajarse ese día.
- Rupertina mató el gusanillo apenas salió de la cama.

Expression to be tested:

- **matar el gusanillo** (idm. *drink alcohol on an empty stomach*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Rupertina mató el gusanillo.

Experiment 1 - Background Stories

30

Background story:

- Juan es un político muy astuto.
- Anoche tuvo una cena de beneficencia para anunciar su candidatura a la presidencia.
- Al comienzo del evento, el público estaba tenso, pero Juan rompió el hielo con sus anécdotas.

Expression to be tested:

- **romper el hielo** (idm. *overcome initial social awkwardness*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Juan rompió el hielo.

Experiment 1 - Background Stories

31

Background story:

- Juan es un chico privilegiado. Luis, su mejor amigo, es el juez del tribunal.
- Anoche después de unos tragos en el bar, la policía detuvo a Juan por conducir bajo los efectos del alcohol.
- Como siempre, su amigo intervino. Luis le sacó las castañas del fuego a Juan evitando que lo arrestaran.

Expression to be tested:

- **sacar las castañas del fuego** (idm. *help someone from a risky situation*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Luis le sacó las castañas del fuego a Juan.

Experiment 1 - Background Stories

32

Background story:

- Pedro y Petunia son novios.
- Ayer mientras corría para tomar su autobús, La mochila de Petunia se rompió y la computadora golpeó el piso rompiéndose en pedazos.
- Pedro le sacó la pata del barro a Petunia cuando le regaló una computadora nueva para sus clases.

Expression to be tested:

- **sacar la pata del barro** (idm. *help someone*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Pedro le sacó la pata del barro a Petunia.

Experiment 1 - Background Stories

33

Background story:

- Luis es un chico muy cómico. Siempre anda contando historias divertidas.
- Ayer mientras contaba una retahíla de chistes, dijo uno que ofendió a Teresa.
- Luis le sacó la piedra a Teresa finalmente con sus chistes.

Expression to be tested:

- **sacar la piedra** (idm. *annoy someone excessively*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Luis le sacó la piedra a Teresa.

Experiment 1 - Background Stories

34

Background story:

- Ana es muy sarcástica.
- Anoche Ana fue a recoger a su hermano a la escuela porque había nevado.
- Cuando vio a su hermano resbalarse en la acera, Ana soltó el trapo en lugar de ayudarlo.

Expression to be tested:

- **soltar el trapo** (idm. *burst out laughing or crying*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Ana soltó el trapo.

Experiment 1 - Background Stories

35

Background story:

- Luis es un político muy astuto.
- El gobernador renunció luego de que la prensa revelara los actos de corrupción de su campaña.
- Luis tendió las velas y anunció su candidatura a la gobernación.

Expression to be tested:

- **tender las velas** (idm. *take advantage of a situation*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Luis tendió las velas.

Experiment 1 - Background Stories

36

Background story:

- Ana y Pedro son personas muy atractivas.
- Anoche Pedro se encontró con Ana en la fiesta de beneficencia.
- Luego de terminar la cena, Pedro le tiró los tejos a Ana y la invitó a tomar unas copas.

Expression to be tested:

- **tirar los tejos** (idm. *flirt with someone*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Pedro le tiró los tejos a Ana.

Experiment 1 - Background Stories

37

Background story:

- Miguel es un chico muy holgazán.
- Anoche después de llegar de la universidad, se echó en el sofá a ver televisión.
- Miguel tocó la vihuela toda la noche.

Expression to be tested:

- **<u>tocar la vihuela</u>** (idm. *do nothing*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Miguel tocó la vihuela.

Experiment 1 - Background Stories

38

Background story:

- Luis es un estudiante de postgrado.
- Durante todo el verano, Luis no pudo pasar mucho tiempo afuera porque tenía mucho que estudiar para sus exámenes calificativos.
- Luis se quemó las pestañas durante todo el verano.

Expression to be tested:

- **quemarse las pestañas** (idm. *study a lot*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Luis se quemó las pestañas.

Experiment 1 - Background Stories

39

Background story:

- Pedro es un estudiante de administración.
- Ayer le fue muy mal en su entrevista de trabajo.
- Apenas el entrevistador comenzó a hacerle preguntas difíciles, Pedro tiró la toalla a mitad de la entrevista y se fue.

Expression to be tested:

- **tirar la toalla** (idm. *give up*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Pedro tiró la toalla.

Experiment 1 - Background Stories

40

Background story:

- Juan es hombre de negocios.
- Ayer las acciones de la compañía donde trabaja Juan perdieron mucho valor en la bolsa.
- Luego de es cuchar la noticia, Juan se apretó el cinturón, canceló la subscripción del cable y comenzó a cocinar en casa.

Expression to be tested:

- **apretarse el cinturón** (idm. *to reduce expenses due to lack of funds*)

Someone asks you:

- ¿Qué pasó?

You'll summarize the event using the expression underlined:

- Pasó que Juan se apretó el cinturón.

Experiment 1 - Background Stories

Idiomatic Phrase	Background Story: Literal	Background story: idiomatic
1. aplastar la oreja	José es un espía de la CIA	José es un estudiante graduado
	Anoche logró infiltrase en la casa del embajador de Rusia	Este semestre, José está tomando 20 créditos. Cuando llega el viernes, está exhausto
	Oculto fuera de la habitación, José aplastó la oreja, cuidadosamente, contra la pared para tratar de escuchar la conversación	Apenas llego a casa, José aplastó la oreja para recuperar el sueño perdido
2. beber los vientos	A Ana le encanta salir a tomar con sus amigos	Ana es una chica muy apasionada
	Anoche Ana y su amigo Eduardo fueron al bar juntos	Anoche mientras estaba en sentada al lado de una fogata, comenzó a conversar con Eduardo
	Eduardo está tomando antibióticos ahora, así que Ana bebió las cervezas por Eduardo	Casi inmediatamente, Ana bebió los vientos por Eduardo
3. cantarle las cuatro verdades	Miguel es un cantante famoso. Su amigo Nicolás es el director del coro	Miguel es un político muy astuto. Nicolás es su opositor político más importante
	Anoche durante el ensayo, Miguel finalmente completó su sección sin ningún error	Anoche durante el debate hubo un momento de intensa discusión sobre política internacional
	Miguel le cantó las cuatro estrofas a Nicolás cuando comenzó el segundo acto	Miguel le cantó las cuatro verdades a Nicolás cuando le dijo que era un político corrupto
4. comerse la olla	Luisa es una crítica gastronómica famosa	Luisa es una estudiante de derecho
	Ayer visitó el restaurante que recién recibió una estrella Michelin	Ayer pasó todo el día en la biblioteca Memorial preparándose para el examen, pero no estaba segura si había estudiado el material correcto
	Luisa se comió la paella, ¡no podía parar!	Luisa se comió la olla, ¡no podía parar!
5. cruzar la mirada	Juan y Ana son estudiantes de primer año de la universidad	Ana y Juan son personas muy sociables
	Ayer Juan y Ana aparcaron el carro en un estacionamiento cercano	Ana decidió ir a la fiesta a las 6 pm. Juan también decidió ir a la fiesta a las 6 pm

Experiment 1 - Background Stories

	Ana y Juan cruzaron la calle para llegar a la residencia	Ana y Juan cruzaron la mirada apenas llegaron en la fiesta
6. dar la lata	Luisa es la hermana mayor de Juancito	Luisa es la hermana mayor de Juancito.
	Anoche Luisa estaba preparándose para ir a una cita, pero no lograba decidir qué conjunto llevar	Anoche mientras Luisa trataba de terminar su ensayo, Juancito le hacía pregunta tras pregunta a Luisa.
	Juancito le dio la falda a Luisa para que se viera genial	Juancito le dio la lata a Luisa para que continuara prestándole atención
7. estirar la pata	Teresa es una instructora de yoga	Teresa es una señora mayor
	Ayer le sorprendió mucho descubrir que era el primer día de yoga de todos los participantes	El médico le diagnosticó cáncer pulmonar cuando fue a hacerse un chequeo anual hace unos meses
	Comenzó la clase muy pausadamente. Primero **Teresa estiró la pierna**. Comenzó con la izquierda, luego la derecha	Las cosas empeoraron rápidamente. Anoche **Teresa estiró la pata**
8. recoger las velas	Pedro es un marinero muy diestro	Pedro es un político corrupto.
	Anoche llegó al puerto de La Guaira luego de haber completado una vuelta al mundo	Ayer el senado votó a favor de impugnar a Pedro unánimemente
	Pedro **recogió las velas** después de llegar al puerto	**Pedro recogió las velas** después de escuchar el resultado de la votación y renunció a su cargo
9. matar el gusanillo	Rupertina es entomofóbica	Rupertina es una ejecutiva de ventas con una cuenta muy importante ese día
	Anoche mientras trataba de dormirse, Rupertina vio un gusanillo escurrirse sobre las mantas.	Rupertina sabía que sólo con ginebra sería capaz de relajarse ese día.
	Rupertina **mató el gusanillo** apenas lo vio	Rupertina **mató el gusanillo** apenas salió de la cama
10. romper el hielo	Juan es un conductor de maquinaria pesada	Juan es un político muy astuto
	Anoche hubo una tormenta de lluvia congelante sobre todo	Anoche tuvo una cena de beneficencia para anunciar su

Experiment 1 - Background Stories

	Wisconsin	candidatura a la presidencia
	Todo su camión estaba encapsulado en hielo. Para poder entrar a su camión, Juan **rompió el hielo** con un destornillador	Al comienzo del evento, el público estaba tenso, pero Juan **rompió el hielo** con sus anécdotas
11. sacar las castañas del fuego	Juan es un chef de pastelería muy famoso. Su amigo Luis es su asistente de pastelería	Juan es un chico privilegiado. Luis, su mejor amigo, es el juez del tribunal
	Anoche después de cerrar la tienda, Juan y Luis comenzaron a preparar una receta nueva con castañas en lugar de nueces	Anoche después de unos tragos en el bar, la policía detuvo a Juan por conducir bajo los efectos del alcohol
	Mientras Juan preparaba la masa, Luis **le sacó las castañas del fuego** a Juan	Como siempre, su amigo intervino. Luis **le sacó las castañas del fuego** a Juan evitando que lo arrestaran
12. sacar la pata del barro	Pedro es un pastor de cabras. Petunia es una de sus cabras	Pedro y Petunia son novios
	Ayer mientras exploraba el río, Petunia se quedó atascada con una de las patas en el barro	Ayer mientras corría para tomar su autobús, su mochila se rompió y la computadora golpeó el piso rompiéndose en pedazos
	Pedro **le sacó la pata del barro a Petunia** cuando la encontró al final del día	Pedro le sacó **la pata del barro a Petunia** cuando le regalo una computadora nueva
13. sacar la piedra	Luis es un padre muy cariñoso. Su hija se llama Teresa	Luis es un chico muy cómico. Siempre anda contando historias divertidas
	Ayer mientras estaban en el parque, Luis notó que Teresa cojeaba cada vez que trataba de dar un paso	Ayer mientras contaba una retahíla de chistes, dijo uno que ofendió a Teresa
	Luis descubrió que había una piedra dentro del zapato, así que Luis **le sacó la piedra** a Teresa del zapato	Luis **le sacó la piedra** a Teresa finalmente con sus chistes
14. soltar el trapo	Ana es una persona muy ordenada	Ana es muy sarcástica
	Anoche Ana comenzó a limpiar la cocina pasando un trapo húmedo por encima de todas	Anoche Ana fue a recoger a tu hermano a la escuela porque había nevado

Experiment 1 - Background Stories

	las superficies	
	De repente, notó algo moviéndose en el trapo, Ana **soltó el trato** cuando vio una araña escurriéndose hacia su mano.	Cuando vio a su hermano resbalarse en la acera, Ana **soltó el trapo** en lugar de ayudarlo
15. tender las velas	Luis es un navegante muy diestro	Luis es un político muy astuto
	Anoche soplaba el viento en dirección al oeste. Los cielos estaban despejados.	El gobernador renunció luego de que la prensa revelara los actos de corrupción de su campaña
	Luis **tendió las velas** y zarpó en dirección al próximo puerto	Luis **tendió las velas** y anunció su candidatura a la gobernación
16. tirar los tejos	Ana y Pedro son personas muy abnegadas	Ana y Pedro son personas muy atractivas
	Anoche los dos estaban construyendo casas para los indigentes de su ciudad	Anoche Pedro se encontró con Ana en la fiesta de beneficencia
	Mientras Ana estaba en el techo, Pedro le **tiró las tejas** para que las instalara en el techo	Luego de terminar la cena, Pedro **le tiró los tejos** para invitarla a tomar unas copas
17. tocar la vihuela	Miguel es un músico muy talentoso	Miguel es un chico muy holgazán
	Anoche mientras estaba visitando a sus amigos de la universidad, le pidieron que tocara	Anoche después de llegar de la universidad, se echó en el sofá a ver televisión
	Miguel **tocó la vihuela** toda la noche	Miguel **tocó la vihuela** toda la noche
18. quemarse las orejas	Luis es un estudiante de postgrado	Luis es un estudiante de postgrado
	Ayer Luis pasó el día en la playa recogiendo especímenes, pero se olvidó de ponerse un sombrero	Durante todo el verano, Luis no pudo pasar mucho tiempo afuera porque tenía mucho que estudiar para sus exámenes calificativos
	Al final del día, Luis **se quemó las orejas**	Luis **se quemó las pestañas** durante todo el verano
19. tirar la toalla	Pedro es un chico muy desordenado	Pedro es un estudiante de administración
	Ayer por la tarde fue a la	Ayer le fue muy mal en su

Experiment 1 - Background Stories

	piscina con sus amigos	entrevista de trabajo
	Apenas llegó a casa, Pedro **tiró la toalla** en el medio de la cocina.	Apenas el entrevistador comenzó a hacerle preguntas difíciles, **Pedro tiró la toalla** en el medio de la entrevista y se fue.
20. apretarse el cinturón	Juan es un chico muy a la moda	Juan es un padre de familia
	Ayer Juan compró un pantalón de una marca muy lujosa	Ayer las acciones de la compañía donde trabaja Juan perdieron mucho valor en la bolsa
	Luego se ponerse el pantalón nuevo, Juan **se apretó el cinturón**, se puso la camisa y salió al bar	Luego de escuchar la notica, Juan **se apretó el cinturón**, canceló subscripción del cable y comenzó a cocinar en casa.

Experiment 1 - Background Stories

Idiomatic Phrases

1. aplastar la oreja (idiom. *dormir*)
2. beber los vientos (idiom. *enamorarse*)
3. cantarle las cuatro verdades (idiom. *ser brutalmente honesto*)
4. comerse la olla (idiom. *pensar o preocuparse demasiado*)
5. cruzar la mirada (idiom. *establecer contacto con la mirada brevemente*)
6. dar la lata (idiom. *molestar excesivamente*)
7. estirar la pata (idiom. *morir*)
8. recoger las velas (idiom. *salir súbitamente*)
9. matar el gusanillo (idiom. *beber alcohol en ayunas*)
10. romper el hielo (idiom. *establecer contacto social*)
11. sacar las castañas del fuego (idiom. *ayudar en una situación de riesgo*)
12. sacar la pata del barro (idiom. *ayudar*)
13. sacar la piedra (idiom. *molestar excesivamente)*
14. soltar el trapo (idiom. *reír o llorar repentinamente*)
15. tender las velas (idiom. *tomar ventaja de una situación*)
16. tirar los tejos (idiom. *coquetear*)
17. tocar la vihuela (idiom. *no hacer nada*)
18. quemarse las pestañas (idiom. *estudiar mucho*)
19. tirar la toalla (idiom. *rendirse*)
20. apretarse el cinturón (idiom. *reducir los gastos por escasez*)

Appendix D

Exit Poll - Idiomatic Expressions

Thank you for lending your time to my investigation. The section below lists all the idiomatic phrases you were asked to work with during the interview. Some of the phrases might have been unfamiliar to you, if you had never heard of it, mark it with a 1. If you use the phrase yourself or hear it being used frequently, mark it with a 5. It is okay to say you are familiar with the phrase even if your particular variety uses a slightly different version of the phrase such as 'comerse la cabeza" instead of "comerse la olla".

* Required

1. Email *

Indicate the level of familiarity you have with the idiomatic expression

2. (1) aplastar la oreja / doblar la oreja (idm. 'dormir') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

3. (2) beber los vientos por alguien (idm. 'enamorarse') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

4. (3) cantarle las cuarenta/las cuatro verdades a alguien (idm. 'ser brutalmente honesto') * 243

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

5. (4) comerse la olla/la cabeza (idm. 'pensar o preocuparse en exceso') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

6. (5) cruzar la mirada (idm. 'hacer contacto visual brevemente') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

7. (6) dar la lata (idm. 'molestar o fastidiar excesivamente') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

8. (7) estirar la pata (idm. 'morir') * 244

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

9. (8) recoger las velas (idm. 'salir súbitamente') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

10. (9) matar el gusanillo (idm. 'beber alcohol especialmente en ayunas') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

11. (10) romper el hielo (idm. 'superar la incomodidad inicial del trato social') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

12. (11) sacar las castañas del fuego por alguien (idm. 'ayudar a alguien especialmente si pone en riesgo al que ayuda') * 245

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

13. (12) sacarle la pata del barrio a alguien (idm. 'ayudar') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

14. (13) sacarle la piedra a alguien (idm. 'molestar excesivamente') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

15. (14) soltar el trapo (idm. 'reír o llorar repentivamente) *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

16. (15) tender la velas (idm. 'aprovecharse de una situación favorable') * 246

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

17. (16) tirar los tejos (idm. 'coquetear') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

18. (17) tocar la vihuela (idm. 'no hacer nada') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

19. (18) quemarse las pestañas (idm. 'estudiar mucho') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

20. (19) tirar la toalla (idm. 'rendirse') * 247

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

21. (20) apretarse el cinturón (idm. 'reducir gastos por escasez de recursos') *

Mark only one oval.

	1	2	3	4	5	
Never heard of it before	◯	◯	◯	◯	◯	I use it myself or have heard it being used

22. If you would like to comment about any aspect of the investigation, share your thoughts below

Appendix E

Dear fellow UW-Madison student,

I am conducting a study on idiomatic expressions in Spanish. As a student in the Department of Spanish and Portuguese here at UW-Madison, I believe your native proficiency could help me expand our understanding of idiomaticity. In particular, I am looking at how (or if) the idiomatic interpretation of a sentence relates systematically to the literal interpretation of the same sentence.

I would like to ask for your participation in phase 1 of my study, which will be conducted in a quiet room in Van Hise Hall at a time that is convenient for you. The task you will complete consists of (1) reading a set of background stories and (2) producing a summarizing statement. Each background story is a few sentences long. You will be audio-recorded producing that summarizing statement. Your participation will take approximately 30 minutes to 45 minutes to complete.

To give you an example of what you will be required to do, consider the Spanish idiom *sacar las castañas del fuego por alguien*. There will be two background stories: one that is consistent with the literal reading and one that is consistent with the idiomatic reading. Your task is to read to yourself steps (a), (b), and (c); then you will be recorded reading aloud the **boldfaced** sentence in step (d) below. You may be asked to read one of the two stories or both for the set of idioms I'm investigating.

Idiomatic reading	Literal reading
(a) Juan es un chico privilegiado, su padre es uno de lo jueces del tribunal supremo de justicia.	(a) Juan es un chef de pastelería muy famoso, su amigo Luis es su asistente de pastelería.
(b) Él siempre ha dependido de sus padres para resolver cualquier problema que se le pueda presentar	(b) Toda la ciudad sabe de Juan, especialmente por sus pasteles de nueces.
(c) Anoche después de unos tragos en el bar, la policía detuvo a Juan por conducir bajo los efectos del alcohol.	(c) Anoche, después de cerrar la tienda, Juan y su amigo Luis comenzaron a preparar una receta nueva con castañas en lugar de nueces.
(d) **Como siempre, su padre intervino y le sacó las castañas del fuego a Juan.**	(d) **Mientras Juan preparaba la masa, Luis le sacó las castañas del fuego a Juan.**

As an incentive for your participation, I will provide coffee and pastries the day of the recording and your name will be entered into a drawing for a $50 Amazon gift card. The drawing will take place once phase 2 is completed. Please reply to this email to confirm your willingness to participate in the study. I will contact you to work out the details. If you have any questions, please do not hesitate to ask.

Sincerely,

Erwin Lares
lares@wisc.edu
Ph.D. Candidate
Department of Spanish & Portuguese UW Madison

Appendix F

	counterfactual: targets inceptive telos point incompletive: targets terminative telos point	start-of-event: targets inceptive telos point end-of-event: targets terminative telos point	interruptive: targets a t_i / $t_{inceptive} < t_i < t_{terminative}$ repetitive: targets the entire event
expression estirar la pata	casi v-ed	took -time to v	stopped v-ing
	casi estiró la pata	tomó 10 minutos en estirar la pata	paró de estirar la pata
Literalmente, la expresión **estirar la pata** ocurre cuando un animal extiende una de sus extremidades. En situaciones informales o en broma, la pata se puede utilizar para referirse a la pierna de una persona.	Hablando de su mascota Firuláis, ¿puede decir Pedro que "Firuláis casi estiró la pata" para indicar una **situación literal** donde Firuláis tenía que estirar la pata (para que le revisarán una herida), pero el dolor era tan intenso que no pudo ni siquiera comenzar a estirarla? → nunca ocurrió el evento.	Hablando de su mascota Firuláis, ¿puede decir Pedro que "Firuláis tomó 10 minutos en estirar la pata" para indicar una **situación literal** donde Firuláis tenía que estirar la pata (para que le revisarán una herida), pero el dolor era tan intenso que Pedro tuvo que calmarlo por 10 minutos antes de que comenzara a estirar la pata? → pasaron 10 minutos y comenzó el evento	Teresa paró de entender la pierna: Teresa comenzó a extender la pierna, a mitad de estirarla, paró de hacerlo. → el evento se interrumpió en el medio
	Hablando de su mascota Firuláis, ¿puede decir Pedro que "Firuláis casi estiró la pata" para indicar una **situación literal** donde Firuláis comenzó a estirar la pata, pero no logró terminar de hacerlo? Es decir, comenzó a estirarla, pero el dolor le impedía estirarla completamente. → comenzó el evento, nunca culminó	Hablando de su mascota Firuláis, ¿puede decir Pedro que "Firuláis tomó 10 minutos en estirar la pata" para indicar una **situación literal** donde Firuláis comenzó a estirar la pata, pero como le dolía tanto la retraía una y otra vez, así que el proceso de estirar la pata completamente duró 10 minutos. → comenzó el evento, luego de 10 minuto, concluyó	Teresa paró de extender la pierna: Teresa estiró la pierna una vez, pero luego de haber estirado la pierna, decidió no hacerlo más. → se interrumpió la posible iteración del evento
Figurativamente, **estirar la pata** es una expresión que significa "morir".	Hablando de su mascota Firuláis, ¿puede decir Pedro que "Firuláis casi estiró la pata" para indicar una **situación figurada** donde pasó algo que podría haber causado la muerte de Firuláis, pero terminó no ocurriendo, así que Firuláis sigue vivo? Es decir, no llegó a morir. → nunca ocurrió el evento	Hablando de su mascota Firuláis, ¿puede decir Pedro que "Firuláis tomó 10 minutos en estirar la pata" para indicar una **situación figurada** donde Firuláis fue atropellado por un automóvil y después de transcurridos 10 minutos, Firuláis falleció. → pasaron 10 minuto y comenzó el evento	~~Teresa paró de morir: Teresa comenzó a morir, a mitad de morir, paró de hacerlo.~~ ~~→ el evento se interrumpió en el medio~~
	Hablando de su mascota Firuláis, ¿puede decir Pedro que "Firuláis casi estiró la pata" para indicar una **situación figurada** donde comenzó a morir, pero no logró terminar de morir? → comenzó el evento, nunca culminó	Hablando de su mascota Firuláis, ¿puede decir Pedro que "Firuláis tomó 10 minutos en estirar la pata" para indicar una **situación figurada** donde Firuláis fue atropellado por un automóvil, comenzó a morir y luego de 10 minutos falleció. → comenzó el evento, luego de 10 minutos, concluyó.	~~Teresa murió una vez, pero luego de haber muerto, decidió no hacerlo más~~ ~~→ se interrumpió la iteración del evento~~

	counterfactual: targets inceptive telos point incompletive: targets terminative telos point	start-of-event: targets inceptive telos point end-of-event: targets terminative telos point	interruptive: targets a t_i / $t_{inceptive} < t_i < t_{terminative}$ repetitive: targets the entire event
	casi v-ed	took -time to v	stopped v-ing
soltar el trapo Literalmente, la expresión **soltar el trapo** ocurre cuando alguien deja caer un trapo al suelo.	casi soltó el trapo	tomó 10 minutos en soltar el trapo	paró de soltar el trapo
	¿Puede alguien decir que "Pedro casi soltó el trapo" para indicar una **situación literal** donde Pedro iba a soltar soltar un trapo pero a último momento cambió de opinión y no lo hizo? Es decir, ni siquiera ocurrió. → nunca ocurrió el evento.	¿Puede alguien decir que "Pedro tomó 10 minutos en soltar el trapo" para indicar una **situación literal** donde luego de 10 minutos, Pedro dejó que el trapo se cayera. → pasaron 10 minutos y comenzó el evento	~~Ana paró de soltar el trapo: Ana soltó el trapo, a mitad de soltar el trapo, paró de hacerlo.~~ → el evento se interrumpió en el medio
	¿Puede alguien decir que "Pedro casi soltó el trapo" para indicar una **situación literal** donde Pedro comenzó a soltar el trapo, pero no logró terminar de hacerlo? Es decir, comenzó a soltarlo pero a mitad de soltarlo algo o alguien interrumpió a Pedro por lo que no logró terminar de soltarlo. → comenzó el evento, nunca culminó	¿Puede alguien decir que "Pedro tomó 10 minutos en soltar el trapo" para indicar una **situación literal** donde Pedro comenzó soltar el trapo y le tomó 10 minutos en terminar de hacerlo. → comenzó el evento, luego de 10 minuto, concluyó	Ana paró de soltar el trapo: Ana soltó el trapo una vez, pero luego de haberlo soltado, decidió no hacerlo más. → se interrumpió la posible iteración del evento
Figurativamente, **soltar el trapo** es una expresión que significa, "echarse a reír".	¿Puede alguien decir que "Pedro casi soltó el trapo" para indicar una **situación figurada** donde Pedro iba a echarse a reír, pero justo antes, un ataque cardíaco fulminante evitó que se riera. Es decir, nunca llegó a reírse. → nunca ocurrió el evento.	¿Puede alguien decir que "Pedro tomó 10 minutos en soltar el trapo" para indicar una **situación figurada** donde luego de 10 minutos de escuchar una historia divertida, Pedro comenzó a reírse. → pasaron 10 minutos y comenzó el evento	~~Ana paró de echarse a reír: Ana se echó a reír, pero a mitad de echarse reír, paró de hacerlo.~~ → el evento se interrumpió en el medio
	¿Puede alguien decir que "Pedro casi soltó el trapo" para indicar una **situación figurada** donde Pedro comenzó a reírse pero a mitad de reírse pero no logró terminar de hacerlo? Es decir, se echó a reír, pero algo o alguien interrumpió a Pedro por lo que no terminó de reírse. → comenzó el evento, nunca culminó	¿Puede alguien decir que "Pedro tomó 10 minutos en soltar el trapo" para indicar una **situación figurada** donde al escuchar un chiste, Pedro comenzó a reírse. Luego de 10 minutos, Pedro terminó de reírse. → comenzó el evento, luego de 10 minuto, concluyó	Ana paró de echarse a reír: Ana se echó a reír una vez, pero luego de haberse echado a reír, decidió no hacerlo más. → se interrumpió la posible iteración del evento

	counterfactual: targets inceptive telos point incompletive: targets terminative telos point casi v-ed	start-of-event: targets inceptive telos point end-of-event: targets terminative telos point took -time to v	interruptive: targets a t_i / $t_{inceptive} < t_i < t_{terminative}$ repetitive: targets the entire event stopped v-ing
tender la velas Literalmente, la expresión **tender la velas** ocurre cuando al soplar buen viento, se despliegan las velas de una embarcación para que estas atrapen el viento y se pueda navegar.	casí tendió las velas	tomó diez minutos en tender la velas	paró de tender las velas
	¿Puede alguien decir que "Pedro casi tendió las velas" para indicar una **<u>situación literal</u>** donde Pedro tenía planeado desplegar las velas pero a último momento cambió de opinión y no lo hizo? Es decir, ni siquiera comenzó a hacerlo. → nunca ocurrió el evento.	¿Puede alguien decir que "Pedro tomó 10 minutos en tender las velas" para indicar una **<u>situación literal</u>** donde luego de 10 minutos (quizá tenía que ponerse equipo de seguridad), Pedro comenzó a desplegar las velas. → pasaron 10 minutos y comenzó el evento	Luis paró de desplegar las velas: Luis comenzó a desplegar las velas, pero a mitad de desplegarlas, paró de hacerlo. → el evento se interrumpió en el medio
	¿Puede alguien decir que "Pedro casi tendió las velas" para indicar una **<u>situación literal</u>** donde Pedro comenzó a desplegar las velas, pero no logró terminar de hacerlo? Es decir, comenzó a desplegarlas, pero algo o alguien interrumpió a Pedro por lo que no logró terminar de hacerlo. → comenzó el evento, nunca culminó	¿Puede alguien decir que "Pedro tomó 10 minutos en tender las velas" para indicar una **<u>situación literal</u>** donde Pedro comenzó a desplegar las velas y le tomó 10 minutos terminar de hacerlo. → comenzó el evento, luego de 10 minuto, concluyó	Luis paró de desplegar las velas: Luis desplegó las velas una vez, pero luego de haberlas desplegado decidió no hacerlo más. → se interrumpió la posible iteración del evento
Figurativamente, **tender las velas** es una expresión que significa, cuando se dice de una persona, "aprovecharse de una situación ventajosa".	¿Puede alguien decir que "Pedro casi tendió las velas" para indicar una **<u>situación figurada</u>** donde Pedro tenía planeado aprovecharse de una situación ventajosa, pero a último momento, cambió de opinión y no lo hizo? Es decir, ni siquiera comenzó a hacerlo. → nunca ocurrió el evento.	¿Puede alguien decir que "Pedro tomó 10 minutos en tender las velas" para indicar una **<u>situación figurada</u>** donde donde luego de 10 minutos (quizá Pedro observaba con detenimiento los que sucedía), Pedro comenzó a aprovecharse de la situación ventajosa. → pasaron 10 minutos y comenzó el evento	~~Luis paró de aprovecharse de la situación ventajosa: Luis comenzó a aprovecharse de la situación ventajosa, pero a mitad de aprovecharse, paró de hacerlo.~~ → el evento se interrumpió en el medio
	¿Puede alguien decir que "Pedro casi tendió las velas" para indicar una **<u>situación figurada</u>** donde Pedro comenzó a aprovecharse de una situación ventajosa, pero no logró terminar de aprovecharse de la situación? Es decir, comenzó a aprovecharse, pero algo o alguien interrumpió a Pedro por lo que no terminó de hacerlo. → comenzó el evento, nunca culminó	¿Puede alguien decir que "Pedro tomó 10 minutos en tender las velas" para indicar una **<u>situación figurada</u>** donde Pedro comenzó a aprovecharse de una situación favorable y luego de transcurridos 10 minutos, Pedro logró hacerlo. → comenzó el evento, luego de 10 minutos, concluyó	Luis paró de aprovecharse de la situación ventajosa: Luis se aprovechó de la situación ventajosa una vez, pero luego de haberse aprovechado de la situación, decidió no hacerlo más. → se interrumpió la posible iteración del evento

	counterfactual: targets inceptive telos point incompletive: targets terminative telos point	start-of-event: targets inceptive telos point end-of-event: targets terminative telos point	interruptive: targets a t_i / $t_{inceptive} < t_i < t_{terminative}$ repetitive: targets the entire event
	casi v-ed	took -time to v	stopped v-ing
quemarse las pestañas ss Literalmente, la expresión **quemarse las pestañas** ocurre cuando las pestañas de una persona se queman por contacto con el fuego.	casi se quemó las pestañas ¿Puede alguien decir que "Pedro casi se quemó las pestañas" para indicar una **situación literal** donde Pedro tenía planeado quemarse las pestañas (en uno de esos desafíos online) pero a último momento cambió de opinión y no lo hizo? Es decir, ni siquiera comenzó a hacerlo. → nunca ocurrió el evento. ¿Puede alguien decir que "Pedro casi se quemó las pestañas" para indicar una **situación literal** donde Pedro comenzó a quemarse las pestañas, pero no logró terminar de quemárselas? Es decir, comenzó pero después de quemarse las primeras pestañas, el dolor era tan intenso que no pudo terminar de hacerlo. → comenzó el evento, nunca culminó	tomó diez minutos en quemarse las pestañas ¿Puede alguien decir que "Pedro tomó 10 minutos en quemarse las pestañas" para indicar una **situación literal** donde luego de 10 minutos, comenzó a quemarse las pestañas con un encendedor. → pasaron 10 minutos y comenzó el evento ¿Puede alguien decir que "Pedro tomó 10 minutos en quemarse las pestañas" para indicar una **situación literal** donde Pedro comenzó a quemarse las pestañas y le tomó 10 minutos terminar de hacerlo porque tuvo que quemar cada una de las pestañas individualmente. → comenzó el evento, luego de 10 minutos, concluyó	paró de quemarse las pestañas Luis paró de quemarse las pestañas: Luis comenzó a quemarse las pestañas, pero a mitad de quemarse las pestañas, paró de hacerlo. → el evento se interrumpió en el medio Luis paró de quemarse las pestañas: Luis se quemó las pestañas una vez, pero luego de haberse quemado las pestañas, decidió no hacerlo más → se interrumpió la posible iteración del evento
Figurativamente, **quemarse las pestañas** es una expresión que significa "estudiar con ahínco".	¿Puede alguien decir que "Pedro casi se quemó las pestañas" para indicar una **situación figurada** donde Pedro tenía planeado estudiar con ahínco, pero a último momento, cambió de opinión y no lo hizo? Es decir, ni siquiera comenzó a hacerlo. → nunca ocurrió el evento. ¿Puede alguien decir que "Pedro casi se quemó las pestañas" para indicar una **situación figurada** donde Pedro comenzó a estudiar con ahínco, pero algo o alguien interrumpió a Pedro por lo que no logró terminar de hacerlo. → comenzó el evento, nunca culminó	¿Puede alguien decir que "Pedro tomó 10 minutos en quemarse las pestañas" para indicar una **situación figurada** donde donde luego de 10 minutos, Pedro comenzó a estudiar con ahínco. → pasaron 10 minutos y comenzó el evento ¿Puede alguien decir que "Pedro tomó 10 minutos en quemarse las pestañas" para indicar una **situación figurada** donde Pedro comenzó a estudiar con ahínco, luego de 10 minutos, Pedro terminó de hacerlo. → comenzó el evento, luego de 10 minuto, concluyó	Luis paró de estudiar con ahínco: Luis estudió con ahínco una vez, pero a mitad de estudiar con ahínco, paró de hacerlo. → el evento se interrumpió en el medio Luis paró de estudiar con ahínco: Luis estudió con ahínco una vez, pero luego de haber estudiado con ahínco, decidió no hacerlo más → se interrumpió la posible iteración del evento

Appendix G

Walkthrough with the *quemarse las pestañas* and the *almost* adverbial

Quemarse las pestañas was also paired up with the *almost* adverbial. The sentences for *quemarse las pestañas* are shown in (39).

(39) a. Pedro casi se quemó las pestañas
Pedro almost se.CL burn.PRET the eyelashes
Lit: "Pedro almost burned his eye lashes"

b. Pedro casi se quemó las pestañas
Pedro almost se.CL burn.PRET the eyelashes
idm: "Pedro almost studied hard"

The literal interpretation of *quemarse las pestañas* and the almost adverbial (39) result in two distinct readings: a counterfactual and an incompletive. *Quemarse las pestañas* interpreted literally is an accomplishment. By contrast, when interpreted idiomatically, *quemarse las pestañas* tests out as an activity, with only the counterfactual reading being available.

The data gathering followed the same steps I detailed above for the expression *estirar la pata*. The four scenarios for *quemarse las pestañas* are detailed below:

1. Introducing the expression
    - Literalmente, la expresión **quemarse las pestañas** ocurre cuando las pestañas de una persona se queman por contacto con el fuego. (Literally the expression **burn one's eyelashes** occurs when a person's eyelashes are burned by contact with fire.)
    - Figurativamente, **burn one's eyelashes** es una expresión que significa "estudiar con ahínco." (Figuratively, **eyelash burning** is a expression that means "study hard")
2. The four crafted scenarios are listed below, although participants interacted with

them in random order. The response collected from the participant was the same regardless of the expression or the interpretation. It consisted of a statement of agreement or disagreement as to whether the expression can be used to describe each situation.

1. **Intended use: literal interpretation & counterfactual reading**
    - ¿Puede alguien decir que "Pedro casi se quemó las pestañas" para indicar una situación literal donde Pedro tenía planeado quemarse las pestañas (en uno de esos desafíos online) pero a último momento cambió de opinión y no lo hizo? Es decir, ni siquiera comenzó a hacerlo.
    - Can someone say that "Pedro almost burned his eyelashes" to indicate a literal situation where Pedro had planned to burn his eyelashes (one of those online challenges) but at the last minute changed his mind and didn't? That is, he didn't even start to do it.
2. **Intended use: literal interpretation & incompletive reading**
    - ¿Puede alguien decir que "Pedro casi se quemó las pestañas" para indicar una situación literal donde Pedro comenzó a quemarse las pestañas, pero no logró terminar de quemárselas? Es decir, comenzó pero después de quemarse las primeras pestañas, el dolor era tan intenso que no pudo terminar de hacerlo.
    - Can someone say that "Pedro almost burned his eyelashes" to indicate a literal situation where Pedro started burning his eyelashes, but couldn't finish burning his eyelashes? That is, he started it but after burning a few eyelashes the pain was so intense that he couldn't finish it.
3. **Intended use: idiomatic interpretation & counterfactual reading**
    - ¿Puede alguien decir que "Pedro casi se quemó las pestañas" para indicar una situación figurada donde Pedro tenía planeado estudiar con ahínco, pero a último momento, cambió de opinión y no lo hizo? Es decir, ni

siquiera comenzó a hacerlo.

- Can someone say that “Pedro almost burned his eyelashes” to indicate a figurative situation where Pedro had planned to study hard, but at the last minute, he changed his mind and didn’t do it? That is, he didn’t even start to do it.

4. **Intended use: idiomatic interpretation & incompletive reading**
    - ¿Puede alguien decir que “Pedro casi se quemó las pestañas” para indicar una situación figurada donde Pedro comenzó a estudiar con ahínco, pero algo o alguien interrumpió a Pedro por lo que no logró terminar de hacerlo.
    - Can someone say that “Pedro almost burned his eyelashes” to indicate a figurative situation where Pedro began to study hard, but something or someone interrupted Pedro so he could not finish doing it.

Walkthrough with the *soltar el trapo* and the *takes-x-time* adverbial

For the expression *soltar el trapo*, the same basic process was followed, as shown in (40). I describe the details of the interactions below. As before, the collected response was whether or not the participant believed the statement could be used to accurately describe the scenario.

(40) a. Pedro tomó 10 minutos en soltar el trapo
Pedro take.PRET 10 minutes in drop.INF the rag
Lit: "Pedro took 10 minutes to drop the rag"

b. Pedro tomó 10 minutos en soltar el trapo
Pedro take.PRET 10 minutes in drop.INF the rag
idm: "Pedro took 10 minutes to burst out laughing"

1. Introducing the expressions
    - Literalmente, la expresión soltar el trapo ocurre cuando alguien deja caer un trapo al suelo. (Literally, the expression **drop the rag** occurs when someone drops a cloth on the ground.)

- Figurativamente, soltar el trapo es una expresión que significa, “echarse a reír.” (Figuratively, **drop the rag** is an expression that means “to burst out laughing.”)

2. The four crafted scenarios are listed below, although participants interacted with them in random order. The response collected from the participant was the same regardless of the expression or the interpretation. It consisted of a statement of agreement or disagreement as to whether the expression can be used to describe each situation.
    1. **Intended use:literal interpretation & inceptive reading**
        - ¿Puede alguien decir que “Pedro tomó 10 minutos en soltar el trapo” para indicar una situación literal donde luego de 10 minutos, Pedro dejó que el trapo se cayera.
        - Can anyone say that “Pedro took 10 minutes to drop the rag” to indicate a literal situation where after 10 minutes, Pedro let the rag fall to the ground.
    2. **Intended use: literal interpretation & terminative reading**
        - ¿Puede alguien decir que “Pedro tomó 10 minutos en soltar el trapo” para indicar una situación literal donde Pedro comenzó soltar el trapo y le tomó 10 minutos en terminar de hacerlo.
        - Can someone say that “Peter took 10 minutes to drop the rag” to indicate a literal situation where Peter started to drop the rag and it took him 10 minutes to finish.
    3. **Intended use: idiomatic interpretation & inceptive reading**
        - ¿Puede alguien decir que “Pedro tomó 10 minutos en soltar el trapo” para indicar una situación figurada donde luego de 10 minutos de escuchar una historia divertida, Pedro comenzó a reírse.
        - Can someone say that “Pedro took 10 minutes to drop the rag” to indicate

a figurative situation where after 10 minutes of listening to a funny story, Pedro began to laugh.

4. **Intended use: idiomatic interpretation & terminative reading**
    - ¿Puede alguien decir que “Pedro tomó 10 minutos en soltar el trapo” para indicar una situación figurada donde al escuchar un chiste, Pedro comenzó a reírse. Luego de 10 minutos, Pedro terminó de echarse a reír.
    - Can someone say that “Pedro took 10 minutes to drop the cloth” to indicate a figurative situation where upon hearing a joke, Pedro began to laugh. After 10 minutes, Pedro finished bursting out laughing.

For the expression *tender las velas*, the same basic process was followed, as described in (41). I describe the details of the interaction below. As before, the collected response was whether or not the participant believed the statement matched the scenario described.

(41) a. Pedro tomó 10 minutos en tender las velas
Pedro take.PRET 10 minutes in unfurl.INF the sails
Lit: "Pedro took 10 minutes to unfurl the sails"

b. Pedro tomó 10 minutos en tender las velas
Pedro take.PRET 10 minutes in unfurl.INF the sails
idm: "Pedro took 10 minutes to take advantage of the situation"

1. Introducing the expressions
    - Literalmente, la expresión **tender las velas** ocurre cuando al soplar buen viento, se despliegan las velas de una embarcación para que estas atrapen el viento y se pueda navegar. (Literally, the expression **unfurl the sails** occurs when a favorable wind blows, the sails of a boat are unfurled so they can catch the wind and one can sail)
    - Figurativamente, tender las velas es una expresión que significa, cuando se dice de una persona, “aprovecharse de una situación ventajosa.” (Figuratively, **unfurl the sails** is an expression that means, when said of a person, “to take

advantage of a favorable situation")

2. The four crafted scenarios are listed below, although participants interacted with them in random order. The response collected from the participant was the same regardless of the expression or the interpretation. It consisted of a statement of agreement or disagreement as to whether the expression can be used to describe each situation.

    1. **Intended use: literal interpretation & inceptive reading**
        - ¿Puede alguien decir que "Pedro tomó 10 minutos en tender las velas" para indicar una situación literal donde luego de 10 minutos (quizá tenía que ponerse equipo de seguridad), Pedro comenzó a desplegar las velas?
        - Can someone say that "Pedro took 10 minutes to unfurl the sails" to indicate a literal situation where after 10 minutes (maybe he had to put on safety gear before), Pedro began to unfurl the sails.
    2. **Intended use: literal interpretation & terminative reading**
        - ¿Puede alguien decir que "Pedro tomó 10 minutos en tender las velas" para indicar una situación literal donde Pedro comenzó a desplegar las velas y le tomó 10 minutos terminar de hacerlo?
        - Can anyone say that "Peter took 10 minutes to unfurl the sails" to indicate a literal situation where Peter started to unfurl the sails and it took him 10 minutes to finish doing so?
    3. **Intended use: idiomatic interpretation & inceptive reading**
        - ¿Puede alguien decir que "Pedro tomó 10 minutos en tender las velas" para indicar una situación figurada donde luego de 10 minutos (quizá Pedro observaba con detenimiento los que sucedía), Pedro comenzó a aprovecharse de la situación ventajosa?
        - Can someone say that "Pedro took 10 minutes to unfurl the sails" to indicate a figurative situation where after 10 minutes (perhaps Pedro was

carefully observing what happened), Pedro began to take advantage of the favorable situation.

4. **Intended use: idiomatic interpretation & terminative reading**
    - ¿Puede alguien decir que “Pedro tomó 10 minutos en tender las velas” para indicar una situación figurada donde Pedro comenzó a aprovecharse de una situación favorable y luego de transcurridos 10 minutos, Pedro logró hacerlo?
    - Can someone say that “Pedro took 10 minutes to unfurl the sails” to indicate a figurative situation where Pedro began to take advantage of the favorable situation. After 10 minutes, Pedro finished taking advantage of the situation.

5.1 Appendix H

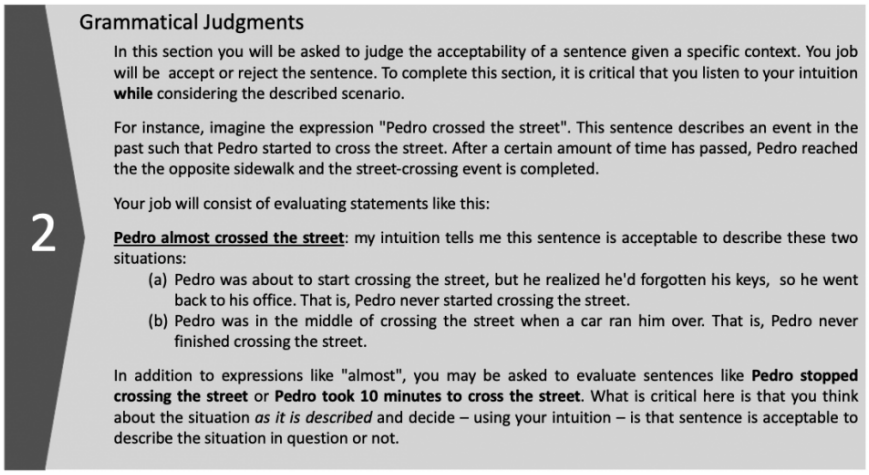


Figure 42: Introduction to Grammatical Judgments