Proceedings of the 21st Meeting of the Texas Linguistics Society

Editors:
Barrett Hamp
Taylor Joyce
Lee Orfila
Venkata S Govindarajan

February 11–12, 2022

Message from the Editors

The 21st meeting of the Teas Linguistics Society (TLS 2022) was held Feb 11–12, 2022, at the University of Texas at Austin. Presentations came from all areas of linguistics, but this year's conference placed a special focus on language identity and language learning.

Many thanks to our keynote speakers: Fernando Llanos (The University of Texas at Austin), Joseph Hill (Rochester Institute of Technology), and Wesley Leonard (University of California, Riverside). Many thanks also to our sponsors at the University of Texas at Austin: the Events Co-sponsorship Board, Department of Linguistics, Department of Asian Studies, Department of Germanic Studies, Department of French and Italian, Department of Middle Eastern Studies, Department of Spanish and Portuguese, Department of Latin American Studies, Department of Native American and Indigenous Studies, Department of Women and Gender Studies, and College of Liberal Arts. Finally, thanks to our presenters and especially those who chose to include a paper in the proceedings; we hope this publication helps your research find an even wider audience.

For more information about the conference, including a listing of all presentations and abstracts, see the conference website at `http://tls.ling.utexas.edu/2022/`. To cite these proceedings, use: Hamp, B., Joyce, T., Orfila, L. and Govindarajan, V. (Eds). (2022). Proceedings of the 21st Meeting of the Texas Linguistics Society.

Signed, the editors:

Barrett Hamp, Taylor Joyce, Lee Orfila, Venkata S Govindarajan
Department of Linguistics,
The University of Texas at Austin

Contents

The Syntactic Role of Ideophones in English

Meredith Belloni
Tulane University
mbelloni@tulane.edu

Abstract

Ideophones are particularly evocative words used to communicate a particular sensory perception or experience. So far, the study of ideophones has focused on their form rather than their function. This paper aims to study ideophones as they function in the syntax of English. I review my preliminary observations about the construction, variation, and common contexts of use of English ideophones, then demonstrate how ideophones can function syntactically and provide common syntactic constructions for all ideophone types. Special attention is paid to patterns within and across the hierarchical organization of ideophones. I provide supporting evidence for the inverse relationship between expressiveness and syntactic integration, originally proposed by Dingemanse and Akita (2017). I argue that the degree to which an ideophone must be overtly performed, rather than simply spoken, is essential in understanding its markedness, lexification, and syntactic distribution. Also explored in this paper are English examples which create new questions for the classification of ambiguous ideophones. Implications for future directions of research are discussed.

1 Introduction

The term ideophone refers to a class of marked and uniquely depictive words which are used to evoke a particular sensory experience via an expressive and imitative phonology (Akita & Dingemanse, 2019). In English, the obvious examples are onomatopoeias which are ideophones for sounds, such as *bang* or *snap*. Ideophones can also evoke senses other than sound, though there is no colloquial name for these. Despite this, English speakers use ideophones all the time in their everyday speech to depict, to evoke, and to enhance the sensory qualities of their speech (Akita & Dingemanse, 2019; Dingemanse, 2013).

The study of ideophones has been in progress for over one hundred years but has primarily been done only in languages in which ideophones are plentiful. Languages

with fewer ideophones, including English, have been entirely left out. This leaves an obvious gap in our understanding of ideophones in general and how they may function specifically in those languages where they are less common. Nuckolls (1999, page 225) even claims that ideophones are "conspicuously underdeveloped in standard average European languages" and calls this a puzzle for linguists, despite the fact that ideophones have hardly been studied in any Indo-European languages. In this analysis, I aim to fill in some of this gap in the literature by examining how ideophones function in the syntax of English. I will review the existing literature on ideophones, with particular attention to the hierarchy of ideophones coined by Dingemanse (2012), the markedness of ideophones, the potential relationship between markedness and syntax, the syntactic function of ideophones in other languages, and iconicity in syntax. I will then present the preliminary work I have done on ideophones in English. For my analysis, I will demonstrate the syntactic properties of English ideophones at each tier of the hierarchy and explore the patterns that exist across tiers. I will also test these patterns to determine if they are useful for resolving the ambiguity of potential, but untested, ideophones.

2   Literature review

Thus far, the study of ideophones has mainly been conducted in African and Asian languages, in which they are plentiful. Though, even in these languages with larger quantities of ideophones, they are disdained by speakers and thought of as somewhat less than full words (Childs, 2001). Despite this, ideophones are regularly used in speech, particularly in social speech contexts and in storytelling/narrative contexts (Childs, 2001; Dingemanse, 2013; Nuckolls, 1999). Ideophones can be characterized by their two most distinctive traits— markedness and depictiveness.

2.1   Markedness

Ideophones are almost always marked both in their construction and in their presentation within speech Newman (2001). Despite this, ideophones are words, by definition, regardless of their markedness and are made from mostly the same phonetic content and phonological rules as the prosaic part of the language. Markedness is not a necessary or sufficient features of ideophones but, rather, is a prototypical feature of them as a lexical class (Akita, 2009; Akita & Dingemanse, 2019; Childs, 1994; Dingemanse, 2012). This variation in markedness leads to some words seeming more ideophonic than others, with less ideophonic words like *chatter*, *splash*, or *slush* passing as normal, prosaic words despite their depictive nature (Akita, 2009).

The imitative and highly expressive nature of ideophones allows them to use

features that other words in their language normally wouldn't; examples in English include the onset in *vroom* or the coda on *buzz* (Akita & Dingemanse, 2019). Ideophones can also be marked morpho-syntactically, being particularly susceptible to processes like lengthening and reduplication (Akita & Dingemanse, 2019; Newman, 1989). The prominence of markedness among ideophones can lead to misconceptions about how set apart they are from the prosaic part of a language (Newman, 2001); Newman (1989) points out that, in Hausa, ideophones are still identifiable as words of the language and not some other language and that they do follow some of the rules, such as only having sounds that are already part of Hausa's phonological inventory. But there is counterevidence in a language like Yoruba, in which the only syllabic nasals in the language appear in its ideophones (Nuckolls, 1999). Ideophones exist to depict and must do so while satisfying both the need to be distinct from the rest of the language and the need to be a recognizable part of that language (Childs, 2014).

2.2 Depictiveness

The ability to depict, rather than to describe, is another key aspect of the ideophone identity. Depiction is described as a mode of language in which relies on mimesis and imitation (Clark & Gerrig, 1990; Dingemanse & Akita, 2017). Lexical depiction uses non-arbitrary mappings of form and meaning, with sound symbolic elements which are less discreet than the phonological content of prosaic words (Akita & Dingemanse, 2019). Experimental work (Dingemanse et al., 2016; Howard, 2019) has shown that the sound symbolism of ideophones can cross languages and that the iconicity of ideophones relies not only on phonetic segments but also on prosody. It is this combination of iconicity and arbitrariness that makes an ideophone both uniquely depictive and a word in its language.

Thompson and Do (2019) lay out a methodology for testing ideophones to determine whether a pattern of phonemes across related words is rooted in iconicity or is simply systematic due to a shared etymology. This methodology also allows for cross-linguistic comparison, so we can begin to understand how sounds become iconic and how universal ideophonic constructions are or how iconic ideophonic constructions pattern across languages of the world.

2.3 Hierarchy of ideophones

Dingemanse (2012) proposes a hierarchical relationship among ideophones which exists across all languages in which ideophones are present. At the lowest level is the most common type of ideophone, those which evoke sounds. In English, these are our onomatopoeias like *honk* or *ding*. This level represents an entry point for

3

a language's ideophonic inventory. If a language has no ideophones for sounds, they will have no ideophones further up the hierarchy. An example of one such language would be Navajo (Dingemanse, 2012). The second tier of the hierarchy is for ideophones which evoke movement. These ideophones are often also associated with sound and can serve to evoke both, but the inclusion of a movement elevates them to the higher tier. An example in English would be *boing* which evokes both the sound of a spring being released and the associated movement of the release. The next tier up is for visual patterns with ideophones like *bling-bling* and *teensy-weensy*, the next for other sensory perceptions such as texture or smell such as *gloopy* or *fuzzy-wuzzy*, and lastly, the highest tier contains ideophones for inner feelings and cognitive states like *blah*, *coo-koo*, and *hunky-dory* (Dingemanse, 2012).

The nature of the hierarchy is such that inner feelings and cognitive states are the rarest type of ideophone, and any language with ideophones at the tier will have ideophones at every preceding level. The tiers in the hierarchy and their ordering most likely results from a complex combination of how our cognitive system processes sensory perceptions, how common and salient each type of perception is, other ways each perception could be communicated, and how easy each perception is to translate into an evocative lexical item (Dingemanse, 2012).

2.4   Ideophones in syntax

Dingemanse (2012) argues that research on ideophones has been too focused on form, not enough on function, but there is still some research on the syntactic role of ideophones. Nuckolls (1999) lists several studies in various languages which identify the parts of speech ideophones may occupy in the languages mentioned, most of which are intensifiers. In some African languages, there are some ideophones types which are restricted to certain sentence types (Newman, 1968).

Dingemanse and Akita (2017) put forward a theory that there is an inverse relationship between the depictive-ness of an ideophone and its potential for syntactic integration. They refer to the semiotic difference between description and depiction in which description relies on the use of arbitrary word-meaning pairs to communicate but depiction relies on a link between form and meaning, often involving physical embodiment. The authors review evidence from ten language and analyze Japanese corpus data as a case study to explain their hypothesis. It is well known from previous work that ideophones are more syntactically independent than most words, able to form complete utterances on their own and typically marked when used in larger syntactic constructions (Dingemanse, 2012; Newman, 2001). Dingemanse and Akita (2017) make the explicit connection between this lack of integration and the level of expressiveness conveyed by the ideophone. Their analysis found that the level of grammatical integration was lower for more expressive

4

ideophones.

Dingemanse and Akita (2017) define expressiveness as "the degree to which they are foregrounded as distinct from other items" (p. 505) which does not significantly differ from common definitions of markedness, particularly as it applies to ideophones (Dingemanse, 2012; Haspelmath, 2006). They then add, "for instance by special intonational or phonational features" (Dingemanse & Akita, 2017, pg 505) which specifies some of the ways in which ideophones can be marked. Expressiveness has been defined in several other ways over the course of the ideophone literature, including definitions related to affective content (Baba, 2003; Samarin, 1970), experiential semantics (Blench, 2013; Klamer, 2002), and the iconic form-meaning mappings typical of ideophones (Diffloth, 1980; Dingemanse & Akita, 2017). These senses of expressiveness all differ from the definition provided by Dingemanse and Akita (2017) which focuses on the construction rather than the function of ideophones.

2.5 Ideophones in English

A primary analysis of ideophones in English has revealed a few observations, some of which will be further explored in this paper. One feature that stands out across the types of ideophones is a pattern in their construction. Many of the ideophones of English have in common a reduplicative construction in which the second word is a minimal pair or near minimal pair to the first, either an ablaut reduplication or a rhyming reduplication. Examples of this pattern include *pitter-patter, splish-splash, zigzag, heebie-jeebies, teeny-tiny,* and *fuzzy-wuzzy*. The pattern works when the component words have one, two, or three syllables, but two syllable component words were most common. Some of the ideophones that have this construction work as independent words with only one of their component parts, like *fuzzy* or *tiny*. But others, like *helter-skelter* or *murmur* only work as a pair. Even in the instances where a component part may stand alone, usually only one of the component parts can manage it. The word *wuzzy* cannot stand alone, nor can *splish*, while *fuzzy* and *splash* are fine on their own. Though, some cases, such as *flitter-flutter* or the exact reduplications like *bling-bling* or *vroom-vroom*, either component can stand alone.

The reduplication pattern is also interesting in how it may affect meaning and use for some ideophones. For instance, the ideophone *bang* can evoke any variety of loud noises made by different kinds of sources, but reduplicated to *bang-bang*, it almost solely evokes gunfire.

Another notable commonality among the ideophones of English is that almost all of them begin with a consonant. Every place and manner of articulation accepted in English's prosaic phonology is represented among the word-initial sounds of the ideophones, but the only vowel to appear word-initial is in the ideophone *itsy-bitsy*.

It is highly likely that ideophones are used in English in the same contexts as in other languages. Ideophones are useful for being especially expressive when storytelling. But in English, perhaps in part due to the unique properties many ideophones have, ideophones are very commonly found in nursery rhymes. The name of the classic nursery rhyme character, *humpty-dumpty*, follows the English ideophone reduplicative pattern, and historically, it was actually a word used to describe a short and clumsy person (Frankis, 1991). Other examples like *higgledy-piggledy, itsy-bitsy,* and *fuzzy-wuzzy* appear in well-known children's songs and nursery rhymes. It could be interesting to look at this trend historically to determine if ideophones enter the language through nursery rhymes or if existing words are used to create new nursery rhymes. The implication from these connections is that many ideophones, particularly these reduplicative ones, sound childish or silly to a native English speaker (Nuckolls, 1999).

Another common domain of ideophone use is in the visual media of comics. Comic strips and book especially utilize ideophones for sound to create a sensory experience from a one-dimensional comic panel and to immerse the reader in the story. A brightly-colored, bubble-lettered *pow* or *bang* is iconic even outside of comic book readerships. These ideophones are usually paired with a visual component other than that of the story's illustration. They are typically placed in a shape or design which also reflects the sensory experience that the word is meant to convey, very similar to how spoken ideophones often co-occur with gestures to enhance their impact (Dingemanse, 2013).

Ideophones are also often found in song lyrics, particularly in pop music. Songs like Crash by the Dave Matthews Band or Boom Boom Pow by the Black Eyed Peas are examples, in the title alone, of ideophones being utilized in music. In both of these songs, the sound of the titular ideophone is also reproduced musically to enhance the evocation of the sensory experience.

All the ideophones discussed so far are drawn from Standard American English, but regional or ethnic dialects may have additional ideophones. I have found some potential ideophones that are specific to my regional, ethnic dialect of Cajun English. These handful of ideophones are, for the most part, expressive in terms of emotional states. Some of them are even enregistered in Cajun community, appearing in memes and in explicit performances of Cajun English in a parody video (BobbyDotComTV, n.d.). These words are not included in lists of unique Cajun English features. This is in keeping with trends mentioned in the literature of ideophones receiving little attention in language documentation processes (Childs, 2001).

Besides dialectal variation, another way it seems ideophones vary sociolinguistically in English is by age of speakers. One of the ideophones on the list, *yeet*, is a new word which originated from a video that went viral a few years ago but has remained in the lexicon of young people who use it to evoke throwing something a

distance away. Older people are unlikely to have ever heard the word, much less to use it. There are also ideophones that show the opposite trend of age-graded use. To my ears, words like *hunky-dory* and *heebie-jeebies* sound old-fashioned, and in informally consulting with other native English speakers my age, they agree with my perception. I have already covered in the previous section how many ideophones are associated with children due to their use in nursery rhymes. This pattern presents interesting opportunities for future research about the use of ideophones in context.

3   Analysis

3.1   Methods

For this analysis, I have drawn on a list of English ideophones that I have generated, drawing from various sources. Some are from previous work on ideophones that was primarily concerned with other languages but made references to ideophones in English. The majority of the list is comprised of ideophones that I generated from my experience and intuition as a native speaker. While some ideophones are very obvious due to their evocative nature, some can be difficult to discern either because they have been highly lexicalized or because ideophones above the sound tier don't have the obvious phonological relationship to the sensation they are meant to evoke.

To informally confirm the ideophonic status of these more subtle words, I rely on a few different strategies. My first strategy is to give a definition and examples of ideophones to other native speakers (some linguists, some not) to see if their intuitions confirm my own. Another strategy I use is to check to see how that word would fit in contexts where ideophones are common. For example, graphic novels and comic books often use highly stylized ideophones to evoke sensory perceptions for a two-dimensional, still drawing. A word like *twinkle* is significantly lexified, so there could be some doubt about its status, but it is very likely to be used in a comic or graphic novel to depict a visual pattern that the drawing alone cannot convey. Another common context for ideophones in English is nursery rhymes, so "Twinkle Twinkle Little Star" provides another piece of evidence for granting *twinkle* its ideophonic status.

For this analysis, I focus on the way ideophones are used in a conversational context. I am deliberately leaving out the ways ideophones may appear in nursery rhymes, song lyrics, and other non-speech contexts because, while interesting, those cases are likely outliers from the way ideophones function in regular, spoken syntactic constructions.

3.2 Syntactic integration

3.2.1 Sound

The syntactic role an ideophone can take in English depends significantly on how lexicalized the word is. Some ideophones seem easily accepted by speakers as more word-like and are open to morphological transformations which expand their potential syntactic uses. Though some ideophones higher in the hierarchy can be lexicalized, it is much more common among the sound ideophones of the first tier. However, lexicalization does not happen equally across the sound tier. For example, contrast the ideophones *pop* and *vroom*, both of which evoke sounds. It is easy and natural to attach various morphemes to *pop*, creating words like *popped*, *popping*, and so on. Because *pop* accepts these verbal morphologies, it can be used as the verb in syntactic constructions. *Vroom*, on the other hand, does not sound natural with these transformations and cannot function as a verb, giving it a more limited syntactic distribution.

Ideophones on this tier can function as nouns, but have limited distributions. The sentence *There was a [ideophone]* allows almost any ideophone on this tier to act as a noun, but very few other constructions allow it. Phrases like *the screech of the tires* also accept ideophonic nouns, but the semantic content of the arguments must match. Use as a noun is probably more common for less marked ideophones.

Sound-based ideophones can also function as adjectives, but they must be in their gerund form, such as in the phrase *a gurgling stream*. The use of the gerund form helps to evoke the continuous nature of some sounds. Continuousness can also be communicated through reduplication, which is more depictive than the gerund form but also more marked. So speakers have a choice between using the gerund morphology and syntactically integrating the ideophone or reduplicating it and limiting its syntactic position. Ideophones on this tier can also be used as adjectives when they are more overtly performed by the speaker. An example of such a performance would be when a speaker uses the word *swish* and elongates the fricatives to better reproduce the desired effect. Such a performance can be somewhat syntactically integrated, but it will be marked by pauses before and after the ideophone performance.

Common Constructions

- It made a _____ sound. (must be overtly performed or in gerund form)

- It went _____

- _____! It (adverb) [appropriate verb].

- It (adverb) _____.

8

- There was a _____.

- It was (gerund form).

- The _____ of the [source of sound]

- The [animal] says _____. (for animal sounds)

3.2.2 Movement

Ideophones on the second tier are usually fairly easy to syntactically integrate. Their association with movement makes particularly susceptible to representation as verbs. They accept tense markers easily, and sound natural when modified by an adverb. Examples at this tier include *zip*, *flutter*, and *scrunch*.

Common Constructions

- It went _____

- It _____-ed (adverb).

- It was (gerund form).

3.2.3 Visual patterns

Ideophones for visual patterns, the third tier, are fairly limited in their syntactic integration. They are not typically modified, appearing in the sentence context in their original, standard form. Most work only as adjectives or, less frequently, adverbs when used in a sentence. For example, *teeny-tiny* can be used on its own as a descriptor or can be used as an adverb to intensify an adjective with a similar meaning, i.e. *teeny-tiny little [noun]*. In the list of ideophones I have compiled so far, there are two exceptions to this rule for visual ideophones. Both *twinkle* and *glimmer* can be used as nouns (*a twinkle/glimmer in her eye*) and as intransitive verbs. Another exception at this tier is *bling-bling* which is used either as a noun or independently but clearly refers to something visual.

Common Constructions

- It was all _____.

- It looked (adverb) _____.

- The _____ [noun]

3.2.4   Sensory perceptions

On the fourth tier, ideophones for the other sensory perceptions can only be syntactically integrated as adjectives. This makes a good deal of sense for what this category of ideophones is meant to evoke. These ideophones depict perceptions that don't involve movement or sound, so it follows logically that, in English, verbs would not be an accurate representation. Nouns also don't fit sensory perceptions well because the taste, smell, or feel of something is an aspect of an entity, not an entity in and of itself. English tends to use adjectives to describe static or inherent features, so this limitation of sensory ideophones fits the patterns of English and likely differs in languages which use different lexical categories for these concepts. Examples at this tier include *fuzzy-wuzzy* and *gloopy*.

<u>Common Constructions</u>

- It was _____

- It felt/smelled/tasted _____.

- The _____ [noun]

3.2.5   Inner feelings and cognitive states

The ideophones of our final tier are restricted to use at the end of a sentence, regardless of the lexical category the word falls into. The primary categories are adjectives and nouns, which would normally be able to take other places in the sentences, but for some reason, the ideophones only sound natural when sentence-final. More complex sentence constructions could cause deviations from this pattern, though. For example, starting a sentence with a factive clause at the end of which is an ideophone would put the ideophone in the middle of the sentence. But even in such a case, the ideophone is still at least clause-final. Examples for this tier include *helter-skelter*, *meh*, and *higgledy-piggledy*.

<u>Common Constructions</u>

- I was feeling _____.

- I felt _____.

- It was all _____.

- Everything was _____.

- It gave me the heebie-jeebies.

3.3 Patterns across tiers

Looking at the big picture, we can see that some patterns exist across the tiers. Tier one and two are the only ones which allow their ideophones to function as verbs. Ideophones from tier one can work as nouns in certain constructions, and there are a few ideophones from other tiers which can function as nouns as well. Examples of these exceptions include *zigzag*, *glimmer*, and *heebie-jeebies*. Ideophones on the first tier can also work as adjectives under certain constraints, particularly in gerund form or as an overt performance. The ideophones of tier three and four can mostly be used only as adjectives. These tiers represent similar concepts, so it makes sense that they pattern together. The most common constructions for tiers three and four are practically identical. Ideophones on the fifth tier can also work as adjectives but are more limited in their syntactic distribution.

4 Discussion

The observations presented here support Dingemanse and Akita (2017) claim that there is an inverse relationship between syntactic integration and expressiveness, when expressiveness is defined similarly to markedness. *Pop* and *vroom* are equally depictive of their respective sounds. So why then is *pop* easily lexified and syntactically integrated when *vroom* is not? The former fits easily into the phonological rules of English, while the latter does not. The *vr-* consonant cluster is marked in English, so it makes sense then that it does not easily take on more word-like qualities. It is limited to ideophonic status, while relatively unmarked ideophones can flow freely between uses as ideophones and as more prosaic words. This theory is further supported by ideophones in other tiers. For example, ideophones like *bling-bling*, *fuzzy-wuzzy*, and *hunky-dory* that have a reduplicative construction are limited in their syntactic distribution and are significantly marked. This provides evidence that English ideophones behave consistently with ideophones in other languages.

There seems to be a difference between the behavior of morphologically marked ideophones and phonologically marked ideophones. A phonologically marked word like *vroom* is difficult to integrate into a sentence at all. It is almost always surrounded by pauses to separate it from the rest of the utterance. A morphologically marked word like *fuzzy-wuzzy*, while limited in its syntactic distribution, can still be integrated into a sentence without these pauses. Note the differences between the simple sentences *It went vroom* and *It was fuzzy-wuzzy*. My theory is that phonologically marked ideophones require a certain amount of overt performance from the speaker, beyond their inherent phonological features. When we use ideophones, especially those which are phonologically marked, it is more effective

to overtly perform them so that they more closely evoke the desired effect. This overt performance can be composed of features like lengthening, reduplication, intonation, volume, phonational features, gestures, facial expressions, and other forms of physical embodiment. For example, speaker rarely says *boing* flatly but rather lengthens the vowel and coda and uses intonation on the final sound to better mimic the sound of a spring being released. This pronunciation is not required or assumed for any given utterance of *boing* and may vary greatly from speaker to speaker and across contexts. Dingemanse and Akita (2017) reference the performative nature of expressive features but seem to assume a standard performance for each ideophone and fail to consider how the use of expressive features may vary across ideophone types, type of markedness, and across utterances. The overt performance used in the pronunciation of ideophones contributes to their markedness and, thereby, affects their potential for syntactic integration. In order to complete our understanding of the syntactic uses of ideophones, future work should examine ideophones as they are produced in context, exploring the nuance of pronunciation.

4.1 Questions for the existing model

In my preliminary work on ideophones in English, I found a handful of words that have strong ideophonic properties but don't fall neatly into any of the existing tiers of ideophones. *Zigzag* is ambiguous, fitting well into both the movement and visual pattern categories. It can describe a movement in which something goes back and forth on a diagonal, but it can also evoke a visual pattern of diagonally bending lines, like chevron stripes. We can see that *zigzag* works as a verb in a sentence like *He was zigzagging all over the place*. It also works in some visual constructions, such as *the zigzag stripes* or *It looked all zigzaggy*. There are several other examples of this potential dual-category membership, most of which also sound natural in common syntactic constructions for both of their possible tiers. It may not be possible to categorize ambiguous tokens without collecting authentic production data from speakers. And it is possible that some ideophones may be able to represent multiple sensory modalities, either in separate contexts or at the same time within one utterance.

Another ambiguous is *hoity-toity*. There is a strong case for ideophone status based on the construction of the word alone, and it seems to lend itself to co-production with physical embodiment which matches the behavior of other ideophones. But where would it belong in the hierarchy of ideophones? The rhyming reduplicative pattern can fit any tier but is more common in the higher tiers. It fits naturally into a few syntactic constructions such as *He was acting very hoity-toity* and *the hoity-toity new professor*. In these sentences we can see it functioning as an adjective, both in phrase-final and phrase-medial position. Its acceptability

phrase-medial makes it unlikely to be an ideophone for an inner feeling or cognitive state. By its definition, *hoity-toity* refers to a superior attitude or pompousness, yet it matches best with ideophones for visual patterns or other sensory perceptions. One could make the argument that we recognize a hoity-toity attitude visually and categorize it with tier three, but this argument feels weak. This example, among others I have collected (e.g. *ta-daa*, *hoopla*, *hocus pocus*, *flim-flam* etc.), calls into question the completeness of Dingemanse (2012) hierarchy of ideophones. Further data collection and analysis is needed to determine if these words are not, in fact, ideophonic or if the existing model needs to be expanded.

4.2   Future directions and limitations

As ideophones have been almost entirely ignored in English so far, there is a great deal of work that can be done. The analysis presented here is very limited, coming primarily from one native speaker and only aiming to make general observations from which a more full analysis of English ideophones can grow. A larger data set and more empirical methods should be used to review the observations noted in this paper. Native speaker impressions can be valuable in making assessments about the meanings and classifications of ideophones (Dingemanse, 2012), but a more robust analysis would be a better foundation for future research.

Work on the use of ideophones and their role in the context of English could also yield interesting results. The literature in other languages presents a strong case for the study of ideophones as used in narrative descriptions and storytelling, but the study of ideophones in visual and written media, children's language, and music could also be productive, as these different contexts may utilize ideophones in their syntactic roles differently.

More work is needed to understand the process of ideophone lexification. The acceptability of each morpheme in an ideophone will have an impact, but it seems that the whole is greater than the sum of its parts. It is also possible that not every instance of a lexicalized ideophone is actually an ideophone. A primary qualification of ideophones is that they are depictive. If an ideophone becomes highly lexicalized and syntactically integrated, is it still serving that depictive, evocative function? Further research should collect native speaker judgements on the depictiveness of lexicalized ideophones to see whether or not they are still fulfilling that goal. Research using cognitive and neuroscience methods could also further our understanding of how ideophones are processed and categorized in the brain, particularly in comparison to the prosaic words of a language. Another crucial step forward is to analyze actual productions of ideophones by speakers so we can better understand the details of their pronunciation and contextual uses.

The observations and questions presented here are the beginning of the study of

ideophones in English and in any language in which ideophones exist more sparsely. For too long, research has focused on the set of languages in which ideophones are strongly present. This approach cannot lead to a comprehensive understanding of ideophones. Several examples presented here call into question existing ideophone theories. As the study of ideophones expands to include more languages and shifts its focus to include the function of ideophones, we grow closer to a complete understanding of depictive language.

References

Akita, K. (2009). *A grammar of sound-symbolic words in japanese: Theoretical approaches to iconic and lexical properties of japanese mimetics* (Doctoral dissertation). Kobe University. Kobe.

Akita, K., & Dingemanse, M. (2019). Oxford research encyclopedia of lingusitics. In M. Aronoff (Ed.). Oxford University Press.

Baba, J. (2003). Pragmatic function of japanese mimetics in the spoken discourse of varying emotive intensity levels. *Journal of Pragmatics*, *35*, 1861–1889.

Blench, R. (2013). Chadic linguistics. In H. Tourneux (Ed.). Köln: Rüdiger Köppe.

BobbyDotComTV, C. M. (n.d.). Pie-yow comme ca (parody commercial).

Childs, G. T. (1994). Sound symbolism. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.). Cambridge University Press (CUP).

Childs, G. T. (2001). Ideophones. In F. K. E. Voeltz & C. Kilian-Hatz (Eds.). John Benjamins Publishing Company.

Childs, G. T. (2014). Constraints on violating constraints: How languages reconcile the twin dicta of "be different" and "be recognizably language.". *Pragmatics and Society*, *5*(3), 341–354.

Clark, H., & Gerrig, R. (1990). Quotations as demonstrations. *Language*, *66*(4), 764–805.

Diffloth, G. (1980). Studies in mon-khmer and thai phonology and phonetics in honor of e. henderson. In T. L. Thongkum (Ed.). Chulalongkorn University Press.

Dingemanse, M. (2012). Advances in the cross-lingusitic study of ideophones. *Language and Linguistics Compass*, *10*(6), 654–672.

Dingemanse, M. (2013). Ideophones and gesture in everyday speech. *Gesture*, *13*(2), 143–165. https://doi.org/10.1075/gest.13.2.02din

Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. *Glossa: a journal of general linguistics*, *3*(1), 1–30.

Dingemanse, M., & Akita, K. (2017). An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideo-

phones, with special reference to japanese. *Journal of Linguistics*, *53*(3), 501–532. https://doi.org/10.1017/s002222671600030x

Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & Mitterer, H. (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language*, *92*(2), 117–133.

Frankis, J. (1991). Language use and description. In I. T.-B. van Ostade & J. Frankis (Eds.).

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, *42*(1), 25–70.

Howard, A. (2019). Language is arbitrary? i wouldn't be saussure: Multimodal ideophone guessability. *Radboud University Student Linguistics Journal*, *1*, 23–31.

Klamer, M. (2002). Semantically motivated lexical patterns: A study of dutch and kambera expressives. *Language*, *78*(2), 258–286.

Newman, P. (1968). Ideophones from a syntactic point of view. *Journal of West African Languages*, *2*, 107–117.

Newman, P. (1989). Reduplication and tone in hausa ideophones. *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society*, 248–255.

Newman, P. (2001). Ideophones. In F. K. E. Voeltz & C. Kilian-Hatz (Eds.). John Benjamins Publishing Company.

Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology*, *28*, 225–252.

Nuckolls, J. B. (2004). To be or not to be ideophonically impoverished. *Proceedings of the Eleventh Annual Symposium about Language and Society— Austin*, *47*, 131–142.

Samarin, W. (1970). Inventory and choice in expressive language. *Word*, *26*, 153–169.

Thompson, A. L., & Do, Y. (2019). Defining iconicity: An articulation-based methodology for explaining the phonological structure of ideophones. *Glossa: a journal of general linguistics*, *4*(1), 1–40.

Competing verbal constructions with functional TAKE in Bulgarian

Vincenzo Nicolò Di Caro   Luca Molinari
Ca' Foscari University of Venice   University of Warsaw
vincenzo.dicaro@unive.it   l.molinari@uw.edu.pl

Abstract

This paper discusses the preliminary results of an online acceptability judgments questionnaire on some Bulgarian verbal periphrases featuring the functional verb *vzemam* 'take'. The study was conceived to fill a gap in the literature about functional TAKE in Bulgarian, which has been known in previous work since Sandfeld (1900) but is still rather scarce and unsystematic. Three TAKE+V2 constructions are identified: i) a Multiple Agreement Construction featuring the connector *da* (TAKE *da*MAC); ii) a MAC featuring *če* (TAKE *če*MAC); iii) Pseudo-Coordination (of the type TAKE + *i* 'and' + V2). The participants are 157 native speakers (112 F, 45 M) with an age range of 18-80 (M = 43.63; SD = 13.92). The results of the questionnaire confirm the presence and the productivity of these constructions with functional TAKE in present-day Bulgarian. Moreover, they show that these constructions all share a monoclausal structure, but with some structural differences: V1 in the TAKE MACs is mainly restricted to the past tense, and V2 only occurs in the present, while in the *i*PseCo V1 and V2 share TAM features and can appear both in the present and in the past. From a semantic point of view, TAKE *da*MAC specializes for inchoativity, while TAKE *če*MAC for mirativity. The *i*PseCo seems to be able to convey both meanings, but it is least preferred than the TAKE MACs.

1 Introduction[1]

The verb TAKE is found in a number of different periphrastic constructions where, as a functional verb, it can serve different purposes, cross-linguistically (see Ross 2017 for an overview). For example, it can appear in 'Serial Verb Constructions' (henceforth, SVCs; cf. Aikhenvald 2006, 2018), which are sequences of multiple verbs forming a single predicate with a monoclausal structure, generally without any marking of syntactic dependency such as coordination or subordination. In SVCs, which are found in West Africa (cf. (1a)), East Asia, Amazonia (cf. (1b)), Oceania, creoles and other languages, each component can occur on its own. The verbs involved share grammatical categories including tense, aspect, mood, modality, and also a prosodic contour.

(1) a. **Mede** aburow migu msum.
1SG.take corn 1SG.flow water-in

'I pour corn into water.' [*Akan*; Aikhenvald (2006: 40)]

b. Mawina-nuku wasã **wheta** wa-hnaã.
pineapple.TOP let's 1PL.take 1PL.eat

'Let's take and eat the pineapple!' [*Tariana*; adapted from Aikhenvald (2006: 183)]

TAKE in SVCs generally follows three grammaticalization paths (i.e. aspectual, valency-increasing, and pragmatic meaning), but it can also express other meanings. According to Lefebvre (1991: 55), this verb in SVCs implies causation because "the subject of the verb is an Agent performing an action which causes the Theme to undergo a change of location".

In Polish, for example, *wziąć* 'take' can display both a perfective and an inchoative meaning (Andrason 2018: 607-9), but it can also grammaticalize to express pragmatic meanings. In all Finno-Baltic languages, it intensifies another verb (Pulkkinen 1966: 212–3). In Estonian (Tragel 2017: 177), for example, *võtma* 'take' (which is however not very frequent in SVCs) lacks a syntactic object and carries intentional meaning to the following V2.

SVCs can be considered as belonging to a macro-category referred to as 'Multiple Agreement Constructions' (henceforth, MACs; Giusti, Di Caro and Ross 2022), since the two verbs involved share TAM features. MACs is the term we will use to refer to the relevant Bulgarian constructions under analysis.

Another construction in which TAKE occurs as V1 is referred to as 'Pseudo-Coordination' (henceforth, PseCo), since it formally appears as a coordination but syntactically behaves as a monoclausal construction (Giusti, Di Caro and Ross 2022)[2]. PseCo is very common in the Germanic languages, where TAKE can appear together with other V1s such as GO, SIT, STAND and LIE. In these constructions in e.g. Swedish and Norwegian, TAKE can express an inchoative (cf. (2)) or a mirative meaning (cf. (3)).[3]

(2) a. Han **tok** og skrev et dikt.
he take.PST and write.PST a poem
'He wrote a poem.' [Norwegian; Lødrup (2002: 121)]

b. Han **tog** o läste en bok.
he take.PST and read.PST a book
'He started reading a book.' [Swedish; adapted from Wiklund (2007: 118)]

(3) Hun **tok** og kysset ham.
she Take.PST and kiss.PST him
'She (suddenly) kissed him.' [Norwegian; Lødrup (2017: 278)]

The Romance varieties also display instances of PseCo featuring TAKE as V1 (Coseriu, 1966). These have been the object of recent interest (see Masini et al. 2019; Giusti and Cardinaletti 2022 for Italian and some Southern Italo-Romance varieties; Soto Gómez 2021 for Spanish; Mendes and Ruda 2022 for Portuguese; and Bleotu 2022 for Romanian). Two different functions can be identified for TAKE in these constructions: it can either serve an inchoative (cf. (4)) or a mirative function conveying a sense of unexpectedness (cf. (5)). This seems to hold true cross-linguistically (as shown in (2) and (3) for Germanic).

---

[2]It has been discussed in the literature whether PseCo can be considered as an instance of SVC. For references cf. e.g., Déchaine (1993); Cardinaletti and Giusti (2001); Manzini and Savoia (2005); Manzini, Lorusso and Savoia (2017); Cruschina (2013); Del Prete and Todaro (2020); Giusti, Di Caro and Ross (2022).

[3]For the mirative use of PseCo in Scandinavian see, *inter alia*, Wiklund (2008; 2009) and Josefsson (2014).

(4) a. Los viernes después de entrenar siempre **cogemos**   y
        the Fridays after   of train.INF always take.PRS.1PL and
        pedimos   chino.
        order.PRS.1PL Chinese
        'On Fridays, after training, we always take and order Chinese food.'
        [Spanish; Soto Gómez (2021: 47)]

     b. Alle   cinque ha **preso** e   ha cominciato a piovere.
        At-the five   has taken and has started   to rain.INF
        'All of a sudden, it started raining at five.' [Italian; adapted from Giusti
        and Cardinaletti (2022: 48)]

(5) a. **Tomó**   y   se   fué.
        take.PST.3SG and REFL go.PST.3SG
        'He (took and) left!' [Spanish; Coseriu (1966)]

     b. Ha **preso** ed   è partita.
        has taken and is left
        'She (took and) left!' [Italian; Giusti and Cardinaletti (2022: 47)]

In Bulgarian, functional TAKE is involved in a number of verbal periphrases
that have been covered in the literature in a rather unsystematic way. Moreover, the
available literature is not always up-to-date. The study we propose here aims at
filling these gaps by discussing the results of a preliminary quantitative study based
on an online acceptability judgments questionnaire administered to Bulgarian native
speakers.

The rest of the paper is organized as follows: Section 2 provides an overview of
all the MACs in Bulgarian and then focuses on those featuring V1 TAKE; Section 3
presents the study and describes the design of the online questionnaire; Section 4
discusses the data collected, draws the conclusions, and proposes some avenues for
future research.[4]

2   The phenomenon: competing constructions in Bulgarian

Before turning to the constructions with functional TAKE in Bulgarian, let us have a
brief overview of the main periphrastic constructions found in this Southern Balkan

---

[4]Although this paper is the result of joint work by the two authors, for the sake of the Italian
Academy Vincenzo Nicolò Di Caro is responsible for Sections 1 and 3, while Luca Molinari is
responsible for Sections 2 and 4.

Slavic language.

2.1 The canonical *da*MACs

Being part of the Balkan *Sprachbund*, Bulgarian displays one of the typical traits of this linguistic group, namely the lack of the infinitive, which is taken over by subjunctive constructions with tensed verbs (Tomić 2006: 456). Interestingly, tensed V2s are also found in Southern Italo-Romance MACs (cf. Southern Calabrian (6a) and North-Eastern Sicilian (6b) featuring inflected V2s with the Italian infinitival counterparts in (6a') and (6b')), which share said *Sprachbund* feature because of some contact effects with Greek, although displaying the indicative instead of the subjunctive mood.[5]

(6)   a. Vuliti   u viniti   â   me casa?
           want.PRS.2PL u come.PRS.2PL at-the my house

           'Do you want to come to my place?' [Southern Calabrian; adapted
           from De Angelis (2017: 138)]

       a'. Volete   venire   a casa mia?
            want.PRS.2PL come.INF at house my

            'Do you want to come to my place?' [Italian]

        b. Ncuminciau   mi parra   accussì.
            start.PST.3SG mi speak.PRS.3SG so

            'He started to speak this way.' [North-Eastern Sicilian; adapted from
            Ganfi (2021: 10)]

       b'. Cominciò   a parlare   così.
            start.PST.3SG to speak.INF so

            'He started to speak this way.' [Italian]

The subjunctive constructions in Bulgarian are introduced by *da*, which is a polyvalent item in that it carries out several different functions.

Da serves as a grammatical particle for the formation of periphrastic tenses (cf. (7)). Moreover, it can be a modal particle with different shades of meaning:[6]

---

[5] For the 'unpopularity of the infinitive' in Southern Italo-Romance, see Rohlfs (1969: §717). See also Ledgeway (2013) for a discussion on the Greek interference exerted over Southern Italo-Romance varieties.

[6] Hansen, Letuchiy, & Błaszczyk (2016) (*apud* Nicolova 2008) treat *da*-forms as a particular mood with various *irrealis* uses and keep it separated from indicative mood forms.

(i) exhortation/request/order (cf. (8a)-(8b)), (ii) desirability (cf. (8c)), and (iii) conditionality (cf. (8d)) (the examples in (7)-(8) are transliterated in Latin script and adapted from Simov and Kolkovska (2004: ex.1ff.)).

(7) Toj šteše / njama da dojde utre.
he will.PST.3SG will.NEG *da* come.PERF.PRS.3SG tomorrow
'He would have come / won't come tomorrow.'

(8) a. Ti da mălčiš!
you *da* shut-up.IMPF.PRS.2SG
'Shut up!'

b. Da ne si posmjal!
da NEG be.IMPF.PRS.2SG dare.PERF.PST.PRT.ACT.M.SG
'Don't you dare!'

c. Da bjax došla togava.
da be.PST.1SG come.PERF.PST.PRT.ACT.F.SG back-then
'If only I had come back then.'

d. Da znaex, bix mu se
da know.IMPF.IMPERF.1SG would.1SG to-him REFL
obadil.
call.PERF.PST.PRT.ACT.M.SG
'If I had known, I would have phoned him.'

As anticipated above, in MACs *da* is found between V1 and V2 as a connecting element, hence these constructions will be referred to as (canonical) *da*MACs. *Da*-clauses are found as complements of intentional verbs, which include (i) volitives such as *iskam* 'want/wish' (cf. (9a)), (ii) modals such as *umeja* 'be able/can' (cf. (9b)), (iii) causatives such as *zapoviadam* 'order' (cf. (9c)), (iv) inchoatives such as *započvam* 'begin' (cf. (9d)), and (v) intentional verbs such as *planiram* 'plan' (cf. (9e)).

(9) a. Iskam da (mu) pročeta pismoto.
wish.IMPF.PRS.1SG da to-him read.PERF.PRS.1SG letter-the
'I want to read the letter (to him).' [adapted from Tomić (2006: 460)]

b. Ne umee da čete.
NEG can/be-able.3SG da read.IMPF.PRS.3SG
'(S)he cannot read.' [adapted from Tomić (2006: 464)]

c. Zapovjadax   da dojdeš   vednaga.
order.PERF.AOR.1SG da come.PERF.PRS.2SG immediately

'I gave an order that you should come immediately.' [adapted from Tomić (2006: 465)]

d. Započvam   da piša.
start.IMPF.PRS.1SG da write.IMPF.PRS.1SG

'I am starting to write.'

e. Ana planira   da otide   v Amsterdam.
Ana plan.IMPF.PRS.2SG da go.PERF.PRS.3SG in Amsterdam

'Ana is planning to go to Amsterdam.' [adapted from Tomić (2006: 466)]

The *da*MACs do not all have the same properties. Krapova and Cinque (2018) classify the subjunctive constructions featuring *da* in three different categories: (i) non-restructuring infinitive-like constructions (cf. (10)), (ii) Romance type subjunctive constructions (cf. (11)), and (iii) restructuring infinitive-like constructions (cf. (12)).

(10)   Očakvam   <ot vsički> da dojdat / da   sa
refuse.PERF.AOR.1SG to   all *da*   come.PERF.PRS.3PL *da*
pristignali do   6 časa.
are.3PL   arrive.PERF.PST.PRT.ACT.PL by 6   o'clock

'I expect that everybody comes/I expect that everybody has arrived by 6 o'clock' [adapted from Krapova and Cinque (2018: 164)]

(11)   Včera   očakvax   [ti   da si
yesterday expect.PERF.AOR.1SG you *da* are.2SG
rešil   zadačite   do utre],   no
solve.PERF.PST.PRT.ACT.M.SG math-homeworks-the by tomorrow but
sega viždam,   če šte ti   trjabva   cjala
now see.IMPF.PRES.1SG that will to-you need.IMPF.PRES.3SG whole
sedmitsa.
week

'Yesterday I expected that you would do your math homework by tomorrow but now I see that you will need an entire week.' [adapted from Krapova and Cinque (2018: 166)]

(12)   Kosta znae   /   započva sega da
Kosta know/ start.IMPF.PRES.3SG now   da   drive.IMPF.PRES.3SG

šofira.   (*utre).
tomorrow
'Now Kosta knows how/begins to drive (*tomorrow).' [adapted from
Krapova and Cinque (2018: 160)]

Krapova and Cinque (2018) argue that these subtypes of *da*MAC display different properties, first of which is the biclausal nature of the former two constructions (10)-(11) vs. the monoclausal nature of the latter one (12). This syntactic difference derives the fact that monoclausal *da*MACs display strict referential identity between the subject of V1 (which has functional nature) and the subject of V2, in that there is only one subject (cf. the impossibility of having a different subject of the V2 in (13) vs. the possibility of disjoint reference as in (10)-(11)).

(13)   Ivan znae   *(Marija) da pluva.
       Ivan know.IMPF.PRES.3SG Marija   da swim.IMPF.PRES.3SG
       'Ivan can swim (*Maria).' [adapted from Krapova and Cinque (2018: 161)]

Furthermore, while in the biclausal *da*MACs the tense of V1 is independent from the tense of V2 (cf. again (10)-(11)), the monoclausal *da*MACs are defective with respect to the tense of V2, which displays present imperfective (as in (12)-(13)).

2.2   The canonical *če*MACs

Bulgarian also displays the complementizer *če* 'that', which derives from the Indo-European pronoun for the neuter gender (Tomić 2006: 458). In contrast to *da*, *če* introduces indicative subordinates describing real events (cf. (14)), thus presumably encoding *realis* mood (Hansen, Letuchiy and Błaszczyk 2016). We will refer to these constructions as (canonical) *če*MACs.

(14)   Interesno e, *če* tuk   e zapazen edinstveni-jat original
       interesting is *če* here is stored   sole-the   original
       'It's interesting that the only original is stored here.' [adapted from Hansen,
       Letuchiy and Błaszczyk (2016: ex. 132)]

Apart from introducing indicative complements, *če* can occur in adverbial clauses of reason (cf. (15a)) and of result (cf. (15b)). Moreover, it can be used (i) as an adversative conjunction (cf. (16a)), (ii) as a cumulative conjunction (cf. (16b)), (iii) as an element forming independent conjunctions (cf. (16c)). Another noteworthy use is in exclamatory sentences with a modifying function (cf. (17)) (all the examples in (15)-(17) are adapted from Tomić 2006: 458-9).

(15) a. Trăgni   sega, če šte   stane   kăsno!
          depart.PERF.IMP.2SG now   če will become.PERF.PRS.3SG late
          'Go now, because it will be late (if you stay any longer).'

      b. Kupixme   ošte   edin televizor, taka če sega imame tri.
          buy.PERF. AOR.1PL more one   TV so če   now   have.1PL

          three
          'We bought another TV, so that now we have three.'

(16) a. Če, kakvo gi   dărži?!'
          če what   them hold.IMPF.PRS.3SG
          'But, what is keeping them?!'

      b. …no mu   natătruzixa   ošte edin če posle ošte edin
          but   to-him force.PERF.AOR.1PL more one če after more one
          '...but they forced upon him one more, and after that one more...'

      c. kato če li… / makar če…
          as   če Q   / even   če
          'As if…' / 'Although…'

(17) Ama,   če lošo čoveče!
      Ah.EXCL če bad man.DIMIN
      'What a bad little man!'

The canonical *če*MACs introducing an indicative subordinate clause pattern along with the biclausal *da*MACs presented in Section 2.1 as both constructions involve the presence of two distinct clauses. In fact, in the canonical *če*MACs V1 and V2 can have disjoint tense, aspect, and reference (18). Moreover, V2 is independent from V1, and its tense is not defective (19).

(18) a. Nadjavam   se,   če Petăr e
          hope.IMPF.PRS.1SG REFL *če* Petăr is
          zaminal.
          leave.PERF.PST.PRT.ACT.M.SG
          'I hope that Petăr has left.' [adapted from Tomić (2006: 467)]

      b. Radvam   se,   če se   vidjaxme.
          Be-glad.IMPF.PRS.1SG REFL *če* REFL see.PERF.AOR.2PL
          'I am glad that we have met.' [adapted from Hansen, Letuchiy and Błaszczyk (2016: ex. 133)]

c. Ne čuvaš   li, če se   čuvstvam   po
NEG understand.IMPF.PRS.2SG Q *če* REFL feel.IMPF.PRS.1SG in

săštija   način?
same-the way

'Don't you understand that I feel in the same way?' [adapted from
Hansen, Letuchiy and Błaszczyk (2016: ex. 131)]

(19) Petăr smiata,   če Ivan šte kupi /
Petăr think.IMPF.PRS.3SG *če* Ivan will-buy.PERF.3SG

šte kupuva /   e kupil   kăštata.
will-buy.IMPF.3SG is buy.PERF.PST.PRT.ACT.M.SG house-the

'Peter thinks that Ivan will buy/will be buying / has bought the house.'
[adapted from Krapova (2021: 220)]

The verbs which can select a *če*-complement are divided by Krapova (2021: 220)
in four main classes and summarized as follows: (i) propositional attitude/epistemic
verbs (e.g., *mislja* 'think', *smjatam* 'consider'), (ii) verbs of communication (such
as *kazvam* 'say', *tvărdja* 'claim'), (iii) verbs of intellection/cognitive predicates
(e.g., *znam* 'know', *razbiram* 'understand'), and (iv) emotive predicates (such as
săžaljavam 'regret', *radvam se* 'be glad').

2.3 Constructions with functional TAKE

Structures with functional TAKE have been well documented for a great number
of different languages (see Section 1). As for Bulgarian, however, the available
literature is rather scarce and quite unsystematic. A few examples of structures
with functional *vzemam* 'take' in Bulgarian are mentioned in Coseriu (1966) and
Kanchev (2010). The latter author distinguishes two types of constructions with
functional TAKE, giving the two examples reported here in (20).

(20) a. Vze   da piše.
take.PERF.AOR.3SG *da* write.IMPF.PRS.3SG

'He started writing.' [adapted from Kanchev (2010: 41)]

b. Vze   če napisa.
take.PERF.AOR.3SG *če* write.PERF.AOR.3SG

'He unexpectedly wrote.' [adapted from Kanchev (2010: 42)]

Kanchev (2010) himself individuates a semantic difference between the two
sentences, claiming that the construction in (20a) (which we will refer to as TAKE

*da*MAC following Giusti and Cardinaletti, 2022) has inchoative semantics, while the construction in (20b) (henceforth TAKE *če*MAC) expresses surprise and unexpectedness. He further mentions that the TAKE *da*MAC requires an imperfective V2, while the TAKE *če*MAC only allows a perfective V2. However, no further description is provided.

Interestingly, it seems that these TAKE MACs are not the only constructions with functional TAKE in present-day Bulgarian. In a web search we conducted before designing our study, we also found instances of constructions with functional TAKE that look like a PseCo in that V1 and V2 share TAM features and are linked by the conjunction *i* 'and'. For this reason, we will refer to them as *i*PseCo. Some examples are reported in (21).

(21) a. Vmesto da prekara   njakoj i   drug čas v bara,
         instead *da* spend.PERF.PRES.3SG some and other hour in bar-the
         tja vze   i   trăgna   s   men
         she take.PERF.AOR.3SG and go-away.PERF.AOR.3SG with me
         kato opaška.
         as   tail
         'Instead of spending another hour or so at the bar, she took off with me like a tail.' [adapted from Marinov (2010: 112)]

      b. Vzemam   i   trăgvam,   tolkova e lesno!
         take.PRES.1SG and go-away.PRES.1SG so-much is easy
         'I'll take and go, it's so easy! (https://www.facebook.com/watch/?v=2525666294318221)

      c. Eto   kakvo ti   predstoi. Vzemaj   i
         here's what to-you awaits   take.IMPF.IMPER.2SG and
         otstăpvaj!
         step.IMPF.IMPER.2SG
         'This is what awaits you. Take and start!' [SketchEngine, "Bulgarian Web 2012", token 116276468]

As is clear from the translation of the sentences in (21), in the *i*PseCo the verb TAKE is devoid of lexical meaning (as is the case of PseCos cross-linguistically, e.g., in Italian). At a first glance, the construction appears to be either mirative (e.g., (21a)) or exhortative (e.g., (21c)).

The brief overview of the constructions with functional TAKE in Bulgarian proposed here calls for a solid empirical base to support the scarce data found in the literature. Only in this way will it be possible to provide a systematic description

of these verbal periphrases, allowing us to compare them both to the equivalent constructions in other languages and to the canonical MACs (see Section 2.1 and 2.2). More so, to the best of our knowledge, the instances of *i*PseCo we found in our web search have never been discussed in the literature, so they need to be brought to light. To start filling this gap in the literature, we designed a pilot quantitative study for the collection of a solid base of data about the three constructions just presented. This study is to be understood as the first piece of research of this effort to study the syntactic and semantic properties of the constructions with functional TAKE in Bulgarian.

3   The preliminary quantitative study

We checked the acceptability of the constructions described in Section 2 by means of an anonymous online questionnaire. In fact, it was not possible to control for all the available feature combinations regarding the two verbs involved in such a rich verbal system like that of Bulgarian, where verbal morphology encodes tense, mood, and aspect. Moreover, considered the exploratory nature of the study, we wanted the participants to be able to complete the questionnaire in no more than 15 minutes in order to prevent too many of them from abandoning the completion.

For this reason, we limited the V2s tested to the following verbs: GO (18 items), APOLOGIZE (9), LOOK (3), SPEND (3), STAY (3) and THROW (3). In one case, the V2 GO is followed by a third verb, i.e. BUY (3 items) (cf. (22)).

(22)   Kogato e gladna, vzema   i   otiva   da si
       when   is hungry take.IMPF.PRS.3SG and go.IMPF.PRS.3SG *da* REFL
       porǎčva   pica.
       buy.IMPF.PRS.3SG pizza

The imperfective aspect of the constructions was tested in 26 items while the perfective one in 13 items. As for the persons of the paradigm, we focused on 1SG (12 items) and 3SG (15) and we limited the other persons to 3 items each. Finally, as for the tenses, we tested the distinction between present and past. As the latter comes in different types in Bulgarian (aorist, perfect, imperfect, anterior past etc.), we only focused on aorist. Present was tested in 24 items, while aorist in 15.

3.1   The questionnaire

The selection of the relevant syntactic features for the questionnaire was preceded by some previous qualitative research based on interviews to Bulgarian native speakers,

which allowed us to rule out those feature combinations that were less likely to occur and thus less worth exploring. Then, we administered the questionnaire to 157 participants.

The questionnaire contains:

(23)　i　39 items consisting of sentences that feature TAKE *če*MAC, TAKE *da*MAC and *i*PseCo described in Section 2.3, to be judged through a 5-point scale (1 = totally unacceptable, 5 = totally acceptable);

　　　ii　3 items that provide the participants with a context and ask them which construction better describes the situation provided.

The average (un)acceptability of these constructions is expressed in terms of percentages (cf. Figures 2-4) obtained by summing the judgments ranging 4-5 (indicating acceptability) separately from those ranging 1-3 (indicating unacceptability). This sum was repeated for each sentence. The mean of all the resulting sums was calculated for each category of sentences (e.g., all the sentences displaying a TAKE *da*MAC in the present tense) to obtain an average (un)acceptability rate. Some examples of the items in the questionnaire (here transliterated in Latin script) presented above in (i) and (ii) are provided in (24) and (25), respectively:

(24)　a.　Sega vzemaš　　　　če ì　　se　　izvinjavaš!
　　　　　now take.IMPF.PRS.2SG *če* to-her REFL apologize.IMPF.PRS.2SG
　　　　　Intended: 'You've got to apologize to her now!' (*če*MAC)

　　　b.　Sega vzemaš　　　　da ì　　se　　izvinjavaš!
　　　　　now take.IMPF.PRS.2SG *da* to-her REFL apologize.IMPF.PRS.2SG
　　　　　Intended: 'You've got to go and apologize to her now!' (*da*MAC)

　　　c.　Sega vzemaš　　　　i　ì　　se
　　　　　now take.IMPF.PRS.2SG and to-her REFL
　　　　　izvinjavaš!
　　　　　apologize.IMPF.PRS.2SG
　　　　　Intended: 'You've got to apologize to her now!' (iPseCo)

(25)　*Včera Ivan beše v dobro nastroenie. Izvednăž započna da plače.*
　　　Yesterday Ivan was in a good mood. Suddenly he started crying.

　　　a.　Ivan vze,　　　　　　če se　　razplaka.
　　　　　Ivan take.PERF.AOR.3SG *če* REFL cry.PERF.AOR.3SG

　　　b.　Ivan vze　　　　　　da plače.
　　　　　Ivan take.PERF.AOR.3SG *da* cry.IMPF.PRS.3SG

c. Ivan vze i se razplaka.
Ivan take.PERF.AOR.3SG and REFL cry.PERF.AOR.3SG
Intended: 'Ivan went and cried.'

We can now have a look at a description of the sample and the data collected.

3.1.1 The sample

Figure 1 shows the distribution of the sample by age. In the sample, which is within an age range of 18 to 75 (M = 43.63; SD = 13.92), there is a greater concentration of participants aged between 30 and 60.
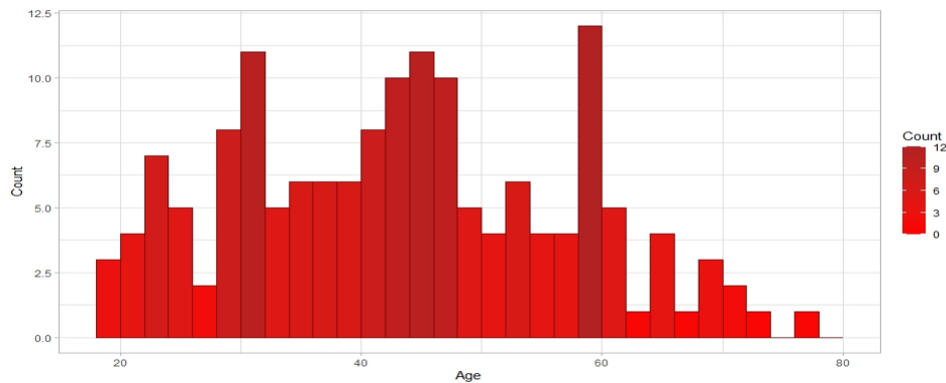

Figure 1: Distribution of the sample by age.

As regards the gender, the sample is unbalanced, with 112 female and only 45 male participants. Finally, as regards the provenance of the sample, 68 participants were from big cities (i.e., from cities with a population greater than 300,000, such as Sofia, Plovdiv and Varna), 37 from medium towns (50,000 < pop. < 300,000), and 52 from small towns (pop. < 50,000).

3.2 The data

The data were collected from August to November 2021. Table 1 shows, for each construction, the percentage of items that have been judged with a 4 or a 5 and have been thus considered as acceptable. Following this criterion, TAKE *da*MAC is the less acceptable construction, with only 24% of 4 or 5.

Construction	% of acceptability
*če*MAC	35%
*da*MAC	24%
*i*PseCo	34.1%

Table 1: Percentage of acceptability for each construction.

As regards the tenses of the verbs involved, we considered only the present and the past indicative, as shown in Table 2.

Construction	Present	Past
*če*MAC	19.7%	65.6%
*da*MAC	15.9%	40.3%
*i*PseCo	34%	34.4%

Table 2: Percentage of acceptability of the constructions according to the tense.

Figure 2 summarizes the results shown in Table 2.

As regards the action type, we divided the items between habitual and non-habitual, with the results shown in Table 3.

Construction	Habitual	Non-habitual
*če*MAC	20.7%	49.4%
*da*MAC	16%	32.1%
*i*PseCo	36.8%	31.4%

Table 3: Percentage of acceptability of the constructions according to the action type.

Figure 3 summarizes the results shown in Table 3.

Finally, as regards the semantic specialization of the constructions, two types are identified: mirative and inchoative. We have further divided the mirative specialization into disapproval and surprise, with the results shown in Table 4.

Before turning in Section 4 to the discussion of the data collected, some considerations are in order. First, given the colloquial nature of the constructions presented above, the English rendition was not always easy to find (cf. e.g., (24)). Second,
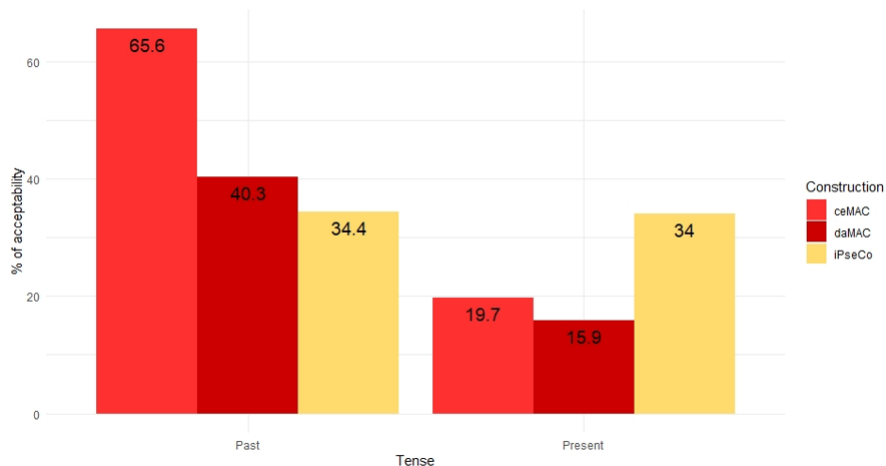
Figure 2: Percentage of acceptability of the constructions according to the tense.

Construction	Inchoative	Mirative (disapproval)	Mirative (surprise)
*če*MAC	19.9%	79.8%	63.7%
*da*MAC	69.6%	11.9%	31.9%
*i*PseCo	10.5%	8.3%	4.4%

Table 4: Percentage of acceptability of the constructions according to the action type.

the relatively low percentages of overall acceptability of the three constructions (cf. Table 1) must be contextualized. Not only the informality of TAKE *če*MAC, TAKE *da*MAC and *i*PseCo surely caused a lower rating of acceptability, but also some features tested in the items (e.g., V1 in the present tense) contributed to boost the percentage of unacceptability.

4   Discussion and conclusions

4.1   Syntactic properties

From a structural point of view, it is interesting to compare the properties of the constructions with functional TAKE with those of the canonical MACs found in Bulgarian (cf. Section 2.1 and 2.2). In this way we can highlight common and deviant features to start capturing the nature of the TAKE constructions and to lead
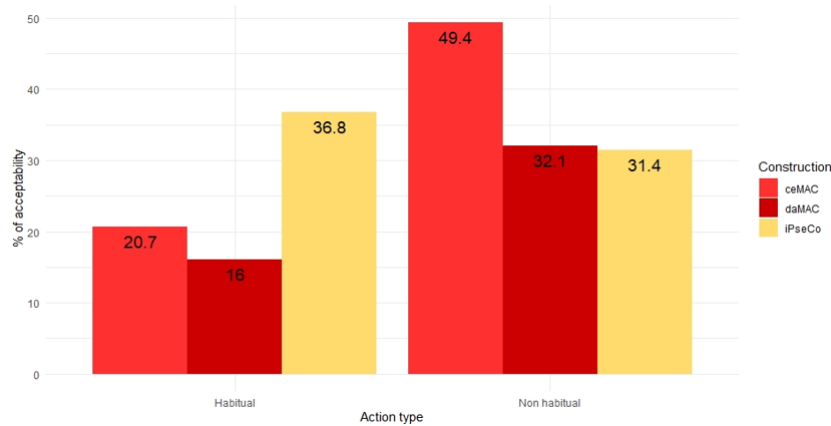
Figure 3: Percentage of (un)acceptability of the *i*PseCo in sentences with a present or a past verb.

to new insights that will suggest further questions for future research.

The relevant structural properties of the canonical MACs are given in Table 5.

The data we collected from the questionnaire (integrated with some exploratory fieldwork and some online research, which preceded the creation of the online survey) allow us to describe the TAKE *če*MAC, TAKE *da*MAC and *i*PseCo in terms of the same features outlined in Table 5 to guarantee maximal comparability of the canonical and TAKE constructions. The structural features of the latter arising from the collected data (that will be discussed in more detail below) are summarized in Table 6.

The comparison between the constructions with functional TAKE and the canonical MACs makes it clear that the former share almost all the features with the canonical monoclausal *da*MAC, deviating from the pattern of the remaining two canonical MACs that are instead biclausal. We thus assume that the three constructions with functional TAKE we investigated have monoclausal nature.

Their monoclausality straightforwardly accounts for the impossibility of having two distinct subjects for V1 and V2, which is common to both the two TAKE MACs and the *i*PseCo. The other features shared are the person paradigm of V1, which is unrestricted, and its class, which instead seems restricted to the verb TAKE.
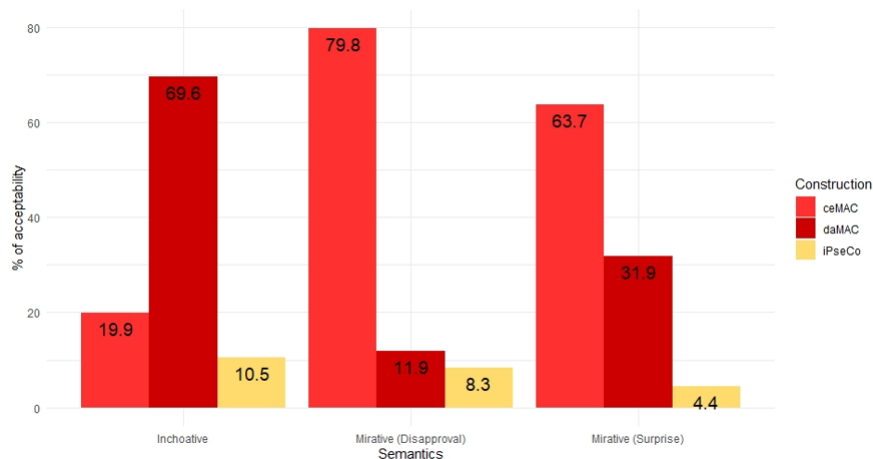
Figure 4: Overall percentage of choice of *i*PseCo, TAKE *da*MAC and TAKE *če*MAC in the three relevant contexts

However, the three TAKE constructions slightly differ in some properties and distribution: the main differences concern the tense of V1 and TAM sharing between V1 and V2.

The two MACs display an overwhelming preference for a past V1 (cf. Figure 2), which points to the fact that these constructions are degraded if TAKE is used in the present tense. Even more restricted is the tense of V2, which can only appear in the present imperfective for TAKE *da*MAC, while it must share TAM features with V1 in the case of TAKE *če*MAC and *i*PseCo. Note that, at least in TAKE *da*MAC, it is the V1 TAKE which provides the reference time for the whole event, the tense of V2 being just an anaphoric form selected by the functional V1.

The *i*PseCo seems instead to be freer in the tense selection of V1. Figure 2 shows that, despite being less accepted than the two TAKE MACs, the *i*PseCo displays a similar acceptability rate for the sentences both in the present and in the past. Moreover, many examples found in online corpora show that this construction is quite productive in the imperative as well (cf. (22c-d)), in line with what Di Caro (2019: 129) reports for Southern Italo-Romance MACs. These data are not sufficient to claim that the tense paradigm of the V1 in the *i*PseCo is *de facto* unrestricted, but they show that the *i*PseCo has a wider distribution than TAKE *če*MAC and TAKE *da*MAC which (almost) exclusively appear in the past. Moreover, the obligatory TAM features between V1 and V2 of *i*PseCo is a feature that holds cross-linguistically for this kind of construction (cf. Section 1). The fact that TAKE

Features	Canonical *če*MAC	Canonical bicl. *da*MAC	Canonical monocl. *da*MAC
Tense and Aspect of V1 and V2	Possibly disjoint	Possibly disjoint	Possibly disjoint
Reference of V1 and V2	Possibly disjoint	Possibly disjoint	Conjoint (only one subject)
Tense of V1	Not restricted	Not restricted	Not restricted
Person of V1	Not restricted	Not restricted	Not restricted
Tense of V2	Not restricted	Not restricted	Restricted (present imperfective)
Person of V2	Not restricted	Not restricted	Same person of V1
Class of V1	Restricted to some classes	Restricted to some classes	Restricted to some classes

Table 5: Summary of the structural features of the Bulgarian canonical MACs.

*če*MAC also displays this feature casts some doubts about its nature and calls for further research.

4.2 Semantic properties

From a semantic point of view the three TAKE constructions behave differently with respect to both the compatibility with habitual actions and the meaning functional TAKE carries in the periphrasis itself. These properties also provide some insight for justifying some of the features discussed in the previous section.

Figure 3 presents the same asymmetry found in Figure 2, namely TAKE MACs behaving in a similar way and differing from the pattern of the *i*PseCo. TAKE *da*MAC and TAKE *če*MAC (to an even greater extent) have a neat preference for non-habitual, single actions. This straightforwardly correlates with their predominant use in the past tense, given the aspect of the V1. In fact, single actions refer to the past, and they are generally expressed in Bulgarian via the perfective form of the aorist. Habitual actions instead require an imperfective verb. Crucially, the imperfective is the only aspect available in the present tense (as the action lacks a result, it cannot be said to be concluded at the speech time).

Features	*če*MAC	*i*PseCo	*da*MAC
Tense and Aspect of V1 and V2	Necessarily conjoint	Necessarily conjoint	Possibly disjoint
Reference of V1 and V2	Conjoint (only one subject)	Conjoint (only one subject)	Conjoint (only one subject)
Tense of V1	Restricted (past tense)	Possibly not restricted	Restricted (past tense)
Person of V1	Not restricted	Not restricted	Not restricted
Tense of V2	Same tense of V1	Same tense of V1	Restricted (present imperfective)
Person of V2	Same person of V1	Same person of V1	Same person of V1
Class of V1	Restricted to the verb TAKE	Restricted to the verb TAKE	Restricted to the verb TAKE

Table 6: Summary of the features of TAKE *če*MAC, TAKE *da*MAC and *i*PseCo.

The same reasoning applies to the *i*PseCo which, unsurprisingly, has a quite similar rate of acceptability with both habitual and non-habitual actions. This goes hand in hand with the occurrence of the *i*PseCo with both the past (perfective) and the present (imperfective).

As for the reading conveyed by functional TAKE, the two MACs operate a very clear division of labors, while the *i*PseCo seems to be broader in its use. The results of the semantic specialization are presented in Figure 4. TAKE *da*MAC has a clear inchoative meaning, indicating the starting point of an action. TAKE *če*MAC specializes instead for at least two shades of mirativity (following DeLancey 1997; Ross 2016), namely the speaker's (i) surprise and (ii) disapproval for the content of the event. The restriction of TAKE *če*MAC to past sentences naturally follows from its semantics: the events it describes, namely unexpected (and often sudden) events which led to a perceivable result, necessarily need to be located in the past. Present (i.e., simultaneous to the speech act) events cannot denote completed actions whose result can trigger a surprise/disapproval reaction by the speaker.

As far as the *i*PseCo is concerned, the distinction is not that clear-cut. First, Figure 4 shows that this construction is least preferred than the TAKE MACs. The contexts investigated only inchoative and mirative semantics in the past, hence

the precise reason of its lower acceptability is still to be understood. Second, in the cases where it is accepted, the *i*PseCo seems to mainly express inchoativity (which would make it compatible also with the present tense) and mirativity with the disapproval connotation. At a first glance, the *i*PseCo can appear semantically redundant, since Bulgarian already has the TAKE MACs to convey the semantics of the *i*PseCo. However, it may be the case that the *i*PseCo makes these semantic nuances available with verbal tenses in which the TAKE MACs are disallowed.

4.3 Conclusions and further perspectives

Given the scarceness of data available in the literature about the relevant constructions, we decided to start investigating their properties collecting data submitting an online questionnaire to native speakers.

From our results we can conclude that TAKE *če*MAC, TAKE *da*MAC and *i*PseCo are attested and used in contemporary Bulgarian, although the former two seem to be more productive, while the latter is not accepted by all speakers. As far as their structure is concerned, all of them pattern along with other monoclausal constructions; the two TAKE MACs are used to describe past events, while the *i*PseCo occurs both in the present and in the past with a similar rate. As for their semantics, the TAKE MACs are compatible with non-habitual actions, while the *i*PseCo can characterize habitual actions as well. Moreover, TAKE *da*MAC specializes for inchoativity, while TAKE *če*MAC for mirativity. The data reveal that the *i*PseCo is mainly inchoative and mirative (with a disapproval flavor). The existence of the *i*PseCo, apparently redundant from a semantic point of view, may be justified by the fact that it makes the construction available with verbal tenses otherwise disallowed.

This piece of research raised some questions to be addressed for future research. First, the monoclausality of TAKE *če*MAC raises the question about the status of the connector *če*, which is considered as a complementizer with full rights. In the case of this construction, instead, it could have a different nature, possibly having a role in the semantics or in the selection of V2. Second, we have to verify the existence of morphemic restrictions (i.e., whether there are any cells of the paradigm of V1 that are not allowed because of non-syntactic reasons). Third, the monoclausal status of these constructions could be further corroborated by investigating the role of the negation (namely, whether the two verbs can be negated separately). Fourth, we must verify whether V1 can project a full argument structure (e.g., take a direct object), as this would say much about its functional nature. Last but not least, the semantics of the *i*PseCo must be further investigated to understand what its exact

meaning is and whether this is dependent on the (imperfective vs. perfective) aspect of V1.

References

Aikhenvald, A. (2006). Serial verb constructions in typological perspective. In A. Y. Aikhenvald & R. M. W. Dixon (Eds.), *Serial verb constructions: A cross-linguistic typology* (pp. 1–68).

Aikhenvald, A. (2018). *Serial verbs*. Oxford University Press.

Andrason, A. (2018). The WZIAĆ gram in polish. a serial verb construction, or not? *STUF - Language Typology and Universal*, *71*(4), 577–629.

Bleotu, A. C. (2022). The properties of the '(a) lua şi x' ('take and x') construction in Romanian: Evidence in favor of a more fine-grained distinction among pseudocoordinative structures. In *Pseudo-coordination and multiple agreement constructions* (pp. 149–168). John Benjamins.

Buchstaller, I., & Khattab, G. (2013). Population samples. In *Research methods in linguistics*. Cambridge University Press.

Cardinaletti, A., & Giusti, G. (2001). "Semi-lexical" motion verbs in Romance and Germanic. In *Semi-lexical categories* (pp. 371–414). De Gruyter.

Coseriu, E. (1966). "Tomo y me voy": Ein problem vergleichender europäischer syntax. *Vox Romanica*, *25*, 13–55.

Cruschina, S. (2013). Beyond the stem and inflectional morphology: An irregular pattern at the level of periphrasis. In *The boundaries of pure morphology: Diachronic and synchronic perspectives* (pp. 262–283). Oxford University Press.

De Angelis, A. (2017). Between Greek and Romance: Competing complementation systems in Southern Italy. In *Language and identity in multilingual Mediterranean settings: Challenges for historical sociolinguistics* (pp. 135–156). De Gruyter.

Déchaine, R.-M. (1993). Serial verb constructions. In J. Jacobs, A. von Stechow, W. Sternefeld, & T. Vennemann (Eds.), *Syntax: Ein internationales handbuch zeitgenössischer forschung [an international handbook of contemporary research]* (pp. 799–825). Walter de Gruyter.

DeLancey, S. (1997). Mirativity: The grammatical marking of unexpected information. *Linguistic Typology*, *1*, 33–51.

Di Caro, V. N. (2019). *Multiple agreement constructions in Southern Italo-Romance. The syntax of Sicilian pseudo-coordination* (Doctoral dissertation). Ca' Foscari University of Venice.

Ganfi, V. (2021). *Diacronia e sincronia del complementatore mi in Siciliano*. LIN-COM.

Giusti, G. (2011). Structural protocols for linguistic awareness enhancing language identity [online] [Available at http://www.crassh.cam.ac.uk/people/profile/giuliana-giusti].

Giusti, G., & Cardinaletti, A. (2022). Theory-driven approaches and empirical advances: A protocol for pseudo-coordinations and multiple agreement constructions in Italo-Romance. In *Pseudo-coordination and multiple agreement constructions* (pp. 35–64). John Benjamins.

Giusti, G., Di Caro, V. N., & Ross, D. (2022). Pseudo-coordinations and multiple agreement constructions: An overview. In *Pseudo-coordination and multiple agreement constructions* (pp. 1–34). John Benjamins.

Hansen, B., Letuchiy, A., & Błaszczyk, I. (2016). Complementizers in Slavonic (Russian, Polish, and Bulgarian). In K. Boye & P. Kehayov (Eds.), *Complementizer semantics in european languages* (pp. 175–223). De Gruyter Mouton.

Josefsson, G. (2014). *Pseudo-coordination in swedish with gå 'go' and the 'surprise effect'*. Lund University.

Kanchev, I. (2010). Семантика, типология и произход на конструкцията вземам/взема че (та, и) + глагол от свършен вид [On the semantics, typology and origin of the construction вземам/взема че (та, и) + perfective aspect verb] [http://www.slav.uni-sofia.bg/index.php/nova-knizhka/611-3-2010]. Съпоставително езикознание*[Contrastive Linguistics]*, *35*(3).

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). *Lexicography*, *1*(1), 7–36.

Kilgarriff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004). *Itri-04-08. The sketch engine* (tech. rep.). Information Technology Research Institute Technical Report Series.

Krapova, I. (1998). Subjunctive complements, null subjects and case checking in Bulgarian. *University of Venice working papers in linguistics*, *8*(2), 73–93.

Krapova, I. (2001). Subjunctives in Bulgarian and modern Greek. In *Comparative syntax of Balkan languages* (pp. 105–126). Oxford University Press.

Krapova, I. (2002). On the left periphery of the Bulgarian sentence. *University of Venice Working Papers*, *12*, 107–128.

Krapova, I. (n.d.). Complementizers and particles inside and outside of the left periphery: The case of Bulgarian revisited. In *Clausal complementation in south slavic* (pp. 161–211). De Gruyter Mouton.

Krapova, I., & Cinque, G. (2018). Universal constraints on Balkanisms. a case study: The absence of clitic climbing. In I. Krapova & B. Joseph (Eds.), *Balkan*

syntax and (universal) principles of grammar (pp. 151–191). De Gruyter Mouton.

Krapova, I., & Karastaneva, T. (2002). On the structure of the cp field in Bulgarian. *Balkanistica*, *15*, 293–322.

Krapova, I., & Petkov, V. (n.d.). Subjunctive complements, null subjects, and case checking in Bulgarian. In *Annual workshop on formal approaches to Slavic linguistics 7: The seattle meeting 1998* (pp. 267–285). Slavic Publications.

Ledgeway, A. (2013). Greek disguised as Romance? the case of Southern Italy. In *Proceedings of the 5th international conference on greek dialects and linguistic theory* (pp. 184–228). Laboratory of Modern Greek dialects, University of Patras.

Lefebvre, C. (1991). *Serial verbs: Grammatical, comparative and cognitive approaches*. John Benjamins.

Lødrup, H. (2002). The syntactic structures of norwegian pseudocoordinations. *Studia Linguistica*, *56*(2), 121–143.

Lødrup, H. (n.d.). Norwegian pseudocoordination with the verb drive "carry on": Control, raising, grammaticalization. In M. Butt & T. H. King (Eds.), *Proceedings of the lfg'17 conference* (pp. 264–284). CSLI Publications.

Manzini, R., Lorusso, P., & Savoia, L. (2017). *a*/bare finite complements in Southern Italian varieties: Mono-clausal or bi-clausal syntax? *Quaderni di linguistica e studi orientali*, *3*, 11–59.

Manzini, R., & Savoia, L. (2005). *I dialetti italiani e romanci. morfosintassi generativa, vol. i: Introduzione – il soggetto – la struttura del complementatore, frasi interrogative, relative e aspetti della subordinazione*. Edizioni dell'Orso.

Masini, F., Mattiola, S., & Vecchi, G. (n.d.). La costruzione "prendere e v" nell'italiano contemporaneo. In B. Moretti, A. Kunz, S. Natale, & E. Krakenberger (Eds.), *Le tendenze dell'italiano contemporaneo rivisitate. atti del lii congresso internazionale di studi della società di linguistica italiana (berna, 6–8 settembre 2018)* (pp. 115–137). Bulzoni.

Mendes, G., & Ruda, M. (2022). Pseudo-coordination and ellipsis. expressive insights from Brazilian Portuguese and Polish. In *Pseudo-coordination and multiple agreement constructions* (pp. 169–190). John Benjamins.

Nicolova, R. (2008). *Bǎlgarska gramatika – morfologija*. Sv. Kliment Oxridski.

Pitsch, H. (2018). Bulgarian moods. *Journal of Slavic Linguistics*, *26*(1), 55–100.

Prete, F. D., & Todaro, G. (2020). Building complex events: The case of Sicilian doubly inflected construction. *Natural Language & Linguistic Theory*, *38*(1), 1–41.

Pulkkinen, P. (1966). *Asyndeettinen rinnastus suomen kielessä (asyndetic coordination in finnish)* (Doctoral dissertation). University of Helsinki. Helsinki.

Rohlfs, G. (1969). *Sintassi e formazione delle parole* (Vol. 3). Einaudi.

Ross, D. (2016). Going to surprise: The grammaticalization of itive as mirative. In *Online proceedings of cognitive linguistics in wrocław web conference*. Polish Cognitive Linguistics Association; University of Wrocław.

Ross, D. Pseudocoordinación del tipo tomar y en eurasia: 50 años después [Pseudocoordination with take and in eurasia: 50 years later] [http://hdl.handle.net/2142/110127]. In: In *Vi congreso internacional de lingüística coseriana: Actualidad y futuro del pensamiento de eugenio coseriu, pontificia universidad católica del perú & universidad del pacífico*. http://hdl.handle.net/2142/110127. Lima, 2017.

Rudin, C. (1986). *Aspects of Bulgarian syntax: Complementizers and wh constructions*. Slavica Publishers.

Sandfeld, J. (1900). *Rumaenske studier i. infimtiv og udtrykkene derfor i rumaensk og balkansprogene, en sammenlignende undersøgelse*. Copenhagen.

Simov, K., & Kolkovska, S. (2004). Interpretacija na da-konstrukciite v opornata frazova gramatika. *Slavistikata v načaloto na XXI vek - tradicii i očakvanija*, 153–162.

Soto Gómez, J. F. (2021). *Pseudocoordination in spanish. a two construction analysis* (Master's thesis). Ca' Foscari University of Venice.

Tomić, O. (2006). *Balkan sprachbund morpho-syntactic features*. Springer.

Tomić, O. Mood and negation in balkan slavic (G. Zybatow, L. Szucsich, U. Junghanns, & R. Meyer, Eds.). In: *Formal description of slavic languages: The fifth conference, leipzig 2003* (G. Zybatow, L. Szucsich, U. Junghanns, & R. Meyer, Eds.). Ed. by Zybatow, G., Szucsich, L., Junghanns, U., & Meyer, R. Frankfurt am Main: Peter Lang, 2008, 461–477.

Tragel, I. (2017). Serial verb constructions in Estonian. In *Argument realisation in complex predicates and complex events: Verbverb constructions at the syntax-semantic interface* (pp. 169–189). John Benjamins.

Wiklund, A.-L. (2007). *The syntax of tenselessness: Tense/mood/aspect-agreeing infinitivals*. Mouton de Gruyter.

Wiklund, A.-L. (2008). Creating surprise in complex predication. *Nordlyd*, *35*, 163–187.

Wiklund, A.-L. (2009). The syntax of surprise: Unexpected event readings in complex predication. *Working Papers in Scandinavian Syntax*, *84*, 181–224.

Reduplication in Urdu-Hindi and other Indo-Aryan Languages: A postdoctoral research at University of North Texas

Riaz Ahmed Mangrio   Sadaf Munshi

Sindh Madressatul Islam University, Karachi   University of North Texas, Denton

riazmangrio@yahoo.com   sadaf.munshi@unt.edu

Abstract

Reduplication is a morphological operation involving complex phonological processes to express new meaning. Inkelas (2005) views reduplication as the double (or multiple) occurrence of a morphological constituent meeting a particular morphosemantic description.Katre (1939) observes onomatopoetic type reduplication in Sanskrit. Reduplication in Urdu-Hindi and other Indo Aryan languages affects nouns, verbs, adjectives, particles e.g. Urdu *vah vah* (for appreciation), quantifiers and even nonsensical words. It conveys general meanings of distribution, emphasis, iteration etc, but also sometimes specific meaning e.g. *rʌŋ brʌŋgi* 'colourful'. This study is a comparative analysis of reduplications looking at morpho-phonological features and the meaning variation in Urdu-Hindi mainly and then Punjabi, Saraiki/Pothwari, Sindhi and Baluchi. In addition to a) highlighting morpho-phonological and semantic features, the study will also look at b) some negative markers working as infixes and changing semantics like conjunctive particle *-o-* and *-e-* working as infixes and changing semantics c) the difference of semantics expressed by the reduplication of lexical verb essentially requiring converb and an imperfective verb which shows simultaneity. Revisiting Montaut (2008, p. 29) is needed. She does not differentiate lexical verb and imperfective verb for reduplication. The study consists of descriptive analysis of the data with reference to previous studies on reduplication in world languages and in IA languages.

1   Introduction

Reduplication is a linguistic phenomenon in which a linguistic item is doubled or copied in language for some linguistic purposes. Generally, the doubling or copying is called duplication in which all or part of one linguistic constituent is repeated or doubled to produce a completely new constituent, serving different

semantic functions, and occurring at many levels from phonologically partial lexical reduplication to the repetition of the whole phrases and sentences. If complete word is copied and pasted along the base word, e.g. *ləmba ləmba* 'very tall' in Urdu-Hindi, it is called total reduplication; if part of a word is copied and pasted alongside the base word, e.g. Urdu-Hindi *khal val* 'skin and the like', it is called partial reduplication or echo reduplication. Kachru (1990, p. 62) calls it echo compounding, as it is the repeated form of a word with a different initial consonant e.g. *ləmha* 'moment'→*ləmha-b-ləmha* 'moment-with-moment'. According to Montaut (2008, p. 18), echo reduplication is 'a phonetic alternation of the base words'. As will be seen, it shows various interesting phonological features.

The term reduplication is used when such doubling is discussed with reference to a particular linguistic level. Therefore, duplication at the level of words, phrases, and sentences is called lexical reduplication, phrasal reduplication and syntactic reduplication respectively (Hurch, 2005). Lexical reduplication is based on regular words or lexical items and is also known as substantial reduplication (Olsson, 2015). One constituent in this form usually carries meaning while the other just adds something more to the meaning of the first one and does not appear alone. Morphological reduplication- also termed as expressive or formal reduplication (Abbi, 1992; Lu, 2017) and occurs in most of the world languages Nintemann, 2016, deals with the onomatopoeic expressions, imitations, sound symbolism, mimic words.

In simple words, reduplication is a morpho-phonological process by which a word or a part of a word is copied and affixed to the original word, the 'base'. Repetition of an entire word e.g. *hə̃sʈe* 'laugh' → *hə̃sʈe hə̃sʈe* 'laughingly' (U) involves total reduplication. The repetition of part of a word is called partial reduplication e.g. *aʈa* 'flour' → *aʈa vaʈa* 'flour and the like'. If there is a change in the onset, like this, it is called echo compounding (Kachru, 1990, p. 62) or echo reduplication (Bögel et al., 2008), which is more complex than simple or partial reduplication. Echo reduplication e.g. *fancy schmancy* is very common cross linguistically and appears to be contagious aerial phenomenon in Indo-Aryan and Dravidian languages especially. The mainstream term used is 'melodic overwriting' (Inkelas, 2005).

Hurch (2005: 2) considers that the question of reduplication described in phonological or morphological terms is *vaexata quaestio* 'a disputed question', as it is linked to the status of morphology in grammar; the grammatical function of reduplication e.g. plurality, intensification or a derivation by compounding is counterbalanced by the fact that most reduplication can be described entirely in phonological terms. In Morphological Doubling Theory (MDT), by Inkelas (2005), reduplication is viewed as the double (or multiple) occurrence of a morphological constituent meeting a particular morphosemantic description. The reduplicant and the base

are both a part of a construction, which also embodies semantic and phonological generalizations. Base and reduplicant are distinctive meaningful constituents of a compound and may be separated from each other with an infix in between. Bauer (2001) terms such an infix an interfix, which actually connects the two constituents of a compound. Urdu-Hindi and other IA languages frequently show this feature in reduplications. Following subsections give a brief sketch of reduplications in world languages and then Indo Aryan languages.

1.1 Reduplication in the world languages

Researchers have discussed various phonological, morphological, semantic and grammatical features in the reduplication of world languages. For example, Keenan & Polinsky (1998:571) consider that the reduplication in Malagasy is suffixal in nature and analyse that the stress on the base segment falls on the copy in the reduplication. In contrast, Hannahs (2004) emphasizes that this reduplication is rather infixal arguing that Keenan & Polinsky's claim is crucially based on which bits are identified as base and which are identified as copy. If the copy is infixed to the left of the structure copied, rather than suffixed to its right, the entire reduplication process can be expressed more easily and with fewer stipulations. Hannahs (2004) highlights that Malagasy attenuative reduplication consists of a maximally bi-syllabic copying of the syllable bearing the main stress and a syllable to its right, and infixation of this copy to the left of the syllable bearing main stress. Malagasy reduplicant to him is closer to the base than to the input.

	base form	reduplicated form
(1)	vovó 'barking'	vovòvó 'bark occasionally'
	lèhibé 'big'	lèhibèbé' biggish'

Noonan (1992: 175) notes that a southern dialect of Uganda language, Lango, has an emphatic syntactic construction which repeats the verb. The first copy of the verb is inflected normally. The second copy, however, appears in what is called the gerund form; it "is given a high tone and preceded by à- and followed by -â..."

(2) a. àbínô   àbín´ˆ   àwó'ró
       1SG.COME.PERF come.GER yesterday
       'I did come yesterday'

    b. ákô   òmyE›lò   àmyE@lâ
       1SG.COME.PERF come.GER yesterday
       'The girl just danced'

Thurgood (1997) devises templates and rules to show that Bontok heavy syllable reduplication is linked with two phonological processes, gemination and metathesis. Gemination occurs to make a reduplicant heavy; metathesis in order to avoid a glottal stop in the coda. Golston and Thurgood (2003) disagree to any grammatical role in the reduplication of Bontok and claim that reduplication is tied to specific morphemes or sets of morphemes. They find three Bontok reduplicants of different prosodic size with light (CV) syllabic for intensive sense, heavy (CVC) syllable for progressive and a foot (two syllables) for repetitive meaning. Bontok has an extensive amount of verb reduplication and a small amount of noun reduplication- with the three size reduplicants, depending on the meaning e.g. *ka kamaNan* 'hurry a lot'. As they note, a similar feature is found in Chumash also. Final VC reduplication in Chumash seems to involve an echo word-finally. Individual morphemes are important both in Bontok and Chumash reduplication, which means that lexicon dominates grammar. Reduplication also marks inflectional categories e.g. a nominal reduplication, in Tarok, expresses 3rd person possession.

Unseth (2003) notes frequent reduplication of antepenultimate consonant (the last two consonants) in Semitic languages. Looking at geographical linkages, between Persian and Semitic languages, Persian reduplication is also notable. As Persian is the language of mixed nature and features, and so most of the reduplications come in the form of a phrase consisting of a Persian word part *-ma-* and an Arabic word e.g. *t̪əɣd̪ir-məɣd̪ir* 'fate/luck and the like' (also used in Pashto- an IA language). The Persian reduplications can be further classified into true and quasi ones. In true reduplication, both words are real and meaningful. In quasi-reduplication, one of the words is meaningful and the other is meaningless. Some examples of true reduplication in Persian are: *xert-o-pert* 'odds and ends', *tʃʌrt̪-o-pʌr* 'nonsense', *tʃʌrin̪d̪ pʌrin̪d̪* 'grass eating animals and birds', *āb-o-t̪āb* 'much detail'. The quasi-reduplications are *zʌn-o-mʌn* 'wife', *d̪avā-mavā* 'argument', *t̪ʌlā-mʌlā* 'jewelry' and *rʌxt-o-pʌxt* 'Items of clothing'. In general, reduplication in Persian is mainly used in the sense of mockery of words with non- Persian origins.

Persian has heavy influence on Urdu-Hindi and other Indo Aryan languages. The focus of this study is to descriptively discuss mainly Urdu-Hindi reduplication at morphological level with no theoretical debate. Then, there is some description of reduplicative structures in some other Indo Aryan languages like Sindhi, Punjabi, Baluchi and Saraiki. However, it is not a comparative study. Persian is a big source of lexical items in Urdu-Hindi, exhibiting rich morphological structures and various functions which we observe also in other Indo Aryan languages. Therefore, the complex issues of reduplication seem even more complex in Urdu-Hindi. There are three major things to be discussed a) Reduplication with an internal vowel change

in the base b) The use of native and Persian negative markers in reduplication c) The use of a converb with the reduplication of lexical verb. This is followed by some discussion on Montaut (2008).

2 Reduplication in Indo Aryan languages

It is important for understanding reduplication in Indo Aryan languages to give its brief historical overview.

2.1 Reduplication in IA languages- A historical perspective

Katre (1940) highlights some morpho-phonological features in Middle Indo Aryan (MIA) and New Indo Aryan (NIA) reduplication. Focusing on Sanskrit, Pali, Prakrit, Marathi, Gujrati, Hindi and Nepali, he considers a comparative etymology of MIA and NIA languages important to study their verbal bases, nominal derivatives, non-Aryan substratum, jingle words, echo words e.g. *ku ku* 'cuckoo sound' and *tʃĩ tʃĩ* 'chirping' and other types in Urdu-Hindi. In MIA reduplications, mainly vocatives and vocative phrases are recurrent in the same or modified rhyming forms in which sometimes lengthening of vowels is a key aspect of phonology. He observes that onomatopoeia is one of the major features of reduplication particularly in Sanskrit; it affects single consonant, vowel or group of vowels. Katre (1940) notes that, in classical and Vedic Sanskrit, some indeclinable compounds like *nakha nakhi* 'nail to nail' and compound verbs, such as e.g. *piba* 'drinking/who or what drinks' → *piba piba* 'drink, drink' (stress) also exist in MIA reduplications. Gender markers e.g. *kuli* (neuter nominative/accusative/vocative singular N/A) and *kulam* (masculine acc N/A) exhibit a form of reduplication as *kulikulam*, although it is a compound. The reduplication of imperative verbs e.g. *khadata* 'eat' produce feminine *khadata modata* 'continual eating and rejoicing'. Like other morpho-phonological features, Urdu-Hindi reduplicative gender marking follows MIA.

Looking at a large number of examples in Marathi and Gujrati, Katre (1940) assumes that reduplication is a popular form of word choice in the NIA languages. One frequent feature noted in the NIA reduplication is the use of an infix *-o-* e.g. *kaːn-o-kaːn* 'ear to ear' and *raːt-o-raːt* 'by night/in the night'. Bashir and Conners (2021) also highlight some infixes, like *-aː-* and *-e-* in all the NIA languages e.g. *gərm-a-gərəm* 'fresh/very hot', *maːl-aː-maːl* 'rich/replete/full' in Hindi and *kan-e-kan* 'ear to ear' in Gujrati. Although the examples of *-a-* also exist- e.g.

gərm-a-gərəm and *ma:l-a:-ma:l*, the two infixes *-o-* and *-e-* are more frequent, in Urdu-Hindi, working as conjunctive particles with giving the semantics of *ɔr* 'and' and genitives *ka/ki/ke* 'of'.

Similarly, the reduplication of sound beating and object sounds are also explicit e.g. *kʰən kʰəna:na* 'bangle sound', *pʰəʈ pʰəʈa:na:* 'to flap', *pʰəʈ pʰəʈa:na:* 'to flutter', which are commonly used in Urdu-Hindi today as well. There is no definite rule, for the reduplication of such jingle words, but Katre (1940) views that an inference can be helpful to understand such structures in languages and the dialects.

Besides iterative compound nouns, numerals, pronouns and prepositions, there are adjectival reduplications formed with preposition e.g. *krlakrlam* 'done and undone'.

Some NIA verbal triplications or multiplicative verbs- as termed by Katre, e.g. Marathi *ghu ghu* 'the hoot or cry of the owl, pigeon, and of the bird hola'→ *ghu ghu ghu* 'have (something)' and *guma* 'still, quietly, silently; in the stray, missing, or lost state' → *guma guma guma* 'astray' are also noteworthy. The phonological changes in all of them e.g. addition of bilabial nasal */m/* in the reduplicant here are visible.

Katre (1940) looks at reduplications generally in IA languages, not specifically in Urdu-Hindi. Therefore, it is important to study the modern Urdu-Hindi reduplications looking at also some examples in other Indo-Aryan languages.

2.2 Reduplicative features in Urdu-Hindi and IA languages

Urdu-Hindi is a prestige language in IA family in the sense that it is commonly understood and used as lingua franca throughout the Indian subcontinent. There has not been a comprehensive study on its reduplication, particularly with the focus on the description of the reduplicative structures, although there are some smaller general observations available by linguists. A good descriptive study first is important to start a theoretical debate on reduplication in Urdu-Hindi.

Modern linguists seem to observe the same reduplicative features, as seen by Katre (1940) generally in IA languages, although there may be some dissimilarities. Urdu-Hindi reduplication exhibits interesting morpho-phonological changes. Both vowel and consonantal changes are visible. Some of the general reduplicative features and strategies in Urdu-Hindi include 1) phonological changes in reduplicant 2) insertion of an infix generally and a negative marker particularly in between base

and reduplicant and 3) a converb being part of reduplicant and thus reduplication being complete, as seen below:

(3)

	a.	dʰ**iː**l		+	dʰ**aː**l		=	dʰ**iː**l dʰ**aː**l
		'loose'						'loose/relaxation'
	b.	kaːm	(m)	+	**naː** + kaːdʒ		=	kaːm-**na**-kaːdʒ
		'work/job'						'no job/nothing to do'
	c.	ro		+	ro + **kər**		=	ro ro **kər/ke**
		'weep/cry'						'by crying'

The first process of reduplication shows the revoweling in the reduplicant i.e. insertion, deletion as well as the substitution of a vowel, as long */a/* substituting long */i/* in the reduplicants (3a). Secondly, like other infixes, e.g. *-a-, -o-* and *-e-*, there is an insertion of a negative marker *-na-* (3b). This means a reduplicant requires an infix to connect with its base. Thirdly, a converb *k@r* or a particle *ke* (3c) is obligatory along with reduplicant to connect with base. In all three cases, the changes in reduplicants are important.

The first form of reduplication that we observe here shows the complete modification of base word *dʰ**i:***l* 'loose' in its reduplicant *dʰ**a:***l*. As there is an internal change of vowel, this is also called revoweling. This could mean that there are various "internally reduplicating" processes of nouns, pronouns, adjectives, verbs and adverbs etc exhibiting insertion, deletion and alternation of a vowel in the reduplications, e.g. *ɖər* 'fear' → *ɖər ɖʊr* 'fear' (with emphasis). Some examples may be seen below.

A. Nouns and adjectives with onset alternation (*m-* versus *ʈ-*). There are also other possibilities of alternations, which include both single vowels and consonants, not necessarily at the onset position e.g. *dʰ**i:***l dʰ**a:***l* 'relaxation' and *maŋ ʈaŋ* 'borrowing'.

B. Nouns and adjectives with an affix, whose analysis is open – prefixed to reduplicant, or suffixed to base, or some kind of interfix in Bauer's (2001) words. For example, see the reduplication of nonsensical words e.g. *gũ-m-gũ* '(someone) not being decisive'

C. The insertion of converb in the reduplication of a lexical verb e.g. *bʰag bʰag-kər* 'due to running' and in the echo reduplication of a lexical verb e.g. *bən ʈʰən-kər* 'being well dressed' are also some interesting reduplicative processes that produce a variety of semantics. The converb *kər* shows the cause of an action in the former, and the completion of a job in the latter. It is replaceable with the particle

ke.

For the insertion or substitution of an initial consonant, Indo-Aryan languages generally use labial fixed onsets /v/, /p/, /pʰ/, /b/ or /m/. The use of labial onsets in the reduplicants seems to be similar to the rhyming forms discussed by Ghaniabadi (2005) in Persian. He terms the reduplication pattern as *m & p* echo reduplication, the most common and productive type of reduplication in Persian. His analysis is based on the assumption that *m & p*-echo reduplication represents an instance of Alderete et al.'s (1999) "morphological fixed segmentism" whereby fixed segments generally exhibit the following properties of affixing morphology. They may:

  i) form marked structures and be in contrast with other fixed segments.

 ii) be left-aligned, right-aligned or infixed.

iii) alternate by suppletion or allomorphy

Although the above properties are seen in Urdu-Hindi and other Indo Aryan languages, there is some variation also, as can be noted.

D. There is a presence of labial fixed onsets /v/, /p/, /pʰ/, /b/ or /m/ in the reduplicants rather than merely the *m & p* that Ghaniabadi (2005) finds in Persian.

E. In the case of the reduplication of past participle, i.e. a deverbal adjective, there is often affixation of some negative markers with reduplicant, as in (3b). There are certain other examples of negative markers functioning as infixes in reduplicative structures.

F. In the case of the reduplication of a lexical verb, there must be suffixation of the converb *kər* or particle *ke* with reduplicant, as in (3c). This suffixation shows not only cause of resulting event taking place but also various other interpretations- e.g. completion of job, as exemplified above. They are highlighted in the following sections.

G. The base and reduplicant must not necessarily have identical segments.

  i) Every segment in the base may or may not be present in the reduplicant.

 ii) Every segment in the reduplicant does not necessarily have a correspondent in base.

iii) Reduplicant is not necessarily syllabified the same as the base, i.e. the two forms must not necessarily have the same number of syllables.

As compared to total reduplication, partial reduplication or echo reduplication is more frequent in Urdu-Hindi. In addition to revoweling, some consonantal changes in reduplicant can also be seen. Something that contradicts Katre (1940) presently is the absence of total triplication in Urdu-Hindi or other IA languages, as shown below:

(4)
a.	d̪o two.num	+	d̪o Reduplicant	=	d̪o d̪o R	(for appreciation) (Numeral)
b.	vɑh exclamation (pleasure)	+	vɑh Reduplicant	=	vɑh vɑh R	exclamatory appreciation)
c.	oh exclamation (sadness)	+	oh Reduplicant	=	oh oh R	(exclamation for sorrow/shock)

(5)
a.	roʈi bread.N	+	ʃoʈi Reduplicant	=	roʈi ʃoʈi R	(N)
b.	pejɑːz onion.N	+	vejɑɑːz Reduplicant	=	pejɑɑːz vejɑɑːz R	(N)
c.	rʌŋ colour.N	+	brʌŋ Reduplicant	=	rʌŋ brʌŋi R	(A)

(6)
a.	ɖʰiːl credulous. straight.m	+	ɖʰɑːl Reduplicant	=	ɖʰiːl ɖʰɑːl R	'loose/ relaxation' (A)
b.	ʈʰiːk right.fine.Z	+	ʈʰɑːk Reduplicant	=	ʈʰiːk ʈʰɑːk R	'okay'
c.	ɖɑnɑ grain.N	+	ɖuːnɑ Reduplicant	=	ɖɑnɑ ɖuːnɑ R	'grain' (N)
d.	nʌŋgɑ naked.nude.A	+	nʌŋgəm Reduplicant	=	nʌŋgəm nʌŋgɑ R	(A)

(7)
a.	d̪ekh-i + ən + d̪ekh-i see.pst.f.V neg. Redup.	=	d̪ekh-i ən-d̪ekh-i R	'not properly judged' (Neg.V)
b.	vəqt̪ + be + vəqt̪ time.N neg Redup.	=	vəqt̪-be-vəqt̪ R	'no specific time'(Neg.N)
c.	kəbʰi + nɑ + kəbʰi sometime.Adv neg Redup.	=	kəbʰi-nɑ-kəbʰi R	'at least sometimes (Neg.Adv)

	a.	roṭæ	+	roṭæ	=	roṭæ roṭæ	'while crying' (imp V)
		imp V		Reduplicant		R	
	b.	mərṭæ	+	mərṭæ	=	mərṭæ mərṭæ	'while dying' (imp V)
(8)		imp V		Reduplicant		R	
	c.	ro	+	ro + kər	=	ro ro kər	'because of dying'
		imp V		Reduplicant		R	(imp V)
	d.	mər	+	mər + kər	=	mər mər kər	'hardly' (imp V)
		imp V		Reduplicant		R	

(9)	a.	sɑmne	+	ɑmne	=	ɑmne sɑmne	'in front of each other' (Adv)
		in front.adv		Reduplicant		R	

(10)	a.	mɪjaõ	+	mɪjaõ	=	mɪjaõ mɪjaõ	'cat sound'
		cat voice		Reduplicant		R	

Although present total and partial reduplications seem to be descendants from MIA and NIA, the examples of total triplications- as seen by Katre (1940) in NIA, do not exist in Urdu-Hindi or even any other Indo Aryan language being discussed presently. There may be some examples of partial triplications *ṭaẽ ṭaẽ fɪʃ* 'finished/all gone' and *aẽ baẽ ʃaẽ* 'a talk not to the point', in which the base is not meaningful lexical item. Reduplication is however frequent and it is the reduplication of bi-syllabic words. Partial or echo reduplication is more frequent than total reduplication. There is no category change generally, and nouns, adjectives, verbs, adverbs, numerals and exclamations remain the same in their reduplications. However, a possibility is seen e.g. in *rʌŋ brʌŋgi* (5c).

Phonological changes in reduplicant occur in the forms of both revoweling or alternation of vowels (6a-6c) and consonantal change (5a-c & 6d). Revoweling is mainly seen in the word medial position in the reduplicant e.g. *ṭʰiːk ṭʰaːk* 'okay'. A reduplicant appearing with revoweling sometimes shows alliteration, as seen in this case.

The key consonantal change observed is a substitution of the first syllable onset in the reduplicant disregarding what the base is. It is often either voiceless palatal fricative /ʃ/ e.g. *taːla: ʃaːla:* 'lock and the like' or voiced labio dental fricative /v/ e.g. *taːla: vaːla:*. This is also important to note which of the two is preferred if there is a choice to be made. The answer lies in the dialectical preferences. Urdu-Hindi speakers living in the areas where it is considered to be native language or mother tongue e.g. Karachi or Hyderabad prefer /v/ and those living in Punjab and Kashmir where it is not mother tongue in general prefer /ʃ/. Indo-Aryan languages generally use labial fixed onsets i.e. /v/, /p/, /pʰ/, /b/, /m/, e.g. *həlka pʰʊlka* 'very light'

in Urdu-Hindi and an alveo palatal fricative /ʃ/ e.g. *raːɳɖ ʃaːɳɖ* 'sports' in Saraiki. This is a general observation. A distinctive quantitative study is however needed and proposed in this regard, as this is not the focus of the present study.

In addition to revoweling or a merely a consonantal change, the other important changes in the reduplicant includes the second syllable insertion of reduplicant marker, complete modification of reduplicant itself and even an addition of a nonsensical lexical item as a reduplicant, as seen below:

	a.	d̪ʰəkkɑ	'a push'	(N)	→	d̪ʰəkkəm d̪ʰəkkɑ	'push'
	b.	nʌŋga	'naked'	(A)	→	nʌŋgəm nʌŋga	'completely naked'
	c.	kʰʊlla	'open'	(A)	→	kʰʊlləm kʰʊlla	'very openly'
	d.	bʰukɑ	'hungry'	(A)	→	bʰukəm bʰukɑ	'very happy'
	e.	sukʰɑ	'dried'	(A)	→	suːkʰəm sukʰɑ	'very dried/ very slim'
	f.	mehnga	'costly'	(A)	→	mehngəm mehnga	'very expensive'
(11)	g.	baːʰg	'run'	(V)	→	baːʰgəm baːʰg	'by running fast'
	h.	ʃor	'noise'	(N)	→	ʃor ʃərabɑ	'too much noise'
	i.	kɑlɑ	'black'	(A)	→	kɑlɑ kəloʈɑ	(sarcastic) 'black'
	j.	d̪uːr	'far'	(Ad)	→	d̪uːr d̪ərɑz	'far away'
	k.	mɑr	'beat/kill'	(V)	→	mɑr d̪ɑɽ	'beating/ killing'
	l.	lʌmbɑ	'tall'	(A)	→	lʌmbɑ t̪ʌɽəŋga	'very tall'

Note that (11a) is a noun while the examples (11a-f) are mainly adjectives. Although they begin with different initial consonants, their word final phonetic forms are all same, and end with long /A:/. Therefore, their reduplicant endings are also the same. The word final vowel /A:/ is dropped and the reduplicant marker -@*m* is inserted, disregarding whether the base is noun (11a) or an adjective (11b-f). Looking at the examples, which are part of the lexicon, other similar combinations including even a verb- as in (11g), are also possible. All reduplications (11a-11g) disregarding any base category show that reduplicant appears even before the base.

Examples (11h-j) present a different appearance of reduplicant, which is not changed phonetically at only word final position but also complete modified form of the base. The key point in the reduplicant now is the sharing of a couple of sound

segments with the base. Consider *ʃor ʃəraba* and *kala kəloʈa*. The sound segments present are palatal fricative */ʃ/* and rhotic */r/* in the former and velar stop */k/* and lateral */l/* in the latter. The reduplicant is modified form of the base sharing two sound segments.

A reduplicant in Urdu-Hindi and other Indo Aryan languages may also have entirely a different phonetic form- as in (11k-l), that may look like a lexical item but actually a nonsensical word. Kiparsky (1986) proposes that total reduplication is akin to compounding, while partial reduplication is akin to affixation. This refers to complete or partial copy and paste of base in the reduplicant. However, it is only copy and paste of phonetic form. The examples in (11k-l) contradict Kiparsky's proposal in this regard, as reduplicant has nothing in common phonetically with the base. It is semantically fulfilling its role, not phonetically- an additional point of discussion, not the focus here.

Among other morphological changes, disregarding phonetics, one may see the insertion of an infix in between base and reduplicant. The infixal morpheme may be either merely a vowel like *-e-* or *-o-* (which also give some semantics) or a consonant e.g. a voiced bilabial stop like *-b-* in *rʌŋ-b-rʌŋg* 'colourful' or voiceless velar stop *-k-* in *puʈ-k-puʈ* 'good and bad son' or it may even be a negative marker.

Bauer (2003) terms an infix in a compound as interfix and describes it as follows, "A special kind of infix that appears between the two elements of a compound is an interfix. This is found in many of the Germanic languages e.g. *tag + reise* → *tag-e-reise* 'day's journey'." The use of such an infix in Urdu-Hindi is a Persian borrowing. In this regards, Naim (1999) states that it is the most commonly used Persian grammatical feature. Bogel et al (2008) and David et al (2009) discuss *-e-* as *ezafah* 'increase/addition' that theoretically it can only join Perso-Arabic loanwords, but in spoken usage it is occasionally used with Indic words as well. The *-e-* expresses possession with gender/number morphology by representing the genitive marker *ka / ki / ke* and the agentive particle *vala*.

The formation of new Urdu-Hindi compounds taking such an infix like *-e-* e.g. *əd̪aləʈ-e-ʊzma* 'supreme court' or *-o-* e.g. *hʊsən-o-dʒəmal* '(kinds of) beauties' is often seen. Mangrio et al. (2013) discuss five functions of the two infixes for their representation of genitive marker *ka/ki/ke (-e-)* and conjunctive particle *(-o-)* in the Urdu-Hindi endocentric compounds. Unlike *-e-*, the infix *-o-* is in fact a phonetic substitution of *-v-* that is often used to connect two nouns or two adjectives, e.g. *Ahmed-v-Hamid*, and conveys the conjunctive sense of *Or* 'and'/'an addition to something'. Historically, *-v-* is a reduced and shorter form of *ɔr* 'and', like many

other Sanskrit origin words, and the development is described in Paniniyan grammar
. Although they perform important functions in Urdu-Hindi compounds, the two
infixes *-e-* or *-o-* are not the necessary part of reduplication. However, they do have
place in the reduplicative process. Therefore, examples like *maːl-o-maːl*, 'very rich'
and *raṭ-õ-raṭ* 'late night' in Urdu-Hindi are often seen. The nasal infix *-õ-* is not
frequent and used in the reduplication of oblique nouns.

An infix inserted in between base and reduplicant may also be a negative marker.
Various negative markers are used in Urdu-Hindi compounds but their use in redu-
plication is yet to be discovered. Mangrio (2016) discusses three sources of Urdu
negative markers i.e. native, Persian and Arabic loans. Although Arabic loan nega-
tive markers often seen in the Urdu derivations, they have no role in reduplication.
The reduplicative role of the native and Persian loan negative markers is discussed
ahead.

3   Negative markers in the Urdu-Hindi reduplication

There are three sources of negative markers i.e. Native, Persian loans and Arabic
loans in Urdu morphological formations, as discussed by Mangrio (2016). Of the
three sources, the Arabic loans have no role in any sort of reduplication. Persian
loans *na- be-* and *kəm-* and native *ən-* are inserted in the process of reduplication
mainly as infixes. The insertion and function of negative markers *ən, kə* (native)
na, be, and *kəm* (ʔersian) in the reduplications of adjective, adverb and verb is
often seen in Urdu-Hindi. Various semantic functions can be observed in different
structures. From examples like *puːṭ-k-puːṭ* (Urdu-Hindi), contrastive meaning is
visible. There are also manners of speaking, stylistic and expressive use e.g. na in
the reduplication of Urdu-Hindi adverb of time *kəbʰi kəbʰi* 'sometimes' → *kəbʰi
na kəbʰi* 'at least sometimes'. Bashir and Conners (2021) suggest that in such a case
of reduplication, with a negative marker, the meaning becomes indefinite. Although
this semantics of indefiniteness and uncertainty is also seen in the reduplication of
Urdu-Hindi past participles, e.g. *ḍekha ənḍekha* 'not well seen', the structure also
shows attributive function and thus is an adjective:

		a.	ḍekhi 'saw' (f)	+	ən + ḍekhi	=	ḍekhi ən-ḍekhi	'not well
			see.pst.f.V		neg Redup.		R	judged' (V)
(12)		b.	vəqt	+	be vəqt	=	vəqt-be-vəqt	'no specific
			time.N		neg Redup.		R	time' (Neg.N)
		c.	kəbʰi	+	nɑ + kəbʰi	=	kəbʰi-nɑ-kəbʰi	'at least
			sometime.adv		neg Redup.		R	sometimes'
								(Neg.Adv)

3.1 Infixal role of negative markers in reduplication

In this role, a negative marker is just connecting the two constituents of a reduplication-base and reduplicant e.g. *kəbʰi-na-kəbʰi* 'at least sometimes'. Such a reduplication functions as an adverbial or an impersonal pronoun. Take *na* to see how it is connected with base and reduplicant with what semantics. Although phonologically the same, Urdu-Hindi has got *na* from two different sources, Persian and Indic. Both are different grammatically and orthographically. Indic *na* 'no' is used as a noun with the meaning of negation or refusal for something. Persian *na* is used as an affix- mainly as a prefix, and is used as a negative marker. It attaches to mostly adjectives and then adverbs, and produces antonyms e.g. *ehl* 'competent' and *na-ehl* 'incompetent'. In reduplication, Persian *na* interconnects base and reduplicant. Consider the examples.

		a.	kəbʰi	nɑ	kəbʰi	'at least	Adverbial
			sometime/	-no/not-	sometime	sometimes'	(Adverb of time)
			ever				
		b.	kəhĩ:	nɑ	kəhĩ:	'at least	Adverbial
(13)			somewhere	-no/not-	somewhere	someone'	(Adverb of place)
		c.	koi	nɑ	koi	'at least	Impersonal
			someone	-no/not-	someone	someone'	Pronoun
		d.	kɪsi:	nɑ	kɪsi:	'at least	Impersonal
			someone	-no/not-	someone	someone'	Pronoun

The Persian *na* is not obligatory for a reduplication. It is optional. The semantics in both the cases is nevertheless different. Reduplication with *na* casts some doubt and expresses some sense of indefiniteness or at least e.g. *kəbʰi-na-kəbʰi* 'at least sometime'. In Bashir and Conners' (2021) words, nothing definite is said. Examples like *kəbʰi kəbʰi* 'sometimes' and *kəhĩ: kəhĩ:* 'somewhere' are very common. These are two parallel structures. So we are not simply removing *na:*, but the opposite. There is now some sense of definiteness with respect to some event.

Semantic expression may be little different with or without the presence of another Persian negative marker *-be-* which is used as an infix between base and reduplicant. Infix is termed as interfix, if used between two constituents of a compound (Bauer, 2005). An intfixal use of a negative marker in reduplication is shown below.

(14)   a. vəqt̪-be-vəqt̪ 'regardless of any time'
        b. bɑːt̪-be-bɑːt̪ 'without any reason'

These structures are particular usages of Urdu and do not occur in Persian. They are different from those in (16). Unlike the reduplication in the absence of -na-, the one without *-be-* is grammatical e.g. *vəqt̪ vəqt̪ ki bɑːt̪* 'a matter of specific time' or *bɑːt̪ bɑːt̪, pər narɑːz* 'angry on every matter' which are commonly used and a part of daily language. Semantically, *-be-* expresses complete absence of something mentioned in the noun. Persian marker be attaches to certain nouns e.g. *vəqt̪* 'time' here and derives a meaningful structure *be-vəqt̪* 'not at right time'. When this derivation is added to base *vəqt̪* 'time', it derives a compound word *vəqt̪-be-vəqt̪* 'regardless of time', which is however a reduplicative structure.

Persian negative markers also independently produce reduplication. The last Persian negative marker *kəm* 'few/little' is though used with negative meaning sometimes, it is actually a quantifier. It is used as a modifier of a noun e,g. *kəm əqəl* 'fool' or a verb e.g. *kəm bolna* 'to speak a little'. The examples like *kəm əqəl* 'someone with low IQ' as against *əqəl mənd̪* 'wise' are common to express negative meaning. Therefore, this quantifier has also been taken as a negative marker. According to grammatical functions, as exemplified, the formation is considered as an adjective/adverb. It however requires an interfix in the reduplication process to connect base and reduplicant, as seen below.

(15)   a. k@m @z k@m
        b. k@m se k@m

Note that *kəm* requires an ablative case marker *se* or *əz* i.e. *kəm əz kəm* or *kəm se kəm* both are used as interfix and expressing the same semantics 'at least'. There is some sense of definiteness. In contrast with *na*, which is inserted as an interfix in between base-reduplicant e.g. *kəbʰi-na-kəbʰi* 'at least sometimes', *kəm* is grammatically used itself as an adjective or an adverbial and takes an interfix to produce the process of reduplication. Comparing both, there may be some sense of doubt or indefiniteness in the former, as in *kəbʰi-na-kəbʰi* unlike the latter which expresses some definiteness, as in *kəm əz kəm* or *kəm-se-kəm*. The sense of indefiniteness can also be seen in the native *-ən-*.

(16)   a. ḍekhɑ -ən- ḍekhɑ '(something) seen or unseen (m)

       b. kəhiː -ən- kəhiː '(something) said or unsaid' (f)

       c. dʒɑːnɑː -ən- dʒɑːnɑː '(something) known or unknown'

The three examples are not the typical instances of reduplication, as the structures like *ən-ḍekhaː* 'unseen', *ən-kəhi* 'unsaid' and *ən-dʒaːna* 'unknown' are parallel negative structures standing with their positive counter parts side by side. However, the combination shows reduplication morphologically. They are commonly used in daily life.

Among all negative markers, except for *kəm-* which produces itself the process of reduplication, a negative marker e.g. *-na-* and *-ən-* generally plays as an infix in a reduplicative process of noun, pronoun, adjective and adverb. The same can be seen in the reduplicative process of a past participle verb e.g. *ḍekha-ən-ḍekha* 'not well seen', with a sense of indefiniteness as per Bashir and Conners (2021). Note that base and reduplicant ending with *-a* and *-i* refer to masculine and feminine form respectively. Gender and number marking shown by adjective and verb also is a feature of Indo-Aryan languages. Semantics produced in the verbal reduplication is definiteness in the absence of a negative marker but indefiniteness in its presence. Without its presence, the verbal reduplication may or may not exist, e.g. *ḍekha ḍekha* is possible and expresses some familiarity for something but **kəhiːkəhi* is impractical and does not exist.

Past participle verb is the only form that may take infix. The other two verbal forms i.e. imperfective and lexical verbs do not require it. Verbal reduplication in the three distinctive forms is interesting semantically. It is highlighted in the next section.

3.2   Reduplication of verbs: PP, lexical & imperfective forms

Reduplicative process takes place in the three forms of verbs i.e. past participle, lexical, and imperfective verbs as well as a combination of intransitive and transitive verb. There are however semantic variations in the three forms. The reduplication of past participle verb with or without an infix is already discussed much. Therefore, the focus of this section is on the reduplication of lexical and imperfective verbs. The role of converb is important.

In contrast with a negative marker- used as an infix, a converb *kər* or particle *ke* functions as an adverbial suffix with the reduplication of a lexical verb (20b-d) and thus derives adverbs of manner. A reduplication of lexical verb is entirely different

from the reduplication of a past participle verb or that of an imperfective verb. There are interesting semantic features in the three different cases. Consider the following.

(17)

a.	soæ 'slept'	+	soæ	=	soæsoæ		'while sleeping'	
	pp.V		Redup.		R		(Past Participle)	
b.	ro 'weep/cry'	+	ro + kər	=	ro ro kər		'by crying'	
	lex V		Redup.converb		R		(LV+Converb)	
c.	sʊn 'listen/hear'	+	sʊn kər	=	sʊn sʊn kər		'by listening'	
	lex.V		Redup.converb		R		(LV+Converb)	
d.	mər 'die'	+	mər kər	=	mər mər kər		'hardly'	
	lex.V		Redup.converb		R		(LV+Converb)	
e.	sʊnte	+	sʊnte	=	sʊnte sʊnte		'while listening'	
	imp.V		Redup.		R		(Imp Verb)	
f.	bhagt̪e	+	bhagt̪e	=	bhagt̪e bhagt̪e		'while running'	
	imp.V		Redup.		R		(Imp Verb)	

A past participle form is a deverbal adjective. Generally, a reduplication of a past participle verb expresses adjectival sense. However, *soæsoæ* (Montaut, 2008) is modifying verb and so it is functioning as an adverb here. The reduplication of a lexical verb along with converb *kər* expresses cause of an event. But the reduplication of an imperfective verb exhibits simultaneity or duration in which main event occurs. There are sometimes interesting semantic features in the latter two cases. Bashir and Conners (2021) agree to the adjectival meaning, but they note that a past participle verb can also give cause and effect meaning depending on the verb expressing the main event. Urdu-Hindi is *bæt^hæbæt^hæ* expresses the causal meaning for the resulting event. However, this meaning with a participle verb is not frequent, as compared to one by the combination of a converb with a lexical verb reduplication, in which the verb or its reduplication alone has no role for causal meaning. So, *k^ha k^ha kər* 'by eating' in Urdu-Hindi reflects causal meaning for a resulting event expressed by another verb.

The reduplication of a lexical form does not give any sense without the insertion of the converbal *kər*. There may be various interpretations in such reduplication, although the dominant one is cause and effect or manner of happening e.g. *t̪əʈʌp t̪əʈʌp kər məra* 'died in agony'. The semantics of the converb however depends on the verb with whose reduplication it attaches to, thus giving the sense of cause, in one case, tool in another and manner in the other case e.g. *mar mar kər* 'hardly/very delayed'. It is similar to the Kharia oblique marker *te*, which also marks the adverbial function of the secondary predicate forming *masdar-* as noted by Peterson and Maas (2009), when the reduplication of lexical base e.g. *ter-ter* '(the act of) giving'

does not do the job.

There are semantic variations specifically seen in the examples where negative markers and converb are used. Negative markers and converb give various interpretations. Bashir and Conners (2021) note that Punjabi, Saraki and Hindko, exhibit a wide range of reduplicative processes of noun, adjective, verb, adverb to convey manners of speaking, stylistic and expressive uses.

This is also interesting to note that a verbal reduplication in imperfective and past participle forms is frequent both in nominative and oblique forms of verbs e.g. *voh phəl k^{h}aṭa k^{h}aṭa aja* 'He came while eating fruit' and *voh phəl k^{h}aṭe k^{h}aṭe bola* 'He spoke while eating fruit'. However, this is dependent on the use of an argument. The reduplication of an oblique imperfective verb links it with the event expressed with the main verb. There is no link between the reduplication of a direct case imperfective verb and the verb expressing the main event. It only focuses on the attention of an agent. The difference between the reduplication of lexical verb requiring converb *kər* and imperfective verb is yet to be seen. It is highlighted in the next section.

3.3   Place of converb in the reduplication of lexical verb

Montaut (2008: 28) does not differentiate between the semantics expressed by the reduplication of a lexical verb, which essentially requires the converb *kər* 'do', and the reduplication of an imperfective verb which shows simultaneity. Consider the following.

(18)   a.   wania   bhag bhag-kər ṭhək gəi
             Wania.NOM run   RDP-CV tired go.PERF
             'Wania got tired because of (the continuous) running.'
             (kər shows cause of resulting event)

       b.   wania   bhag-ṭe bhagṭe ṭhək gəi
             Wania.NOM run-IMP RDP   tired go.PERF
             'Wania got tired because of (the continuous) running.'
             (Only simultaneity connects the two events)

The converb *kər* (21a) functions as an adverbial. It shows the event of continuous running *bhag bhag* as the cause of the resulting event *ṭhək gəi* 'got tired'. It is obligatory for the semantic fulfilment with the reduplication of a lexical verb.

No relation between the events of running and being tired in (21b) are seen. Only simultaneity links them.

The converb *kər* is however not restricted to showing cause of an event. There may be further interpretations. The examples like *ro ro-kər* 'by crying/after expressing pain' and *mər mər-kər* 'hardly' invite even more questions concerning various such interpretations, as the first *ro ro-kər* shows the high degree of sorrow and the second *mər mər-kər* expresses high degree of struggle. Three different interpretations of converb with the reduplication of three different lexical verbs force us to see the hidden semantics.

3.4   General and specific meanings: Revisiting Montaut (2008)

After looking at morpho-phonological and semantic features generally seen in the languages, we can now see some specific semantics in Urdu-Hindi reduplication. Montaut (2008) present some examples which apparently don't show any distinction of semantics. Consider the examples below:

(19)   a.   soye-soye mar gayâ ??? soye mar gayâ
           slept-slept die went   slept die went
           He died in his sleep/ In his sleep, he died   (Montaut, 2008)

       b.   tumhârî shikâyat sunte-sunte   (sun-sunkar) main ûb   gayâ thâ
            your   complaint hearing-hearing (-CP)   I   bore go   ppft
            I was fed up listening to your complaint   (Montaut, 2008)

       c.   mər 'die' + mər + kər (converb) → mər mər kər 'hardly'

The verb to 'sleep' is used in its past participle form. A past participle verb and its reduplication function as an adjective. The reduplication *soye soye* 'while sleeping' shows, another event *mɑr gɑyâa* 'died' occurring during the event of sleeping. There is no link between the two events.

The two forms *sunte-sunte* and *sun-sun kɑr/kər* in (22b) are also in fact two different semantic representations, which (Montaut, 2008) do not consider. A reduplication of an imperfective verb e.g. *sunte-sunte (sun̪tæ-sun̪tæ)* 'while consistent listening' actually shows simultaneity. The imperfective form further expresses that there is no specific duration of time spent in an action, e.g. listening in this case. So, the event of boring occurs during the unspecified duration of event of listening.

Further, to say, the converb *kər* generally expresses cause of something to happen. However, example (22c) shows no such meaning. It expresses the manner of doing something. The reduplication conveys a specific semantics of hardly. This is not possible without the combination of converb with merely the verb i.e. *mər kər* 'by dying', which just expresses lexical semantics i.e. something done by dying. The reduplication conveys a general sense of emphasis; it also carries a specific meaning, as in this case.

The reduplication of lexical verb essentially requires converb *kər* or the particle *kæ*, which gives various interpretations e.g. cause in this case *bhag bhag kər/ke* 'because of excessive running'. The reduplication of imperfective verb dominantly shows simultaneity, which means an event expressed by a main verb occurs during another event expressed by reduplicative verb e.g. *bhagʈe bhagʈe kəha* 'said during running'. Semantic difference in the verbs from other IA languages supports the arguments.

In addition to the reduplication of lexical verb or imperfective verb, some intransitive and transitive verbs together produce some sense of reduplication, e.g. in Potwohari *rədʒ rədʒa ke* 'after eating fully'. This combination also takes a converb *kər* or the particle *ke*, sometimes. However, this is not obligatory. All Indo Aryan languages generally show this combination with semantic difference.

3.5 Reduplication of transitive and intransitive verbs

The reduplication of an intransitive verb shows the reduplicant as its modified transitive form e.g. *tʃhʊp tʃhʊpa* 'hiding'. Except for such an example, no other reduplication has distinctive meaningful base-reduplicant combination, whether produced by phonetic alternation of a vowel in the base, or by insertion or by substitution of a consonant. The reduplicant *tʃhʊpa* 'cover' is a distinctive verb with transitive meaning, it is actually the modified form of the base *tʃhʊp* 'hide', an intransitive verb. This is only minutely highlighted here, looking at the length of the work already. A debate is needed in this regard. A reduplicant is always dependent on its base. It is also a process of creating new meaning. Therefore, the examples like *tʃhʊp tʃhʊpa* 'hiding' should be discussed in detail.

Some compounds have two synonymous constituents e.g. *kuːʈ fəræb* 'lies' in Saraiki and other languages. Terming them as doublets, Bashir and Conners (2021) see semantic reduplication in which two synonyms or synonym variants appear together in a quasi-compounding process. However, such examples are only hybrid

compounds. They are not reduplications. Hybrid compounds are the combinations of two synonyms coming from two different sources e.g. *kuːʈ* 'lie' (Sindhi) and *fəræb* 'deception' (Persian).

4   Reduplication in other Indo-Aryan languages

This section will have a glance at reduplication in some major Indo Aryan languages, in Pakistan, which include Marwari, Punjabi, Sindhi, Baluchi and Pashto. These languages commonly having Sanskrit as mother language do show a large number of lexical items in common with phonetic differences. Further, they also share a large number of loanwords from Arabic and Persian. Therefore, the same three features of Urdu-Hindi reduplication are seen i.e. a) morpho-phonological and semantic change b) the role of negative markers as infixes c) semantic difference between the reduplication of lexical verb requiring converb and the reduplication of an imperfective verb mainly showing simultaneity. Some evidence against Montaut (2008; 29) is available in these languages.

Among the examples of morpho-phonological changes, let us begin with *ʈʰa ʈʰa* which is an onomatopoetic expression to convey the gesture of celebration and is common in Punjabi, Marwaʈi, Sindhi Saraiki etc. Similarly, *vah vah* 'great' is another expression commonly used to appreciate somebody for having done something great.

First to see is the morpho-phonological and semantic change. Punjabi and Pothowari *kʊkkəʈ bəkkəʈ* 'chicken and the like' and *gʌɖɖi ʃəɖɖi* 'car and the like' can be good examples. The second example is used as *gʌɖɖi vəɖɖi* in Sindhi but *gaɖi vaɖi* in Marvaʈi. The key difference in the use of vowel between Marvaʈi and Punjabi is that Marvaʈi tends to use long a as compared to Punjabi which tends to use schwa -ə-, which is also frequent in Sindhi. Otherwise, many things are the same in the two languages. Note also that there is no doubling of voiced retroflex -ɖ- in Marvaʈi. Baluchi *tʃãh d̪ãh* 'tea and the like' is used as *tʃãh vãh* in Saraiki. Note that dental -ɖ- is strange as compared to labiodental -v-, which is common in IA languages. This is not an exemption, but Baluchi actually tends to show some phonetic contrast as compared to other languages e.g. *sʌfaːï d̪ʌfaːï* 'cleanliness and the like' and *kɪraːo d̪ɪraːo* 'rent and the like'

Besides, the phonetic changes, the reduplicant appearing before the base is visible in all other Indo Aryan languages e.g. *kəlləm kəlla* 'very lonely' in Punjabi

and Saraiki, but *khʊlləm khʊlla* in Marvaɽi.

All Indo Aryan languages use infixes in between base and reduplicant. Conjunctive particle *-o-* in the reduplication of noun or an adjective to convey distributive or emphatic sense e.g. *sah-o-sah* 'out of breath' in Sindhi, Punjabi, Sariki and Marvaɽi is commonly visible. The two infixes i.e. conjunctive particle *-o-* or genitive marker *-e-* are not the necessary part of reduplication, but they do have some place in the reduplicative process in the languages other than Urdu-Hindi. Therefore, many other examples like *maːl-o-maːl*, 'very rich' and *raṭ-õ-raṭ* 'during night' are present in Marvaɽi, Sindhi and Saraiki also. The nasal infix *-õ-* is only used in the reduplication of oblique nouns. It is not very frequent, though, it can be seen as the case above.

Like that in Urdu-Hindi, the presence of infixes in between base and reduplicant can be in the form of negative markers as well. The semantic expressions are also not different. There are also manners of speaking, stylistic and expressive use e.g. *na* in the reduplication of Punjabi adverb of time *kədi kədi* 'sometimes' → *kədi na kədi* 'at least sometimes'. With a minute phonetic change the adverb is used in Marvaɽi as *kədei kədei* 'sometimes' → *kədei na kədei* 'at least sometimes'. Note that the same Persian negative marker *-na-* is a frequent infix of other IA reduplications. Bashir and Conners (2021) suggest that in such a case of reduplication, with a negative marker, the meaning becomes indefinite. There are Sanskrit based negative markers also e.g. *pʊṭ-k-pʊṭ* 'obedient/disobedient son' in Urdu-Hindi with a contrastive meaning is used the same in Marvaɽi but becomes *pʊṭṭər-k-pʊṭṭər* in Punjabi.

Finally, converb *ke* replacing Urdu-Hindi *kər* is common and frequent with lexical verbs in other Indo Aryan languages, e.g. Punjabi and Marwarhi, with different interpretations depending on the verbs it is used with. Like Urdu *kʰa* 'eat', Punjabi, Marwarhi and Sindhi *kʰa* 'eat' is also used as imperative and past participle verbs. The expression *kʰa kʰa ke* 'by eating' in Punjabi reflects causal meaning for a resulting event expressed by another verb. The reduplication of lexical verb *mər* 'die' requires the coverbial particle *ke* i.e. *mər mər ke* 'hardly', which is the same as that in Urdu-Hindi. The reduplication of other lexical verbs in combination with the converbial *ke* e.g. *bhag bhag ke* 'by running' (Punjabi), *ro ro ke* 'by crying' (Punjabi/Saraiki/Pothwari) and *kɪl kɪl ke* 'forcefully' (Pothwari) very often show manner of an action, not necessarily cause of an event. Bashir and Conners (2021) note that the adjectival meaning of a past participle verb is frequent; it can also give cause and effect meaning depending on the verb expressing the main event e.g. in Punjabi *o bæh bæh ke təŋ pæh gəja* 'he got tired of sitting'. The converbial *ke* is the same as Urdu-Hindi *ke* semantically and in fact a substitute of the converb *kər*. Therefore, it reflects causal meaning for the resulting event *təŋ pæh gəja* '(he)

got tired' in the present case. The light verb final ending -*a* in *gəja* is expressing masculine form of the pronoun *o*.

Although the reduplication of an intransitive verb with its transitive reduplicant e.g. *rədʒ + rədʒa* 'to make stomach full' → *rədʒ + rədʒa* ke 'to have eaten much' in Punjabi, Marwarhi, Pothwari, Saraiki and other IA languages is also frequent, it is not the focus of this study. Therefore, it is only proposed for future researchers.

5 Conclusion

The study has tried to discuss reduplicative structures in Urdu-Hindi and other Indo Aryan languages. There are three points which have been the focus: a) Some morpho-phonological aspects like insertion, substitution and deletion have been seen. b) The role of negative markers in Urdu-Hindi reduplication was the second important point seen. The reduplication of Urdu past participles, for example, *ḍekha əṇḍekha* 'not well seen' as an adjective with the semantics of indefiniteness highlights a different view against the interpretation of iteration and adverbial senses. The insertion of a negative marker is itself important to see, as there are other negative markers also discussed. c) The role of converb in the reduplication of lexical verb is the third important point previously not discussed. Montaut (2008: 28) does not differentiate between the semantics expressed by the reduplication of a lexical verb and that of an imperfective verb. The reduplication of lexical verb requires the converb *kər* 'do' that shows a direct cause and effect relation between the two events expressed by the reduplicated verb and the main verb.

The research is beneficial to Pakistan both theoretically and practically. It will make an important contribution through its descriptive approach to the theoretical linguistics research in Pakistan, thus benefiting research for many other Pakistani languages. Being an aspect of morphological research, it will also be helpful in the areas like computational linguistics and machine translation.

References

Abbi, A. (1985). Reduplicative structures: A phenomenon of the South Asian linguistic area. *Oceanic Linguistics Special Publications*, *20*, 159–171.

Abbi, A. (1992). *Reduplication in South Asian languages: An areal, typological, and historical study*. Allied Publishers Pvt. Ltd, India.

Alderete, J., Beckman, J., Benua, L., Gnanadesikan, A., McCarthy, J., & Urbanczyk, S. (1999). *Linguistic Inquiry*, *30*(3), 327–364.

Bashir, E., & Conners, T. J. (2019). *A descriptive grammar of Hindko, Panjabi, and Saraiki* (Vol. 4). De Gruyter Mouton.

Bauer, L. (2003). *Introducing linguistic morphology*. Edinburgh University Press.

Bögel, T., Butt, M., Hautli, A., & Sulger, S. (2008). Developing a finite-state morphological analyzer for Urdu and Hindi. In T. Hanneforth & K.-M. Würzner (Eds.), *Finite-state methods and natural language processing* (pp. 86–96). Potsdam University Press.

Bresnan, J. (2001). *Lexical-functional syntax*. Wiley.

Bresnan, J., & Nikitina, T. (2003). On the gradience of the dative alternation. *Stanford University*.

Butt, M. (2010). Tense and aspect in Urdu.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Durani, N. (2007). *Typology of word and automatic word segmentation in urdu text corpus* (Master's thesis). National University of Computer and Emerging Sciences. Lahore, Pakistan.

(ed.), J. B. (1982). *The mental representation of grammatical relations*. MIT Press.

Ghaniabadi, S. (2005). M/p-echo reduplication in colloquial Persian: An optimality theory analysis. *Aspects of Iranian Linguistics*.

Golston, C., & Thurgood, E. (2003). Reduplication as echo: Evidence from Bontok and Chumash. In R. Laury & P. J. Mistry (Eds.), *Perspectives in linguistics: Papers in honor of p.j. mistry*. Indian Institute of Language Studies.

Hannahs, S. (2004). Malagasy infixing reduplication. *Durham Working Papers in Linguistics*.

Harrison, K. D. (2000). *Topics in the phonology and morphology of Tuvan*. Yale University.

Harrison, S. (1973). Reduplication in micronesian languages. *Oceanic Linguistics*, *12*(1/2), 407–454.

Harrison, S. (1976). *Mokilese reference grammar*. The University Press of Hawai'i.

Hurch, B. (2005). *Studies on reduplication* (Vol. 28). De Gruyter.

Inkelas, S. (2005). Morphological doubling theory: Evidence for morphological doubling in reduplication. In B. Hurch (Ed.), *Studies on reduplication* (pp. 65–88). De Gruyter.

Inkelas, S. (2014). Non-concatenative derivation: Reduplication. In *The oxford handbook of derivational morphology* (pp. 169–189). Oxford University Press.

Inkelas, S., & Downing, L. J. (2015). What is reduplication? typology and analysis part 1/2: The typology of reduplication. *Language and linguistics compass*, *9*(12), 502–515.

Kachru, Y. (1990). Hindi-Urdu. In *The major languages of South Asia, the Middle East and Africa* (2nd ed.). Routledge.

Katre, S. M. (1939). Apropos epic iyat. *Bulletin of the Deccan College Post-Graduate and Research Institute*.

Keane, E. (2005). Phrasal reduplication and dual description. In *Studies on reduplication* (pp. 239–261). De Gruyter.

Keenan, E., & Polinsky, M. (n.d.). Malagasy morphology. In A. Spencer & A. Zwicky (Eds.), *The handbook of morphology* (pp. 563–624). Wiley.

Khan, M. K. (2012). *Pashto phonology: The relationship between syllable structure and word order* (Doctoral dissertation). University of Azad Jammu and Kashmir. Muzaffarabad.

Lu, M. (2017). *A study of reduplication in old Chinese with reference to the Erya* (Doctoral dissertation). Lieden University. Leiden.

Mangrio, R. A., & Ali, G. (2013). Functions of infixes in the Urdu endocentric compounds. *Kashmir Journal of Language Research*, *16*(1).

McCarthy, J. (1982). Prosodic structure and expletive infixation. *Language*, *58*(3), 574–590.

McCarthy, J. (2002). *A thematic guide to optimality theory*. Cambridge University Press.

McCarthy, J., & Prince, A. (2001). *Prosodic morphology: Constraint interaction and satisfaction* (Vol. 3). Rutgers University Center for Cognitive Science, RuCCS Technical Report.

Montaut, A. (2008). Reduplication and echo-words in Hindi. In *Annual review of south asian languages and linguistics* (pp. 21–62). Mouton De Gruyter.

Naim, C. M. (1999). *Introductory urdu* (Vol. 1). South Asia Language & Area Center, University of Chicago.

Nintemann, J. (2016). *A comparative analysis of forms and functions of reduplication in Kenyan languages* (Doctoral dissertation). Universität Bremen.

Noonan, M. (1992). *A grammar of Lango*. De Gruyter.

Olsson, L. (2015). *Form and function of reduplicated nouns in japanese* (Master's thesis). Stockholm University. Stockholm.

Peterson, J., & Maas, U. (2009). Reduplication in Kharia: The masdar as a phonologically motivated category. In B. Hurch & V. Mattes (Eds.), *Reduplication: Diachrony and productivity* (pp. 207–237). Springer.

A plausible history for Micronesian possessive classifiers. (1988). *Oceanic Linguistics*, *27*(1/2), 63–78.

Prince, A., & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar* (Vol. 2). Rutgers University Center for Cognitive Science, RuCCS Technical Report.

Ramasamy, M. D. (2011). *Topics in the morphophonology of Standard Spoken Tamil (SST): An optimality theoretic study* (Doctoral dissertation). Newcastle University. Newcastle upon Tyne.

65

Rubino, C. (2011). Reduplication: Form, function and distribution. In B. Hurch (Ed.), *Grammar, comparative and general-reduplication* (pp. 11–29). Mouton De Gruyter.

Schmidt, R. L. (1999). *Urdu: An essential grammar*. Routledge.

Singh, R. (2005). Reduplication in modern hindi and the theory of reduplication. In *Studies on reduplication* (pp. 263–282). De Gruyter.

Thurgood, E. (1997). Bontok reduplication and prosodic templates. *Oceanic Linguistics*, *36*(1), 35–148.

Zoll, S. I. C. (2005). *Reduplication: Doubling in morphology* (Vol. 106). Cambridge University Press.

Bien / Bem bonita' o 'muy / *muito* bonita Flavors of intensity and emphasis in Spanish and Portuguese

Piero Visconte
The University of Texas at Austin
`visconte.piero@utexas.edu`

Abstract

With the evolution from Latin to Romance languages, both synthetic (-*issimus, -errimus*) and analytical (*muy/*muito, *bien/bem*) formulas have been implemented to express the degree of the adjective (Wang, 2013). This paper focuses on the process of grammaticalization of the intensifier *bien/bem* + *adjective* in Caribbean Spanish (CS) and Brazilian Portuguese (BP), proposing a quantitative analysis to show that these intensifiers behave as extreme degree quantifiers in the spontaneous speech data of these two varieties. The use of *bien/bem* in these vernaculars is particularly relevant for two reasons: syntactically, their value change from adverbial to intensifying (Pastor & Armstrong 2016; Lívio and Howe 2020); semantically, they "boost" the meaning of the adjectives that follow (Tagliamonte, 2008, p. 361), although they may also retain their original meaning of manner adverbs (Kanwit et al., 2017). All the data collected in Puerto Rico is couched in Bybee (2011) "Usage-Based Theory". This theoretical framework incorporates the basic insight that usage influences linguistic structure (Bybee and Beckner 2015). Participants are Dominican Spanish-speakers (DS) (2), Cuban Spanish-speakers (CuS) (2), Puerto Rican Spanish-speaker (PRS) (2), and Brazil Portuguese (BP) (6), and have been living in Puerto Rico for ten years. Findings show a higher use of the intensifier *bien/bem* in the speakers of both varieties. Moreover, there are contradicting existing statements on the use of intensifiers related to gender and educational level (Madero 1983; Salvador 1987; Arjona 1991; Ito and Tagliamonte 2003). Since the intensifiers *bien/bem* are typical of spontaneous oral production, this paper suggests that subjectivity is a fundamental factor to consider. As stated by Athanasiadou (2007, p. 554), intensification is a "means of indexing the speakers' perspective", which could trigger interesting cases of language accommodation by Brazilians due to the Spanish-speaking environment in which they live.

1 Introduction

Intensifiers constitute a long and exciting topic in linguistics (Pahta 2006 for a cross-linguistically overview). They may be defined as "any device that scales a quality, whether up or down or somewhere between the two" (Bolinger 1972: 17). Bolinger (1972) classified these degree adverbs into four categories: boosters ('very nice'), compromisers ('kind of interesting'), diminishers ('not so / that / too big'), and minimizers ('not very common'). However, as argued by Quirk et al (1972: 439), "variation in the use of intensifiers and their gradient nature make any attempt to establish clear-cut divisions ill-advised".

In Latin, the expression of superlation used to be done through the morphological strategy *-issimus/-a/-um*, although it came to coexist with analytical constructions, in which the forms *maxime* or *plurimum + adjective* were introduced (Rohlfs 1968; Bruyne 1980, 1986; Pinkster 1990; Ledgeway 2012). The main difference between the two strategies was that the synthetic one served for absolute superlation, while the analytical one was more productive for relative superlation (Lara-Bermejo 2016).

The evolution to Romance languages caused the rise of analytical constructions with *multus*, which did not regain the synthetic form until the 16th century, thanks to the influence of Italian (Lara-Bermejo 2016). *multus* suffered two types of palatalizations: the first one *multus > muito > mucho > much*, in which Lara-Bermejo (2016: 226) speaks of the "época afrancesada de la apócope extrema" ('Frenchified era of the extreme apocope'); and the second one *multus > muito > muit > mui / muy*, in which the same diphthong was produced in both Spanish and Portuguese (Alvar and Pottier 1987). Indeed, according to some scholars (Coromines and Pascual 1980; Serradilla 2005), *mucho* coexisted with *muito* in pre-literary Castilian Spanish.

The form *multus + adjective* was considered the cultured one, as opposed to *bene + adjective* (*"bene fortis"*) (Serradilla 2005: 367), an oral and colloquial form used to imitate and parody the uneducated speech of lower-middle class (Serradilla 2006). Despite reflecting the social status of the speakers (Serradilla 2006), the two forms coexisted, overlapped, and began to be interchanged (Serradilla 2007; Wang 2013).

Following this pattern of discourse-related linguistic changes in the development of intensifiers (Lívio and Howe 2020), this paper focuses on the process of grammaticalization of the intensifier *bien/bem + adjective* in Caribbean Spanish (CS) and Brazilian Portuguese (BP)– two Romance varieties whose structural features have been considered related on several occasions (Holm et al. 1999; Guy 2017)–as

68

in (1) and (2).

(1) Mi nueva casa es *bien bonita.*   (PRS M1)
     My new house is *well/so pretty.*

(2) Acho que esta paisagem é bem bonita.   (BP F3)
     I think this landscape is well/so beautiful.

In examples (1) and (2), *bien/bem* are used to "boost" (Tagliamonte 2008: 361) the meaning of the adjectives that follow (bonita, pretty/beautiful). In accordance with some scholars (Cunha and Cintra 2001; Salazar-García 2008), when standing in front of adjectives that carry a gradient effect, intensifiers *bien/bem* act in a similar way to *muy/*muito ('very') or *tan/tão* ('so'), as illustrated in (3a-b) and (4a-b). What makes them unique is their load of "flavors of intensity and emphasis" (Sanchez-Mendes 2021) that gives them a pragmatically enhanced interpretation.

(3) a. Está *muy* elaborado, pero repito varias cosas (very sophisticated). (DS M1)
     b. No final do mês, estou sempre *muito* apressada (very rushed) (BP F1)

(4) a. Es que mi amiga es tan *presentá[da]* (so nosy)   (CuS F2)
     b. Passear de manhã me faz *tão feliz* (so happy)   (BP F1)

The variability of intensifiers makes them ideal candidates for the study of language change in progress, including processes of grammaticalization and the resulting semantic layering (Hopper & Traugott 2003: 122; Ito and Tagliamonte 2003: 277). As stated by Tagliamonte (2012: 88), "the mechanisms of grammaticalization offer diagnostics for interpreting trends in the evolution of forms". Among these mechanisms, the processes of *desemanticization* (loss of semantic content or *"semantic bleaching"*) and *decategorialization* (loss of morphosyntactic properties) stand out for their relevance in the study of intensifiers (Hopper and Traugott 2003; Kanwit et al. 2017; Tagliamonte 2008). In fact, as argued by Hopper and Traugott (2003), the forms of grammaticalization not only arise in the passage from a lexical word to a functional one, but also in the reinterpretation of a functional one to another of the same category.

This is the case of the evolution of *bien/bem*, which have gradually lost their former semantic connotation (*desemanticization*) and have even getting to be interchanged with the cultured form multu + adjective (*decategorialization*) (Gómez-Torrego 2000; Serradilla 2008), reflecting the degree of interaction that a speech community establishes with a particular form.

Both Hopper (1991) and Tagliamonte (2008) suppose the existence of a phase in which there was an overlap of intensifying forms (*very/well* in English). The more grammaticalized an intensifier becomes over time, the broader its range of placement will be, due to the *desemanticization* and *decategorialization* processes that nowadays lead to placing *bien/bem* with adjectives of both positive (*bueno*, 'good') and negative (*difíceis*, 'hard') evaluation in Spanish and Portuguese varieties (Gutiérrez-Rexach and González-Rivera 2014: 62-63), as in (5) and (6).

(5)   El concierto de Gilbertito Santa Rosa estuvo a otro nivel, *bien bueno*.(CuS F3)
       Gilbertito Santa Rosa's concert was at another level, so good.

(6)   Esses anos foram *bem difíceis*.   (BP F6)
       Those years were well hard.

The primary objective of this study is to provide a quantitative analysis to explain the use of *bien/bem* in the spontaneous speech data of CS and BP. On the one hand, these manner adverbs behave as intensifiers of extreme degree quantifier (*super, hiper, totalmente, etc.*, 'super, hyper, totally, etc.), after having undergone a process of grammaticalization (Gutiérrez-Rexach and González-Rivera 2014; Pastor and Armstrong 2016; Lívio and Howe 2020). On the other hand, *bien/bem* assume the function of intensifier in contrast to the canonical *muy/*muito ('very') (Tagliamonte 2008) as shown in (3) and (4), although they can also retain their original positive meaning when implemented with their original role of manner adverbs (Kanwit et al. 2017).

The paper is organized as follows: Section 2 offers a brief analysis of the background literature on intensifiers. Section 3 introduces the study, describing the data and the participants, as well as the methodology. In Section 4, the results are discussed. Finally, section 5 summarizes the main findings and provides the concluding remarks.

2 Background literature

Interest in intensifiers began over a century ago with the work of Stoffel (1901), Borst (1902), and Fettig (1934) on the English language. More recently, a number of scholars have focused their research on the rapid development and variation of intensifiers (Bolinger 1972; Lorenz 1999; Paradis 2000; Ito and Tagliamonte 2003; Serradilla-Castaño 2005, 2008; Tagliamonte 2008; González-Díaz 2008; Kanwit et al. 2017), exploring their structure within the framework of grammaticalization to analyze their sociolinguistic determinants, as well as their diatopic distribution (Macaulay 2006; Nevalainen 2008; Rissanen 2008).

For instance, the association between the variation and change in the use of intensifiers and the variables of age and gender has led Ito and Tagliamonte (2003) to note that *really* is replacing *very* in the younger generations of English speakers. As regards to Spanish, Serradilla-Castaño (2006) stated that *bien* is used more frequently in Latin American Spanish compared to Peninsular varieties. This author also added that the intensifier is more common in oral registers, in the speech of women, and it is used by the lower-middle class, especially with affirmative modality. In addition, genre has also been considered (i.e., Tagliamonte and Roberts 2005 on TV show *Friends*), as well as the dichotomy between educated and popular speech in Mexico (Madero-Kondrat 1983 on the educated speech in Mexico City; Salvador 1987 on the popular speech in Mexico City; Arjona 1990 on Mexican popular speech).

Regarding the diatopic distribution of intensifiers, the existing literature on the topic provides numerous works that span from Canadian (Tagliamonte 2008) to Australian and New Zealand English (Sowa 2009). In Spanish, beyond the already cited works on the Mexican variety, the studies on intensifiers cover both the Peninsular variety (Serradilla-Castaño 2006, 2007; Wang 2013; Zieliński 2013; Pastor and Armstrong 2016) and the Latin American (Sedano 2002 on Venezuelan Spanish; Kanwit et al. 2017 on the comparison between Peninsular and Argentinean Spanish), as well as the Caribbean (Brown and Cortés-Torres 2013, González-Rivera and Gutiérrez-Rexach 2013, on Puerto Rican Spanish).

Other than Spanish, several scholars (see Salazar-García 2008: 717-718 for an overview) in Romance linguistics have examined polysemy and/or grammaticalization of intensifiers in Italian (*tutto* and *assolutamente*) (De Cesare 2003) and in French (*tout + adjective*) (Anscombre 2009), as well as in Portuguese (*bem/muito/todo*) (Quadros-Gomes and Sanchez-Mendes 2015). Regarding the latter, the analysis of linguistic corpora by Lívio and Howe (2020) deserves a spe-

cial mention for providing a complete and innovative account of the behavior of intensifiers in Portuguese, both within and between dialects (Angolan, Brazilian, European, and Mozambican varieties). Lastly, other proposals have studied degree adverbs in other languages, such as Dutch (Klein 1998) and Japanese (Tsujimura 2001), as well as Afrikaans (Berghoff et al. 2020).

With respect to Spanish and Portuguese, in recent years the syntactic and semantic aspects of intensifier variation have catalyzed the focus of analysis on the category of intensifiers (Alves 2006; Gomes 2011; Gutiérrez-Rexach & González-Rivera 2014; Pastor & Armstrong 2016).

3   The present study

To measure the variation of intensifiers *bien/bem* in CS and BP, this study is based on the analysis of a corpus of twelve sociolinguistic interviews carried out between 2018 and 2019 in the metropolitan area of the Puerto Rican capital city, San Juan. These are spontaneous recordings in both formal (university campus) and informal settings (bars, houses, parks). Participants are Dominican Spanish-speakers (DS) (2), Cuban Spanish-speakers (CuS) (2), Puerto Rican Spanish-speaker (PRS) (2), and Brazil Portuguese (BP) (6), and have been living in Puerto Rico for ten years (Table 1). In addition to the degree of formality of the conversation, gender, age, and level of education have been taken into consideration. Therefore, every vernacular has an equal number of male and female participants, for a total of six men and six women ranging from 18 to 48 years old.

The data represents roughly six hours of recordings taken from twelve separate interviews and it is couched in Bybee's (2011) "Usage-Based Theory". As the name indicates, this theoretical framework incorporates the basic insight that usage has an effect on linguistic structure (Bybee and Beckner 2015). Being the intensifiers *bien/bem* typical of spontaneous oral production, they will favor the use of a structure rather than the competence of the speaker. Therefore, patterns, frequency of occurrence, variation, and change are all considered direct and valuable evidence about cognitive representation.

	CARRIBBEAN SPANISH (CS)			BRAZILIAN PORTUGUESE (BP)		
	PRS	DRS	CuS			
MALE SPEAKERS	M1, 20	M2, 25	M3, 36	M4, 18	M5, 43	M6, 48
FEMALE SPEAKERS	F1, 29	F2, 23	F3, 40	F4, 32	F5, 38	F6, 45

Table 1: Study participants (dialect variety, gender, age)

4   Results and discussion

This section illustrates the results of the data extracted from the sociolinguistic interviews in CS and BP and it is developed in two main parts. First, both the cross-dialectal and the intra-dialectal distributions of *bien/bem* are measured, as well as a number of variables (informants' gender, age, and level of education, as well as the degree of formality of the spontaneous speech) to show the frequency rates of *bien/bem* in front of adjectives with a function of intensifier. Second, to show the common path of grammaticalization intensifier *bien/bem* are going through, the top five most frequent collocations of *bien/bem* for the two varieties under study have been extracted and analyzed, coming after Evison's (2010) comparison of the adjective categories.

As can be seen from Figure 1, the total amount of tokens extracted from the interviews and relative to the intensifiers *bien/bem* and *muy/muito* reflects similar cross-dialectal results in both CS (216) and BP (212). Specifically, while the Lusophones interviewed show a balanced alternation in the use of the intensifiers *bem* (41%) and *muito* (59%) followed by an adjective, the Hispanophones tend to express intensification of the degree of adjective through *bien* (67%) more than with *muy* (33%).

Additionally, analyzing the three varieties of the Hispanic Caribbean, a clear preference for the *bien + adjective* form is noted both in Puerto Ricans (75%–25%) and Cubans (67% - 33%), while Dominicans present a balance between the use of the two intensifying forms (Figure 2).

Concerning the gender variable, findings indicate that female speakers use the intensifier *bien* more in PRS, CuS and BP, but not in DS (Figure 3). Further studies
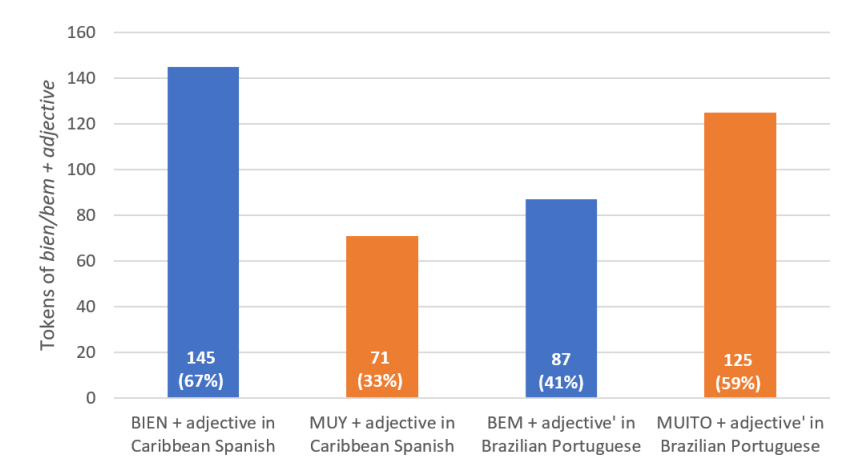
Figure 1: Cross-dialectal distribution of '*bien/bem*' and '*muy/muito*' + *adjective* in Caribbean Spanish and Brazilian Portuguese.
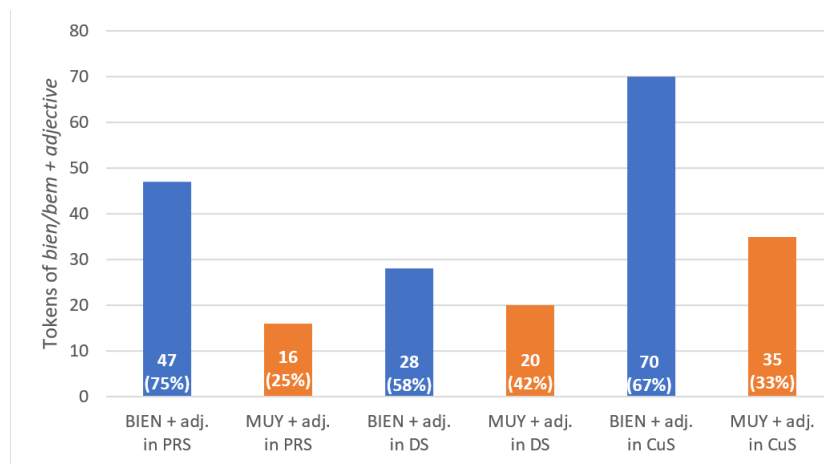

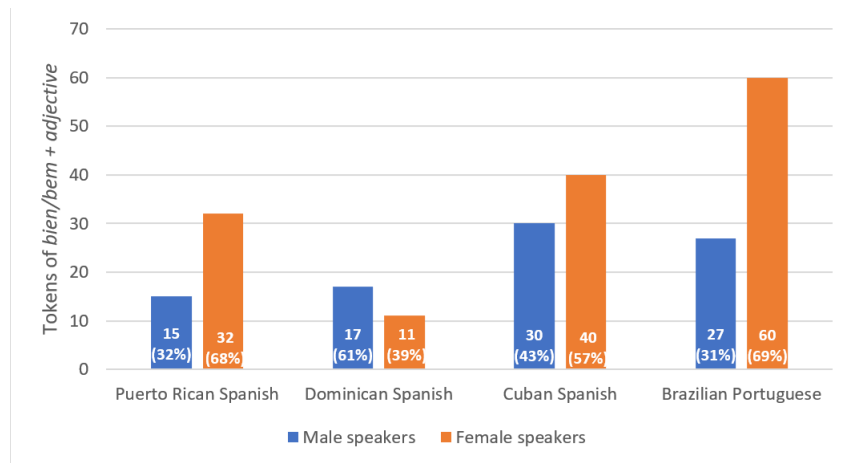Figure 2: Intra-dialectal distribution of 'bien + adjective' in Caribbean Spanish.

Figure 3: Intensifier *bien/bem* + adjective according to the speakers' gender.

are needed on DS, including more participants to measure whether and to what extent the construction *bien + adjective* is increasing.

When taking into account other variables such as age, level of education, and use of *bien/bem* followed by an adjective, different results have been found (Figure 4).

Looking at the data, there is no obvious relationship between the age of the speaker and the use of the intensifiers analyzed. Likewise, no correlation has been reported between the use of intensifiers and some extralinguistic variables, such as the age and educational level of the speakers analyzed. Although a greater number of participants would undoubtedly offer more empirical results in the future, the results of this work challenge the findings of other studies, which have broadly labeled the structure *bien + adjective* as typical among members of the lower cultural level (Madero 1983; Salvador 1987; Arjona 1991; Ito and Tagliamonte 2003). Indeed, university students (M1, F1, M2, F2, M3, F3) were interviewed at the main campus of the University of Puerto Rico (Río Piedras campus) and they do not show a greater use of intensifiers *bien/bem*. On the other hand, as can be seen in Figure 4, two out of six older informants (M5 and M6, respectively 43 and 48 years old) showed less use of *bien/bem + adjective*, even having been interviewed in more informal places (their own houses, bars, and parks).

As mentioned above, the second part of this section focuses on the analysis of
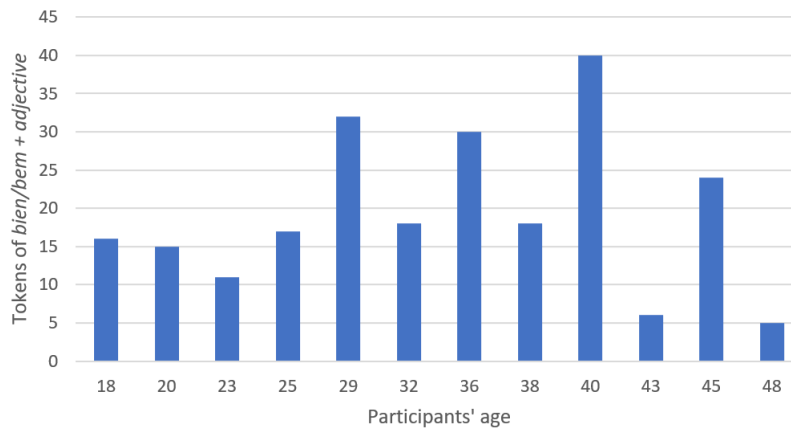
Figure 4: Age, level of education, and use of *bien/bem* + adjective.

the top five most frequent adjectives that follow *bien/bem* to show the common path of grammaticalization of these intensifiers. Following Evison's (2010) comparison of the adjective categories, the present study considers the significant associations between the intensifiers *bien/bem* and the evaluation of the modified adjective (positive, negative, and neutral) in the two vernacular languages taken into analysis.

As shown in Figure 5, CS and BP share a high number of adjectives (three out of five): *bueno* and *bom* ('good'), *lindo/bonito* ('nice'), *diferente* ('different'). Likewise, it can also be observed that in the five adjectives that co-occur in both vernaculars, *bien* is collocated with positive adjectives (*bueno/bom* and *lindo* or *bonito*) (7a-b), as well as negative (*difícil* or *complicado* and *ruim*) (8a-b) and neutral ones (*diferente*) (9a-b).

(7)   a. Las excursiones en Dominicana son **bien buenas**, tú verás.   (DS M2)
           Excursions in the Dominican Republic are really good, you'll see.

      b. Acho que a música daqueles anos é **bem boa**.   (BP F4)
           I find the music of those years really good.

(8)   a. Hacer los papeles [migratorios] en Italia tiene que ser **bien difícil**. (DS F1)
           Doing the immigration papers in Italy has to be way too difficult.

b. No começo, minha vida foi bem sacrificada, **bem ruim**. (BP F5)
In the beginning, my life was really sacrificed, extremely bad.

(9) a. Nosotros tenemos el fufú, pero es **bien diferente** del mofongo. (CuS F3) We have fufú, but it is well different from mofongo.

b. Aqui é **bem diferente**, ou assim os meus pais falam. (BP M4)
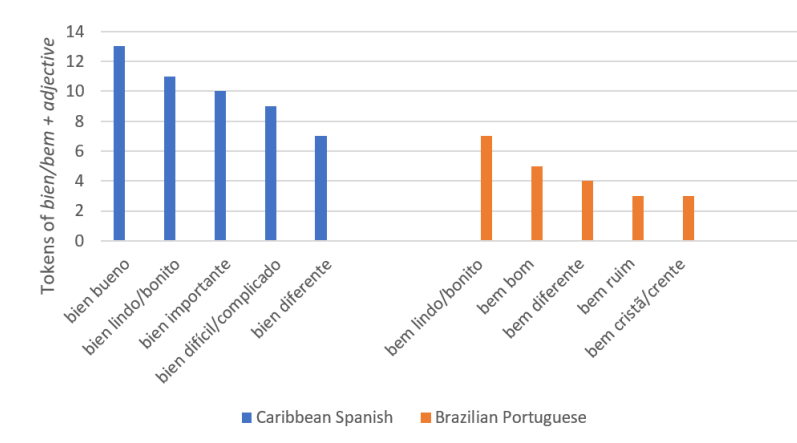Living here it is well different, or so my parents say.

Figure 5: Age, level of education, and use of *bien/bem* + adjective.

Moreover, it is interesting to note that some oral data provides examples of *bien/bem* before an adjective that is already in its superlative form, based on their semantics (10a-b).

(10) a. Los precios están bien buenos y dicen que la comida es bien exquisita. (CuS F3)
The prices are so good, and they say that the food is well exquisite.

b. Tenho uns filhos bem maravilhosos. (BP F6)
I have some really wonderful children.

This type of grammatical redundancy could be the result of the fact that it is an analysis of spontaneous conversations, therefore subjectivity could play a

fundamental role when measuring the frequency of the intensifiers. As stated by Athanasiadou (2007: 554), intensification is in fact a "means of indexing the speakers' perspective", which could explain the use of *bien/bem + superlatives*, as in (10a-b).

A further look in the transcriptions also shows that speakers do make use of other intensifying strategies in both varieties, such as the use of adjective reduplication (*bien bien, bien pero que bien, bem bem*) as in (11a-c):

(11)   a. "Fresa y chocolate" es bien buena... y *bien, bien cubana.*   (CuS F3)
          "Fresa y chocolate" is really good... and it is so so Cuban.

       b. Mi cumple este domingo pasado estuvo *bien pero que bien gufiau.* (PR M1)
          My birthday this past Sunday was so so good.

       c. O teste foi bem forte, ou seja, bem ruim... foi *bem bem ruim.* (BP M4)
          The test was so strong, I mean, so bad... it was so so bad.

5   Conclusion

This study has provided a quantitative analysis to discuss the grammaticalization path of *bien/bem* in the spontaneous speech of Caribbean Spanish (CS) and Brazilian Portuguese (BP), a topic that has been recently drawing attention of semanticists in recent years. At the syntactic level, these manner adverbs behave as intensifiers of extreme degree quantifier (*super, hiper, extremadamente, etc.,* 'super, hyper, extremely, etc.'), after having undergone a process of grammaticalization (Gutiérrez-Rexach and González-Rivera 2014; Pastor and Armstrong 2016; Lívio and Howe 2020). On the other hand, at a semantic level, *bien/bem* has assumed the function of intensifier in contrast to the canonical *muy/muito* ('very') (Tagliamonte 2008), but they can also retain its original positive meaning when implemented with its original role of manner adverbs (Kanwit et al. 2017).

Some scholars have argued that both periphrases *muy/muito + adjective* and *bien/bem + adjective* are well-established, with the latter widely preferred in Latin American and Caribbean varieties compared to the European varieties (Serradilla-Castaño 2006; Brown and Cortés-Torres 2013; Lívio and Howe 2020; etc.). This

assumption has been confirmed by this work, which corroborate the trend in Latin America and the Caribbean in the use of *bien/bem* as intensifiers, highlighting a balanced distribution between the two vernaculars analyzed (Table 2).

Other proposals have analyzed the sociolinguistic determinants behind the use of *bien/bem* as intensifiers, claiming that they are used more by women, as well as members of the lowest cultural level (Madero 1983; Salvador 1987; Arjona 1991). The present study challenges and contradicts these statements (Table 4 and Table 5), showing that the variation in the use of intensifiers is not inversely proportional to the age of the speakers. Similarly, this work does not show a clear difference in the informants depending on their gender, level of education or on the level of formality of the interviews conducted.

Lastly, this study suggests that the balanced alternation between *bien/bem* + *adjective* and *muy/muito* + *adjective* may be due to semantic nuances that can be traced back to the subjectivity of the speaker (Serradilla-Castaño 2008).

References

Anscombre, J. C. (2009). Des adverbes d'énonciation aux marqueurs d'attitude énonciative : Le cas de la construction "tout" + adjectif. *Langue Francaise*, *161*, 59–80.

Athanasiadou, A. (2007). On the subjectivity of intensifiers. *Language Sciences*, *29*, 554–565.

Berghoff, R., Nouwen, R., Bylinina, L., & McNabb, Y. (2020). Degree modification across categories in afrikaans. *Linguistic Variation*.

Bermejo, V. L. (2017). El superlativo absoluto en el español peninsular del siglo xx. *Rilce-revista De Filologia Hispanica*, *34*, 225–239.

Bolinger, D. (n.d.). Degree words.

Brown, E. L., & Cortés-Torres, M. (2013). Puerto rican intensifiers: Bien/muyvariables. *Selected proceedings of the 6th Workshop on Spanish Sociolinguistics*, 11–19.

Bybee, J. (2011). *Language, usage and cognition*. Cambridge University Press.

Bybee, J., & Beckner, C. (2009). Usage-based theory. In B. Heine & H. Narrog (Eds.), *The oxford handbook of linguistic analysis* 2*ed.*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199677078.013.0032

Castaño, A. S. (2006). El proceso de gramaticalización en las perifrasis de superlativo absoluto. *Actas del VI Congreso Internacional de Historia de la Lengua española: Madrid, 29 de septiembre-3 octubre 2003*, 1123–1134.

Castaño, A. S. (2008). Una diferencia sociolingüística en el uso de las fórmulas superlativas en español medieval. *Discurso y sociedad II: nuevas contribuciones al estudio de la lengua en un contexto social*, 597–609.

Castaño, S., & Maria, A. (2016). " bien"+" adjetivo" como perifrasis de superlativo en español: Particularidades semánticas y sintácticas.

Celso ferreira da cunha and luís filipe lindley cintra. (2011). *Nova gramática do português contemporâneo. 2. ed.*

Corominas, J. P., & Pascual, J. A. G. (1980). Diccionario crítico etimológico castellano e hispánico.

De Mello, H. R., Holm, J. A., & Lorenzino, G. A. (1999). Diferentes grados de reestructuración en dos lenguas vernáculas: El español caribeño y el portugués brasileño. *Diferentes grados de reestructuración en dos lenguas vernáculas: el español caribeño y el portugués brasileño*, 43–60.

de Bruyne, J. (1980). Acerca de la traducción de -ísimo. *LEA 2.1*, 27–37.

de Bruyne, J. (1986). Onomástica y elativos en '-ísimo'. *Anuario de Lingüística Hispánica*, 9–20.

de Cesare, A. M. (2003). Una funzione del tutto particolare, quella di assolutamente e simili : Structure linguistiques et fonctions communicatives. *Revue Romane*, *38*, 179–214.

Evison, J. (2010). What are the basics of analysing a corpus. In A. O'Keeffe & M. McCarthy (Eds.), *The routledge handbook of corpus linguistics* (pp. 122–35).

Gomes, A. P., & Mendes, L. S. (2015). Degree modification in brazilian portuguese and in karitiana. *Revista Virtual de Estudos da Linguagem*, *13*, 5–32.

González-Díaz, V. (2008). Recent developments in english intensifiers: The case of very much. *English Language and Linguistics*, *12*, 221–243. https://doi.org/10.1017/S1360674308002608

Gutiérrez-Rexach, J., & Gonzalez-Rivera, M. (2014). Adverbial elatives in puerto rican spanish. *Trabajo presentado en el Coloquio de gramática generativa (CGG) 24*.

Guy, G. R. (2017). The african diaspora in latin america: Linguistic contact and consequences. *Language Contact in Africa and the African Diaspora in the Americas*.

Hopper, P. J., et al. (1991). On some principles of grammaticization. *Approaches to grammaticalization*, *1*, 17–35.

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*. Cambridge University Press.

Iglesias, M. A. (1991). Estudios sintácticos sobre el habla popular mexicana.

Ito, R., & Tagliamonte, S. A. (2003). Well weird, right dodgy, very strange, really cool: Layering and recycling in english intensifiers. *Language in Society*, *32*, 257–279.

Kanwit, M., Teran, V. R., & Sarrió, S. P. (2017). Un fenómeno bien curioso: New methods for analyzing variable intensification across four dialects of spain and argentina. *Studies in Hispanic and Lusophone Linguistics*, *10*, 259–295.

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, *81*, 345–381.

Klein, H. (1998). Adverbs of degree in dutch and related languages.

Kondrat, M. M. (2013). La gradación del adjetivo en el habla culta de la ciudad de méxico. *Anuario de letras*, 71–118.

Ledgeway, A. (2012). From latin to romance: Morphosyntactic typology and change.

Lívio, C. C., & Howe, C. C. (2020). Intensification in portuguese: A cross-dialectal study of muito and bem. *Hispania*, *103*, 475–487. https://doi.org/10.1353/hpn.2020.0109

López, M. A., & Pottier, B. (1983). Morfología histórica del español.

Lorenz, G. (1999). Adjective intensification - learners versus native speakers.a corpus study of argumentative writing.

Macaulay, R. K. S. (2006). Pure grammaticalization: The development of a teenage intensifier. *Language Variation and Change*, *18*, 267–283.

Nevalainen, T. (2008). Social variation in intensifier use: Constraint on -ly adverbialization in the past?1. *English Language and Linguistics*, *12*, 289–315.

Pahta, P. (2006). This is very important: A corpus study of amplifiers in medical writing. *Advances in medical discourse analysis: Oral and written contexts*, *357381*.

Paradis, C. (2000). It's well weird: Degree modifiers of adjectives revisited: The nineties. *Corpora Galore*.

Pastor, A., & Armstrong, G. (2016). Una "buena" manera de hablar acerca de grados: Bien con adjetivos en español. *Studies in Hispanic and Lusophone Linguistics*, *9*, 401–427.

Pinkster, H., & Mulder, H. (1990). Latin syntax and semantics.

Quirk, R. (1974). A grammar of contemporary english. *The Modern Language Journal*, *58*, 139.

Rissanen, M. (2008). From 'quickly'to 'fairly': On the history of rather1. *English Language & Linguistics*, *12*(2), 345–359.

Salazar-García, V. (2008). Degree words, intensification, and word class distinctions in romance languages. *Studies in Language*, *32*, 701–726.

Salvador, F. S. (1987). La gradación adjetiva en el habla popular de ciudad de méxico. *Actas del I Congreso Internacional sobre el español de América*, 419–430.

Sedano, M. (2002). Este tema es muy/bien interesante. *Archivo de filologa aragonesa*, *59*, 859–873.

Serradilla Castaño, A. M., et al. (2005). Evolución de la expresión del grado superlativo absoluto en el adjetivo: Las perifrasis sustitutivas del superlativo sintético en español antiguo. *Cauce*.

Sowa, J. (2009). 'sweet as!': The intensifier as in new zealand and australian english. *English Today*, *25*, 58–61.

Tagliamonte, S. A. (2008). So different and pretty cool! recycling intensifiers in toronto, canada1. *English Language and Linguistics*, *12*, 361–394.

Tagliamonte, S. A. (2011). Variationist sociolinguistics: Change, observation, interpretation.

Tagliamonte, S. A., & Roberts, C. (2005). So weird; so cool; so innovative : The use of intensifiers in the television series «friends». *American Speech*, *80*, 280–300.

Torrego, L. G. (2000). Gramática didática del español.

Tsujimura, N. (2001). Degree words and scalar structure in japanese. *Lingua*, *111*, 29–52.

Wang, C. (2013). *Las fórmulas superlativas en el español de los siglos xviii y xix* (Doctoral dissertation). Madrid, Universidad Autónoma de Madrid.

Zieliński, A. (2013). Evolución semántico-sintáctica del sufijo superlativo-isimo en castellanoo. *Romanica Cracoviensia*, *13*(1).