A *CORPUS* DATA OF WEAK DEFINITES IN BRAZILIAN PORTUGUESE

Thaís Maíra Machado de SÁ
PhD student - PosLin/Universidade Federal de Minas Gerais (UFMG)

Maria Emília SARAMAGO
Undergraduate student - Universidade Federal de Minas Gerais (UFMG)

Maria Luiza CUNHA LIMA
Universidade Federal de Minas Gerais (UFMG)

RESUMO

Este artigo tem por objetivo descrever a ocorrência do definido fraco (ex. Ana foi ao hospital), introduzido por Carlson e Sussman (2005), em corpus do português brasileiro (PB). Foram analisadas 400 ocorrências de 31 palavras que podem apresentar a leitura fraca em PB (ex. o hospital). Observamos se a palavra é determinada por um artigo definido, em seguida se leitura do DP é fraca (Carlson e Sussman, 2005), forte - ou regular - (Russell, 1905) ou genérica (Carlson, 2005). Além da leitura, analisamos a função sintática do DP (sujeito, objeto, adjunto). Como resultado, trazemos a distribuição dos definidos fracos em PB, além de realizarmos uma análise mais detalhada sobre os que ocupam a posição de sujeito.

ABSTRACT

The purpose of this paper is to describe the occurrence of the weak definite (e.g. Ana went to the hospital), introduced by Carlson and Sussman (2005), in Brazilian Portuguese (BP) corpus. Four-hundred occurrences of 31 words, which may present the weak reading in BP (e.g. the hospital), were analyzed. It was observed if the word was determined by a definite article, and then if the DP reading was weak (Carlson and Sussman, 2005), strong - or regular - (Russell, 1905), or generic (Carlson, 2005). The DP's syntactic function - subject, object, adjunct – were also analyzed. As a result, we present the weak definite distribution in BP and perform a more detailed analysis about those that function as subjects.

PALAVRAS-CHAVE

Definido fraco, Definitude, Estudo de corpus, Genérico.

KEYWORDS

Weak definite, Definiteness, Corpus analysis, Generic expression.

Introduction

In this paper, we aim to describe the occurrences of weak (Carlson and Sussman, 2005), regular (Russell, 1905), and generic (Carlson, 2005) definites in a Brazilian Portuguese (henceforth BP) *corpus*. The debate around the proper semantics of weak definites (or even their existence as an independent use of the definite article) has been extremely productive in the last few years (Carlson e Sussman, 2005; Aguilar-Guevara and Zwarts, 2011; Schwarz, 2012; inter alia). Part of this debate stems from the difficulty in determining in which semantic and syntactic environments those definites appear. Some descriptions are based on distributional hypothesis. For instance, Schwarz's (2012: 4) hypothesis of semantic incorporation is grounded on the idea that weak definite interpretations are available only for verb objects (or adjuncts), whereas Carlson (2013) believesa weak definite can occur in the subject position - albeit very rarely. Therefore, a clear er idea of the weak definite distribution may bring important additions to the discussion concerning definiteness.

The best way to map this distribution is observing the phenomenon in *corpus,* since it provides an empirical and ecological analysis in which the lexical items are taken from the linguistic environment where they had occurred (c.f. Sardinha, 2004, Kennedy, 1998). Moreover, the position the weak definite occupies in sentences (subject, object, or adjunct) can inform restrictions for the realization of a weak reading as well as shed light into what syntactic environments may favor this reading. For that

reason, we have searched the *ptTenTen corpus*, available on *Skecth Engine[1]*, in an attempt to answer the following questions: in which structures do noun phrases determined by an article appear? Is there a correlation between the weak definite reading and the role it has in the sentence?

We initially chose 50 words that can receive a weak interpretation[2] in BP. After a cursory investigation, we selected 31 words and examined the first 400 occurrences of those words. All occurrences of Noun Phrases whose determiners were single definite articles were categorized into three categories - regular, weak, and generic -, and afterwards we tagged those NPs as subjects, objects or adjuncts. Our observations confirm that itis indeed possible to have a weak definite as a subject, although in a very restricted number and in specific constructions. For the purposes of the present paper, we have decided to focus on these items, because of their potential contribution to the present discussion.

In section 1, we present distinctions amongst weak, regular and generic definites, and review relevant recent literature. In Section 2, we establish the goals of the *corpus* analysis, and describe the methodology in Section 3. The results are shown in Section 4, and we discuss the relevance of our data to the ongoing debate on the characterization of the weak definite in Section 5.

1. Weak, regular, and generic definites

(1) - Where is Anna?
   - She went to *the bathroom*.
   - And what about John?
   - He went to *the bathroom* too.

---

[1] *Sketch Engine*is a *Corpus* Query System. It is available in http://www.sketchengine.co.uk/

[2] We extracted those words from lists of typically weak definites (for instance, as presented in Carlson and Sussman, 2005; Carlson et al, 2006; Aguilar-Guevara e Zwarts, 2011; Schwarz, 2012) that also can be read as weak in BP.

When one reads dialogue (1), the nominal expression *the bathroom* may be interpreted as co-referential. Nonetheless, it is equally possible to interpret the two instances as referentially independent. In BP, the latter is the most frequent reading (probably so in English, as well)[3]. In contrast, the nominal expression *the auditorium* cannot be referentially disjoint in the following Example (2):

(2)　　- Where is Anna?
　　　　- She went to *the auditorium.*
　　　　- And what about John?
　　　　- He went to *the auditorium* too.[4]

The highlighted expressions in (2) necessarily refer to the same entity. This usage is the typical *regular* DP as classically presented by Russell (1905), showing what the author called the *uniqueness* property. Uniqueness is defined as the attribute which the quantification of the noun phrase "the X" (X as any noun expression) entails the existence of one and only one entity in the world that could be the reference of X. *The auditorium* in (2) is uniquely identifiable and therefore different instances of the same expression in that context necessarily point to the same referent.

Carlson and Sussman (2005) proposed that cases that lack the uniqueness property, such as (1), constitute a different type of definite

---

3 In BP:
　　　　- Onde está Anna?
　　　　- Ela foi ao banheiro.
　　　　- Onde está John?
　　　　- Ele também foi ao banheiro.
4 In BP:
　　　　- Onde está Anna?
　　　　- Ela foi ao auditório.
　　　　- Onde está John?
　　　　- Ele também foi ao auditório.

phrase, which they named as *weak definites*[5].

A main characteristic of weak definites (Carlson and Sussman, 2005; Carlson et al., 2006; Carlson et al., 2013) is that they are lexically restricted. Even within the same construction (e.g. verb to go + the nominal phrase), only *the bathroom* (1) can be read as weak; on the other hand, *the auditorium* (2) demands a regular interpretation. Furthermore, the authors (Carlson and Sussman, 2005; Carlson et al., 2006; Carlson et al., 2013) suggest that weak definites do not usually appear in subject position[6], appearing within constructions like *go to the hospital/bathroom*. As the authors claim (2013:14), "Weak definites are further restricted by the need to co-occur with, or be "governed by," certain other lexical items - verbs and prepositions". Those characteristics and some other experimental data (Carlson et al., 2013; Klein et al.; 2013) led the authors to propose an "incorporation" analysis, in which the noun itself lacks the uniqueness, while providing a notion of "(cultural) familiarity"[7].

What the authors call "incorporation" (their quotation marks) is semantic in nature, not to be confused with what is usually called incorporation in syntactic/morphological theories. One of the central characteristics of incorporation is the so-called semantic enrichment[8], or an enriched interpretation (presented by Goldberg, 1995). The fact that weak definites generally present semantic enrichment is an important argument used by Carlson et al. (2013) to support the "incorporation" analysis.

---

[5] Carlson and Sussman (2005) were not the first to use the name *weak definites*. Poesio (1994) questioned the Russellian uniqueness (1905) and Heim's familiarity (1982) affirming that in sentences like "John got these data from *the student of a linguist*", there is no need to have familiarity with *the student* to understand the sentence or the need of a single individual to be characterized as *the student of a linguist*. He named that kind of definites as *weak definites*. Carlson and Sussman went further on, stating that the weak definites lack uniqueness.

[6] As we shall see, our results show that, albeit rarely appearing as subjects, is very important to understand under which circumstances that occurrence is possible.

[7] Term employed by Carlson et al (2013:19)

[8] Refer to Donazzan (2013) for an alternative view.

Schwarz (2012) has also claimed that weak definites were part of an incorporation process. According to the author (p.15), "weak definites are definites appearing in verb phrases that denote kinds of events". He affirms that weak definites are associated with typical activities. Schwarz also highlights that weak definites only appear as objects of certain verbs and/or prepositions. The difference between the analyses in Carlson et al. (2013) and Schwarz (2012) is that the latter defends the predicate has a 'regular' individual as its argument "to combine semantically with kind-denoting terms" (p.16).

An alternative analysis[9] is proposed by Aguilar-Guevara e Zwarts (2011, 2013). The authors claim the weak definite denotes a kind, not an individual, being a generic DP. They state weak definites and generic definites would be "different faces of same phenomenon" (2011:15). Unlike Schwarz (2012) and Carlson et al. (2013), Aguilar-Guevara and Zwarts declare that the noun denotes kind, not only the incorporated VP[10]. Therefore, this would be the realization of a third type of definite: the generic one. This generic definite, illustrated by *the radio* in (3) is productive in BP as well as in English.

(3)   Marconi invented *the radio[11]*.

The analysis in Carlson et al. (2013) is ultimately difficult to distinguish from Aguilar-Guevara and Zwarts', but the authors emphasize a different focus on their investigation. They defend that the most important aspect

---

[9] The cursory review presented here does not contemplate other possible explanations, for example, Beyssade's (2013), which suggests that weak definites are always types and never tokens. For further research on this hypothesis, refer to Beyssade and Oliveira (2013).

[10] Aguilar-Guevara and Zwarts (2013) say: "One important reason why we do not pursue a pseudo-incorporation analysis is that this would not straightforwardly allow us to attribute a meaningful role to the presence of the definite article in weak definites. Another reason is that this analysis does not permit us to capture the parallelism between generic definites and weak definites."

[11] Example by Carlson (2005:2).

is to delineate clearly what is the compositional role fulfilled by the definite article in those constructions (p. 19):

> Also, in a way the incorporation analysis we are assuming makes it very hard to distinguish it from the "kinds" analysis discussed above. This is because the meaning of the incorporated bare nominal form is "generic" in the sense that it does not have the capacity for individual reference. So the issue for us is primarily the compositional role of the definite article, and not primarily about generic interpretations. (authors' emphasis)

Similarly, in our work, we endeavor to remain relatively agnostic in what is the correct theory; however, we aim to contribute to a better understanding of the definite article compositional role. When categorizing our *corpus* instances, we conservatively decided to hold on to Carlson's initial characterization of weak definites, but remaining open to adopt any alternative explanations. Consequently, we stuck to an initial distinction between generic and weak definites: we categorized our tokens into prototypical generic definites (Carlson, 2005), and initially categorized tokens as weak ones where there was the possibility of a weak reading. Examples (4) and (5) below illustrate this approach. Provisionally, we have decided that examples like (4) - where there is a clear reference to a kind or a whole category - are to be called generic definites, whereas instances similar to (5) - where there is an individual that is not uniquely identifiable - were labelled weak.

(4)   Em relação   ao   foguete espacial, *o   avião*   anda devagar.
Gloss: In   relation to the rocket   spatial   *the airplane* goes slow.
        Compared to a spacecraft, airplanes fly slowly.

(5)   No dia seguinte cedo, pegamos   *o avião*   de volta a Salvador.
Gloss: In day next   early took(pl1stpearson) *the*   *airplane* to back to SALVADOR.
        Early on the following day, we took the flight back to Salvador.

As expected, the distinction is not always easy to be made and there are plenty of examples that defy the initial criteria. One of the most common cases is the one where there is a clear generic context, but the proper DP categorization is somewhat more doubtful. For instance, in Example (6): is *the airplane* a weak or generic definite?

(6)   Se o   radar Mostrar que *o*   *avião*   se   desviou mais de   alguns
GLOSS: IF THE RADAR SHOW(INF) THAT *THE* *AIRPLAIN* CL.3SG DIVERTED MORE THAN FEW

    quilômetros, ou   graus,   do   plano de   voo,   a   primeira resposta
    KILOMETERS, OR   DEGREES, OF THE PLAN   OF   FLIGHT, THE   FIRST   ANSWER

    é   um   controlador da   FAA tentar o   contato via   radio.
    is   a   controller   of the FAA try(inf) the   contact BY   RADIO.

If the radar shows that the airplane has been diverted for more than a few kilometers, or degrees, from the initial flight plan, the first response is an attempt from the FAA's controller to contact (the pilot) via radio.

A possible criterion to decide between the alternatives is to use a property of weak definites observed by Aguilar-Guevara (2011), Schwarz (2012) and other authors: a pronoun referring back to a weak definite always present a unique reading, as shown in (7):

(7)   Bill is in *the hospital*$_p$ and John is, too. It$_i$ has an excellent heart surgery department.[12]

---

[12] Example by Schwarz (2012:4)

Example (7) shows that a DP that may initially be read as weak has its interpretation decidedly steered towards a regular one when it is referred back by a pronoun (*it* in the Example). We suggest, in contrast, that the same does not happen with generic definites. Compare examples 8 (b and c) below. A pronoun whose antecedent is a generic still refers to a kind. If the generic definite is re-introduced by a pronoun, the generic reading is maintained. In (8a) *the guitar* is introduced as a clear kind reference. If a pronoun refers to it in a following sentence (as in 8b), *the guitar* still has a generic reading.

(8a)   *O   violão*   é o   instrumento mais presente da   música popular brasileira.

Gloss: *The guitar* is the instrument most present ^OF ^THE music   popular Brazilian.

The guitar is the most popular instrument in Brazilian folk music.

(8b) And one reason is that *it* is cheaper than most of instruments.

Whereas, a sentence that does not allow a generic reading (as in 8c) fails to make an appropriate reference to the antecedent:

(8c) *And *it* is broken.

For instance, if we retrieve Example (6) above, and create a continuation with an anaphoric pronoun for it (as in 6b) the regular interpretation is enforced:

(6)   If the radar shows that *the airplane* has been diverted for more than a few kilometers.
(6b)   *It* was kidnaped.

This distinction in Example (6) (although, by all means, not conclusive) is an indication of a greater similarity between weak definites and generic ones, and was the criterion operationally adopted here for the categorization.

In addition, by categorizing definite expressions and looking into their distribution, we intended to evaluate their semantic behavior in different structures, which, in turn, may help to understand the incorporation matter, especially in subject occurrences.

2. Methodology

2.1 The *ptTenTen corpus*

The *corpus ptTenTen* used for this study has a database of more than 2.7 million words, and was developed by Kilgarriff et al (2014)[13]. The *corpus* was built searching the language productions through the Web, due to the Internet richness of informal and speech-like genres.

2.2 Materials

We extracted the first 400 hundred occurrences of the selected 31 words, regardless of determination. Then, we have selected from the subset of words the ones preceded by a definite article, with or without a preposition[14]. Nouns preceded by indefinite articles, determined by pronouns, bare nominal expressions, and any case with nouns not determined by definite article were excluded.

We also excluded words that were part of a title or a proper name, an idiom or metaphors and, also, items that functioned as adjectives in specific contexts. We did not include any words that deviated from

---

[13] Kilgarriff and his team developed Sketch Engine.

[14] In BP, the definite article and several prepositions can be contracted (Cunha and Cintra, 2001). For Example, the preposition "de" and the masculine definite article "o" are realized as "do".

the expected form – probably due to typographical errors – or from the expected meaning. Lastly, we have focused only on singular nominal expressions. The final item count included in the present study is 2196.

2.3 Procedures

Each item was extracted and analyzed in its clause. All the procedures and tests were decided prior to the beginning of the analysis. Every occurrence was analyzed by, at least, two researchers, and the tests described below were used to make the final decision.

We proceeded to categorize each item as weak, regular, and generic. Simple tests were applied to each token to help with the label attribution. Weak definites were operationally defined here as expressions that refer to individuals, but lack a uniquely identifiable referent (9); regular definites had uniquely identifiable referents (10); and the generics were the ones referring to kinds (11).

(9)   Maria   Silva, de   56   anos,   foi   parar   no   hospital.

Gloss: MARIA   SILVA   OF   56   YEAR SOLD   WENT   STOP(INF)   INTHE   HOSPITAL.

Mara Silva, 56 years old, ended up in the hospital.

(10)   Em   Jenin,   a   ocupação   israelense   cercou   o   hospital

Gloss: In   JENIN   THE   OCCUPATION   ISRAELI   SURROUNDED   THE   HOSPITAL

e   expulsou   a   imprensa.

and   EXPELLED   THE   PRESS.

In Jenin, the Israeli occupation surrounded the hospital and expelled the press.

(11)	Na	segunda	metade	do	século	XIX	a	morte	passa	a
Gloss:	In	second	HALF	OF	CENTURY	XIXTH	THE	DEATH	PASS	TO
	não	acontecer	mais	em	casa,	ao	alcance	dos	olhos	da
	NOT	HAPPEN(INF)	MORE	IN	HOME	WITHIN THE	REACH	OF THE	EYES	FROM THE
	família,	mas	sim,	*no*	*hospital.*					
	FAMILY	BUT	YES	IN THE	HOSPITAL.					

In the second half of the XIXth century, death doesn't happen at home anymore, under the eyes of the family, but in the hospital.

The sloppy identity[15] test was applied to differentiate weak and regular definites. In this case, only weak definites are felicitous with sloppy identity, as the contrast between (9a) e (10a) shows.

(9)a. Mara Silva, 56 years old, ended up in the hospital. João, 36 years old, did too. (different hospitals acceptable)

(10)a. In Jenin, the Israeli occupation surrounded the hospital and expelled the press. The Syrian occupation did too (must be the same hospital).

The test proposed in the introduction was used in controversial decisions between generic and weak definites. The testuses a pronoun and analyzes its consequence to the referent identity. In cases where the presence of the pronoun generated referent individualization, we decided to categorize the expression as weak, otherwise as a generic. The comparison between (9b) and 11(b) illustrates once again the test.:

(9)b. Mara Silva, 56 years old, ended up in the hospital. It was very dirty.

---

[15] Sloppy Identity happens in contexts where the identity of the pronoun in an elided VP is not identical to the antecedent VP. For further information, refer to Merchant and Simpson (2012).
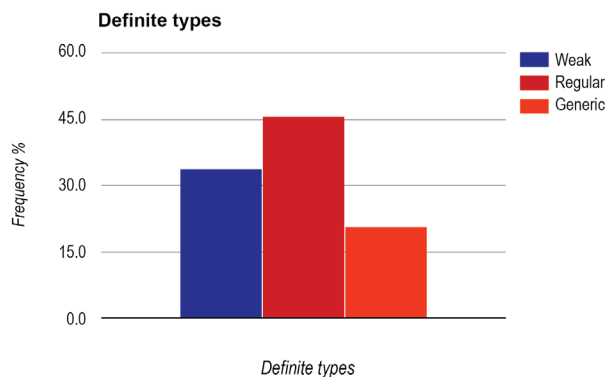
(11)b. *The second half of the XIXth century, death doesn't happen at home anymore, under the eyes of the family, but in the hospital. It was very dirty.

In the next stage, we further classified each token according to its syntactic role into subject, object, adjunct, or other. To dispel any questions on specific cases where it is debatable if one token is an object or an adjunct, we have turned to Luft (2010). The "other" classification was chosen when the expression had any other syntactic function, e.g. passive agent.
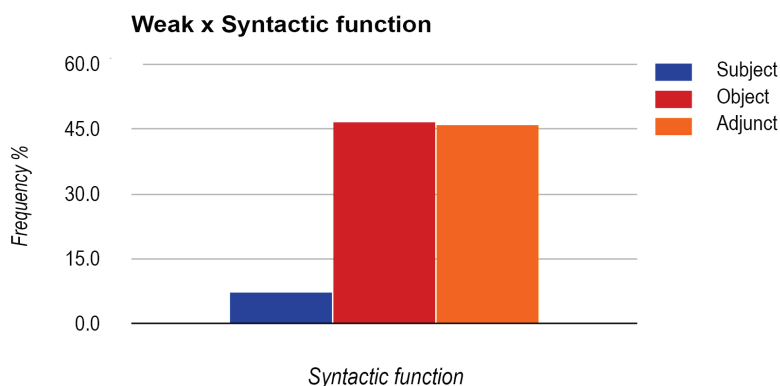
3. Results

We observed the distribution presented in Graphic 1 that shows that the regular reading is significantly more frequent than the others, 45,6% ($X^2 = 205.2568$, df = 2, p-value $< 0.01x10^{-14}$), as expected. An interesting finding is that, according to the categorization criteria employed here, weak DPs occur 33,7%, significantly more ($X^2 = 68.5059$, df = 1, p-value $< 0.01x10^{-14}$) than the generic ones, 27,5%.
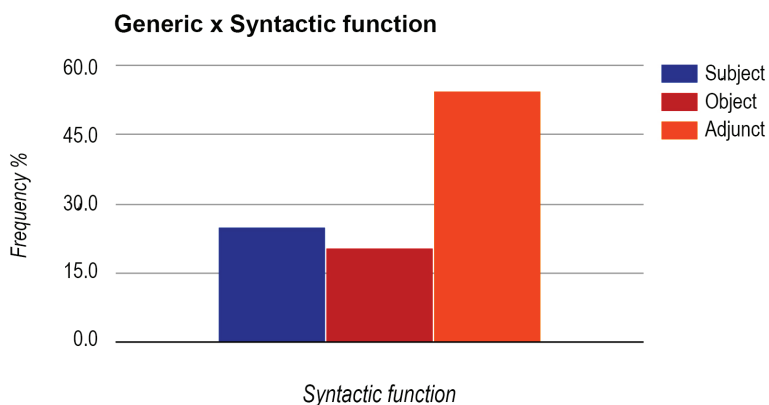
GRAPHIC 1: Frequency of definite types

Afterwards, we examined the relation between syntactic function and the classification of definites. It is no surprise that adjuncts are the most productive category, since one sentence can have only one subject and a maximum of two objects, but an open number of adjuncts. That is the reason why adjuncts present the largest number of occurrences in all categories. Interestingly, however, weak definites (Graphic 2) appear as adjuncts (45,7%) as much as objects (46,6%), (statistical test: $X^2 = 0.0544$, df = 1, p-value = 0.8156). As described in the introduction, Carlson et al. (2006) were correct in saying that the weak definites seldom appear as subjects - only 7,2% of the occurrences (Graphic 2), significantly less than the other categories ($X^2 = 212.5607$, df = 2, p-value $< 0.01 \times 10^{-14}$).

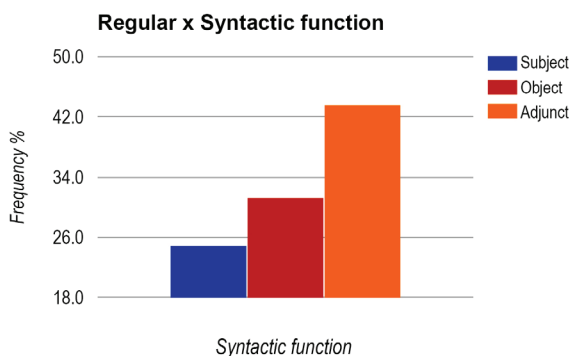GRAPHIC 2: Weak x Syntactic Function

Weak x Syntactic function

Generics (Graphic 3) are more uniformly distributed between subject (25,1%) and object (20,3%), being adjuncts a significant half of the data (54,6%).This difference is confirmed by the chi-square test ($X^2 = 85.6232$, df = 2, p-value $< 0,01 \times 10^{-14}$). The same overall pattern is observed with regular definites (Graphic 4), presenting a significant majority of adjuncts (43,7%) ($X^2 = 52.7934$, df = 2, p-value $< 0,01 \times 10^{-10}$), followed by objects (31,3%), and subjects (25%).

GRAPHIC 3: Generic x Function

Generic x Syntactic function

Syntactic function

GRAPHIC 4: Regular x Syntactic function

Regular x Syntactic function

Syntactic function

The most important finding in this distribution pattern is the occurrence of a small amount of subjects that are weak definites. From a universe of 45 of these, 42 (93.3%) have a similar structure, exemplified in (6) above and repeated here:

(6)	Se	o	radar	mostrar	que	*o*	*avião*	se	desviou	mais	de	alguns
Gloss:	IF	THE	RADAR	SHOW(INF)	THAT	*THE*	*AIRPLAIN*	CL.3SG	DIVERTED	MORE	THAN	FEW

	quilômetros,	ou	graus,	do	plano	de	voo,	a	primeira	resposta
	KILOMETERS,	OR	DEGREES,	OF THE	PLAN	OF	FLIGHT,	THE	FIRST	ANSWER

	é	um	controlador	da	FAA	tentar	o	contato	via	radio.
	IS	A	CONTROLLER	OF THE	FAA	TRY(INF)	THE	CONTACT	BY	RADIO.

If the radar shows that the airplane has been diverted for more than a few kilometers, or degrees, from the initial flight plan, the first response is an attempt from the FAA's controller to contact (the pilot) via radio.

In those cases, it is disputable if we are dealing with a weak or a generic definite. The individualization test we devised here suggests that we should subscribe to a weak interpretation[16]. In a remainder of cases, admittedly very few (7%) of weak definites are in subject position, and it is more difficult to waive the weak interpretation in favor of a generic one. Observe Example (12):

(12)	*O*	*avião*	partia	do	Rio	às	17h.
Gloss:	*THE*	*AIRPLANE*	TOOK OF	FROM THE	RIO	AT	5 P.M.
	The airplane took off from Rio at 5p.m.						

In this case, the context makes it clear that one is referring to no specific and unique airplane, but one that is linked to an airplane path, although not identifiable.

---

[16] We are debating this in more detail in the final discussion.

4. Discussion

Both Carlson et al. (2013) and Schwarz (2012) incorporation analyses are possibly supported by the examples of weak definites in object and adjunct roles. But the presence of weak definites in subject position by itself calls into question their proposal as presented in recent works, since subjects cannot, by definition, be incorporated into VPs. Our characterization of examples similar to (12) and (6) as weak - though based on a test that has some limitations, at best. In theory, it is possible to defend another characterization (at least for cases comparable to Example (6)), and, instead, dispense with the individualization test as an adequate way to single out weak definites altogether.

In our opinion, the problem we face here does not lay solely on the test adequacy to one or other structure, but on the very subjective nature of these tests themselves. To adequately untangle different interpretations of definite expressions, and even to establish if there are actual different interpretations, we need to use measures that are more objective, and gather real world (ideally experimentally controlled) data.

The work presented here is relevant because it brings real data to the table and creates at least one starting point for a deeper investigation. We believe that future experimental work with these particular subject occurrences can be important to shed light on the definiteness problem, and minimally help to discard some inadequate explanations (for instance, VP incorporation), or refine the definition of weak definites.

Further work with the *corpus* remains to be done. One possible avenue is to better examine the role of weak definite adjuncts within the incorporation framework. Considering that adjuncts do not take part in the verb argument structure and are optional, how are they incorporated into VPs? Are there specific kinds of adjuncts that allow this incorporation or are they linked to some verbal characteristic?

We hope to have shown that more empirical forms of investigation can be made and do bring important evidence to this type of phenomena. Furthermore, we hope to continue pursuing this line of work and stimulating others to follow it as well.

References

AGUILAR-GUEVARA, A.; ZWARTS, J. **Weak definites and reference to kinds**. Proceedings of SALT, Ithaca, NY: CLC Publications. 2011, v.20, p. 179-196.

_____. **Weak definites refer to kinds**. In: BEYSSADE, C.; OLIVEIRA, R. P. de. Weak definites across Languages. Recherches linguistiques de Vincennes. 2013, V.42, p.33-60.

BASSO, R. **Telicidade e detelicização:** semântica e pragmática do domínio tempo-aspectual. Dissertação de mestrado. Campinas: Unicamp, 2007.

BEYSSADE, C.; OLIVEIRA, R. P. de. **Weak definites across Languages**. Recherches linguistiques de Vincennes. 2013, V.42

CARLSON, G. **Generic Reference**. In: BROWN, Keith. The Encyclopedia of Language and Linguistics (Second ed.). Amsterdam: Elsevier, 2005.

CARLSON, G. and SUSSMAN, R. **Seemingly indefinite definites**. In: KEPSER, Stephan; REIS, Marga. Linguistic evidence: empirical, theoretical, and computational perspectives. Berlin: Mouton de Gruyter, 2005. p. 71–85.

CARLSON, G., et al. **Weak Definite Noun Phrases**. In: DAVIS, C.; DEAL, A. R. and ZABBAL, Y. Proceedings of NELS 36. Amherst: GLSA, 2006. p. 179-196.

CARLSON, G., et al. (2014).**Weak definites as a form of definiteness: Experimental investigations**. In: BEYSSADE, C.; OLIVEIRA, R. P. de. Weak definites across Languages.Recherches linguistiques de Vincennes. 2013, V.42, p.11-32

CHUNG, S.; LADUSAW, W. **Restriction and Saturation**. Cambridge, MA: MIT Press, 2004.

CUNHA, C.; CINTRA, L. **Nova Gramatica do Português Contemporâneo**. Rio de Janeiro: Nova Fronteira, 3ª ed.2001.

DAYAL, V. **Bare NPs, reference to kinds, and incorporation**. In: MATTHEWS, T.; STROLOVITCH, D. Proceedings of SALT. Cornell University, Ithaca, NY: CLC Publications. 1999, v.9, p.34-51.
_____. **Hindi pseudo incorporation**. Natural Language and Linguistic Theory 29/1. Rutgers University. 2011, p.123-167 .

DONNELLAN, K. S. **Reference and Definite Descriptions**. In: The Philosophical Review. 1966, Vol. 75, No. 3, p.281–304.

DOWTY, D. R. **Aktionsarten, NP semantics, and the structure of events**. Paper presentedat the joint Association for Symbolic Logic/Linguistic Society of America Conference on Logic and Natural Language, Stanford University, 1987.

FARKAS, D.; de SWART, H. **The Semantics of Incorporation**. Chicago: CSLI, 2003.

GOLDBERG, A. E. **Constructions:** A Construction Grammar Approach to Argument Structure. Chicago: The University of Chicago Press, 1995.

HEIM, I. **The Semantics of Definite and Indefinite Noun Phrases**. In: HANKAMER, J. Outstanding dissertations in Linguistics. New York & London: Garland Publishing Inc., 1982.

KENNEDY, G. **An introduction to *corpus* linguistics.** New York: Longman, 1998.

KILGARRIFF, A. et al. **PtTenTen:** A *corpus* for Portuguese lexicography. In: SARDINHA, T.B.; FERREIRA, T. Working with Portuguese corpora. New York: Bloomsbury Publishing, 2014.

LUFT, C. P. **Dicionário prático de regência verbal**. São Paulo: Ed. Ática, 2010

MERCHANT, J.; SIMPSON, A. (eds) **Sluicing:** Cross-Linguistic Perspectives. New York: Oxford, 2012.

POESIO, M. **Weak definites**. In: HARVEY, M.; SANTELMANN, L. Proceedings of the Fourth Conference on Semantics and Linguistic Theory: Cornell University, Ithaca, NY: CLC, 1994. p. 282-299.

ROBERTS, C. **Uniqueness in definite noun phrases**. Linguistic and philosophy. Netherlands: Kluwer Academic Publishers, 2003.

RUSSELL, B. **On Denoting**. In: Mind, New Series. 1905, 14, p.479-493.

SARDINHA, T.B. **Linguística de *Corpus***. São Paulo: Manole, 2004.

SCHWARZ, F. **How weak and how definite are Weak Definites?** (in press). Aceito para publicação em Weak Referentiality. (Revised draft, December 21 de 2012)

STRAWSON, P. **On Referring**. In: Mind, New Series.1950, 59, p.320-344.