

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)**ScienceDirect**

Procedia - Social and Behavioral Sciences 95 (2013) 195 – 202

**Procedia**  
Social and Behavioral Sciences

5th International Conference on Corpus Linguistics (CILC2013)

# Portuguese Future / Future Preterit and Present / Preterit Indicative: A Collocational Comparison

Shintaro Torigoe\*

*Gunma Prefectural Women's University, Kaminote 1395-1, Tamamura-cho, Saba-gun, Gunma, 370-1193, Japan*  
*Tokyo University of Foreign Studies, Asahi-cho 3-11-1, Fuchu-shi, Tokyo, Japan*

---

## Abstract

This study examines the collocational similarity between the Future/Future Preterit and the Present/Preterit Indicative in Portuguese. From syntactic approach, Bello (1847) claimed that in nominal completive subordinate clause the Future and the Future Preterit forms collocate with the same matrix verbs which follow the Present and Preterit Indicative tense forms, rather than the Subjunctive verb forms. In Torigoe (2011), the author attempted to show a corpus-based counterevidence to Bello's intuition-based structural explanation. As the result, he verified the statistical similarity in global collocational tendency between the Future/Future Preterit and the Present/Preterit Indicative. On the other hand, he identified some matrix verb groups that predominantly and significantly collocate with one of the morphological groups in subordinate clause. In the present study, the author discusses further, analyzing the additional corpus data, the Corpus Brasileiro, and comparing with the result of Torigoe (2011) to seek more credible findings. The result suggests that, as in Torigoe (2011), the global collocational tendency is statistically similar among two morphological groups, as well as there are similar matrix verb groups which predominantly collocate with one another, though there is some minor difference in its lexis.

© 2013 The Authors. Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/4.0/).  
Selection and peer-review under responsibility of CILC2013.

*Keywords:* Portuguese; future; future preterit; conditional; mood and modality; corpus analysis; collocation; corpus brasileiro; Sketch Engine

---

---

\* Corresponding author. Tel.:  
E-mail address: [shintarao@yahoo.co.jp](mailto:shintarao@yahoo.co.jp)

## 1. Introduction

This study examines the collocational (or colligational) similarity between Portuguese Future/Future Preterit and Present/Preterit Indicative when they occur in a subordinate clause with the matrix verb in a nominal clause. The background section reviews some relevant studies, mainly in Spanish grammar<sup>1</sup>, and outlines the research question. The author then summarizes his earlier corpus-based study on this issue (Torigoe 2011), describing its procedure, its findings, and its relevance to the present study, which seeks to improve upon the last. The majority of the discussion pertains to the author's replication of the preceding work. In that section, another Portuguese large corpus is analysed, and the findings from two studies are compared. Finally, the discussion addresses some limitations from both studies.

## 2. Background: Future and Future Preterit 'Indicative'

In Spanish and Portuguese grammar, the so-called Future has historically been classified as 'Future Present', 'Simple Future' or, 'Future Indicative'. Until recently, however, the Future Preterit had been defined as 'Subjunctive' by Nebrija (1492) and Barros (1540) and then as 'Subjunctive', 'Potential', or 'Conditional'. Currently, the Spanish and the Brazilian Portuguese grammar classify this form as Future Preterit Indicative, following the Real Academia Española (1973, cited in Castronovo 1989) and the Nomenclatura Gramatical Brasileira (1967, cited in Cunha & Cintra 2007). In European Portuguese, however, the form continues to be classified as Conditional.

In the semantic field, there are many claims that Future and Future Preterit tense (Conditional) forms in general index modality as much as future-time reference (e.g. Lyons 1977; Fleischman 1982; Bybee, Pagliuca, & Parkins 1994; Givón 1994; Palmer 2001). Furthermore, a French grammarian, Yvon (1952, in Fleischman 1982) and some grammarians in Spanish (e.g. Alarcos Llorach 1980, 1994; Deguchi 1986; Terasaki 1998, etc.) have hypothesized that the Future and the Future Preterit are the non-past and the past tense in a third mood that is neither Indicative nor Subjunctive. In contrast, Bello (1847) and some followers claimed the indicativity of the Future and the Future Preterit. He claimed that in a nominal completive subordinate clause, the two verb forms collocate with the same matrix verbs that follow the Present and the Preterits (Perfect and Imperfect) Indicative tense forms, rather than the Subjunctive verb forms. For example, *parecer* normally follows either the Present and the Preterit or the Future and the Future Preterit in completive clause (as in (1)-(4)), but not the Subjunctive forms. In contrast, the verb *dudar* (in Portuguese, *dividir*) exclusively follows the Subjunctive forms (as in (5) and (6)), but does not follow the Future and the Future Preterit (as in (7) and (8)).

- (1). Parece                      que        llueve.  
 seem<sub>-pres-ind-3-sg</sub>        that        rain<sub>-pres-ind-3-sg</sub>  
 (It seems to be raining.)
- (2). Parece                      que        anoche        llovió.  
 seem<sub>-pres-ind-3-sg</sub>        that        last night        rain<sub>-pres-pret-3-sg</sub>  
 (It seems to have rained last night.)
- (3). Parece                      que        mañana        lloverá.  
 seem<sub>-pres-ind-3-sg</sub>        that        tomorrow        rain<sub>-fut-ind-3-sg</sub>  
 (It seems to rain tomorrow.)

---

<sup>1</sup> I intuitively assume here the similarity between the Portuguese Future and the Spanish counterpart in their usage and in the collocational issue treated in this paper, but I know that the cross-linguistic verification should be done

- (4). Ayer me pareció que hoy llovería.  
 yesterday me seem<sub>-pres-pret-3-sg</sub> that today rain<sub>-fut.pret-ind-3-sg</sub>  
 (Yesterday, it seemed that it would rain today.)
- (5). Dudo que continúen todavía las negociaciones.  
 doubt<sub>-pres-ind-1-sg</sub> that continue<sub>-pres-subj-3-pl</sub> still the negotiations  
 (I doubt that they still continue the negotiation.)
- (6). Dudé que continuasen todavía las negociaciones.  
 doubt<sub>-pret-ind-1-sg</sub> that continue<sub>-impf-subj-3-pl</sub> still the negotiations  
 (I doubted that they still continued the negotiation.)
- (7). \*Dudo que continuarán todavía las negociaciones.  
 doubt<sub>-pres-ind-1-sg</sub> that continue<sub>-fut-ind-3-pl</sub> still the negotiations  
 (I doubt that they will still continue the negotiation.)
- (8). \*Dudé que continuarían todavía las negociaciones.  
 doubt<sub>-pres-ind-3-sg</sub> that continue<sub>-fut.pret-ind-3-pl</sub> still the negotiations  
 (I doubted that they would still continue the negotiation.)  
 (Bello (1847/1988), reordered, numbered, grossed\*, underlined, and asterisked by the author)  
 \*pres=present, pret=preterit, impf=imperfect, fut=future, fut.pret=future preterit, ind=indicative,  
 subj=subjunctive, 1=first person, 3=third person, sg=singular, pl=plural

This classical explanation was so convincing that it pushed the third mood hypothesis out of the mainstream (Resnick 1984, Castronovo 1989).

However, the author doubts Bello's explanation. Some matrix verbs seem to prefer the Future and the Future Preterit to the Present and the Preterits Indicative, and others vice versa. If this collocational tendency is determined not by future-time reference but by other factors such as propositional modality indicated through matrix verb lexis, as in the case of Subjunctive constructions, then the Future and the Future Preterit will be shown to be a mood rather than tense. Moreover, Bello's claim does not seem to be evidenced by any analysis on authentic language use, since the analysis of quantitative, usage-based, and authentic data in grammar studies had been lacking in his days, and in fact they remained so until the past two decades. As McEnery, Xiao, and Tono (2006) point out, observations that solely rely on nativespeaker's intuition may not reliable. For these reasons, in Torigoe (2011), the author attempted to show some counterevidence to Bello's explanation by analysing a large corpus of authentic language use data. In Spanish and Portuguese grammar, the so-called Future has historically been classified as 'Future Present', 'Simple Future' or 'Future Indicative'. However, the Future Preterit had been defined as 'Subjunctive' by Nebrija (1492) and Barros (1540) and then, until recently, as 'Subjunctive', 'Potential', or 'Conditional'. Currently, the Spanish and the Brazilian Portuguese grammar classify this form as Future Preterit Indicative, following the Real Academia Española (1973, cited in Castronovo 1989) and the Nomenclatura Gramatical Brasileira (1967, cited in Cunha & Cintra 2007). In European Portuguese, however, the form continues to be classified as Conditional.

### 3. Previous work: Torigoe (2011)

#### 3.1. Methods

In a prior study, the author analysed the Cetempúblico and the Cetenfolha, the pre-set corpora available in a powerful concordance tool called the Sketch Engine. These are corpora of Portuguese and Brazilian newspapers (*Público* and *Folha de São Paulo*) and are originally constructed by LINGUATECA. Of the total 200 million Portuguese words in these corpora, only about 66 million of them are available through the Sketch Engine. The Sketch Engine was selected because of its iterativity in query filtering, which enables step-wise complex search options such as listing matrix verbs in completive clause constructions. First of all, the author searched the 'node

morphemes' that are the Future/Future Preterit and the Present/Preterit Indicative forms. Because of the poor POS annotation in the corpora (the only tag attributed to verbs is 'V'), regular expressions were used to search all verbs by inflectional suffix. The interchangeability between some regular Present Indicative and other regular Present Subjunctive forms meant that searching would have to be limited to Present/Preterit forms in the five most frequent and irregular verbs in the corpora: *ser* ('to be'), *ter* ('to have'), *estar* ('to be'), *poder* ('to be able to'), and *fazer* ('to do'). The author then ran concordances for the Future/Future Preterit and the Present/Preterit 'pattern' (Hunston and Francis 1999), and filtered them for collocations with the conjunction *que* ('that') on left 4-2 nodes and left 2-1 nodes, supposing the existence of subjects or adverbs. Periods, commas, and hyphens were removed from left 3-1 nodes in the former and on 2-1 nodes in the latter. The former case was designated '4-node' and the latter '2-node'. Finally, collocation lists by lemma were produced on 6-5 nodes for 4-node and 4-2 nodes for 2-node. The illustrations of the procedure are as follows:

(9).  $V_{\text{main}}$  (xx) que (xx) (xx) xx  $V_{\text{fut/f,pret} | \text{pres/pret}}$   
 Matrix verb on left 6-5nodes conjunction on left 4-2 nodes "node morpheme"

(10).  $V_{\text{main}}$  (xx) que (xx)  $V_{\text{fut/f,pret} | \text{pres/pret}}$   
 Matrix verb on left 4-2nodes conjunction on left 2-1 nodes "node morpheme"

This procedure generated four collocation lists, 4-node and 2-node lists of the Future/Future Preterit pattern and of the Present/Preterit pattern. The collocation lists contain the frequency, the MI score, and the log-likelihood score of each word. The author then removed all non-verb words, and extracted top-100 verb lexis.

### 3.2. Findings

Results verified, on the one hand, the statistical similarity in global collocational tendency between the Future/Future Preterit and the Present/Preterit Indicative pattern in each 'node' scope. The correlation scores were  $r=.98$  in 4-node and  $r=.91$  in 2-node. On the other hand, a number of lexical items predominantly collocate with one pattern or the other. The Figure 1 and 2 below simply visualize the findings.

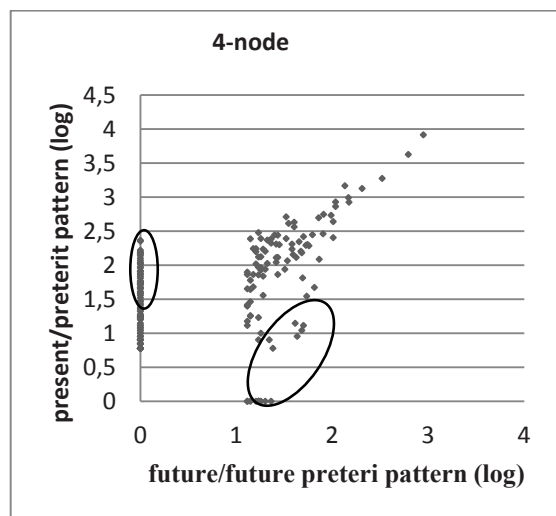


Figure 1. Frequency plot of matrix verbs (4-node)

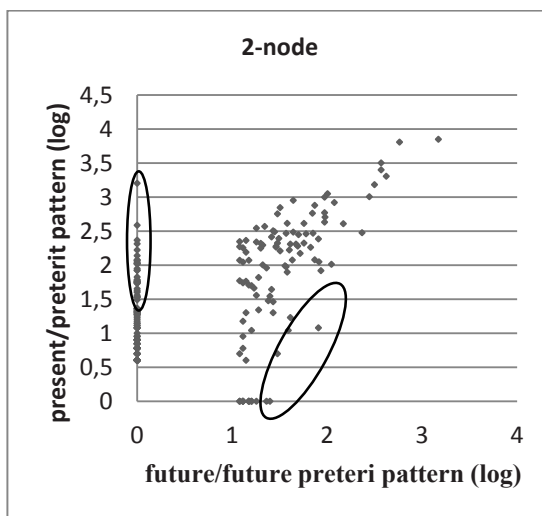


Figure 2. Frequency plot of matrix verbs (2-node)

To identify these vocabularies, the matrix verbs were then filtered according to the following criteria: more than 20 in frequency, more than 3 in MI score, more than 3.83 in log-likelihood score (cf. Dunning 1993, Hunston 2002, McEnery et al. 2006), on both 4-node and 2-node. As the result, the author found that 21 matrix verbs prefer the

Future forms: verbs of presenting new information (*adiantar, anunciar, avisar, indicar, sugerir*), of estimation and evaluation (*avaliar, elaborar, estimar*), of decision (*decidir, definir, determinar*), of anticipation (*antecipar, calcular, prever*), of supposition and imagination (*imaginar, supor*), of promise or guarantee (*assegurar, garantir, prometer*), and *afirmar* and *reafirmar*. On the other hand, 38 verbs prefer the Present and Preterit Indicative including verbs of assertion and explanation (*concluir, explicar, insistir, mostrar, verificar*, etc.), of perception and memory (*compreender, esquecer, lembrar, notar, recordar-se, sentir*, etc), and of response (*arrepender, concordar, responder*).

### 3.3. Limitations

Torigoe (2011) showed some counterevidence to Bello's traditional but convincing account. However, findings were drawn exclusively from a newspaper register and from a corpus of limited size. For reliability and generalizability of the findings, they need to be reexamined and retested using a larger corpus. Moreover, the judgments on matrix verbs are somewhat subjective, and a more objective approach is in order.

## 4. Present study

In the present study, we discuss a new set of corpus data, the recently constructed Corpus Brasileiro, which is the first corpus of more than a billion words in Brazilian Portuguese and which has been newly added to Sketch Engine's default corpora. Findings based on this much larger corpus are compared to those in Torigoe (2011), and implications of the comparison are explored.

### 4. Methods

The Corpus Brasileiro is much larger and more general than the Cetempúblico/Cetenfolha. Compared to the limited availability of the Cetempúblico/Cetenfolha on the Sketch Engine, the Corpus Brasileiro is fully available. In addition, although it is limited to the Brazilian variety of Portuguese, its composition of 24 written and five spoken genres promises opportunities for richer insight.

The search protocol is similar to the one in the previous study: the author began by searching 'node morphemes' and; filtered these by collocation with the conjunction *que* on left 4-2 nodes and 3-1 nodes (4-node and 2-node); removing periods. Punctuation was removed, and lists of collocations on left 6-5 nodes and 4-2 nodes were extracted for each pattern. However, because the POS annotation in the Corpus Brasileiro is much poorer than that of the Cetempúblico/Cetenfolha, especially on the Future Preterit forms and because there are so many 'noise' in the data such as the word *história* ('history') mistakenly tagged as "V", these forms were also limited to *ser, ter, estar, poder, and fazer*. From each collocation list, 300 matrix verbs were selected, in contrast to the 100 verbs in the previous study.

To determine the matrix verbs' preference for one pattern or the other, Pearson's chi-square test ( $\alpha=.05$ ) was adopted. For this, the author used a Microsoft Excel<sup>®</sup> VBA Macro file, appended in Ishikawa, Maeda, and Yamazaki (2010), which enables word-by-word comparison of the differences in collocational preference within each subcorpus, or, in this study, within the Future/Future Preterit or the Present/Preterit 'pattern' (Hunston & Francis, 1999). This provides a more systematic and objective judgment, comprehensively calculated by the subtotals and the expectations of all matrix verbs in all patterns together. In contrast, Torigoe (2011) employed the simple sum of the partial calculations (each score having been calculated separately for each pattern) and the judgments of the author. To avoid overestimation, matrix verbs with an observed frequency or an expected frequency of less than five were excluded from 'significant' lexis.

### 4.2. Findings

A total of 516 verbs in 4-node and 529 verbs in 2-node were analysed in the new study. As in Torigoe (2011), the global collocational tendency is statistically similar for the two morphosyntactic patterns. The correlation scores are

$r=0.93$  in 4-node and  $r=0.92$  in 2-node. However, as in the previous study, there are certain matrix verbs which predominantly collocate with the Future/Future Preterit pattern and others which prefer the Present/Preterit pattern. Figures 3 and 4 show that many matrix verbs predominantly collocate with the Future/Future Preterit (the plots on or near the horizontal baseline) or with the Present/Preterit pattern (the plot on or near the vertical baseline).

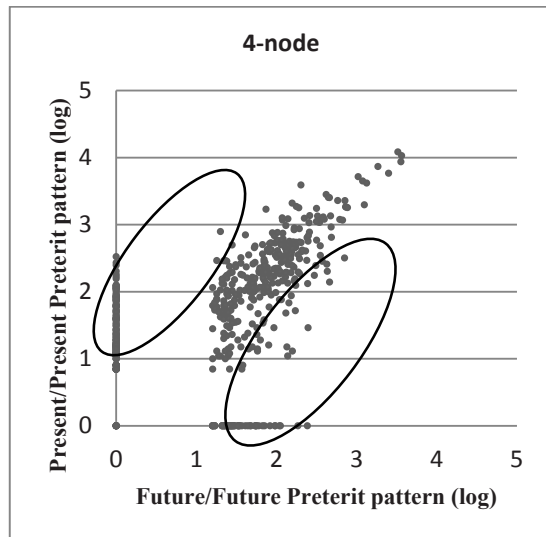


Figure 3. Frequency plot of matrix verbs (4-node)

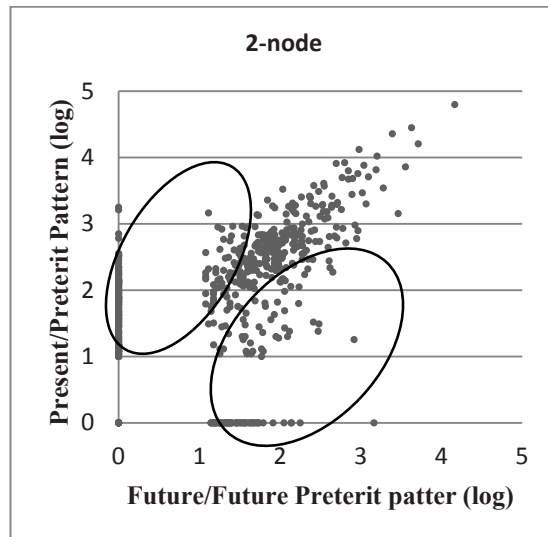


Figure 4. Frequency plot of matrix verbs (2-node)

Then these verbs are identified by the significant difference between two patterns calculated by chi-square test. In 4-node, 262 words, and in 2-node, 132 words statistically prefer the Future/Future Preterit pattern. Similarly, nine verbs in 4-node and 92 verbs in 2-node significantly collocate with the Present/Preterit pattern. The author then selected the matrix verbs which significantly collocate with each pattern in both two node scopes, then, filtered the verbs that do not tend to compose nominal completive clauses (e.g. *buscar* ‘to look for’ or *enviar* ‘to send’) or the verbs which, in required ‘node’ scopes, are used as nouns or participles (e.g. *parlamentar* ‘to parley’ or *participar* ‘to participate’). As a result, 51 verbs collocate with the former pattern, and only eight verbs with the latter.

Compared with Torigoe (2011), in which 21 verbs collocated with the former and 38 verbs with the latter, this study shows strikingly more matrix verbs which prefer the Future/Future Preterit pattern and fewer verbs which prefer the Present/Preterits pattern. This suggests that, in the present study, much of the lexis which collocated with the Present/Preterit pattern in the previous study is calculated as a ‘general group’ which collocates with both patterns. Only two of these lexical items are analysed as significant collocations with the Future/Future Preterit pattern in this study. However, the relatively lower subtotal of verb frequency in the Future/Future Preterit pattern than those in Present/Preterit pattern might lead the chi-square method to overestimate the collocational significance of the former pattern.

Among the two studies, the matrix verb lexis in one or the other pattern somewhat overlap. The group of verbs preferring the Future/Future Preterit pattern in this study is quite similar to that in the previous study: Of the 21 verbs in Torigoe (2011)’s findings, 14 verbs are confirmed in the present study (see the appendix). Although only eight verbs are shown to collocate with the Present/Preterit pattern in this study, more than the half of them were also found in the previous study. Meanwhile, it is worth noting that, if we limit ourselves to the in 2-node results, 92 verbs statistically prefer this pattern, 21 of which overlap with the 38 lexical items found in the previous study.

In addition, some verbs are found which could or should follow the Subjunctive in completive clause, such as *esperar* (to hope), *evitar* (to evade), *imaginar* (to imagine), *implicar* (to imply), *negar* (to negate), *obter* (to obtain), *proponer* (to propose). Based on the KWIC concordance lines and sample texts, the author confirmed that the majority of matrix verbs follow the Subjunctive, but some follow the Future/Future Preterit, to some extent. This is an area worthy of further exploration.

## 5. Limitations

The findings of the present study, as well as those of Torigoe (2011), are partial ones, for they focus only on the case of nominal completive clauses. The Future/Future Preterit and Present/Preterit are also intermixed in adverbial and relative adjectival clauses. In order to support and generalize the third mood hypothesis, it will be necessary to clarify the cases in these clauses and synthesize them all.

Furthermore, the two studies are mainly based on a quantitative approach that moves “from pattern to lexis” (Hunston and Francis 1999). A more precise understanding may be gained through intensive qualitative analysis, which might offer complementary confirmation by moving ‘from lexis to pattern’ and at the same time avoid ‘noise’ in data. In addition to these linguistic perspectives, pragmatic approaches and sociocultural analysis have a role to play as well, for they offer insights on how context and strategy impact the use and choice of the Future/Future Preterit and the Present/Preterit forms.

## Acknowledgements

This study is supported by Japan Society for the Promotion of Science.

## Appendix A. Matrix verbs which significantly collocate with each pattern (Y: the verb collocated with the same pattern in the previous study.)

Future/Future Preterit pattern	Torigoe (2011)	Future/Future Preterit pattern	Torigoe (2011)	Present/Preterits pattern	Torigoe (2011)
acreditar		estabelecer		constatar	Y
acrescentar		estimar	Y	demonstrar	Y
adiantar	Y	evitar		esquecer	Y
admitir		garantir	Y	mostrar	Y
advertir		identificar		observar	
alegar		imaginar	Y	perceber	
alertar		implicar		registrar	
anunciar	Y	indicar	Y	verificar	Y
apontar		informar			
apostar		julgar	Present/Preterit		
acreditar		levantar			
acrescentar		negar			
apurar		obter			
avaliar	Y	pretender			
avisar	Y	prometer	Y		
comunicar		propor			
confirmar		provocar			
constar		queixar			
Crer		reafirmar	Y		
decidir	Y	realizar			
declarar		resolver			
definir	Y	seguir			
determinar	Y	sentir	Present/Preterit		
discutir		significar			

divulgar	sinalizar	
dizer	supor	Y
esperar	tentar	

---

## References

- Alarcos Llorach, E. (1980). *Estudios de Gramática Funcional del Español* (tercera edición). Madrid: Editorial Gregos.
- Alarcos Llorach, E. (1994). *Gramática de la Lengua Española*. Madrid: Espasa Calpe.
- Barros, J. de. (1540). *Gramática da Língua Portuguesa*. In J. de Barros. (1971). *Gramática da Língua Portuguesa*. Lisboa: FLUL.
- Bello, A. (1847/1988). *Gramática de la Lengua Castellana*, destinada al Uso de los Americanos (1). Madrid: Arco Libros.
- Bybee, J., Perkins, R., and Pagliuca, W. (1994). *The Evolution of Grammar – Tense, Aspect, and Modality in the Languages of the World*. Chicago: the University of Chicago Press.
- Castronovo, B. J. (1989). The Strange History of -ría Form. *Hispania* 72, 378-384.
- Cunha, C. e Cintra, L. F. L. (2007). *Nova Gramática do Português Contemporâneo*; 4ª edição Revista e Ampliada. Rio de Janeiro: Lexikon.
- Deguchi, A. (1986). Un nuevo modo "presuntivo" en sustitución de los tiempos futuro y condicional (written in Japanese). *Linguística Hispánica*, 9, pp.19-34.
- Dunning, T. (1993). Accurate Methods for the Statistic of Surprise and Coincidence. *Computational Linguistics*, 19 (1), 61-74.
- Fleischman, S. (1982/2009). *Future in Thought and Language*. Cambridge: Cambridge University Press.
- Ishikawa, S.; Maeda, T.; and Yamazaki, M. (2010). *An Introduction in Statics for Language Research* (written in Japanese). Tokyo: Kuroshio Publisher
- Givón, T. (1994). Irrealis and the Subjunctive. *Studies in Language* 18(2), 265-337.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. and Francis, G. (1999). *Pattern Grammar*. Amsterdam: John Benjamin.
- Lyons, J. (1977). *Semantics* 2. Cambridge: Cambridge University Press.
- McEnery, T.; Xiao, R.; and Tono, Y. (2006). *Corpus-Based Language Studies; an Advanced Resource Book*. London: Routledge.
- Nebrija, A. de. (1492/1980). *Gramática de la Lengua Castellana*. In Antonio Quilis (Ed). (1980) *Gramática de la Lengua Castellana*. Madrid: Editora Nacional.
- Palmer, F. R. (2001). *Mood and Modality* (2<sup>nd</sup> edition). Cambridge: Cambridge University Press.
- Resnick, M. C. (1984). Spanish Verb Tenses: Their Names and Meanings. *Hispania* 67, 92-99.
- Terasaki, H. (1998). *Estructura de Gramática Española* (written in Japanese). Tokyo: Daigaku-Shorin.
- Torigoe, S. (2011). A Similaridade Colocacional entre Futuro e Futuro do Pretérito, e Presente e Pretéritos do Indicativo do Português (written in Japanese). *Romance Studies*, vol. 44, 41-50.
- Corpus Brasileiro: <http://corpusbrasileiro.pucsp.br/x/>
- LINGUATECA: <http://www.linguateca.pt/>
- Sketch Engine: <https://the.sketchengine.co.uk/>