PROPBANK-BR

This corpus presents a layer of semantic role labels over the syntactic trees of the Brazilian portion of corpus Bosque, the manually revised subcorpus of Floresta Sintá(c)tica[1] (Affonso et al, 2002), parsed by Palavras (Bick, 2000). Bosque has 9368 sentences and 4213 of them correspond to the Brazilian portion (extracted from the journal Folha de São Paulo of 1994). The annotation tool adopted is SALTO (Burchardt et al, 2008).

Read more in:

Magali Sanches Duran, Sandra Maria Aluísio "Propbank-Br: a Brazilian Portuguese corpus annotated with semantic role labels". In the Proceedings of the 8th Symposium in Information and Human Language Technology, October 24-26, Cuiabá/MT, Brazil.

Contact: magali.duran@gmail.com

DETAILS ABOUT CORPUS ANNOTATION

The set of semantic role labels is identical to the ones created by Propbank (Palmer et al. 2005).

LABEL	CONTENT
arg0	agent or causer
arg1	Patient, experiencer or theme
arg2	defined in the frames files* of each verb
arg3	idem
arg4	idem
arg5	idem
argm-adv	adverbial
argm-cau	cause
argm-dir	direction
argm-dis	discourse
argm-ext	extension
argm-loc	locative
argm-mnr	manner
argm-neg	negation
argm-pnc	purpose
argm-prd	secondary predication
argm-rec	reciprocal
argm-tmp	time

* Frames files are under construction and will be made available as soon as possible.

Following Propbank guidelines (http://verbs.colorado.edu/~mpalmer/projects/ace/PBguidelines.pdf) each sentence of the original corpus (Bosque corpus) was repeated as many times as the

---

[1] http://linguateca.pt

number of main verbs it contains. This strategy allowed annotating each verb of the sentence in a different instance.

All the verbs that have an auxiliary or semi-auxiliary role were not annotated. In the near future, these verbs will be automatically annotated with special tags: "argm-aux-mod" for modals, "argm-aux-tmp" for temporal, "argm-aux-asp" for aspectual and "argm-aux-pas" for auxiliaries of passive voice constructions (see Table of auxiliaries in the Appendix).

The annotation of semantic role labels is found in the corpus under <frames>.

ADDITIONAL ANNOTATION

Aiming at providing extra features for machine learning (ML) and input for other NLP tasks, extra sentence annotation is provided, as follows:

1. SENTENCE FLAGS

    a. ELIPSE = one argument of the verb annotated is elliptic in the clause and the annotation assign the role label to the argument of another clause of the same sentence, which corresponds to the elliptic argument. This type of occurrence may present some difficulties for machine learning, so that one may opt for suppressing these sentences from training corpus or using the sentence flag "ELIPSE" as a feature.

    b. CORREF = one argument of the verb annotated is represented by a pronoun that co-refers to a constituent of another clause in the same sentence. However, as almost always the co-referred element is not annotated as a constituent in the parser output, the role label is assigned to the pronoun. These sentences may be suppressed from training corpus or used to make a pre-processing step of co-reference resolution.

    c. OCULTO = the subject of the annotated verb has been omitted. In Portuguese, unlike in English, the subject may be omitted and inferred from verbal inflection. As well, these sentences may be used for a pre-processing step aiming at providing an artificial subject inferred from the verb inflection.

    d. INDETERMINADO = the subject of the annotated verb has been omitted and is not inferable from verbal inflection.

2. SENTENCE FLAGS PARAMETERS THAT MAY BE USED TO CREATE SUBCORPUS FOR OTHER NLP TASKS

    a. MWE: label used to identify sentences containing multi-word expressions.

    b. CP: label used to identify sentences containing complex predicates

c. "SE" particle annotation. The "se" in Portuguese is a very ambiguous lexical item. The parser output is successful in distinguishing "se" conjunction from "se" pronoun. The "se" pronoun has seven functions (see table below) and only three of them may support a semantic role label (SE-REF-OD, SE-REF-OI and SE-REF-REC).

The sentences containing "se" pronoun were annotated with the respective function of this particle in the sentence. These sentences may be used to learn how to disambiguate "se" functions in order to improve the parser output.

Label	Content	Example
SE-REF-OD	Reflexive pronoun (Direct Object)	Ele se feriu
SE-REF-OI	Reflexive pronoun (Indirect Object)	Ele se deu um presente
SE-REC	Reciprocal pronoun	Eles se encontraram
SE-PAS	Particle used to construct passive voice	Vendem-se casas
SE-IND	Particle that indicates subject indeterminacy	Concordou-se com tudo.
SE-EXP	Expletive particle	Acabou-se a festa.
SE-VPR	Particle that integrates the verb	Apaixonou-se e arrependeu-se.

d. INFINITE CLAUSES. All the infinite clauses were annotated with their function (Substantive, Adjective or Adverb). Moreover, adverbial infinite clauses received semantic annotation according to their meaning. The set of sentence flags parameters used to annotate these clauses is presented in the following table.

SENTENCE FLAG	Content
GERUNDIO-ADJ	gerund infinite clause Adjective
GERUNDIO-SADV-Causal	gerund infinite clause Adverbial-cause
GERUNDIO-SADV-Concessiva	gerund infinite clause Adverbial-concessive
GERUNDIO-SADV-Condicional	gerund infinite clause Adverbial-conditional
GERUNDIO-SADV-Consecutiva	gerund infinite clause Adverbial-result
GERUNDIO-SADV-Modal	gerund infinite clause Adverbial-manner
GERUNDIO-SADV-Temporal	gerund infinite clause Adverbial-time
INFINITIVO-SADJ	infinitive infinite clause Adjective
INFINITIVO-SADV-Causal	infinitive infinite clause Adverbial- cause
INFINITIVO-SADV-Concessiva	infinitive infinite clause Adverbial-concessive
INFINITIVO-SADV-Condicional	infinitive infinite clause Adverbial-conditional
INFINITIVO-SADV-Consecutiva	infinitive infinite clause Adverbial-result
INFINITIVO-SADV-Final	infinitive infinite clause Adverbial-purpose
INFINITIVO-SADV-Temporal	infinitive infinite clause Adverbial-time
INFINITIVO-SSA	infinitive infinite clause Substantive-appositive
INFINITIVO-SSCN	infinitive infinite clause Substantive-noun complement
INFINITIVO-SSOD	infinitive infinite clause Substantive-direct object
INFINITIVO-SSOI	infinitive infinite clause Substantive-object of a preposition
INFINITIVO-SSP	infinitive infinite clause Substantive-predicative
INFINITIVO-SSS	infinitive infinite clause Substantive-subject
PARTICIPIO-SADJ	past participle infinite clause Adjective
PARTICIPIO-SADV-Causal	past participle infinite clause Adverbial-cause
PARTICIPIO-SADV-Concessiva	past participle infinite clause Adverbial-concessive
PARTICIPIO-SADV-Condicional	past participle infinite clause Adverbial-conditional
PARTICIPIO-SADV-Temporal	past participle infinite clause Adverbial-time

APPENDIX

TABLE OF AUXILIARIES

VERB	PREPOSITION	AUXILIATED INFINITE FORM	AUXILIAR FUNCTION	ROLE LABEL TO BE ASSIGNED
ser		particípio	passiva	Argm-aux-pas
estar		particípio	passiva	Argm-aux-pas
haver		particípio	temporal	Argm-aux-tmp
ir		infinitivo	temporal	Argm-aux-tmp
ter		particípio	temporal	Argm-aux-tmp
ameaçar		infinitivo	modal	Argm-aux-mod
conseguir		infinitivo	modal	Argm-aux-mod
dever		infinitivo	modal	Argm-aux-mod
ficar	de	infinitivo	modal	Argm-aux-mod
haver	de	infinitivo	modal	Argm-aux-mod
haver	que	infinitivo	modal	Argm-aux-mod
hesitar		infinitivo	modal	Argm-aux-mod
hesitar	em	infinitivo	modal	Argm-aux-mod
necessitar		infinitivo	modal	Argm-aux-mod
poder		infinitivo	modal	Argm-aux-mod
precisar		infinitivo	modal	Argm-aux-mod
procurar		infinitivo	modal	Argm-aux-mod
querer		infinitivo	modal	Argm-aux-mod
saber		infinitivo	modal	Argm-aux-mod
ser	de	infinitivo	modal	Argm-aux-mod
tentar		infinitivo	modal	Argm-aux-mod
ter	de	infinitivo	modal	Argm-aux-mod
ter	que	infinitivo	modal	Argm-aux-mod
acabar		gerúndio	aspectual	Argm-aux-asp
acabar	de	infinitivo	aspectual	Argm-aux-asp
acabar	por	infinitivo	aspectual	Argm-aux-asp
andar		gerúndio	aspectual	Argm-aux-asp
cessar	de	infinitivo	aspectual	Argm-aux-asp
chegar	a	infinitivo	aspectual	Argm-aux-asp
começar	a	infinitivo	aspectual	Argm-aux-asp
começar	por	infinitivo	aspectual	Argm-aux-asp
continuar		gerúndio	aspectual	Argm-aux-asp
continuar	a	infinitivo	aspectual	Argm-aux-asp
correr	a	infinitivo	aspectual	Argm-aux-asp
costumar		infinitivo	aspectual	Argm-aux-asp
dar	para	infinitivo	aspectual	Argm-aux-asp
dar	de	infinitivo	aspectual	Argm-aux-asp
deixar	de	infinitivo	aspectual	Argm-aux-asp
desatar	a	infinitivo	aspectual	Argm-aux-asp
estar		gerúndio	aspectual	Argm-aux-asp
estar	para	infinitivo	aspectual	Argm-aux-asp
estar	para	infinitivo	aspectual	Argm-aux-asp
estar	por	infinitivo	aspectual	Argm-aux-asp
ficar		gerúndio	aspectual	Argm-aux-asp
ir		gerúndio	aspectual	Argm-aux-asp
ousar		infinitivo	aspectual	Argm-aux-asp
parar	de	infinitivo	aspectual	Argm-aux-asp

passar	a	infinitivo	aspectual	Argm-aux-asp
permanecer		infinitivo	aspectual	Argm-aux-asp
pôr-se	a	infinitivo	aspectual	Argm-aux-asp
recomeçar	a	infinitivo	aspectual	Argm-aux-asp
sair		gerúndio	aspectual	Argm-aux-asp
seguir		gerúndio	aspectual	Argm-aux-asp
terminar		gerúndio	aspectual	Argm-aux-asp
tornar	a	infinitivo	aspectual	Argm-aux-asp
vir		gerúndio	aspectual	Argm-aux-asp
vir	a	infinitivo	aspectual	Argm-aux-asp
viver		gerúndio	aspectual	Argm-aux-asp
voltar	a	infinitivo	aspectual	Argm-aux-asp