The FrameCorp Project:
A preliminary report

Rove Chishman, Anderson Bertoldi, Maristela Kohlrausch
Universidade do Vale do Rio dos Sinos
Graduate Program in Applied Linguistics
93022000 São Leopoldo, RS, Brazil
rove@unisinos.br, andersonbertoldi@yahoo.com, makohlrausch@gmail.com

The FrameCorp Project can be described as an initiative to construct a semantically annotated corpus of Brazilian Portuguese with frame information. This work is a first experiment to discover some problems that FrameCorp Project will face. The corpus of our research, named SUMM IT, is composed by 50 texts, totalizing 17.123 tokens. Other associated researchers have still added anaphoric and rhetorical relations. Our task is to add a layer of frame information. As a characteristic of these texts, they are composed by many sentences in reported speech. Then, we decided to start annotation by Statement frame. The Statement frame describes a situation in which a Speaker talks about a Topic or sends a Message through a Medium. The verbs of Statement have been studied in other approaches. Levin (1984) treats these verbs as Communication Verbs, creating some classification for then, like Verbs of Transfer of a Message, Verbs of Manner of Speaking, and Verbs of Instrument of Communication. Neves (2000) classifies the verbs of Statement as Verbs of Elocution. One of the objectives of this study is to verify whether the FrameNet approach captures the semantic richness of these verbs. After have chosen the frame “Statement” to annotate, the next step was finding Portuguese translation equivalents for the entries listed in the FrameNet database. All the lexical units evoking Statement frame were manually identified.

After reaching the final list of the lexical units evoking Statement frame, we started to compare the syntactic and semantic combinatorial possibilities between English and Portuguese. The study of the lexical unit dizer, say in English, has showed that the lexical units of Statement have in Portuguese a similar behavior to those in English. Through the use of SALTO annotation tool, we could certify that the FrameNet database fruitfully be used to annotate Portuguese sentences. We found in our corpus 7 different lexical units evoking the Statement frame, all of them having an English translation equivalent in FrameNet. Statement verbs seem to be a critical class. While there are verbs semantically more neutral, like “say”, others are less neutral. There are verbs that lexicalize the mean by which the statement was made, like “write”. Others express a positive evaluation, like “affirm”, or a negative, like “deny”. We have seen in the corpus analyses that these verbs can be exchanged for enunciative reasons. Even though, this is not the objective of FrameNet, one of the solutions for this enunciative phenomenon could be the creation of semantic relations to represent elocution verbs.