

# **Project resources**

Bernardo Severo  
2014-10-29

# Table of Contents

Machine translation of phrasal verbs .....	3
JeuxDeMots Português .....	3
Conference comparable corpus .....	3
Semantic Role Labeling - Brazilian Portuguese .....	3
VOAR - A Visual and Integrated Ontology Alignment Environment .....	3

This page contains a repository of freely available resources generated in the context of the CAMELEON project. The resources can be used for research purposes under the terms of [GNU GPL](#) license.

## Machine translation of phrasal verbs

We evaluated the difficulty in translation English phrasal verbs (e.g. *give up*, *take off*) into French using a standard Moses SMT system. We focused on transitive, split occurrences (e.g. *take my shirt off*) and compared hierarchical and phrase-based models. Please click on the resources below or contact us if you need further information.

- [Annotation guidelines, BLAST tool and data to annotate](#)
- [Annotated data](#)
- [Tools and scripts used to create and process annotation data](#)
- [Phrasal verb synsets in Wordnet](#)
- [Article describing the experiments, published in MUMTTT 2013](#) - How hard is it to automatically translate phrasal verbs from English to French?

This work was carried out by A. Kobzar, under the supervision of C. Ramisch and L. Besacier

## JeuxDeMots Português



The Portuguese version of the serious lexical game JeuxDeMots is online! Go to the website and try it out now!

[JeuxDeMots Português](#)

Also, do not forget to like us on [Facebook](#) :-)

## Conference comparable corpus

This is a corpus generated using the labels of ontologies describing the conference domain. The labels were combined and used as Google queries for retrieving web documents, which were further downloaded, cleaned and parsed. The corpus is aimed at supporting ontology research and is available in English, Portuguese and French.

- [Go to GIT HUB repository page](#)
- [Download version 0.1 - raw corpus and sample of annotated corpus](#)

Corpus dimensions:

Language	Documents	Sentences	Word tokens	Word types
English	10,127	1,119,971	15,852,659	459,501
Portuguese	5,342	641,452	12,876,344	405,623
French	5,154	595,938	9,482,156	362,548

## Semantic Role Labeling - Brazilian Portuguese

Check [this page](#) .

## VOAR - A Visual and Integrated Ontology Alignment Environment

VOAR is a new web-based environment for ontology alignment visualization and manipulation. Built in order to provide a centralized e optimized resource for a series of alignment operations.

Within this graphical environment, users can manually create/edit correspondences and apply a set of operations on alignments (filtering, merge, difference, etc.). VOAR allows invoking external ontology matching systems that implement a specific alignment interface, so that the generated alignments can be manipulated within the environment. Evaluating multiple alignments together against a reference one can also be carried out, using classical evaluation metrics (precision, recall and f-measure). The status of each correspondence with respect to its presence or absence in reference alignment is visually represented.

In the process of research and development of such solution, it was important the collaboration with Cassia Trojahn dos Santos, improved with the in place coordination provided by travelling to Brazil. As result, a paper describing the resource was published and presented on the 9th International Conference on Language Resources and Evaluation (LREC 2014): [VOAR: A Visual and Integrated Ontology Alignment Environment](#) .

The VOAR software is publicly available in the following URL: <http://voar.inf.pucrs.br>