VISL – Multilingual Dependency Parser

1. BASIC INFORMATION
    1. Tool name: *VISL – multilingual dependency parser*
    2. Overview and purpose of the tool: Web application parser for several European languages (English, Spanish, French, Italian, Swedish, Norwegian, German, etc.)
    3. A short description of the algorithm: *VISL's own Constraint Grammar systems are inspired by Eckhard Bick's PALAVRAS parser for Portuguese (Bick 2000), and use, as a novelty, subclause function, generalized dependency markers and semantic prototype tags. For most languages, a lexicon based morphological analyzer provides input to the first CG level, while the output of the last CG-level can be converted into syntactic tree structures by specially designed Phrase Structure Grammars (PSG's), using syntactic functions, not words, as terminals. Other, hybrid combinations are, however, feasible. Thus, the French system uses PoS information from a probabilistic tagger.*

    Constraint Grammar (CG) is a methodological paradigm for Natural Language Parsing (NLP). Linguist-written, context dependent rules are compiled into a grammar that assigns grammatical tags ("readings") to words or other tokens in running text. Typical tags address lemmatisation (lexeme or base form), inflexion, derivation, syntactic function, dependency, valency, case roles, semantic type etc. Each rule either adds, removes, selects or replaces a tag or a set of grammatical tags in a given sentence context. Context conditions can be linked to any tag or tag set of any word anywhere in the sentence, either locally (defined distances) or globally (undefined disances). Context conditions in the same rule may be linked, i.e. conditioned upon each other, negated or blocked by interfering words or tags. Typical CG's consist of thousands of rules, that are applied set-wise in progressive steps, covering ever more advanced levels of analysis. Within each level, safe rules are used before heuristic rules, and no rule is allowed to remove the last reading of a given kind, thus providing for a hight degree of robustness.
2. TECHNICAL INFORMATION
    1. Software dependencies and system requirements
        Browser and internet connection.
    2. Installation
        n.a.
    3. Execution instructions
        Go to http://beta.visl.sdu.dk/visl/en/parsing/automatic/dependency.php , introduce an English paragraph and press the Go! button.
    4. Input/Output data formats
        a. Input data formats
            Plain text.
        b. Output data formats
            Lemattized, Pos-tagged, syntactically annotated text.
    5. Integration with external tools
        Fully self-contained.

3. CONTENT INFORMATION
    1. a test input file
        "The task of extracting translation equivalents from bilingual corpora has been approached in different manners, according to the degree of parallelism between the source and target parts of the corpora involved."
    2. the output file
        The [the] ART S/P @>N #1->2

task [task] N S NOM @SUBJ> #2->4
of [of] PRP @N< #3->2
extracting [extract] <mv> V PCP1 @ICL-P< #4->0
translation [translation] N S NOM @<ACC #5->4
equivalents [equivalent] N P NOM @N< #6->5
from [from] PRP @<PIV #7->4
bilingual [bilingual] ADJ POS @>N #8->9
corpora [corpus] N P NOM @P< #9->7
has [have] <aux> V PR 3S @FS-STA #10->0
been [be] <aux> V PCP2 AKT @ICL-AUX< #11->10
approached [approach] <mv> V PCP2 PAS @ICL-AUX< #12->11
in [in] PRP @<ADVL #13->12
different [different] ADJ POS @>N #14->15
manners [manners] N P NOM @P< #15->13
, [,] PU @PU #16->0
according to [according=to] PRP @<ADVL #17->12
the [the] ART S/P @>N #18->19
degree [degree] N S NOM @P< #19->17
of [of] PRP @N< #20->19
parallelism [parallelism] N S NOM @P< #21->20
between [between] PRP @<ADVL #22->12
the [the] ART S/P @>N #23->24
source [source] N S NOM @P< #24->22
and [and] KC @CO #25->24
target [target] N S NOM @>N #26->27
parts [part] N P NOM @P< #27->24
of [of] PRP @N< #28->27
the [the] ART S/P @>N #29->30
corpora [corpus] N P NOM @P< #30->28
involved [involve] <mv> V PCP2 PAS @ICL-N< #31->30
. [.] PU @PU #32->0

3. approximation of the time necessary to process the test input file.
   Approx.. 3 seconds.

4. ADMINISTRATIVE INFORMATION

   Contact: Eckhard Bick, eckhard.bick@mail.dk

5. RELEVANT REFERENCES AND OTHER INFORMATION

Bick, Eckhard (2004), http://beta.visl.sdu.dk/pdf/Atala2004.ps.pdf, In: Patrick Paroubek, Isabelle Robba & Anne Vilnat: Méthodes et outils pour lévaluation des analyseurs syntaxiques (Journée ATALA, May 15, 2004). pp. 4-9. Paris: ATALA.

Bick, Eckhard (2010), http://beta.visl.sdu.dk/pdf/frag2009.pdf, In: Calzolari, Nicoletta et al. (eds.), Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC2010 (Malta, May 19-21, 2010). (http://www.lrec-conf.org/proceedings/lrec2010/Summaries/688.html