



Dependency and Constituency in Translation Shift Analysis

M. Sanguinetti
C. Bosco
L. Lesmo

Dipartimento di Informatica,
Università di Torino

{manuela.sanguinetti; cristina.bosco; leonardo.lesmo}@unito.it

Overview

Aim of our research

- to create a syntax-driven alignment system for parallel parse trees

Initial assumption

- the use of syntactic information on dependency relations and on predicative structure provided by annotated corpora can be useful while tackling the alignment task
- structural alignment should take account of **translation shifts**, i.e. the departure from formal correspondence when going from a source to a target text (Catford, 1965)

Aim of this study

- in order to examine whether and to what extent dependencies are able to capture parallelisms, we compared them to a constituency representation

Translation shifts in ParTUT

Structural alignment should take account of translation shifts

In ParTUT we identified three main classes of shifts:

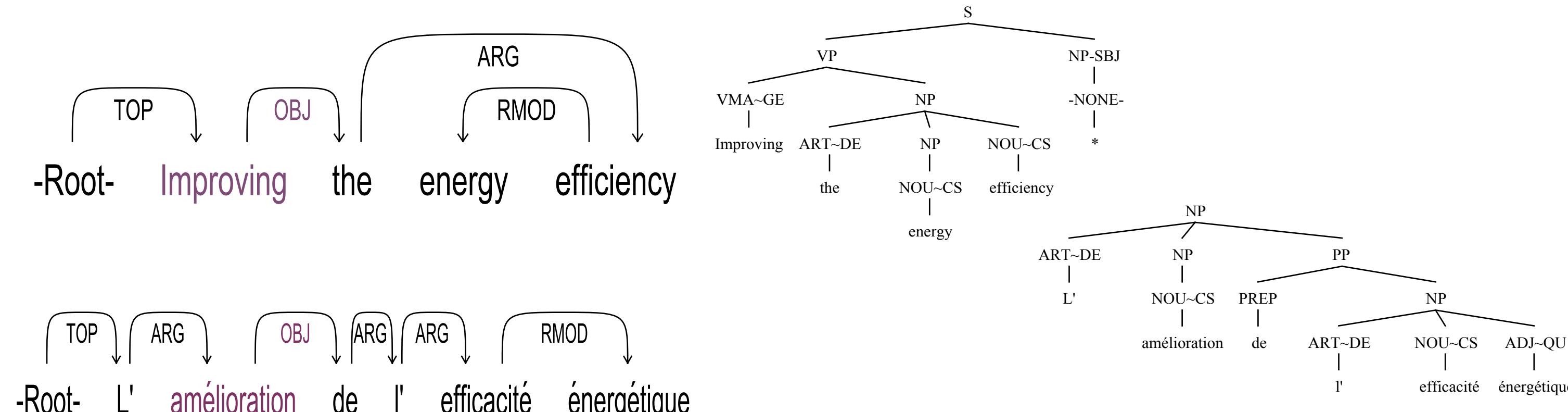
Category shifts (divergences in the Part of Speech)

Structural shifts (divergences in syntactic realizations at many levels, e.g. diathesis)

Semantic shifts (divergences in meaning)

Cross-format comparison

A) Nominalization (Category shift)



EN: Improving the energy efficiency
FR: Amélioration de l'efficacité énergétique

Resource used: ParTUT

a multilingual parallel treebank for Italian, English and French (approx. 89,000 tokens)

<http://www.di.unito.it/~tutreeb/partut.html>

FORMATS USED FOR COMPARISON:

Dependency: **native TUT**

- centered upon the notion of **argument structure** and based on the principles of the **Word Grammar** (Hudson, 1984)
- it exploits **null elements** (pro-drop, equi, long distance dependencies, elliptical structures)
- compound nouns and contracted forms are split and each component is associated to a node

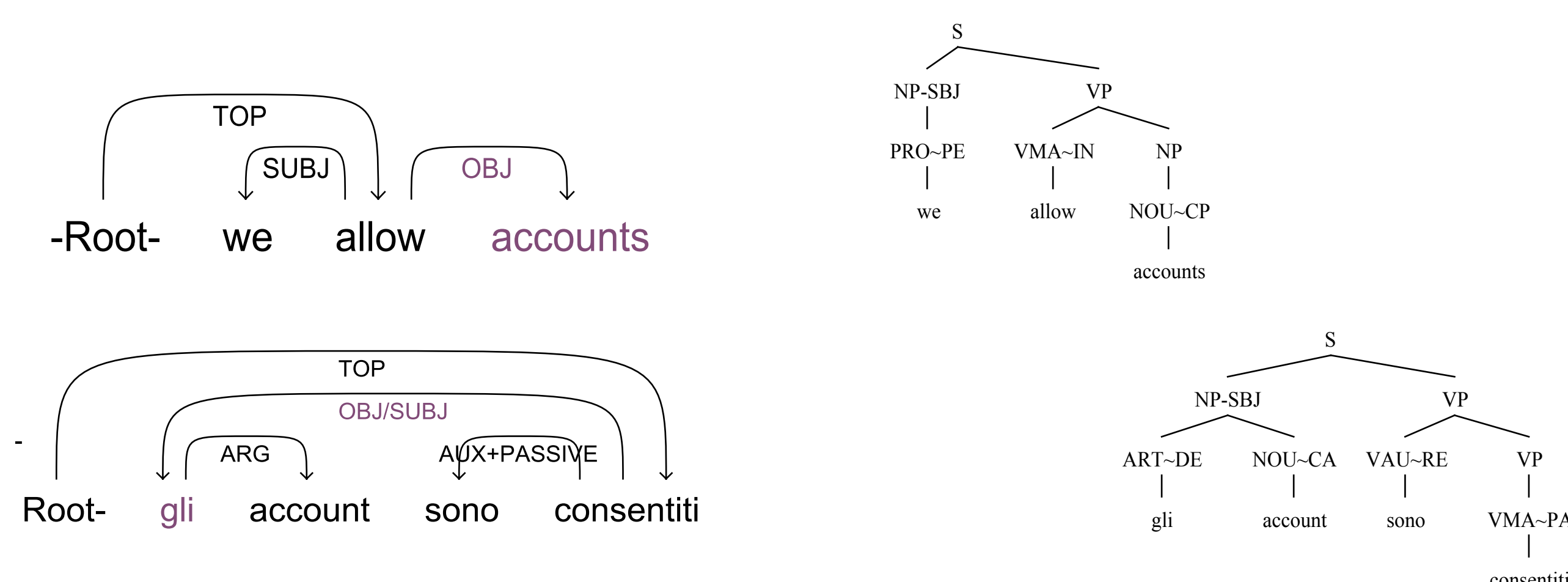
Constituency: **TUT-Penn**

- richer morphological tag set than the standard PTB
- extended inventory of functional relations

Alignment approach

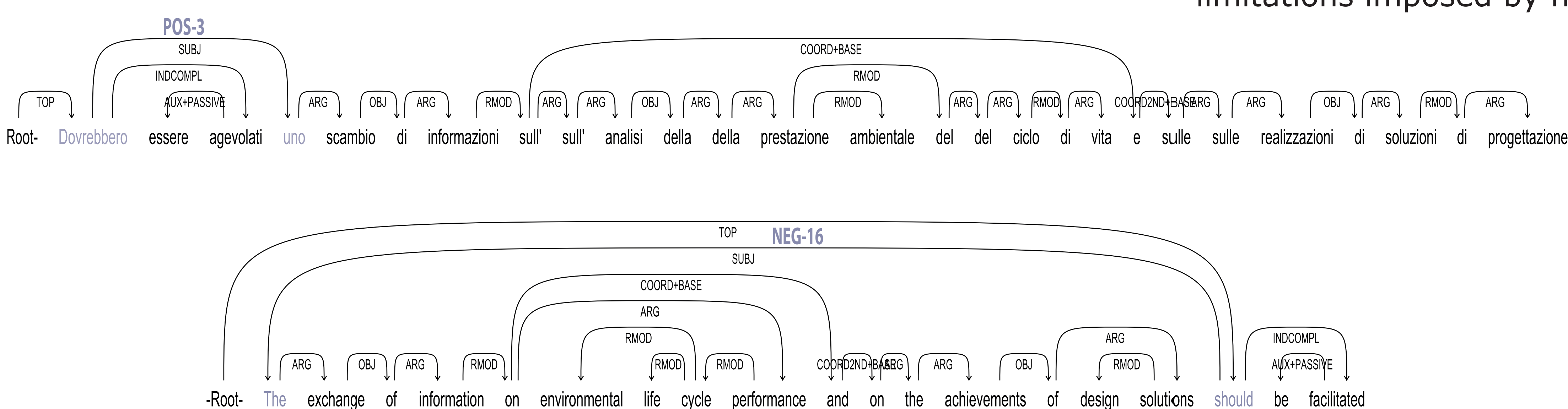
Basic design: lexical mapping and alignment expansion by means of relational information

B) Passivization (Structural shift)



EN: We allow accounts
IT: Gli account sono consentiti

C) Word order and long-distance dependencies (Structural shift)

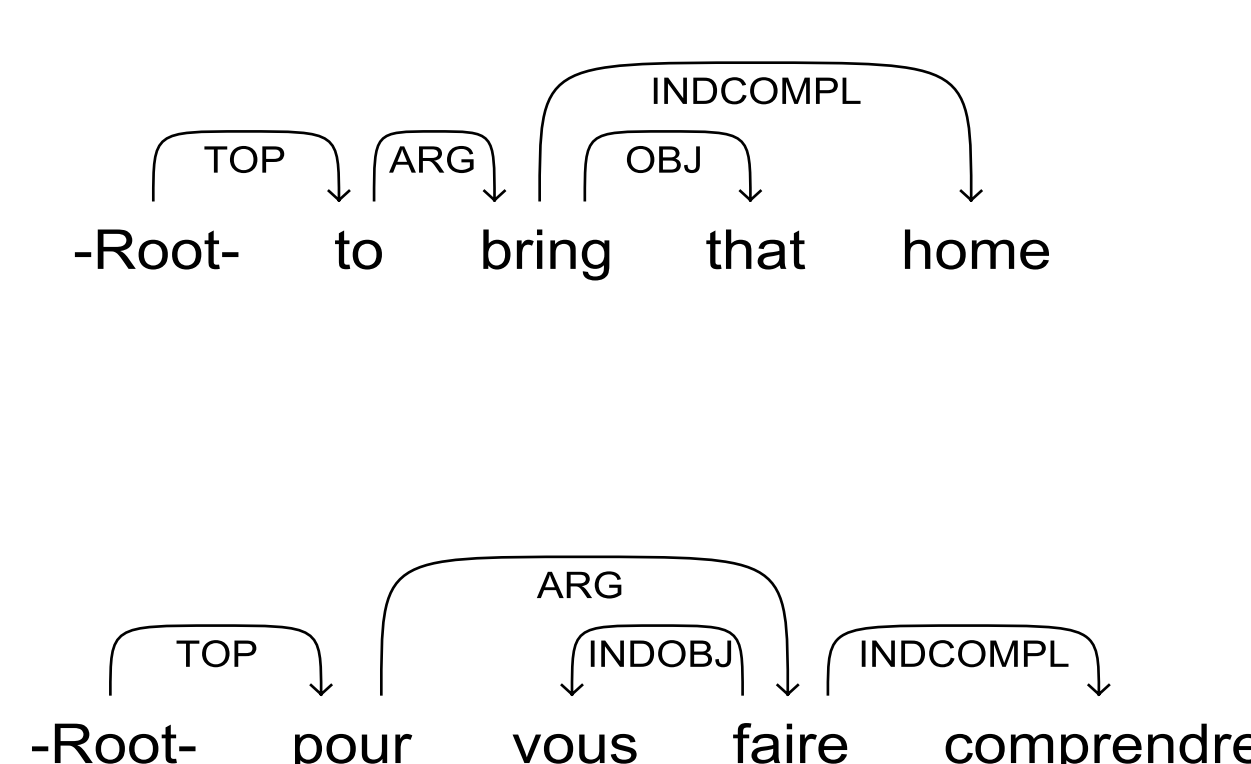


IT: Dovrebbero essere agevolati uno scambio di informazioni sull'analisi della prestazione ambientale del ciclo di vita e sulle realizzazioni di soluzioni di progettazione
EN: The exchange of information on environmental life cycle performance and on the achievements of design solutions should be facilitated

Closing Remarks

Some classes of shifts (e.g. where divergences are due to differences in the idiosyncratic use or the low compositionality) may require the integration of a more abstract notion of substructure which can be partially assimilated to that of constituency subtree, in order to link the entire substructure to its equivalent node.

This seems to us a viable solution that could balance the limits imposed by the format with the useful linguistic information it provides.



EN: To bring that home
FR: Pour vous faire comprendre

There are, however, more tricky cases, where neither lexical mapping, nor structural information could help (see ex. D)

D) Idioms (Structural shift)

