Sharing resources between free/open-source rule-based machine translation systems: Grammatical Framework and Apertium

Grégoire Détrez, Víctor M. Sánchez-Cartagena, Aarne Ranta

Friday 30 May 2014

GF & Apertium

Two machine translation platforms

- Rule based
- Open Source
- actively developed

But using different intermediate representation.

Example of translation in GF

Input

the red cars

Example of translation in GF

Input > English concrete tree

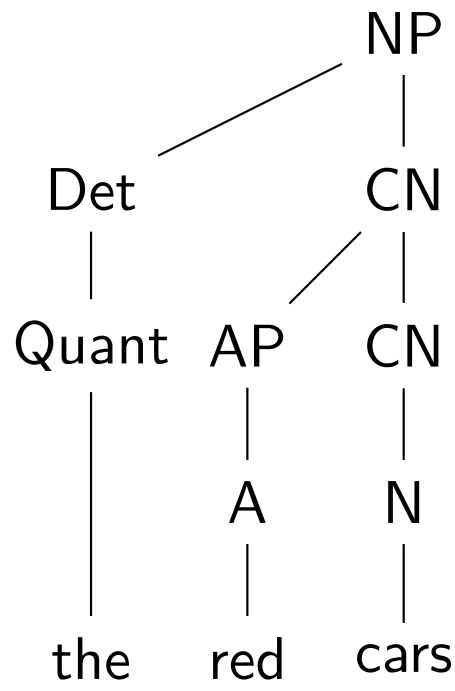


Figure : English concrete tree obtained after parsing

Example of translation in GF

Input > English concrete tree > Abstract syntax tree

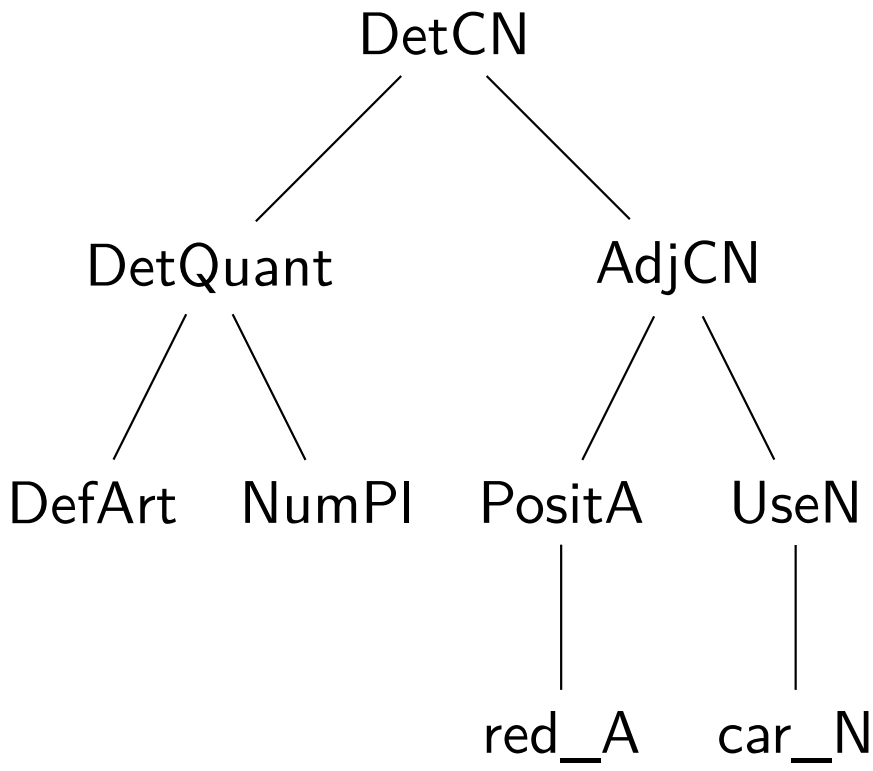


Figure : Abstract syntax tree

Example of translation in GF

Input > English concrete tree > Abstract syntax tree > Spanish concrete tree

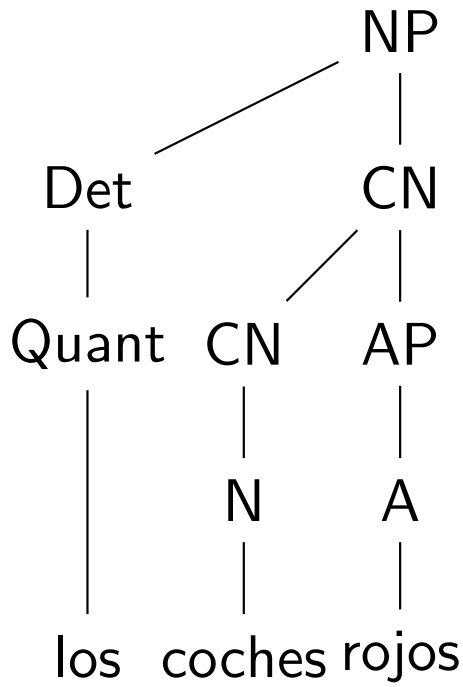


Figure : Spanish concrete tree

Example of translation in GF

Input > English concrete tree > Abstract syntax tree > Spanish concrete tree > Linearization

los coches rojos

Apertium translation

Input

the red cars

Apertium translation

Input > Intermediate representation

the.**det**.def *red*.**adj** *car*.**n**.pl

Apertium translation

Input > Intermediate representation > Transfer

the.**det**.def *red*.**adj** *car*.**n**.pl

Apertium translation

Input > Intermediate representation > Transfer

the.**det**.def *red*.**adj** *car*.**n**.pl

Apertium translation

Input > Intermediate representation > Transfer

el.**det**.def *red*.**adj** *car*.**n**.pl

Apertium translation

Input > Intermediate representation > Transfer

el.**det**.def *rojo*.**adj** *coche*.**n**.m.pl

Apertium translation

Input > Intermediate representation > Transfer

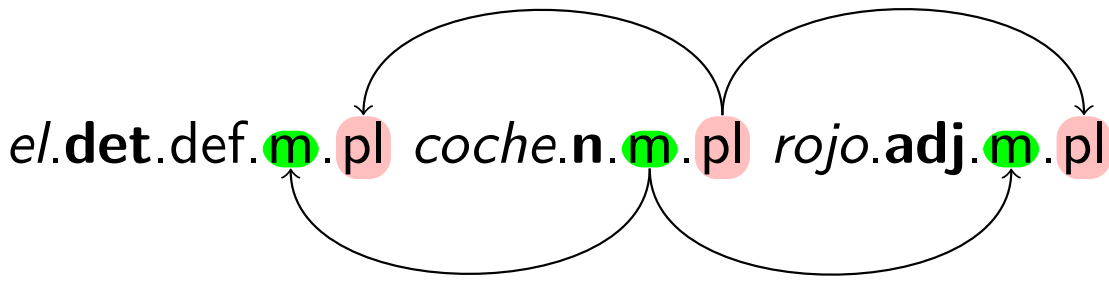
el.**det**.def *rojo*.**adj** *coche*.**n**.m.pl

Apertium translation

Input > Intermediate representation > Transfer

Apertium translation

Input > Intermediate representation > Transfer



Apertium translation

Input > Intermediate representation > Transfer > linearization

los coches rojos

How can we share/reuse resources?

Contributions

- An experiment in reusing resources between free and open source rule based machine translation systems
- A method for creating shallow transfer rules from parallel grammars

Apertium → GF: Sharing lexicon

- Import the lexicon from Apertium to GF

Apertium → GF: Sharing lexicon

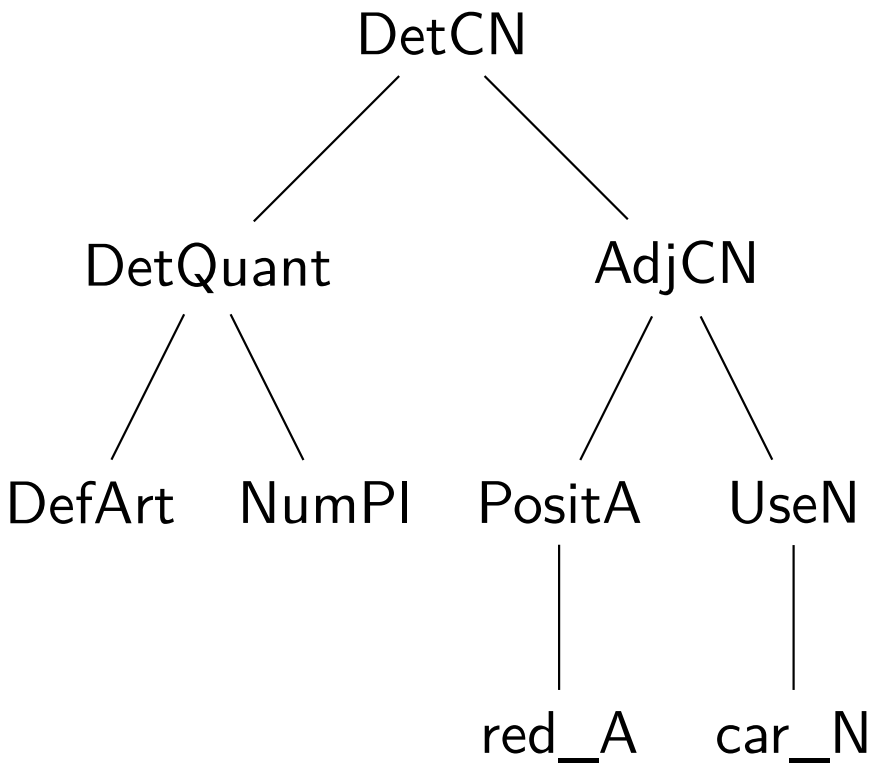
- Import the lexicon from Apertium to GF
- Different format

Apertium → GF: Sharing lexicon

- Import the lexicon from Apertium to GF
- Different format
- Missing information: smart-paradigms

GF $\rightarrow$ Apertium: Apertium rules from the RGL

Abstract syntax tree



GF $\rightarrow$ Apertium: Apertium rules from the RGL

Original linearizations

the red cars $\rightarrow$ *el coches rojos*

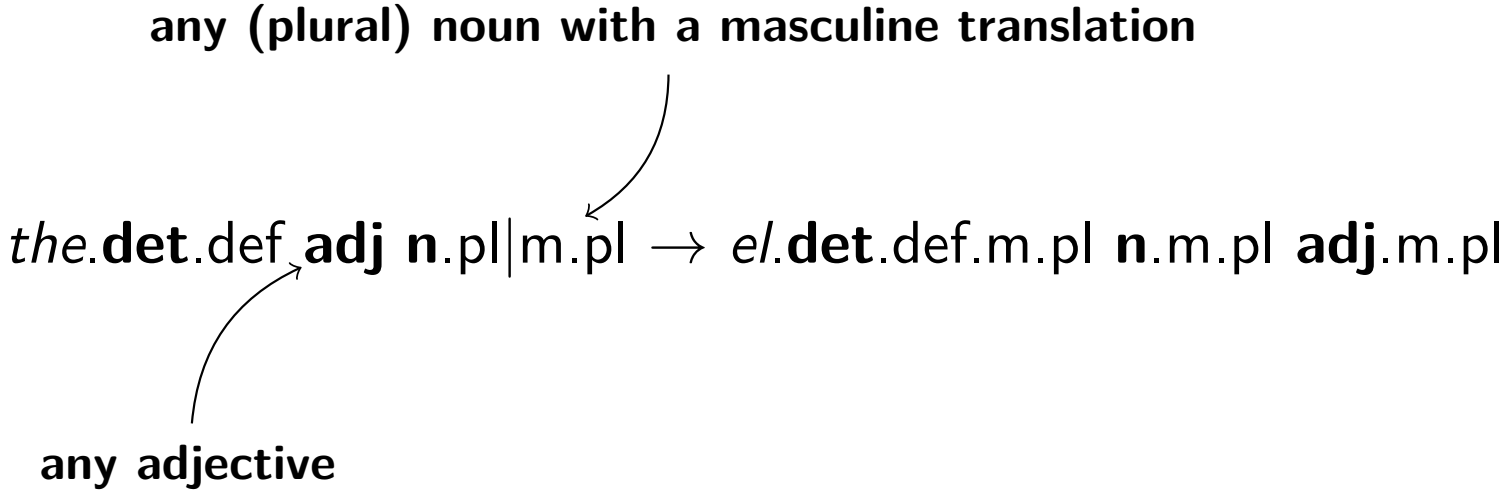
GF $\rightarrow$ Apertium: Apertium rules from the RGL

Replacing tokens with lexical forms

the.**det**.def *red*.**adj** *car*.**n**.pl $\rightarrow$ *el*.**det**.def.m.pl *coche*.**n**.m.pl
rojo.**adj**.m.pl

GF $\rightarrow$ Apertium: Apertium rules from the RGL

Generelizing over nouns and adjectives



GF $\rightarrow$ Apertium: Apertium rules from the RGL

Alignment template

the.**det**.def **adj** **n**.pl|m.pl $\rightarrow$ *el*.**det**.def.m.pl $3.**n**.m.pl $2.**adj**.m.pl

GF $\rightarrow$ Apertium: Apertium rules from the RGL

Alignment template

Evaluation

Corpora:

- *newstest2011A* (1896 sentences) sentences which can be parsed (either fully or partially) by GF (in a reasonable time)
- *newstest2011B* (130 sentences) sentences fully parsed by GF

Systems:

- *sharedApertium* is an Apertium-based system containing the original Apertium lexicon and a set of shallow-transfer rules created from the GF RGL
- *sharedGF* is a GF-based system in which the lexicon has been ported from Apertium
- *Apertium word-for-word* Apertium system using only the bilingual dictionary
- *Apertium* Apertium system with manually created rules

Evaluation

Corpus	System	BLEU	METEOR	TER
A	sharedGF	0.027	0.181	0.847
	sharedApertium	**0.138**	**0.390**	**0.678**
	Apertium word-for-word	0.111	0.368	0.703
	Apertium	0.200	0.443	0.617
B	sharedGF	0.152	0.388	0.703
	sharedApertium	0.148	0.391	0.691
	Apertium word-for-word	0.106	0.361	0.713
	Apertium	0.212	0.451	0.620

Thanks!