

*Parallel Parsing Evaluation
across Europe:*

*the EVALITA-PASSAGE
2009 shared test.*

Patrick Paroubek

Laboratoire pour la Mécanique et les Sciences de l'Ingénieur
Centre National de la Recherche Scientifique

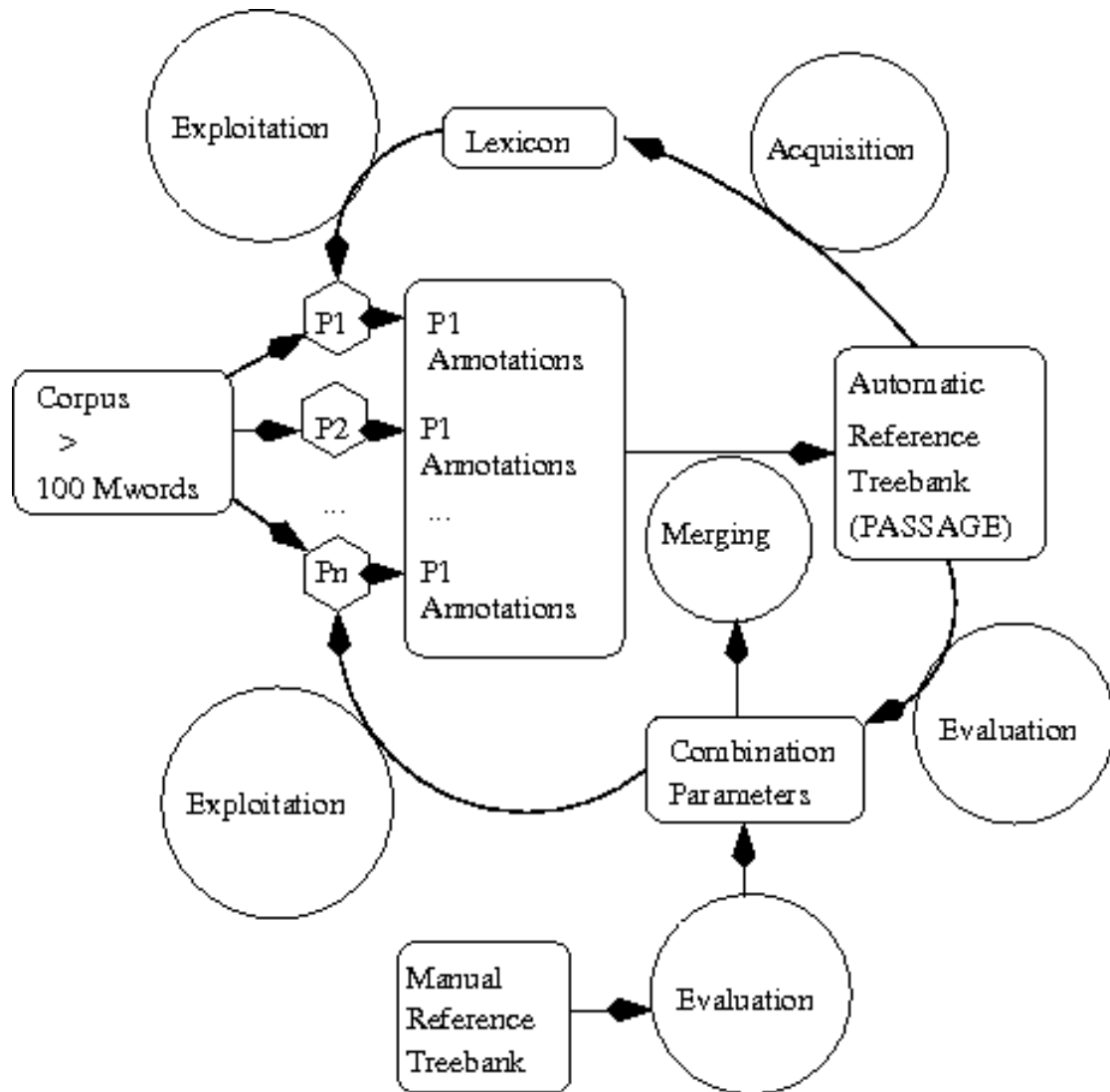
P A S S A G E

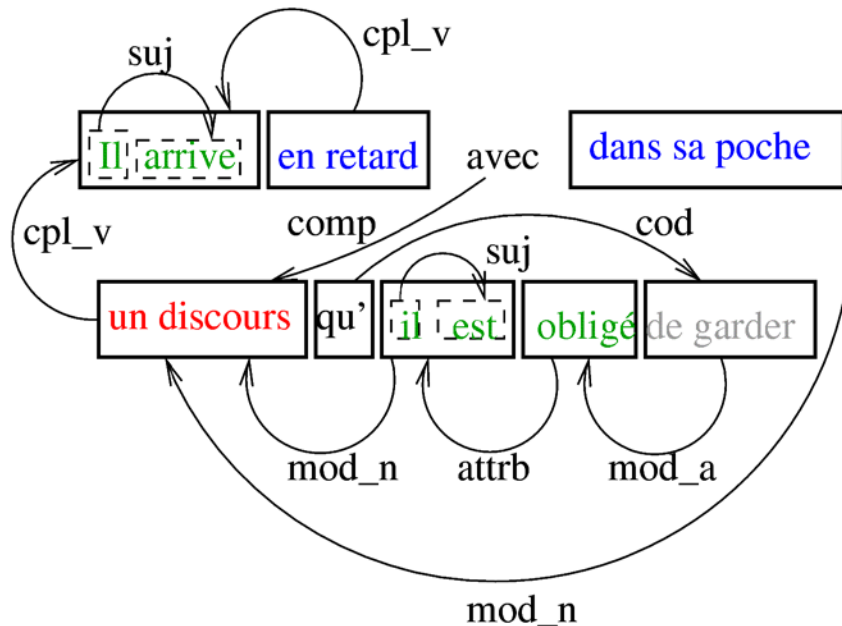
2006
2010

ANR

INRIA

E. de la Clergerie





a chunk

suj : subject
cod : object

attrb : subject attribute
mod_n : noun modifier

a form

cpl_v : verb complement
comp : complementor

mod_a : adjective modifier

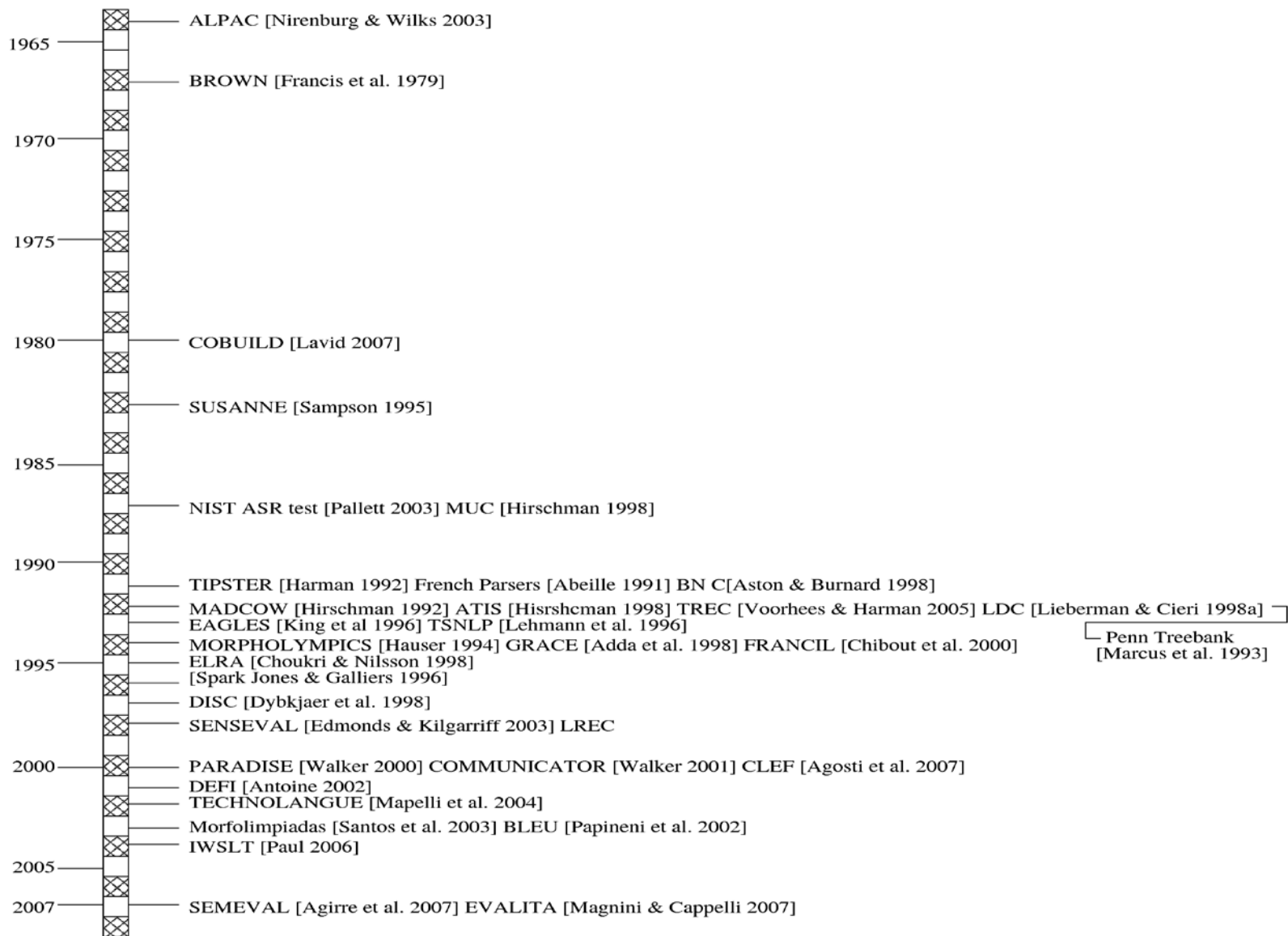
He arrives late with in his pocket a discourse he is compelled to keep

- The following example, taken from Marneffe & Manning (COLING 2008)
 - « I feel like a little kid", says a gleeful Alex de Castro, a car salesman, who has stopped by a workout of the Suns to slip six Campaneris cards to the Great Man Himself to be autographed. » (WSJ-R)

Scheme	Relation
SD	appos(Castro, salesman)
PARC	appos(Alex de Castro, salesman
GR	adjunct(Castro, salesman)
PASSAGE	appos(FIRST:[a gleeful Alex de Castro] GN APPOSED:[a car salesman] GN)

PASSAGE-2 runs (nov. 2009)

- LIC2M-CEA-LIST/LIMA, Gael de Chanlendar, dependency based parser
- INRIA, B. Crabbe
- INRIA/FRMG, De La Clergerie, an hybrid TIG/TAG metagrammar
- LIRMM/SYGMART, J. Chauche, developed at LIRMM,
- LORIA 1, B. Guillaume,
- LORIA 2, C. Cerisara
- SYNAPSE/CORDIAL, D. Laurent.
- TAGMATICA/TAGParser, G. Francopoulo, an extended chunker.
- UNIGE-LATL, N. Luka.
- U.Provence/STP1, S. Rauzy
- XEROX, S. Ait-Mokhtar



Evalita test

jrc_057.txt_0_20000.xml 54 sentences

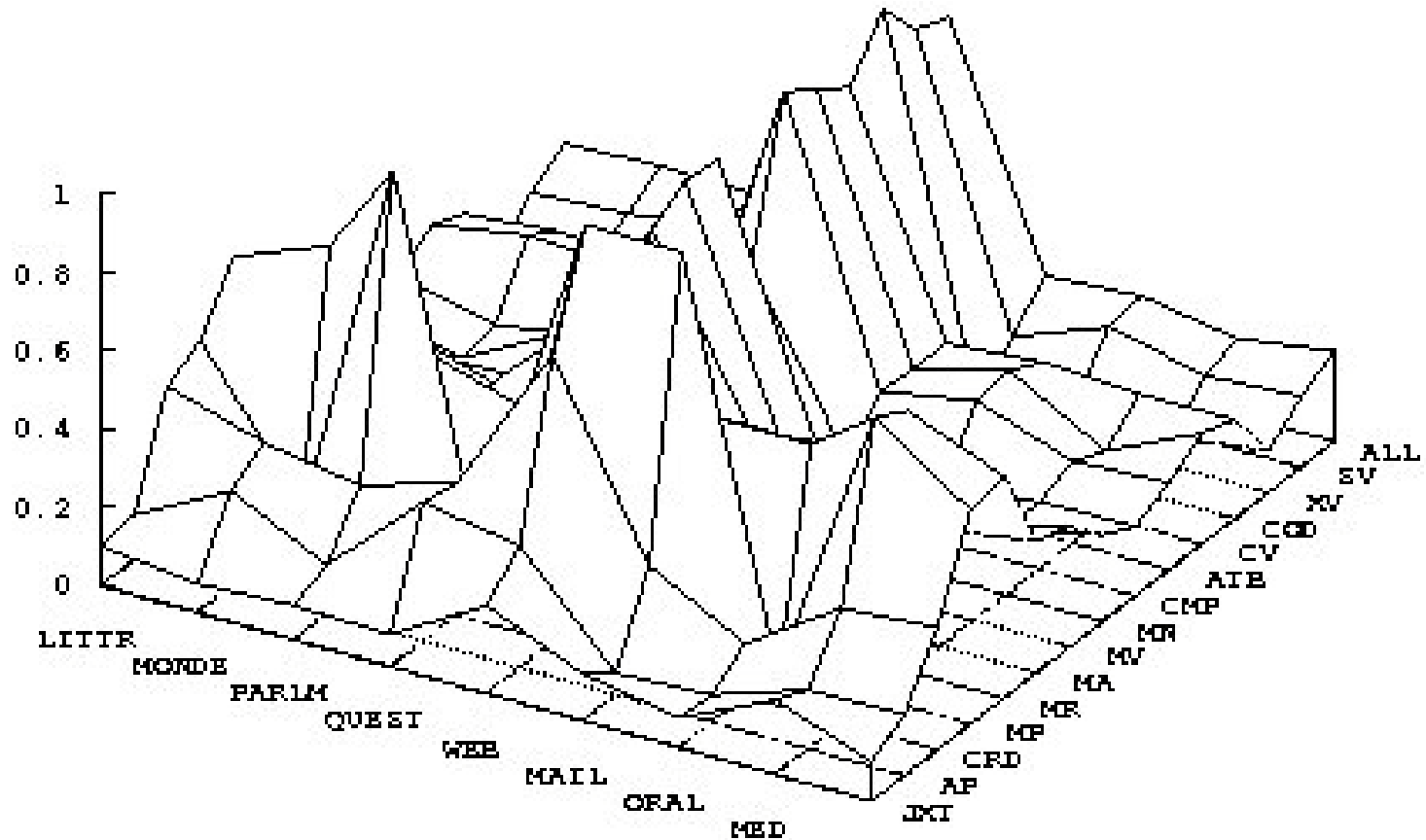
- JRC-ACQUIS 32002D0803 it/2002/jrc32002D0803-it.xml
- JRC-ACQUIS 32002D0804 it/2002/jrc32002D0804-it.xml
- JRC-ACQUIS 32002D0805 it/2002/jrc32002D0805-it.xml
- JRC-ACQUIS 32002D0807 it/2002/jrc32002D0807-it.xml
- JRC-ACQUIS 32002D0811 it/2002/jrc32002D0811-it.xml

evalita dev

- Size of the EVALITA dev corpus 14849 words 218 sentences

difference between recall of union of all participants
and best recall (P3) at EASY campaign
(note it is always positive)

It represent the potential gain in recall for pure combination methods



FLaReNet

- “Fostering Language Resources Network”
- Thematic Network E-Content+
- 0.9 M€EC funding (2008-2011)
- Language Resources (data/tools) and Evaluation for **Automatic Human Language Processing (ICT)**
- Think Tank
 - Launching event (Vienna, February 2009)
 - 2010 event (Barcelona, February 2010)
- 81 institutional members (31 countries), 251 individual members
- Link EU-US : INTEROP-SILT (NSF)

T4ME

- “Technologies for a Multilingual Europe”
 - Proposal of a NoE in FP7 4th ICT Call
 - 3 Pillars
 - Pushing the frontiers (Machine Translation)
 - Open Resource Infrastructure (ORI)
 - On-line LR production, annotation, standardisation, validation, distribution
 - On-line LT evaluation
 - Vision, planning, networking, promotion
 - Kick-off meeting planned in Feb. 2010