Pairing Model-Theoretic Syntax and Semantic Network for Writing Assistance

Jean-Philippe Prost and Mathieu Lafourcade

LIRMM, Université Montpellier 2

CSLP@Context, 2011

The Problem

Syntax/Semantics interface for Property Grammar (PG) through writing assistance

Example

**L'avocat le dossier de son client* (*The lawyer his client's file*)

- Case of (likely) missing word:
  L'avocat `X` *le dossier de son client*
  where `X` is of category `V`
- Expected surface realisation: *plaide* (*pleads*)

Outline

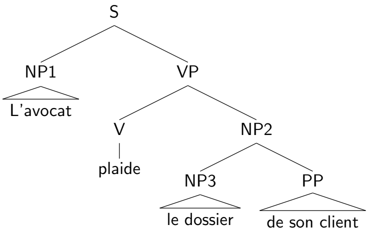
Sketch of the Process

1. Error detection with approximated parse(s)
2. (unrealised) Re-generation by tree transduction
3. Surface realisation with lexical network
   - choice of functional and semantic roles
   - completion message
   - propagation in the network

Property Grammar (Blache, 2001)

Model-Theoretic Semantics for PG (Duchier et al., 2009)

- Models are labelled trees



- The grammar is a constraint system over tree nodes

$$NP : D \prec N$$

$$VP : \triangle V$$

$$NP : N \Rightarrow D, \ldots$$

Model-Theoretic Semantics for PG

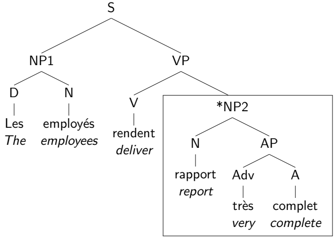
- Strong semantics (i.e. well-formedness)

$\tau : \sigma \models G$

a syntax tree τ is a strong model of property grammar G, with realization σ, iff it is admissible and $R_\tau(\varepsilon) = \sigma$ and $I^-_{G,\tau} = \emptyset$

Model-Theoretic Semantics for PG

- Loose semantics



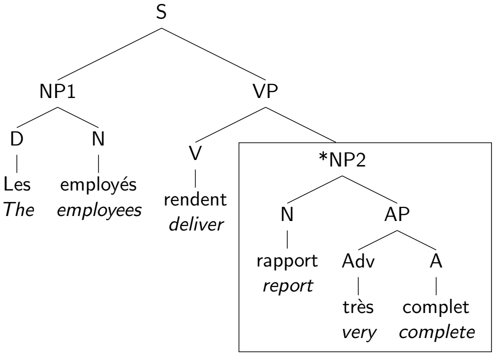
fitness

$$F_{G,\tau} = I^{+}_{G,\tau} / I^{0}_{G,\tau}$$

loose models

$$\tau : \sigma \mid\!\approx G \quad \text{iff} \quad \tau \in \operatorname*{argmax}_{\tau' \in \mathcal{A}_{G,\sigma}} (F_{G,\tau'})$$

Getting Approximated Parses



- Use of robust parsers's combined output as set of models
- Most robust parsers are not capable of deciding about the well-formedness of the input sentence

Characterising Approximated Parses

Characterisation of a model: set of pairs (pertinent instance of property, truth value)

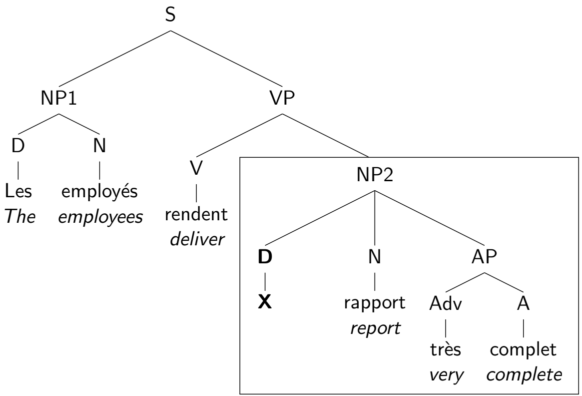
$$
\begin{aligned}
I^0_{G,\tau} &= \{r \in \mathcal{I}_\tau[\![G]\!] \mid P_\tau(r)\} \\
I^+_{G,\tau} &= \{r \in I^0_{G,\tau} \mid S_\tau(r)\} \\
I^-_{G,\tau} &= \{r \in I^0_{G,\tau} \mid \neg S_\tau(r)\}
\end{aligned}
$$

Re-generation by Tree Transduction

- Basic tree operations, where τ is a tree, and c, c_1, c_2 are node labels (i.e. categories):
  - *Node insertion*, denoted by $\tau \downarrow c$
  - *Node deletion*, denoted by $\tau \nmid c$
  - *Node permutation*, denoted by $c_1 \overset{\tau}{\leftrightarrow} c_2$
- Transduction:

Property	*Violated instances*	*Tree operation*
Requirement	$\mathcal{I}_\tau[\![c_0 : c_1 \Rightarrow s_2]\!]$	$\tau \downarrow s_2$
Obligation	$\mathcal{I}_\tau[\![c_0 : \triangle c_1]\!]$	$\tau \downarrow c_1$
Linearity	$\mathcal{I}_\tau[\![c_0 : c_1 \prec c_2]\!]$	$c_1 \overset{\tau}{\leftrightarrow} c_2$
Uniqueness	$\mathcal{I}_\tau[\![c_0 : c_1 !]\!]$	$\tau \nmid c_1$
Exclusion	$\mathcal{I}_\tau[\![c_0 : c_1 \not\Leftrightarrow c_2]\!]$	$\tau \nmid c_1 \cup \tau \nmid c_2$

Re-generated Model



The Lexical Network

rezoJDMFR

- Nodes and directed relations
- Weights and types

Example

cat $\xrightarrow{isa}$ animal   cat $\xrightarrow{loc}$ sofa

cat $\xrightarrow{can}$ pur   cat $\xrightarrow{part}$ claw

- Many relation types including semantic roles
  AGENT, PATIENT, INSTRUMENT
- Other relations
  - Typical location, manner, entailment
  - isa, partof, substance, synonym, antonym, ...

A game for building the network

JeuxDeMots

`http://jeuxdemots.org/`

- Users associate terms given a relation
- A popular consensus filtered by pairs of players
- In 4 years time
  - over $1,200,000$ relations among $230,000$ terms
  - evaluation through a guessing game (AKI, *tip-of-the-tongue*)
    - term found in more than 75% of cases
    - while the typical human score is around 46%

Jeux de mots


DONNER DES IDEES ASSOCIEES AU TERME QUI SUIT :
kaput
Crédits : 863220
Honneur : 137527
941
gripple aviaire
Temps
41 s
30s
OK
épidémie
‹ poulet (volaille)
poulet ›
H1N1
gripe
‹ oiseau (animal)
oiseau ›
maladie
8/13
Dernier terme proposé : épidémie
Ce terme a plusieurs sens ou il en manque ? Demandez de l'aide à vos amis

Jeux de mots

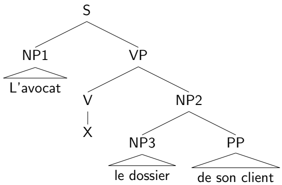

DONNER DES IDEES ASSOCIEES AU TERME QUI SUIT :
Déconnexion
Jouer
gripPE aviaire
Réponses données par kaput : poulet (volaille) - poulet - H1N1 - oiseau - oiseau (animal) - gripPE - maladie - épidémie
Réponses données par BackInBusiness : fièvre - tousser - Asie - H1N1 - AH1N1 - gripPE
AH1N1 - oiseau (animal) - volaille - gripPE espagnole - virus - gripPE
H1N1 - oiseau (animal) - gripPE
Vous gagnez 241 crédits et 5 point(s) d'honneur
J'aime
Soyez le premier de vos amis à indiquer que vous aimez ça.
jackpot
jetons
(in)justice
en savoir +

Completion Message

$$\langle a, \mathtt{:R}, b \rangle$$

$\mathtt{:R}$ denotes an oriented semantic relation, and
a and b its oriented elements.

Gathering Functional and Semantic Roles



VP (Verb Phrase)

$\begin{bmatrix} \text{HEAD} & V \\ \text{SUBCAT} & \langle \text{V.SUBCAT} \rangle \end{bmatrix}$

Obligation	: $\triangle$V
Uniqueness	: $\text{V}_{[\text{main past part}]}!$
	: NP!
	: PP!
Linearity	: V $\prec$ NP
	: V $\prec$ Adv
	: V $\prec$ PP
Requirement	: $\text{V}_{[\text{past part}]} \Rightarrow \text{V}_{[\text{aux}]}$
Exclusion	: $\text{Pro}_{[\text{acc}]} \not\Leftrightarrow \text{NP}$
	: $\text{Pro}_{[\text{dat}]} \not\Leftrightarrow \text{Pro}_{[\text{acc}]}$
Dependency	: V $\begin{bmatrix} \text{ROLE} & \textit{PRED} \\ \text{SUBCAT} & \left\langle \text{ARG} \begin{bmatrix} \text{ROLE} & \textit{OBJ} \mid \textit{P-OBJ} \mid \textit{A-OBJ} \\ \text{CAT} & \textit{NP} \end{bmatrix} \right\rangle \end{bmatrix} \rightsquigarrow \text{NP}[\text{ROLE } \textit{PAT}]$
	: V $\begin{bmatrix} \text{ROLE} & \textit{PRED} \\ \text{SUBCAT} & \left\langle \text{ARG} \begin{bmatrix} \text{ROLE} & \textit{OBJ} \mid \textit{P-OBJ} \mid \textit{A-OBJ} \\ \text{CAT} & \textit{PP} \end{bmatrix} \right\rangle \end{bmatrix} \rightsquigarrow \text{PP}[\text{ROLE } \textit{PAT}]$

- NP2 is in a Patient relationship with V
- NP2's head is inherited from NP3's: *dossier* (*file*)

Message $\sharp 1 = \langle \texttt{X}, \texttt{:PAT}, \textit{dossier} \rangle$

Gathering Functional and Semantic Roles

S (Utterance)

$$
\begin{array}{rl}
\text{Obligation} : & \triangle\texttt{VP} \\
\text{Uniqueness} : & \texttt{NP!} \\
: & \texttt{VP!} \\
\text{Linearity} : & \texttt{NP} \prec \texttt{VP} \\
\text{Dependency} : & \texttt{VP}\begin{bmatrix} \text{ROLE} & \textit{PRED} \\ \text{SUBCAT} & \left\langle \text{ARG} \begin{bmatrix} \text{POS} & \textit{0} \\ \text{ROLE} & \textit{SUBJ} \\ \text{CAT} & \textit{NP} \end{bmatrix} \right\rangle \end{bmatrix} \rightsquigarrow \text{NP}\begin{bmatrix} \text{ROLE} & \textit{AGT} \end{bmatrix}
\end{array}
$$

- NP1 is in an Agent relationship with VP
- by inheritance, NP1's head is *avocat* (*lawyer*)

Message $\sharp 2 = \langle \textit{avocat}, :\texttt{AGT}, \texttt{X} \rangle$

Completion Messages

$$\{\langle \mathtt{X}, \mathtt{:PAT}, \mathit{dossier}\rangle, \langle \mathit{avocat}, \mathtt{:AGT}, \mathtt{X}\rangle\}$$

Propagation Algorithm

The Problem

Given an underspecified input, such as
$\langle avocat, \texttt{:AGT}, \texttt{X}\rangle$ and $\langle \texttt{X}, \texttt{:PAT}, dossier\rangle$
can $\texttt{X}$ be filled in?

- Propagation in the lexical network
  - iterative and globally convergent (3 runs)
  - what is the most activated node for $\texttt{X}$?
    plaider, . . . , étudier
  - what are the most activated nodes for the constraints?
    avocat $\rightarrow$ *avocat>justice*
    dossier $\rightarrow$ *dossier>affaire*

Conclusion and Perspectives

- Near-lack of Syntax/Semantics interface in PG prevents its use for deep processing
- Yet the PG formal properties are well-suited to address Grammar Checking problems
- The linguistic information provided by PG properties (i.e. *characterisation*) allows building a detailed information structure about non-canonical sentences
- Work in progress: completion of a characterised syntax tree through the exploration of a semantic network.