Grammars and Generation

Claire Gardent

CNRS/LORIA Nancy (France)

Joint work with Ben Gottesman, German Kruszewski, Shashi Narayan, Yannick Parmentier, Laura Perez-Beltrachini and Sylvain Schmitz

Grammars and Corpus 2018, Paris
Outline

1. Tree Adjoining Grammar
2. Writing Grammar
3. Improving Grammars
4. Grammar for Language Learning
5. Grammar for NL Interfaces
1. Tree Adjoining Grammar

2. Writing Grammar

3. Improving Grammars

4. Grammar for Language Learning

5. Grammar for NL Interfaces
Tree Adjoining Grammar

A set of trees
- Initial
- Auxiliary

```
NP  S

NP↓ VP

V

VP

VP* ADV
```
Lexicalised Tree Adjoining Grammar

Trees are lexicalised

```
NP
  ↓
John

S
  ↓
NP
  ↓
NP
  ↓
VP
  ↓
V
  ↓
runs

VP
  ↓
VP
  ↓
ADV
  ↓
fast
```
Feature-Based Lexicalised Tree Adjoining Grammar

Tree nodes are labelled with feature-structures

```
NP
  | John
S
  | VP
    | num:N
    | V
    | num:N
    | runs
VP
  | VP*
    | ADV
    | fast
```
Feature-Based Lexicalised Tree Adjoining Grammar with Semantics

Trees are assigned a semantics

\[ \text{NP}_j \]
\[ \text{John} \]
\[ john(j) \]

\[ \text{Se}_1 \]
\[ \text{NP}^x \downarrow \text{VP}^{e_1} \]
\[ \text{num:N} \]
\[ \text{V} \]
\[ \text{num:N} \]
\[ \text{runs} \]
\[ run(e,x) \]

\[ \text{VP}_f \]
\[ \text{VP}^f \star \text{ADV} \]
\[ \text{fast} \]
\[ fast(f) \]
Grammar in Action: Parsing

Jean aime Marie

\[ \text{love}(l, j, m) \land \text{john}(j) \land \text{mary}(m) \]
Grammar in Action: Generation

\[ \text{love}(l,j,m) \land \text{john}(j) \land \text{mary}(m) \]

\[ \text{Jean aime Marie} \]
Grammar-Based Generation

\[
\begin{array}{c}
S \\
\text{NP}^x \quad \text{VP}^{e_1} \\
\text{NP}_j \quad V \\
\text{John} \quad \text{runs} \\
john(j) \quad run(e,x) \\
\text{VP}^f \quad \text{ADV} \\
\text{fast} \quad fast(f)
\end{array}
\]

\[
john(j), \ run(e,j), \ fast(e)
\]
Grammar-Based Generation

\[ \text{S} \rightarrow \text{NP} \downarrow j \rightarrow \text{VP}^e_{e_1} \rightarrow \text{V}[\text{agr:3sg}] \rightarrow \text{runs} \rightarrow \text{run}(e,j) \]

\[ \text{VP} \rightarrow \text{VP}^e \rightarrow \text{ADV} \rightarrow \text{fast} \rightarrow \text{fast}(e) \]

\( \text{john}(j), \text{run}(e,j), \text{fast}(e) \)
Grammar-Based Generation

S
   /     \\  
/ NP               VP\textsubscript{e\textscript{i}}  \
\     \\
John john(j)      \       runs run(e,j)
          \\
\      \                      \\
\       VP\textsubscript{e}  ADV
\      /  \\
\    fast
\      \\
\       fast(e)

John runs
John runs fast
Separating Grammar from Lexicon

Since each tree is lexicalised, the resulting grammar can be very large. In practice, we therefore...
Separating Grammar from Lexicon

Since each tree is lexicalised, the resulting grammar can be very large. In practice, we therefore

- abstract over lexical items in the grammar
Separating Grammar from Lexicon

Since each tree is lexicalised, the resulting grammar can be very large. In practice, we therefore

- abstract over lexical items in the grammar
- use a lexicon to determine which grammar tree is lexicalised/anchored by which lexical items
Separating Grammar from Lexicon

\[
S \\
NP^C \\
VP^{B_1}_B \\
V \\
\text{runs} \\
\text{run}(B,C)
\]
Separating Grammar from Lexicon

S
  NP\(^C\)
  VP\(^{B_1}\)
    V
      runs
      run(B,C)
Separating Grammar from Lexicon

\[ \text{run}(B, C) \Rightarrow R(B, C) \]
Separating Grammar from Lexicon

Semantics: \textit{departure}
Tree: nx0V
Syntax: CanonicalSubject
Anchor: \textit{departs}

Semantics: \textit{arrival}
Tree: nx0V
Syntax: CanonicalSubject
Anchor: \textit{arrives}

...
Which fruit has John eaten?
Which fruit has John eaten?

TAG for French: 6,000 trees
(Leroux, Crabbé and Parmentier 2006)
Creating and Curating Forests of TAG trees

How to write them?

* XMG, a grammar writing environment

How to verify them?

* Error Mining: Using Generation to debug the grammar
XMG

A declarative framework for specifying tree based unification grammars

B. Crabbe, D. Duchier, C. Gardent, J. Leroux and Y. Parmentier
XMG, eXtensible Meta Grammar.
**XMG Tree Fragments**

\[
\begin{align*}
\text{CanonSubj} & \rightarrow \text{N} \overline{\text{R}} \text{V} \overline{\text{W}} \\
\text{CanonObj} & \rightarrow \text{V} \overline{\text{W}} \text{N} \overline{\text{R}} \\
\text{CanonIndirObj} & \rightarrow \text{a} \overline{\text{R}} \\
\text{CanonByObj} & \rightarrow \text{par} \overline{\text{R}} \\
\text{RelatSubj} & \rightarrow \text{N} \overline{\text{R}} \text{V} \overline{\text{W}} \\
\text{WhObj} & \rightarrow \text{V} \overline{\text{W}} \\
\text{WhByObj} & \rightarrow \text{par} \overline{\text{R}} \\
\text{WhIndirObj} & \rightarrow \text{a} \overline{\text{R}} \\
\text{ActiveVerbForm} & \rightarrow \text{V} \overline{\text{B}} \\
\text{PassiveVerbForm} & \rightarrow \text{V} \overline{\text{B}} \text{V} \overline{\text{B}}
\end{align*}
\]
Creating Trees with XMG

(Canonical Subject)  (Active verb morph)  (e.g. the boy sleeps)

(Extracted Subject)  (Active verb morph)  (e.g. the boy who sleeps)
Applying General Principles

\[(Jean)\]  \[(le)\]  \[(lui)\]  \[(donne)\]
Compact Grammar Specifications

293 tree fragments ⇒ 6,000 TAG tree
Large Scale XMG Grammars

FrenchTAG: a Tree Adjoining Grammar for French (Benoît Crabbé)

SemTAG: XMG-based XTAG extended with semantics (Claire Gardent)

FrenchTAG + MWE: a FrenchTAG grammar updated with a number of Multi Word Expressions (Agata Savary)

XMG_GC_metagrammar: a Tree Adjoining Grammar for Guadeloupean Creole (Emmanuel Schang)

Interaction Grammar for French (Guy Perrier)

XMG-based XTAG: a Tree Adjoining Grammar for English based on XTAG (Katya Saint-Amand, Claire Gardent)
Ikota Morphology

Morphological Lexicon: 600+ trees

Denys Duchier, Brunelle Magnana Ekoukou, Yannick Parmentier, Simon Petitjean and Emmanuel Schang

Describing Morphologically-rich Languages using Metagrammars: a Look at Verbs in Ikota

Proceedings of the Workshop on Language Technology for Normalisation of Less-Resourced Languages SaLTMiL 8 - AfLaT, 2012
XMG Extensions

Simon Petitjean, Denys Duchier and Yannick Parmentier.  
XMG 2: Describing Description Languages  
LACL 2016

Laura Kallmeyer’s TreeGrasp ERC Project
Depictive grammar: an LTAG grammar fragment with semantic frames for English depictives (Benjamin Burkhardt)
Spotting Errors

Grammar traversal

Statistical Error Mining
GraDe (Grammar Debugger)

Top-Down Grammar Traversal

Outputs the sentences generated by the grammar

User-Defined parameters control the search to ensure (i) termination and (ii) interesting linguistic coverage.

Claire Gardent and Eric Kow
Spotting overgeneration suspects.
ENLG 2007

Claire Gardent and German Kruszewski
Generation for Grammar Engineering
INLG 2012
GraDe Example: Checking for Coherence

For each grammar rule anchored by a verb, can we find at least one derivation?

```plaintext
family: VERB_FAMILY
cat: s
features: [mod:ind]
max_results: 1
max_adjunctions:
   {N: 1, NP: 0, V:1, VP: 1, ADJ: 0, S: 0}
depth: 5
```
Checking for Grammar Coherence

<table>
<thead>
<tr>
<th>Tree Family</th>
<th>Trees</th>
<th>Fails</th>
<th>Fails/Trees</th>
</tr>
</thead>
<tbody>
<tr>
<td>CopulaBe</td>
<td>60</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>iliV</td>
<td>2</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>n0V</td>
<td>10</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>n0ClV</td>
<td>9</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>n0ClVn1</td>
<td>45</td>
<td>2</td>
<td>4%</td>
</tr>
<tr>
<td>n0ClVden1</td>
<td>36</td>
<td>3</td>
<td>8%</td>
</tr>
<tr>
<td>n0ClVpn1</td>
<td>29</td>
<td>3</td>
<td>10%</td>
</tr>
<tr>
<td>n0Vn1</td>
<td>84</td>
<td>3</td>
<td>3%</td>
</tr>
<tr>
<td>n0Vn1Adj2</td>
<td>24</td>
<td>6</td>
<td>25%</td>
</tr>
<tr>
<td>n0Vn1</td>
<td>87</td>
<td>3</td>
<td>3%</td>
</tr>
<tr>
<td>n0Vden1</td>
<td>38</td>
<td>3</td>
<td>7%</td>
</tr>
<tr>
<td>n0Vpn1</td>
<td>30</td>
<td>3</td>
<td>10%</td>
</tr>
<tr>
<td>iliVcs1</td>
<td>2</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>n0Vcs1</td>
<td>30</td>
<td>23</td>
<td>74%</td>
</tr>
<tr>
<td>n0Vas1</td>
<td>15</td>
<td>10</td>
<td>66%</td>
</tr>
<tr>
<td>n0Vn1Adj2</td>
<td>24</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>s0Vn1</td>
<td>72</td>
<td>9</td>
<td>12%</td>
</tr>
<tr>
<td>n0Vs1int</td>
<td>15</td>
<td>12</td>
<td>80%</td>
</tr>
<tr>
<td>n0Vn1n2</td>
<td>24</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>n0Vn1an2</td>
<td>681</td>
<td>54</td>
<td>7%</td>
</tr>
</tbody>
</table>

Approximately 10% of the verb trees fail to licence a complete derivation.
Syntactic Variants

Which syntactic variants does the grammar generate for a given verb type?

family: n0V
cat: s
features: [mod:ind]
max_results: all
adjunctions:
  {N: 2, NP: 0, V:1, VP: 1, ADJ: 0, S: 0/1}
depth: 5
Example Output

Elle chante (She sings), La tatou chante-t’elle? (Does the armadillo sing? ), La tatou chante (The armadillo sings ), Chacun chante -t’il (Does everyone sing? ), Chacun chante (Everyone sings ), Quand chtante chacun? (When does everyone sing? ), Quand chante la tatou? (When does the armadillo sing? ) Quand chante quel tatou? (When does which armadillo sing? ), Quand chante Tammy? (When does Tammy sing? ), Chante-t’elle? (Does she sing? ) Chante -t’il? (Does he sing? ), Chante! (Sing! ), Quel tatou chante ? (Which armadillo sing? ), Quel tatou qui chtante ..? (Which armadillo who sings ..? ) Tammy chante-t’elle? (Does Tammy sing? ), Tammy chante (Tammy sings ), une tatou qui chtante chante (An armadillo which sings sings ), C’est une tatou qui chtante (It is an armadillo which sings ), ...
Some incorrect cases

*Chacun chante-t’elle?*

*(Everyone sings?)*

Missing agreement constraint between the inverted subject clitic and the subject.

*La tatou qui chante-t’elle?*

*(The armadillo which does she sing?)*

Missing constraint on the inverted subject clitic

(should be disallowed in embedded clauses)
Generation-Based Grammar Analysis

- Can all rules in the grammar be used in at least one derivation?
- Are all possible syntactic realisations of the verb and of its arguments generated and correct?
- Does the grammar correctly capture the interactions between basic clauses and modifiers?
- etc.
Statistical Error Mining

Generate from large corpus of input meaning representations

Divide the input into FAIL and PASS

Use statistics to identify **subtrees (forms)** in the set of inputs which frequently associate with failure and rarely with success

Shashi Narayan and Claire Gardent
Error Mining with Suspicion Trees: Seeing the Forest for the Trees
COLING 2012

Claire Gardent and Shashi Narayan
Error Mining on Dependency Trees
ACL 2012
The most troublesome report is the August merchandise trade effect deficit due out tomorrow.
Error Mining on trees

Input tree $\implies$ FAIL | PASS

Calculate a **suspicion score** for each subtree in the input

Structure the **suspicious subtrees** into a tree
Suspicious Forms

Subtrees of the input dependency trees labelled with lemma, parts-of-speech and/or dependency information
Suspicion Score Metrics

Adapted from ID3 decision tree algorithm

The suspicion score of a form $f$

$$S_{score}(f) = \frac{1}{2} \left( \text{Fail}(f) \ast \ln \text{count}(f) + \text{Pass}(\neg f) \ast \ln \text{count}(\neg f) \right)$$

FAIL score

$$\text{Fail}(f) = \frac{\text{count}(f|\text{FAIL})}{\text{count}(f)}$$

PASS score

$$\text{Pass}(\neg f) = \frac{\text{count}(\neg f|\text{PASS})}{\text{count}(\neg f)}$$
Ranked List of Suspicious Forms

1. (POSS)
2. (NNP, (POSS))
3. (CC)
4. (NN, (POSS))
5. (NN, (NNP, (POSS)))
6. (NN, (NN, (POSS)))
7. (NN, (CC))
8. (NNP, (NNP), (POSS))
9. (NN, (NNP, (NNP), (POSS)))
10. (NN, (NNP, (NNP)))
11. (CC, (JJ))
12. (JJ, (CC))
13. (NNP, (NNP, (POSS)))
14. (NN, (NN), (POSS))
15. (DT, (IN))
16. ...
Tree of suspicious forms

The Right Frontier shows the most important sources of errors
Building the Tree of suspicious forms

The decision tree algorithm recursively partitions the data by

1. selecting the most suspicious form
2. splitting the data into two subsets, a subset of the data that contain that suspicious form (yes) and a subset that does not (no).
Example Suspicion Tree

- **(POSS)** A mismatch between input and grammar representation
  
  (DAD/NN, (JOHN/NNP, (’s/POSS)))
  
  (DAD/NN, (’s/POSS, (JOHN/NNP)))

- **(CC)** conflicting feature values in the grammar of NP coordination

- **(DT, (IN))** POS tag mismatch
  
  *some DT/PRP of the audience*

...
Different Views highlight Different Errors

**(days/NN)**  POS tag assignment error. 
**DAYS/NN mapped to the wrong TAG family**

**(DT, (IN))**  POS tag mismatch 
**some DT/PRP of the audience**
Cases that always fail

Single source of error
Cases that always fail

Several sources of error

- \( (im\text{-VB}) \) – infinitival verbs.
- \( (opr\text{d-TO}) \) – control / raising cases
- \( (im\text{-VB}, (prd\text{-JJ})) \) – adjectival complement
- \( (nmod\text{-TO}, (im\text{-VB})) \) – infinitival is a noun modifier
Cases that sometimes fail

Cardinals lead to generation failure in the contexts shown but not in all context (CD does not occur)
Experiment

Surface Realisation Challenge Dataset (Belz et al. 2011). 26,725 input dependency trees derived from the Penn Treebank

XMG induced FB-LTAG Grammar (K. Alahverdziehva)

Grammar-Based Surface Realiser (Narayan and Gardent, COLING 2012)
Results

Corrections

- 11 rewrite rules (Gen-1, Dt-4, Adv-1, Inf-3, Aux-1 and Final-1),
- 2 grammar corrections and
- a few lexicon updates

<table>
<thead>
<tr>
<th></th>
<th>Input Data</th>
<th>Initial Failures</th>
<th>Final Failures</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-ALL</td>
<td>26725</td>
<td>19280 (72.1)</td>
<td>5157 (19.3)</td>
</tr>
</tbody>
</table>

- Sentence length – min:1, max:134, avg:22
- Coverage: 81.74%, BLEU:0.73 (for the covered data)
1 Tree Adjoining Grammar
2 Writing Grammar
3 Improving Grammars
4 Grammar for Language Learning
5 Grammar for NL Interfaces
Generating Grammar Exercises

Generate sentences

Use the detailed linguistic information output by the generator to select and build exercises

Three types of exercises: FIB, Shuffle and Reformulation

C. Gardent and L. Perez-Beltrachini.
Using FB-LTAG Derivation Trees to Generate Transformation-Based Grammar Exercises.

L. Perez-Beltrachini, C. Gardent and G. Kruszewski
Generating Grammar Exercises.
The 7th Workshop on Innovative Use of NLP for Building Educational Applications, NAACL-HLT Worskhop 2012, Montreal, Canada, June.
Grammar Exercises

Built from a single sentence.

[FIB] Complete with an appropriate personal pronoun.

(S) _Elle adore les petits tatous_
(S) (She loves the small armadillos)
(Q) _____ adore les petits tatous (gender=fem)
(K) elle

[Shuffle] Use the words below to make up a sentence.

(S) _Tammy adore les petits tatous_
(S) (Tammy loves the small armadillos)
(Q) tatous / les / Tammy / petits / adore
(K) Tammy adore les petits tatous.
Grammar Exercises

Built from a pair of syntactically related sentences

[Reformulation] Rewrite the sentence using passive voice

(Q) C’est Tex qui a fait la tarte.
   (It is Tex who has baked the pie.)

(K) C’est par Tex que la tarte a été faite.
   (It is Tex by whom the pie has been baked.)

Active/Passive, NP/Pronoun, Assertion/Wh-Question, Assertion/YN-Question
The *GramEx* framework: generating and selecting sentences to build exercises
Creating a grammar exercise

aime(e, be, bi), bijou(bi), les(bi), betty(be)
Creating a grammar exercise

aime(e,be,bi),bijou(bi),les(bi),betty(be)

Bette aime le bijou.
C'est Bette qui aime les bijoux.
Bette aime les bijoux.
Creating a grammar exercise

**aime(e, be, bi), bijou(bi), les(bi), betty(be)**

**Goal:** Plural form of irregular nouns.  
**Exercise type:** Fill-in-the-blank.

* Bette aime le bijou.  
* C'est Bette qui aime les bijoux.  
* Bette aime les bijoux.
Creating a grammar exercise

*aime(e, be, bi), bijou(bi), les(bi), betty(be)*

*Bette aime le bijou.*  
*C'est Bette qui aime les bijoux.*  
*Bette aime les bijoux.*

**Goal:** Plural form of irregular nouns.  
**Exercise type:** Fill-in-the-blank.

↓

1. Select sentences
   NP[\(\text{num} = \text{pl} \land \text{plural} = \text{irreg}\)]
   \(\land \text{CanonicalOrder}\)
Creating a grammar exercise

*aime(e,be,bi),bijou(bi),les(bi),betty(be)*

Bette aime le bijou.
C'est Bette qui aime les bijoux.
Bette aime les bijoux.

Goal: Plural form of irregular nouns.
Exercise type: Fill-in-the-blank.

1. Select sentences
   NP[\(num = pl \land plural = irreg\)]
   \(\land CanonicalOrder\)

\[
S\]

\[
\begin{align*}
NP && \quad VP \\
\text{Bette} \quad V && \quad NP[\text{num}=\text{pl}] \\
\text{aime} \quad \text{D}[\text{num}=\text{pl}] \\ les && \quad \text{NP}[\text{lemma}=\text{bijou}]
\end{align*}
\]

\{CanonicalObject, CanonicalSubject, ActiveVerb\}
Creating a grammar exercise

\[
aime(e, be, bi), \text{bijou}(bi), \text{les}(bi), \text{betty}(be)
\]

Bette aime le bijou.
C'est Bette qui aime les bijoux.
Bette aime les bijoux.

**Goal:** Plural form of irregular nouns.
**Exercise type:** Fill-in-the-blank.

1. Select sentences
   \[\text{NP}[\text{num} = \text{pl} \land \text{plural} = \text{irreg}]\land \text{CanonicalOrder}\]

2. Process the selected sentence
   \[\text{NP}[\text{num} = \text{pl}] \Rightarrow \text{blank}\]
   \[\text{NP}[\text{lemma} = \text{bijou}] \Rightarrow \text{hint}\]

\{\text{CanonicalObject, CanonicalSubject, ActiveVerb}\}
Creating a grammar exercise

Goal: Plural form of irregular nouns.

Exercise type: Fill-in-the-blank.

1. Select sentences
   
   NP[\text{num} = \text{pl} \land \text{plural} = \text{irreg}] \land \text{CanonicalOrder}

2. Process the selected sentence
   
   NP[\text{num} = \text{pl}] \Rightarrow \text{blank}

   NP[\text{lemma} = \text{bijou}] \Rightarrow \text{hint}

   (Q) Bette aime les _______. (bijou)

   (K) bijoux

\{\text{CanonicalObject, CanonicalSubject, ActiveVerb}\}
Selecting appropriate sentences

GramEx's boolean constraint language: syntax and use

Boolean constraint language

- conjunction, disjunction and negation of morpho-syntactic and syntactic properties

Describes the linguistic requirements imposed by pedagogical goals

- Permits retrieving appropriate sentences from the DB
Selecting appropriate sentences

Some examples

**Pedagogical goal:** Pre/post nominal irregular adjectives

\[\text{Epith} \land \text{flexion: irreg}\]

✓ *Tammy a une voix douce* (Tammy has a soft voice)

X *Tammy a une jolie voix* (Tammy has a nice voice)

**Pedagogical goal:** Prepositions with infinitives; Simple Clause

POBJinf \land CLAUSE

POBJinf \equiv (DE-OBJinf \lor A-OBJinf)

CLAUSE \equiv Vfin \land \neg Mod \land \neg CCoord \land \neg Sub

✓ *Tammy refuse de chanter* (Tammy refuses to sing)

X *Jean dit que Tammy refuse de chanter* (John says that Tammy refuses to sing)
Transformation-based grammar exercises

Finding syntactically related sentences (e.g. active/passive)

(Q) C’est Tex qui a *fait* la tarte.
   (It is Tex who baked the pie.)

X (K) Tex a *fait* la tarte.
   (Tex baked the pie.)

X (K) La tarte a été *faite* par Tex.
   (The pie was baked by Tex.)

X (K) C’est *par* Tex que la tarte sera *faite*.
   (It is Tex who will bake the pie.)

X (K) Est-ce que la tarte a été *faite* par Tex ?
   (Has the pie been baked by Tex ?)

✓ (K) C’est *par* Tex que la tarte a été *faite*.
   (It is Tex by whom the pie was baked.)
Derived and Derivation Tree

(a) Derived tree

(b) Derivation tree
Creating transformation-based grammar exercises

- Define tree filters on pairs of derivation trees
- Retrieve sentences pairs that match those tree filters
Why Derivation Trees?

\[ \alpha_{\text{faire}}:\{\text{Active}, \text{CleftSubj}, \text{CanObj}\} \]
\[ (\text{num:sg}, \text{tse:pst}, \text{mode:ind}, \text{pers:3}) \]

\[ \alpha_{\text{tex}}:\{\text{ProperNoun}\} \]
\[ (\text{fun:subj}, \text{gen:fem}, \text{num:sg}, \text{pers:3}) \]

\[ \alpha_{\text{tarte}}:\{\text{Noun}\} \]
\[ (\text{fun:obj}, \text{gen:fem}, \text{num:sg}) \]

\[ \beta_{\text{la}}:\{\text{DefDet}\} \]
\[ (\text{gen:fem}, \text{num:sg}) \]

Detailed syntactic information
Why Derivation Trees?

More compact than derived trees. Allow fewer and simpler filters.
Derivation Tree Filters

Tree filter types

- **e.g. active/passive**
  - $\alpha_s\{\text{Active}, \text{CleftSubj}, \text{CanObj}\}
  - $\leftrightarrow \bullet_t\{\text{Passive}, \text{CleftAgent}, \text{CanSubj}\}$

- **e.g. NP/Pronoun**
  - $\alpha_s\{\text{CanSubj}\} \leftrightarrow \bullet_t\{\text{CliticSubj}\}$

- **e.g. Assertion/YN-Question**
  - $\emptyset \leftrightarrow \bullet_q\{\text{questionMark}\}$
Meaning Preserving Transformations

Same core meaning (e.g. active/passive)

(Q) C’est Tex qui a fait la tarte. ⇔ (K) C’est par Tex que la tarte a été faite.
(It is Tex who has baked the pie) (It is by Tex that the pie has been baked)

⇔ (K) La tarte a été faite par Tex.
(The pie has been baked by Tex)
Meaning Altering Transformations

Related core meaning: content deleted, added or replaced (e.g. Assertion/Wh-Question)

\[
\begin{align*}
\alpha\text{-dort}: & \left\{ \text{CanSubj} \right\} \\
\alpha\text{-tatou}: & \left\{ \ldots \right\} \\
\beta\text{-chante}: & \left\{ \ldots \right\} \\
\beta\text{-petit}: & \left\{ \ldots \right\} \\
\beta\text{-le}: & \{ \text{defDet} \} \\
\end{align*}
\]

\[
\begin{align*}
\alpha\text{-dort}: & \left\{ \text{whSubj} \right\} \\
\alpha\text{-tatou}: & \left\{ \ldots \right\} \\
\beta\text{-petit}: & \left\{ \ldots \right\} \\
\beta\text{-quel}: & \{ \text{WhDet} \} \\
\text{Quel petit tatou dort?} & \\
\text{Which small armadillo sleeps?} \\
\end{align*}
\]

\[
\begin{align*}
\alpha\text{-dort}: & \left\{ \text{whSubj} \right\} \\
\alpha\text{-tatou}: & \left\{ \ldots \right\} \\
\beta\text{-quel}: & \{ \text{WhDet} \} \\
\text{Quel tatou dort?} & \\
\text{Which armadillo sleeps?} \\
\end{align*}
\]

\[
\begin{align*}
\beta\text{-qui}: & \{ \text{WhPron} \} \\
\text{Qui dort?} & \\
\text{Who sleeps?} \\
\end{align*}
\]

Le petit tatou qui chantera dort.
The small armadillo that will sing sleeps
Evaluation

Correctness

- around 80% of the automatically generated exercises are correct (Manual annotation of a sample of generated exercises)

Productivity

- 52 input formulae ⇒ around 5000 exercises  
  (using SemFraG and lexicon tailored to Tex’s French Grammar vocabulary)

Integration

- Exercises generated by GramEx are integrated in I-FLEG (serious game) and WFLEG (web interface)
Chapter 1: Bonjour!

1.1: Subject pronouns

Grammar topic: Pronoun

Fill in the blank - missing word: Subject pronouns

Fill in the blank with the appropriate subject pronoun. Remplir le trou avec le pronom personnel approprié.

......... adore l’odeur des pesticides

Type your answer here

Time and score

- Question time: 00:00:21
- Exercise time: 00:00:21
- Session time: 00:00:21
- Current Exercise score: 0
- Exercise score in previous session: 3
- Session score: 0
- Session score in previous session: 9
- Best Exercise score: 47 [FLEG.Test]
- Best score: 47 [FLEG.Test]
WFLEG

Tex and Tammy grammar exercises

Amount of tests done: 616
Average score: 43.67%
Average time: 00:01:39

Tests results

Last Tests results

Amount of tests done

Score of the last tests done

Response time of the last tests done
1. Tree Adjoining Grammar
2. Writing Grammar
3. Improving Grammars
4. Grammar for Language Learning
5. Grammar for NL Interfaces
Natural Language Interfaces

I am looking for a car.

- it should be equipped with an equipment
  - with an engine
  - with a diesel engine
- it should be located in a country
  - with an optional feature
  - with a transmission system
- it should be produced by something
  - with an electric engine
  - with a gasoline engine
- it should be sold by a car dealer
  - with a natural gas engine
- it should produce something
  - with a propane engine
Incremental Query Refinement

The user queries the KB using NL

- Possible extensions of the current user query are computed by an automated reasoner
- Each formal extension is then verbalised using NLG
- Grammar based generation is used to convert KB formulae into text
- A statistical module is used to choose the best output

L. Perez-Beltrachini and C. Gardent
Incremental Query Generation

C. Gardent and L. Perez-Beltrachini
A Statistical, Grammar-Based Approach to Micro-Planning
A Statistical Grammar-Based Approach

Input = KB Query

Professor ∩ Researcher ∩ ∃ teach.LogicCourse
∩ ∃ worksAt.AlicanteUniversity

*I am looking for a professor who is a researcher and teaches a course on logic. He should work for Alicante University.*

**Microplanning Task:** Segment, lexicalise, aggregate and realise
A Statistical Grammar-Based Approach

The grammar

- Enforces grammaticality
- Accounts for language variability (paraphrasing)

The Statistical Module (Hypertagger)

- Enforces microplanning choices (fluency)
- Enhances efficiency (speed)
Results: Output quality

Human Evaluation

- 48 input queries
- from 13 knowledge bases (2 not used in training corpus)
- 24 raters
- Online evaluation
- Sliding ruler
- Scale 0-50
- Latin Square design
Results: Output quality

Clarity

Fluency

Template
Hybrid

Symbolic
Hybrid

Human Score

90 / 97
Results: Output quality (BLEU Scores)

<table>
<thead>
<tr>
<th>BLEU Score</th>
<th>Templates</th>
<th>Symbolic</th>
<th>Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>0.6</td>
<td>0.4</td>
<td>0.8</td>
</tr>
<tr>
<td>Generated</td>
<td>0.6</td>
<td>0.4</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Example Output: Sentence Segmentation

3 relations, 4 concepts: 1 sentence
I am looking for a used car whose color should be white, which should be located in a France and whose model should be a toyota 4 runner.

4 relations, 5 concepts: 2 sentences
I am looking for a new car whose exterior color should be beige and whose body style should be a utility vehicle. The new car should run on a natural gas and should be located in a country.

3 relations, 5 concepts: 2 sentences
I am looking for a new car whose body style should be a utility vehicle, an off road. The new car should run on a natural gas and should be located in a country.
Example Output: Syntactic Variation

I am looking for a car dealer located in a country and who should sell a car whose make should be a toyota. The car should run on a fuel and should be equipped with a manual gear transmission system. (Participial)

I am looking for a car dealer who sells a car whose model is a toyota. It should be located in a country. (Sentence with Pronominal Subject)

I am looking for a new car, an off road whose body style should be a utility vehicle. The new car should run on a natural gas and should be located in a country. (Coordinated VP)

I am looking for a car produced by a car make. The car make should be the make of a toyota. The car make should be located in a city and should produce a land rover freelander. (Canonical Declarative Sentence)
Example Output: Aggregation

VP Coordination
NewCar (...) □∃runOn.NaturalGas □∃locatedInCountry.Country
I am looking for a new car (...). This new car (should run on natural gas and should be located in a country)_{VP}. N1 (V1 N1 and V2 N2)

Relative Clause Coordination
CommunicationDevice □∃assistsWith.Understanding
□∃assistsWith.HearingDisability
I am looking for a communication device (which should assist with a understanding and which should assist with a hearing disability)_{RelCl}.
Example Output: Aggregation

NP Coordination
CarDealer ⊓∃sell.CrashCar ⊓∃sell.NewCar
I am looking for a car dealer who should sell (a crash car and a new car)_{NP}.

N-Ary NP Coordination
I am looking for a car equipped with (a manual gear transmission system, an alarm system, a navigation system and an air bag system)_{NP}. 
Summary

Ambiguous Grammar = High Expressivity, Large Search Space

Hypertagging = Making Choices
Thanks!