Automatically Generated Grammar Exercises and Dialogs for Language Learning

Claire Gardent
(Joint work with Marilisa Amoia, Treveur Breteaudiere, Alexandre Denis, German Kruszewski, Cline Moro and Laura Perez-Beltrachini)
CNRS/LORIA, Nancy

February 21, 2014
KU Leuven, Belgium
Outline

Generating Grammar Exercises

▶ Grammar Exercises
▶ Automatic Generation of Grammar Exercises
▶ Results, Evaluation

I-FLEG, a Serious Game for Learning French

▶ Game Scenario
▶ Game Architecture
▶ Automatic Generation of Training Activities

Human-Computer Dialog Systems for Learning French

▶ Free Flowing Dialogs
▶ Constrained Dialogs
Using Surface Realisation to Generate Grammar Exercises

February 17, 2014

C. Gardent and L. Perez-Beltrachini
Using FB-LTAG Derivation Trees to Generate Transformation-Based Grammar Exercises
Proceedings of TAG+11, 2012, Paris, France

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

Built from a single sentence.

[FIB] Complete with an appropriate personal pronoun.

(S) Elle adore les petits tatous
(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
(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
SemTAG based generation for language learning

The *GramEx* framework: selecting sentences and building exercises

1. Sentence generation
   - GraDe

2. Sentence selection
   - *GramEx*
   - query language

3. Exercise generation
   - *GramEx*
   - activity types

Generated Sentences

Grammar (SemTAG) and Lexicon

input constraints

Source sentences

Exercise items
Creating a grammar exercise

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

**Pedagogical goal:** Plural form of irregular

**Exercise type:** Fill-in-the-blank.

1. Select sentences
   \( \Rightarrow \) NP\([num = pl & plural = irreg]\)
   (+canonical order)

2. Process the selected sentence
   NP\([num = pl]\) \(\Rightarrow\) blank
   NP\([lemma = bijou]\) \(\Rightarrow\) hint

(S) Bette aime les bijoux.
(Q) Bette aime les ________ (bijou)
(K) bijoux

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

GramEx’s boolean constraint language: Signature

**morpho-syntactic** (feature structures)
syntactic (tree properties – XMG metagrammar)

```
S
   / \ 
  NP   VP
 /     |
Tammy V[tense=pst mode=ind]
   |
   / \ 
   a   NP[num=sg]
    /     |
    D[num=sg]   NP[num=sg]
     /     |
     une   NP[num=sg gen=f]
      /     |
      voix   Adj[flexion=irreg]
       |
       douce
```

6/19
Selecting appropriate sentences

GramEx's boolean constraint language: syntax and use

- Boolean constraint language:
  - conjunction, disjunction and negation of morpho-syntactic and syntactic properties

- Describe the linguistic requirements of pedagogical goals
  ⇒ linguistic characterization of appropriate source sentences
Selecting appropriate sentences
GramEx's boolean constraint language: an example

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

[ Epith ∧ 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**

POBJinf ∧ CLAUSE

POBJinf ≡ (DE-OBJinf ∨ A-OBJinf)

CLAUSE ≡ Vfin ∧ ¬Mod ∧ ¬CCoord ∧ ¬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 has baked the pie.)

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

X (K) La tarte a été faite par Tex.
(The pie has been 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 has been baked.)
Creating transformation-based grammar exercises

- To identify pairs of sentences that are identical up to a single syntactic transformation:
  - Use the information contained in SemTAG derivation trees
  - Define tree filters on pairs of SemTAG derivation trees
  - Retrieve sentences pairs that match those tree filters
Why SemTAG derivation trees?

- **α-faire:** {Active, CleftSubj, CanObj}
  (num: sg, tse: pst, mode: ind, pers: 3)

- **α-tex:** {ProperNoun}
  (fun: subj, gen: fem, num: sg, pers: 3)

- **α-tarte:** {Noun}
  (fun: obj, gen: fem, num: sg)

- **β-la:** {DefDet}
  (gen: fem, num: sg)

- Detailed syntactic information
- Informational content of the sentence
Why SemTAG derivation trees?

- More abstract description than derived trees
Derivation Tree Filters

Tree filter types

- α\{P_s\} → α\{P_t\}
  - e.g. active/passive
    - s\{Active,CleftSubj,CanObj\} ↔ t\{Passive,CleftAgent,CanSubj\}

- α\{P_s\} → α\{P_t\}
  - e.g. NP/Pronoun
    - s\{CanSubj\} ↔ t\{CliticSubj\}

- α\{P_s\} → α\{P_t\}
  - e.g. Assertion/YN-Question
    - ∅ ↔ q\{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)

\[\begin{align*}
\alpha\text{-faire: } \{\text{Active,CleftSubj,CanObj}\} \\
\alpha\text{-tex: } \{\ldots\} & \quad \beta\text{-avoir: } \{\ldots\} & \quad \alpha\text{-tarte: } \{\ldots\} \\
\downarrow & \quad \downarrow & \quad \downarrow \\
\beta\text{-la: } \{\ldots\}
\end{align*}\]

\[\begin{align*}
\alpha\text{-faire: } \{\text{Passive,CleftAgent,CanSubj}\} \\
\alpha\text{-tex: } \{\ldots\} & \quad \beta\text{-avoir: } \{\ldots\} & \quad \alpha\text{-tarte: } \{\ldots\} \\
\downarrow & \quad \downarrow & \quad \downarrow \\
\beta\text{-la: } \{\ldots\}
\end{align*}\]
Meaning Altering Transformations
Related core meaning: content deleted, added or replaced (e.g. Assertion/Wh-Question)

Le petit tatou qui chantera dort.
The small armadillo that will sing sleeps
Correctness, Productivity, Integration

Manual annotation of a sample of generated exercises

- using SemFraG and lexicon tailored to *Tex’s French Grammar* vocabulary
- around 80% of the automatically generated exercises are correct
- 52 input formulae $\Rightarrow$ around 5000 exercises

Exercises generated by *GramEx* are integrated in I-FLEG (serious game) and WFLEG (web interface)
Welcome!

ALLEGRO is an EU funded INTERREG IV A project which focuses on the development of new technologies for second language learning. Our aim is to exploit research technologies from Natural Language Generation to automatically generate grammar exercises and Learner-Computer Dialog Systems which enable self-practice. While the learner has full autonomy to decide on the exercises to be practiced, the system keeps tracks of the learner’s activities and results. This in turn opens the door for adaptive training systems i.e., systems which promote learning by suggesting new activities based on the learner’s history.

To showcase the power of our technology, we embedded our exercise generator tool both in WFLEG (this web service) and in the IFLEG (Interactive French Learning Game) serious game.

Please select the exercise you want to play with during this session:

Vocabulary exercises, I-FLEG Grammar exercises, Tes and Tammy exercises

W-FLEG Vocabulary exercises

WFLEG includes exercises designed to help learning French vocabulary. The learner is shown an image depicting an object and prompted for its name. All interactions are logged in a database thereby supporting detailed analysis of the learner’s activities. In the future, we plan to use the data to develop adaptive learning systems which make use of a learner’s history to assist the learner in choosing activities likely to enhance his/her progress. The database recording WFLEG activities vocabulary and grammar is common to the IFLEG serious game so that a learner’s activities in both IFLEG and WFLEG can equally be taken into account to analyse his/her progress.

Anyone can play with WFLEG. Register W-FLEG & OpenEFL French Test

I-FLEG Grammar exercises

IFLEG proposes grammar exercises which were automatically generated using Natural Language Generation techniques. The IFLEG grammar exercises can be practiced using the IFLEG serious game where the learner practice by walking through a house, clicking on objects and selecting a training activity related to that object.

IFLEG is a game to help you learn French. Developed by a team of researchers, it is a platform to research how language can be used to teach language. It is an interactive game that will help you learn French. More can be found here.

We provide you with simple yet compelling exercises based on this technology.

Register here to play with our IFLEG grammar based exercises

Tex and Tammy exercises

These exercises follow the curriculum proposed in the Tex and Tammy French Grammar course which is arranged like many other traditional reference grammars with the parts of speech (nouns, verbs, etc.) used to categorize specific grammatical terms (gender of nouns, irregular verbs).

NOTE:
- The original Tex & Tammy is about the epic love story of Tex and Tammy, two teen-agers in America, and their not-so-secret love. In addition to this, we introduce a role, the use of characters including Edouard, a passionate French oral, Joe Dano, a dim-witted guard from College Station, and Gino, a cookman who always getting high and watching the TV on his sofa in the French hometown.

More can be found here.

Register here to play with our Tex & Tammy grammar based exercises
IFLEG, a Serious Game for Learning French

M. Amoia, T. Breteaudiere, A. Denis, C. Gardent and L. Perez-Beltrachini
A Serious Game for Second Language Acquisition in a Virtual Environment
Journal on Systemics, Cybernetics and Informatics, Volume 10, Number 1, 2012.

M. Amoia, C. Gardent and L. Perez-Beltrachini
A serious game for second language acquisition.
Proceedings of the Third International Conference on Computer Aided Education (CSEDU 2011),
Noordwijkerout, The Netherlands.
Game Scenario

The learner moves around a house and clicks on objects.

Object clicks trigger exercises.

All interactions are logged and scores are computed and displayed.

The game consists of four phases:

- Initialisation: Registering, creating an avatar, flying to the IFLEG island.
- Tutorial: Learning to move around.
- Training
- Testing
The Virtual Environment

The I-FLEG game can be played by connecting to the Allegro Island in Second Life. This island is populated with houses of different colors whose content is identical, each of which representing a game unit.
The I-FLEG Architecture
I-FLEG ...

- is embedded in Second Life and provides a 3D virtual environment for immersive situated learning,
- allows several sessions to be played in parallel thereby supporting classroom based teaching,
- is linked to an ontology coding information about the virtual world, thus supporting the automated generation of learning activities,
- integrates a sentence generator which permits producing learning material that is both adaptive and versatile,
- integrates a scoring and a basic error detection and correction mechanism which permit providing the learner with verbal and a numerical feedback,
- maintains a database storing the interactions between learners and system.
Automatic Generation of Training Activities

Starting from an ontology describing the content of the 3D world, the generation module first selects a set of facts and then turns this set of facts into a sentence. From this output sentence, the I-FLEG system will then derive a test item consisting of the exercise itself (e.g., a fill in the blank utterance) and its solution. More specifically, the generation of a test item is parametrized for:

- **Language Content**: the syntactic form the output should have, e.g. a sentence, a noun phrase, etc.
- **Semantic Content**: the set of individuals the test item refers to, generally corresponds to the touched object.
- **Teaching Goal**: the linguistic topic which the test item should illustrate, e.g. adjective morphology.
- **Type of Activity**: the type of test item that should be generated, e.g. vocabulary, fill in the blank, scramble sentence, etc.
Generation Parameters
Semantic Content = [table₁]
Language Content = Sentence
Teaching Goal = “Adjective agreement”
Training Activity = Fill in the blank (FIB)

Generated Test Item
Query = “C’est une table ... (blanc)”
Expected Answer = “C’est une table blanche”

System/User Interaction
System: Please, complete the following sentence.
System: C’est une table ... (blanc).
User: C’est une table blanc.
System: Non, c’est une table blanche.
12 German speaking students from Saarland University played I-FLEG for an hour

The experiment shows

- High practice rate (185 exercises / hour in average)
- Good variety level (different lexical and grammatical exercises tried)
- Good learning rate
High Practice Rate

Number of Interactions by Session

Number of Interactions

TRAINING
EVALUATION
Varied Practice

Number of Exercises by Session

- Varied Practice

- Number of Exercises by Session

- Number of Distinct Exercises

- TRAINING

- EVALUATION
Good Learning Rate
Weakly and Strongly Constrained Dialogs for Language Learning

L.M. Rojas-Barahona, A. Lorenzo, C. Gardent and L. Perez-Beltrachini
Dialog Systems for Language Learning

Chatbot systems

- Free-flowing dialog
  (Shawar and Atwell 2007, Jia 2004)
- Encourages language practice (with a virtual tutor)
- Requires non trivial knowledge of lexicon and syntax

Form-focused dialog Systems

- Restrict the learner answer (Wilske and wolska 2011)
- Provide the learner with an answer template to be filled in by the learner for the dialog to continue
- Focuses on linguistic form (grammar training)
A Dialog Toolbox for Language Learning

Supports both free-flowing and form-focused Man/Machine dialogs

► free-flowing dialog
  Use supervised classification techniques to predict the system dialog move based on learner’s input

► form-focused Man/Machine dialog
  The dialog system provides the learner with a (contextually appropriate) grammar exercise which the learner must solve to construct her answer

Ensures that in both cases, dialogs are restricted to a specific lexical field

Makes use of a finite-state automaton (FSA) describing the range of permissible interactions
FSAs for Language Learning Dialogs

FSA defines the different states in the dialog for each question $Q_X$ ($S =$ system, $P =$ learner)
Dialog in I-FLEG

Explore the house and select a category of exercises when you are ready!
FSAs for Language Learning Dialogs

The FAS models a generic sequence of System Question, Learner Answer, System reaction turns

The **system question** explicitly lists a set of possible answers.

The **learner’s answer** is then classified into four main types.

▶ it selects one of the proposed answers;
▶ it chooses an alternative not mentioned in the system’s question;
▶ it indicates that the learner does not know the correct answer or does not understand the question;
▶ it meets neither of these categories.

2 possible **system reactions**: confirm / inform (i.e., give correct answer)
Example of Free Flowing Dialog

>S: Welcome to my house!, I’m the cat. Do you want to play with me? || ask(ready)
>L: Ok. || answer(ready)
>S: Great! How are you doing? || ask(health)
>L: I’m doing very well, thanks! || answer(health, pref-option)
>S: Awesome! || confirm(health, pref-option)
>S: Come! I want to show you the kitchen! || inform(health)
>S: In the kitchen we prepare lunch. I prepare soup very often. Do you prefer soup or salad? || ask(pref-meal)
>L: I prefer salad. || answer(pref-meal, pref-option)
>S: So do I! || confirm(pref-meal, pref-option)
>S: I also love cheese. || inform(pref-meal)
>S: Where do we usually keep the cheese? || ask(cheese where)
>L: In the fridge || answer(cheese where, pref-option)
>S: You are right || confirm(cheese where, pref-option)
>S: In the fridge || inform(cheese where)
...
>S: Thanks for this little chat! See you soon || quit
Free-Flowing Dialog System

Combines a logistic regression classifier, the FSA and a generation-by-selection algorithm.

1. Produce next **system question** using FSA
2. Send **learner’s answer** to classifier
3. The classifier predicts the system response
4. Check classifier’s prediction and issue System’s response (confirm/inform):
   - “preferred options” and “do not know” learner answers are detected using keyword spotting methods.
     Is the predicted system dialog act consistent with the states in the FS? If not, select a valid dialog move.
   - Is classifier prediction consistent with the prediction made by key word spotting? If not, choose the FSA transition.
   - Issue **System Response**
Data Collection and Classifier Training

Example Tutor-Learner dialogs are collected for training (52 dialogs, 1906 turns)

- They were annotated with dialog moves to support supervised training of the classifier predicting a system answer from a learner input
- The corpus is also used to provide example dialogs for the guided system
Guided Dialog System

Reproduces a dialog in the training corpus but substitute learner's answer with exercises

Correct learner’s answers are annotated with a semantic representation
(Learner Answer, Semantics)

a sentence is generated from this representation
(Generated Sentence, Parse Tree)

Exercises are then generated from the generated sentence and its associated morpho-syntactic information
(Exercises)
### Example Guided Dialog

**Free dialog interaction**

<table>
<thead>
<tr>
<th>S: Vous préférez la soupe ou la salade?</th>
<th>Do you prefer soup or salad?</th>
</tr>
</thead>
<tbody>
<tr>
<td>L: Je préfère la salade.</td>
<td>I am hungry. I prefer salad.</td>
</tr>
</tbody>
</table>

### Guided Dialog Interaction

<table>
<thead>
<tr>
<th>S: Vous préférez la soupe ou la salade?</th>
<th>Do you prefer soup or salad?</th>
</tr>
</thead>
</table>
| **Please answer using the following words:** { je, préférer, le, soupe } | }
Benefits for the Language Learner

Free dialog interaction

- Free uninhibited language practice
- Restricted domain (lexicon)

Guided dialog interaction

- Contextualised Grammar Exercises
Thanks! Questions?