

Génération automatique de texte (Natural Language Generation)

Laura Perez-Beltrachini

Free University of Bolzano-Bozen / LORIA

joint work with Claire Gardent (CNRS/LORIA)

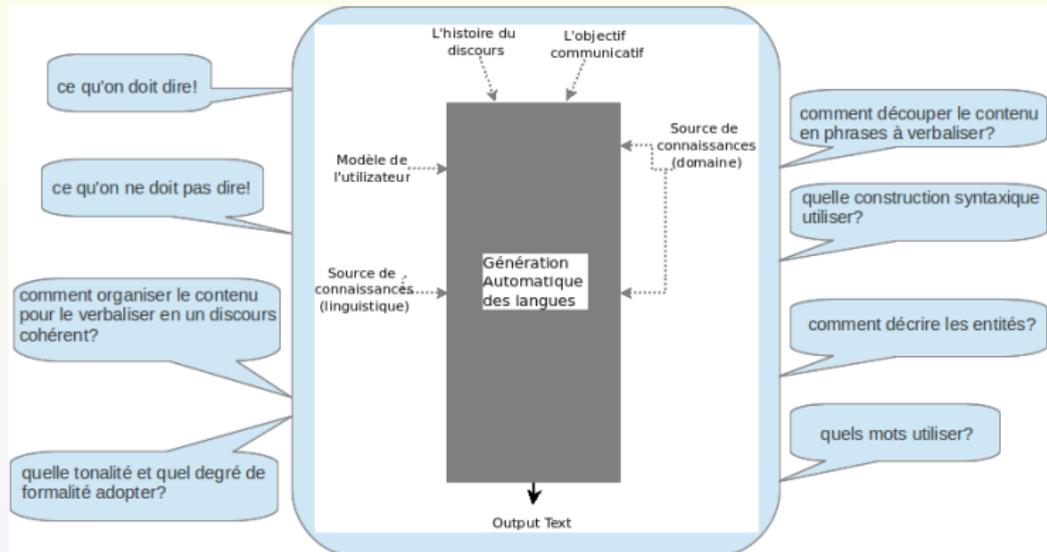
2013

Qu'est ce que la génération ?

- Produire des textes ou de la parole dans une langue humaine.
- A partir de :
 - ▶ Un “but communicatif” e.g., décrire, comparer, expliquer un objet/ un évènement, etc.
 - ▶ Des données :
 - ★ “data-to-text” : base de connaissances, ontologies, base de données, données numériques, etc.
 - ★ “text-to-text” : résumer, simplifier un texte

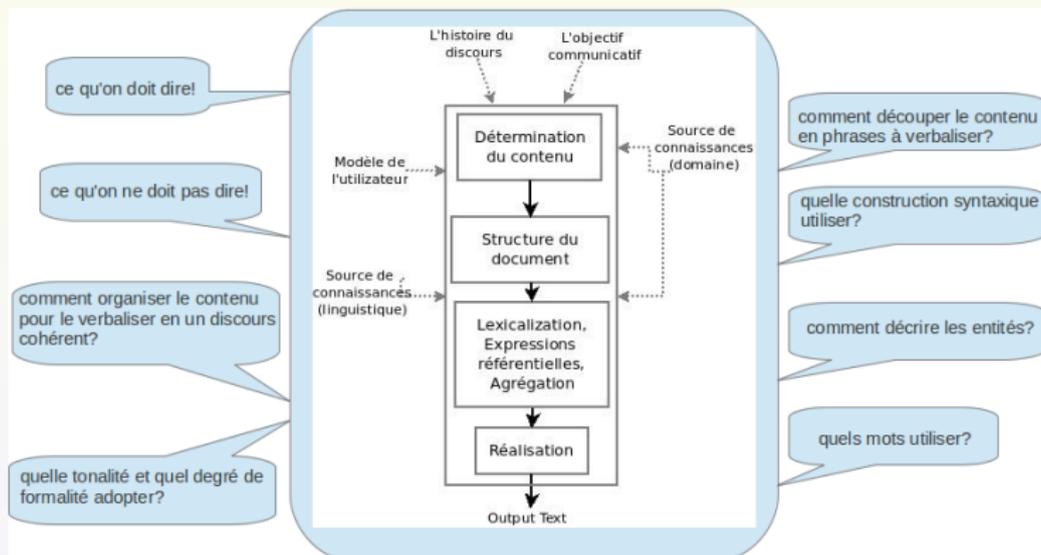
NLG est un problème de choix

Un système NLG devra prendre des décisions



Les modules types d'un système de Génération Automatique

Architecture des Systèmes de Génération



Applications

- Génération de résumés
 - ▶ BabyTalk : Génère un résumé des informations recueillies pour les patients dans un service néonatal de soins intensifs pour les professionnels (médecins, infirmières, anesthésistes, etc) et la famille (public non spécialiste, langage différent, moins techniques)
 - ▶ FoG (Forecast Generator) : génère des bulletins météorologiques bilingues à partir des données météo (vitesse du vent, montant des précipitations etc)
 - ▶ PLANDOC : génère des rapports décrivant les modifications opérées dans un réseau téléphonique
- Génération à partir d'ontologies, bases de connaissances, bases de données
 - ▶ SWAT : verbalisation de bases de connaissances ; permet la visualisation du contenu de ces bases sous forme de texte (plutôt que de formules logiques)
 - ▶ Quelo : verbalisation des requêtes utilisateurs
- Simplification de textes
 - ▶ Pour des lecteurs aphasiques
- Génération des réponses systèmes dans un système de dialogue
- Traduction automatique

Sujets de recherche

- Doctorat (Octobre 2009 / Avril 2013)
 - ▶ Optimisation : développer des algorithmes performants (rapides) pour la génération
 - ▶ Exploitation de la génération automatique de phrases pour l'apprentissage de langues (génération automatique d'exercices de grammaires et de leur solution)
- Post-doc (Avril 2013 - maintenant)
 - ▶ Génération automatique incrementale, génération de requêtes utilisateurs

Notre système de génération automatique de phrases

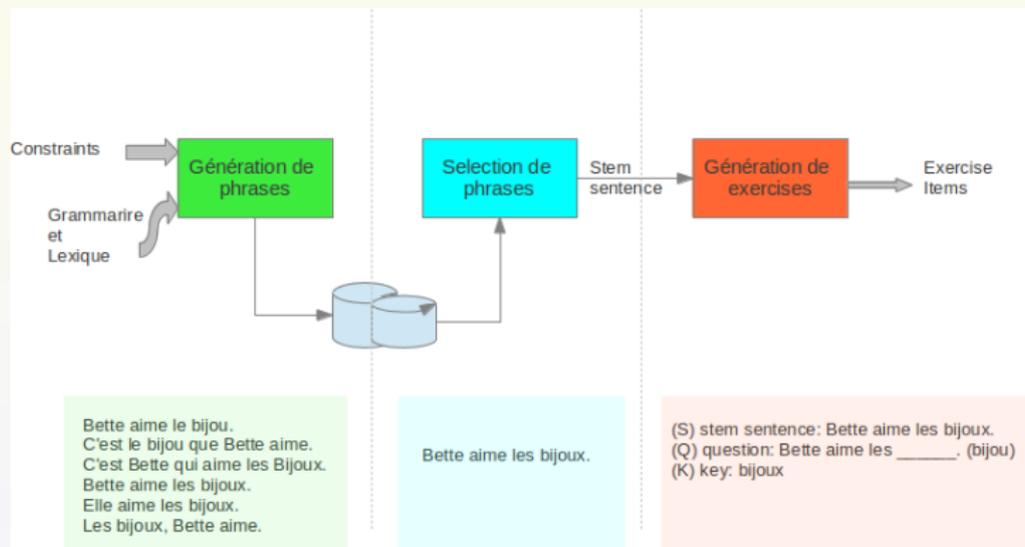


NLG pour l'apprentissage des langues

- **Objectif** générer automatiquement des exercices de grammaires pour les apprenants
- **Idées clés**
 - ▶ **productivité** : certains choix sont sous spécifiés
une ou quelques entrées \Rightarrow plusieurs sorties
 - ▶ **contrôle** : exploiter les informations linguistiques produites par le générateur pour produire des exercices à partir des phrases générées

NLG pour l'apprentissage des langues

GramEx : un environnement pour la génération automatique d'exercices de grammaire [Perez-Beltrachini et al., 2012]



Génération automatique d'exercices de grammaire

Types d'exercices [Perez-Beltrachini et al., 2012]

- les exercices sont produits à partir d'une **phrase sélectionnée (P)** sur la base de ses propriétés linguistiques (e.g., phrase avec adjectifs pour les exercices sur les adjectifs, etc.)

[FIB] Compléter avec le bon pronom personnel.

(P) *Elle adore les petits tatous*

(Q) _____ adore les petits tatous (genre=fem)

(R) elle

[Shuffle] Utiliser les mots suivants pour former une phrase.

(P) *Tammy adore les petits tatous*

(Q) tatous / les / Tammy / petits / adore

(R) Tammy adore les petits tatous.

Génération automatique d'exercices de grammaire

Types d'exercices [Gardent and Perez-Beltrachini, 2012]

- Construits à partir de paires de phrases reliées par une relation syntaxique donnée (e.g., passif/actif)

⇒ (Q) est reliée à (R) par une relation syntaxique

[Reformulation] Mettez la phrase au passif

(Q) *C'est Tex qui a fait la tarte.*

(R) *C'est par Tex que la tarte a été faite.*

- Actif/Passif, GN/Pronom, Assertion/Question, Assertion/YQuestion
oui/non

Un cahier d'exercices sur la web

Exercices organisés par leçon

W-FLEG | Homepage | Vocabulary | F-FLEG exercises | **Tex and Tammy** | Stats & Management | My account | Welcome WFLEG Test | [login](#)

Chapter 1 | Chapter 2 | Chapter 3 | Chapter 4 | Chapter 5 | Chapter 6 | Chapter 7 | Chapter 8 | Chapter 9 | Chapter 10 | Chapter 11 | Chapter 12 | Chapter 13 | [Log out](#)

Chapter 1 : Bonjour!

1.1 : Subject pronouns

Grammar logic : Pronoun

Hello! Try and answer the exercise above!

Fill in the blank - from a set of hint features: Subject pronouns

Fill in the blank with the appropriate subject pronoun given the gen --gender, num --number, pers --person features.

Remplir le trou avec le pronom personnel approprié par rapport aux valeurs données pour gen --genre, num --nombre, pers --personne.

..... (gen:m; num:sg; pers:2) es américain

Type your answer here

[GO!](#)

Need help ?
Below are some links to *Tex & Tammy* original website:

[See *Tex and Tammy* Index](#) | [See Chapter](#) | [See Subchapter](#)

Time and score

Question time.....	00:59:28
Exercise time.....	00:59:28
Session time.....	00:59:28
Current Exercise score.....	0
Exercise score in previous session.....	3
Session score.....	0
Session score in previous session.....	4
Best Exercise score.....	47 [FLEG.Test]
Best score.....	47 [FLEG.Test]

Allegro | | Delivered under the European Union's Support Programme for Culture and Creativity under the Interreg IVA | **Interreg IVA** | | | |

Génération d'exercices dans un jeux sérieux

Le joueur apprend en se déplaçant dans un monde virtuel

[Amoia et al., 2011, Denis et al., 2012, Amoia et al., 2012]



Génération d'exercices dans des dialogues

Context conversationnel [Gardent et al., 2013]



Unterstützt durch den Europäischen Fonds für regionale Entwicklung im Rahmen des Programms Interreg IJK-Großregion. Die Europäische Union investiert in Ihre Zukunft.
Projet cofinancé par le Fonds européen de développement régional dans le cadre du programme Interreg IJK Grande Région. L'Union européenne investit dans votre avenir.



Corpus FLE

Connecté en tant que:

pepa



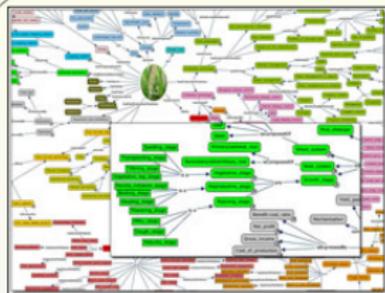
>MANA: Bonjour, je suis le chat. Veux tu jouer avec moi?
>MANA: Entrez le mot manquant:
oui je (vouloir; present tense, indicative) bien
>LEARNER: veux
>MANA: Parfait!
>MANA: Et oui! Merci allons y, comment vas tu?
>MANA: Entrez le mot manquant:
je (aller; present tense, indicative) bien merci
>LEARNER: vais
>MANA: Parfait!
>MANA Viens je vais te montrer la cuisine.
>MANA: Dans la cuisine, on prépare les repas. Je fais souvent des soupe:
tu la soupe ou la salade?
>MANA: Entrez le mot manquant:
je (préférer; present tense, indicative) la soupe
>LEARNER: prefere
>MANA: Presque : préfère.
>MANA: Où est-ce que tu ranges le fromage dans la cuisine d'habitude?
>MANA: Répondre en utilisant les mots suivants:
range / le / réfrigérateur / dans / fromage / on / le
>LEARNER: on le range dans le refrigerateur
>MANA: Non : on range le fromage dans le réfrigérateur.
>MANA: Ou'est ce que tu aimes boire avec ton renas? Moi i'adore le lait.
chat:

Logout

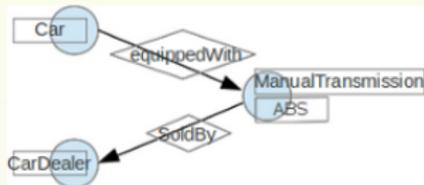
End Dialog

Verbalisation de requêtes guidée par la sémantique

Le projet Quelo



Query Logics



Natural Language Interface (NLG)



Open Query

I am looking for a car.

Scramble Clear

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

Quelo NLJ v2011.07.14-beta

Verbalisation de requêtes guidée par la sémantique

Sujets de recherche

- Lexicalisation des concepts et des relations formant l'ontologie
- Segmentation du contenu à verbaliser en phrases
- Identification des contenus répétitifs pouvant donner lieu à des formulations elliptiques (*Les chiens sont des animaux et les plantes des végétaux* plutôt que *Les chiens sont des animaux et les plantes sont des végétaux*)
- Traitement des “expressions référentielles” : “le chien”, “un chien”, “ce chien”, “il” ?
- Génération incrémentale
 - ▶ le texte généré doit pouvoir être modifié de façon incrémentale au fur et à mesure que l'utilisateur ajoute, enlève ou modifie le contenu de sa requête

Merci beaucoup!
Questions?

References I



Amoia, M., Bretaudiere, T., Denis, A., Gardent, C., and Perez-Beltrachini, L. (2012).
A Serious Game for Second Language Acquisition in a Virtual Environment.
The Journal on Systemics, Cybernetics and Informatics (JSCI), pages 24–34.



Amoia, M., Gardent, C., and Perez-Beltrachini, L. (2011).
A serious game for second language acquisition.

In *CSEDU 2011 (Third International Conference on Computer Aided Education)*, Noordwijkerout, The Netherlands.



Denis, A., Falk, I., Gardent, C., and Perez-Beltrachini, L. (2012).

Representation of linuquistic and domain knowledge for second language learning in virtual worlds.
In *LREC 2012 : Posters*, Istanbul, Turkey.



Gardent, C., Lorenzo, A., LORIA, N., Perez-Beltrachini, L., FUB, B., and Rojas-Barahona, L. (2013).

Weakly and strongly constrained dialogues for language learning.



Gardent, C. and Perez-Beltrachini, L. (2012).

Using FB-LTAG Derivation Trees to Generate Transformation-Based Grammar Exercises.
In *TAG+11 : The 11th International Workshop on Tree Adjoining Grammars and Related Formalisms*, Paris, France.



Perez-Beltrachini, L., Gardent, C., and Kruszewski, G. (2012).

Generating Grammar Exercises.

In *NAACL-HLT 7th Workshop on Innovative Use of NLP for Building Educational Applications*, Montreal, Canada.