

# Méthodes statistiques en traduction automatique

François Yvon

avec l'aide de D. Déchelotte, P. Koehn, P. Langlais, H. Schwenk,  
P. Knight

LIMSI-CNRS & Université Paris-Sud 11

January 20, 2011

# Outline

## Traduire

Anatomie d'un système de traduction statistique

Traduction statistique mot-à-mot et alignements

Les alignements mot-à-mot

Modèles de segments

Décodage et recherche

Traduire la parole

Perspectives

# Rapide point d'histoire

## Contexte :

- ▶ Domaine scientifique complexe faisant intervenir pratiquement tous les aspects du traitement du langage naturel
- ▶ Domaine de recherche ancien (Russe/Anglais)
- ▶ Initialement plutôt approches “classiques”, IA + TAL: utilisation de Parseurs, de règles développées par des humains, ...
- ▶ Initialement, traduction de textes; depuis quelques années, traduction de la parole

# Rapide point d'histoire

## Contexte :

- ▶ Domaine scientifique complexe faisant intervenir pratiquement tous les aspects du traitement du langage naturel
- ▶ Domaine de recherche ancien (Russe/Anglais)
- ▶ Initialement plutôt approches “classiques”, IA + TAL: utilisation de Parseurs, de règles développées par des humains, ...
- ▶ Initialement, traduction de textes; depuis quelques années, traduction de la parole

# Rapide point d'histoire

## Contexte :

- ▶ Domaine scientifique complexe faisant intervenir pratiquement tous les aspects du traitement du langage naturel
- ▶ Domaine de recherche ancien (Russe/Anglais)
- ▶ Initialement plutôt approches “classiques”, IA + TAL: utilisation de Parseurs, de règles développées par des humains, ...
- ▶ Initialement, traduction de textes; depuis quelques années, traduction de la parole

# Rapide point d'histoire

## Contexte :

- ▶ Domaine scientifique complexe faisant intervenir pratiquement tous les aspects du traitement du langage naturel
- ▶ Domaine de recherche ancien (Russe/Anglais)
- ▶ Initialement plutôt approches “classiques”, IA + TAL: utilisation de Parseurs, de règles développées par des humains, ...
- ▶ Initialement, traduction de textes; depuis quelques années, traduction de la parole

# À quoi ça sert ?

Traduire et localiser

## Quelques applications de la TA

- ▶ Grande quantité de pages en langue étrangère  
⇒ inaccessible sans traduction automatique
- ▶ Communauté européenne : nombre croissant de langues officielles ⇒ grand effort de traduction
- ▶ Mondialisation ⇒ multiplication des échanges et rencontres entre locuteurs de langues différentes (conversations, documentations, règlementations, etc)
- ▶ Traduction de parole téléphonique
- ▶ Applications humanitaires et militaires dans des pays étrangers  
(il n'est pas toujours facile de trouver de traducteurs)

La traduction automatique est une technologie émergente en fort développement

# À quoi ça sert ?

Traduire et localiser

## Quelques applications de la TA

- ▶ Grande quantité de pages en langue étrangère  
⇒ inaccessible sans traduction automatique
- ▶ Communauté européenne : nombre croissant de langues officielles ⇒ grand effort de traduction
- ▶ Mondialisation ⇒ multiplication des échanges et rencontres entre locuteurs de langues différentes (conversations, documentations, règlementations, etc)
- ▶ Traduction de parole téléphonique
- ▶ Applications humanitaires et militaires dans des pays étrangers  
(il n'est pas toujours facile de trouver de traducteurs)

La traduction automatique est une technologie émergente en fort développement



# À quoi ça sert ?

Traduire et localiser

## Quelques applications de la TA

- ▶ Grande quantité de pages en langue étrangère  
⇒ inaccessible sans traduction automatique
- ▶ Communauté européenne : nombre croissant de langues officielles ⇒ grand effort de traduction
- ▶ Mondialisation ⇒ multiplication des échanges et rencontres entre locuteurs de langues différentes (conversations, documentations, règlementations, etc)
- ▶ Traduction de parole téléphonique
- ▶ Applications humanitaires et militaires dans des pays étrangers  
(il n'est pas toujours facile de trouver de traducteurs)

La traduction automatique est une technologie émergente en fort développement

# À quoi ça sert ?

Traduire et localiser

## Quelques applications de la TA

- ▶ Grande quantité de pages en langue étrangère  
⇒ inaccessible sans traduction automatique
- ▶ Communauté européenne : nombre croissant de langues officielles ⇒ grand effort de traduction
- ▶ Mondialisation ⇒ multiplication des échanges et rencontres entre locuteurs de langues différentes (conversations, documentations, règlementations, etc)
- ▶ Traduction de parole téléphonique
- ▶ Applications humanitaires et militaires dans des pays étrangers  
(il n'est pas toujours facile de trouver de traducteurs)

La traduction automatique est une technologie émergente en fort développement

# À quoi ça sert ?

Traduire et localiser

## Quelques applications de la TA

- ▶ Grande quantité de pages en langue étrangère  
⇒ inaccessible sans traduction automatique
- ▶ Communauté européenne : nombre croissant de langues officielles ⇒ grand effort de traduction
- ▶ Mondialisation ⇒ multiplication des échanges et rencontres entre locuteurs de langues différentes (conversations, documentations, règlementations, etc)
- ▶ Traduction de parole téléphonique
- ▶ Applications humanitaires et militaires dans des pays étrangers  
(il n'est pas toujours facile de trouver de traducteurs)

La traduction automatique est une technologie émergente en fort développement

# À quoi ça sert ?

Traduire et localiser

## Quelques applications de la TA

- ▶ Grande quantité de pages en langue étrangère  
⇒ inaccessible sans traduction automatique
- ▶ Communauté européenne : nombre croissant de langues officielles ⇒ grand effort de traduction
- ▶ Mondialisation ⇒ multiplication des échanges et rencontres entre locuteurs de langues différentes (conversations, documentations, règlementations, etc)
- ▶ Traduction de parole téléphonique
- ▶ Applications humanitaires et militaires dans des pays étrangers  
(il n'est pas toujours facile de trouver de traducteurs)

La traduction automatique est une technologie émergente en fort développement

# À quoi ça sert ?

Traduire et localiser

## Quelques applications de la TA

- ▶ Grande quantité de pages en langue étrangère  
⇒ inaccessible sans traduction automatique
- ▶ Communauté européenne : nombre croissant de langues officielles ⇒ grand effort de traduction
- ▶ Mondialisation ⇒ multiplication des échanges et rencontres entre locuteurs de langues différentes (conversations, documentations, règlementations, etc)
- ▶ Traduction de parole téléphonique
- ▶ Applications humanitaires et militaires dans des pays étrangers  
(il n'est pas toujours facile de trouver de traducteurs)

La traduction automatique est une technologie émergente en fort développement

# Difficultés de la traduction Automatique

## Pourquoi la traduction est-elle difficile ?

- ▶ Problèmes de morphologie (les langues ne marquent pas toutes les mêmes distinctions)
- ▶ L'ordre des mots varie entre les langues
- ▶ Les mots peuvent avoir plusieurs sens
- ▶ Expressions idiomatiques
- ▶ Résolution des références
- ▶ Tout est-il traduisible ??

# Variations dans l'ordre des mots et des groupes syntaxiques

- ▶ **Anglais** : `adjectif-noun` **et** `sujet-verb-objet`
  - ▶ **Français** : `noun-adjectif`
  - ▶ **Japonais** : `sujet-objet-verb`
  - ▶ **Allemand** : position du verbe “inhabituelle” (rejet en fin dans certaines constructions verbales)
- ⇒ Il faut réordonner” les mots

# Belle marquise...

## Exemples :

- ▶ anglais / japonais
  - ▶ *IBM bought Lotus* ↔ *IBM Lotus bought*
  - ▶ *Reporters said IBM bought Lotus* ↔ *Reporters IBM Lotus bought said*
- ▶ français / anglais
  - ▶ *une voiture neuve* ↔ *a new car*
  - ▶ *une voiture de course* ↔ *a racing car*



# Quelques problèmes de sémantique

## Ambiguïté sémantique: multiplicité des sens d'un mot

- ▶ Anglais : *plant* (arbre ou entreprise); *bank* (banque ou bord d'une rivière)
  - ▶ Français : *allumer* (une cigarette ou le moteur), *couper* (les cheveux (en 4) ou le moteur)
- ⇒ Souvent les sens différents correspondent à des traductions différentes

## Idiomes

- ▶ Expressions poly-léxématiques qu'on ne peut traduire mot par mot (= *non-compositionnelles*)
- ▶ *être au pied du mur* → *To be at the foot of the wall* ?
- ▶ *tenir sa langue* → *keep ones tongue* ?
- ▶ *ne pas mâcher ses mots* → *not to chew ones words* ?

# Quelques problèmes de sémantique

## Ambiguïté sémantique: multiplicité des sens d'un mot

- ▶ Anglais : *plant* (arbre ou entreprise); *bank* (banque ou bord d'une rivière)
  - ▶ Français : *allumer* (une cigarette ou le moteur), *couper* (les cheveux (en 4) ou le moteur)
- ⇒ Souvent les sens différents correspondent à des traductions différentes

## Idiomes

- ▶ Expressions poly-léxématiques qu'on ne peut traduire mot par mot (= **non-compositionnelles**)
- ▶ *être au pied du mur* → *To be at the foot of the wall* ?
- ▶ *tenir sa langue* → *keep ones tongue* ?
- ▶ *ne pas mâcher ses mots* → *not to chew ones words* ?

# Problèmes de morpho-syntaxe

## Utilisation des pronoms

- ▶ Certaines langues autorisent l'omission des pronoms personnels sujet (eg. espagnol, italien)
- ▶ Parfois la forme verbale identifie le bon pronom
- ▶ Mais pas son genre ... (*he*, *she* ou *it*)

## Marques flexionnelles

- ▶ *He is nice* → *Il est beau* vs *She is nice* → *Elle est belle* : accord d'un côté mais pas de l'autre
- ▶ *Ich gebe der Frau einen Buch* → *Je donne à la femme un livre*: marques casuelles d'un côté pas de l'autre
- ▶ Arabe : singulier, pluriel et dual

Traduire est + difficile quand la cible est morphologiquement riche.

# Problèmes de morpho-syntaxe

## Utilisation des pronoms

- ▶ Certaines langues autorisent l'omission des pronoms personnels sujet (eg. espagnol, italien)
- ▶ Parfois la forme verbale identifie le bon pronom
- ▶ Mais pas son genre ... (*he*, *she* ou *it*)

## Marques flexionnelles

- ▶ *He is nice* → *Il est beau* vs *She is nice* → *Elle est belle* : accord d'un côté mais pas de l'autre
- ▶ *Ich gebe der Frau einen Buch* → *Je donne à la femme un livre*: marques casuelles d'un côté pas de l'autre
- ▶ Arabe : singulier, pluriel et dual

Traduire est + difficile quand la cible est morphologiquement riche.

# Problèmes de morpho-syntaxe

## Utilisation des pronoms

- ▶ Certaines langues autorisent l'omission des pronoms personnels sujet (eg. espagnol, italien)
- ▶ Parfois la forme verbale identifie le bon pronom
- ▶ Mais pas son genre ... (*he*, *she* ou *it*)

## Marques flexionnelles

- ▶ *He is nice* → *Il est beau* vs *She is nice* → *Elle est belle* : accord d'un côté mais pas de l'autre
- ▶ *Ich gebe der Frau einen Buch* → *Je donne à la femme un livre*: marques casuelles d'un côté pas de l'autre
- ▶ Arabe : singulier, pluriel et dual

Traduire est + difficile quand la cible est morphologiquement riche.

# Les difficultés de la traduction

- ▶ en: Should this fail to materialise, we should not be surprised if public opinion proves sceptical about Europe, or even rejects it.
- ▶ fr: Faute de quoi comment s'étonner du scepticisme, voire du rejet de l'Europe dans l'opinion publique.
- ▶ ⇒ construction verbale (*proves sceptical...rejects*) en anglais, nominale (*scepticisme...du rejet*) en français

# Les difficultés de la traduction

- ▶ en: The Intergovernmental Conference - to address a third subject - on the reform of the European institutions is also of decisive significance for us in Parliament.
- ▶ fr: Pour nous, en tant que Parlement - et j'aborde là un troisième thème -, la Conférence intergouvernementale sur la réforme des institutions européennes est aussi éminemment importante.
- ▶ ⇒ déplacement majeur de *...for us in parliament*

# Les difficultés de la traduction

- ▶ en: ... because it has made a major contribution to this project.
- ▶ fr: ... parce qu'il a contribué fortement à ce travail.
- ▶ ⇒ construction nominale (*made a major contribution*) en anglais, verbale (*a contribué*) en français



# Quelques approches de la traduction automatique

## Approches :

- ▶ Traduction mot par mot
- ▶ Transfert syntaxique
- ▶ Utilisation d'une "langue" pivot (Interlingua)
- ▶ Approches utilisant des textes déjà traduits
  - Exemple : l'approche proposée dans des traductions effectuées par des humains
  - Exemple : le problème d'acquisition de connaissances
  - Exemple : *corpus machine translation (CMT)*
  - Exemple : *memory-translation-based MT*
  - Exemple : *statistical machine translation (Statistical machine translation)*
- ▶ Systèmes hybrides



# Quelques approches de la traduction automatique

## Approches :

- ▶ Traduction mot par mot
- ▶ Transfert syntaxique
- ▶ Utilisation d'une "langue" pivot (Interlingua)
- ▶ Approches utilisant des textes déjà traduits
  - Utiliser "l'expertise" contenue dans des traductions effectuées par des humains
  - Utiliser le modèle de la machine de traduction (Machinetranslation) :
    - Exemple : *Le système de gestion des données* (source)
    - *The data management system* (target)
  - Approches hybrides (Exemple : *Statistical Machine Translation*)
- ▶ Systèmes hybrides

# Quelques approches de la traduction automatique

## Approches :

- ▶ Traduction mot par mot
- ▶ Transfert syntaxique
- ▶ Utilisation d'une "langue" pivot (Interlingua)
- ▶ Approches utilisant des textes déjà traduits
  - ▶ Utiliser "l'expertise" contenue dans des traductions effectuées par des humains
- ⇒ Minimiser le problème d'acquisition de connaissances
  - ▶ *Example-based machine translation (EBMT)*
  - ▶ *Corpus-, memory- similarity-based MT*
  - ▶ **Approche statistique** (*Statistical machine translation*)
- ▶ Systèmes hybrides

# Quelques approches de la traduction automatique

## Approches :

- ▶ Traduction mot par mot
- ▶ Transfert syntaxique
- ▶ Utilisation d'une "langue" pivot (Interlingua)
- ▶ Approches utilisant des textes déjà traduits
  - ▶ Utiliser "l'expertise" contenue dans des traductions effectuées par des humains
- ⇒ Minimiser le problème d'acquisition de connaissances
  - ▶ *Example-based machine translation (EBMT)*
  - ▶ *Corpus-, memory- similarity-based MT*
  - ▶ *Approche statistique (Statistical machine translation)*
- ▶ Systèmes hybrides

# Quelques approches de la traduction automatique

## Approches :

- ▶ Traduction mot par mot
- ▶ Transfert syntaxique
- ▶ Utilisation d'une "langue" pivot (Interlingua)
- ▶ Approches utilisant des textes déjà traduits
  - ▶ Utiliser "l'expertise" contenue dans des traductions effectuées par des humains
  - ⇒ Minimiser le problème d'acquisition de connaissances
    - ▶ *Example-based machine translation (EBMT)*
    - ▶ *Corpus-, memory- similarity-based MT*
    - ▶ *Approche statistique (Statistical machine translation)*
- ▶ Systèmes hybrides

# Quelques approches de la traduction automatique

## Approches :

- ▶ Traduction mot par mot
- ▶ Transfert syntaxique
- ▶ Utilisation d'une "langue" pivot (Interlingua)
- ▶ Approches utilisant des textes déjà traduits
  - ▶ Utiliser "l'expertise" contenue dans des traductions effectuées par des humains
  - ⇒ Minimiser le problème d'acquisition de connaissances
  - ▶ *Example-based machine translation (EBMT)*
  - ▶ *Corpus-, memory- similarity-based MT*
  - ▶ *Approche statistique (Statistical machine translation)*
- ▶ Systèmes hybrides

# Quelques approches de la traduction automatique

## Approches :

- ▶ Traduction mot par mot
- ▶ Transfert syntaxique
- ▶ Utilisation d'une "langue" pivot (Interlingua)
- ▶ Approches utilisant des textes déjà traduits
  - ▶ Utiliser "l'expertise" contenue dans des traductions effectuées par des humains
  - ⇒ Minimiser le problème d'acquisition de connaissances
    - ▶ *Example-based machine translation (EBMT)*
    - ▶ *Corpus-, memory- similarity-based MT*
    - ▶ *Approche statistique (Statistical machine translation)*
- ▶ Systèmes hybrides



# Quelques approches de la traduction automatique

## Approches :

- ▶ Traduction mot par mot
- ▶ Transfert syntaxique
- ▶ Utilisation d'une "langue" pivot (Interlingua)
- ▶ Approches utilisant des textes déjà traduits
  - ▶ Utiliser "l'expertise" contenue dans des traductions effectuées par des humains
  - ⇒ Minimiser le problème d'acquisition de connaissances
    - ▶ *Example-based machine translation (EBMT)*
    - ▶ *Corpus-, memory- similarity-based MT*
    - ▶ **Approche statistique** (*Statistical machine translation*)
- ▶ Systèmes hybrides

# Quelques approches de la traduction automatique

## Approches :

- ▶ Traduction mot par mot
- ▶ Transfert syntaxique
- ▶ Utilisation d'une "langue" pivot (Interlingua)
- ▶ Approches utilisant des textes déjà traduits
  - ▶ Utiliser "l'expertise" contenue dans des traductions effectuées par des humains
  - ⇒ Minimiser le problème d'acquisition de connaissances
    - ▶ *Example-based machine translation (EBMT)*
    - ▶ *Corpus-, memory- similarity-based MT*
    - ▶ **Approche statistique** (*Statistical machine translation*)
- ▶ Systèmes hybrides

# Traduire avec un dictionnaire

## Traduction mot par mot

- ▶ Utilisation d'un dictionnaire bilingue pour traduire chaque mot dans un texte (de façon indépendante)
- ▶ Facile à mettre en place
- ▶ Le résultat donne une vague idée du sujet du texte
- ▶ Problèmes avec l'ordre, le sens de mots, les expressions, ...

très mauvaise solution

# Traduire avec un dictionnaire

## Traduction mot par mot

- ▶ Utilisation d'un dictionnaire bilingue pour traduire chaque mot dans un texte (de façon indépendante)
- ▶ Facile à mettre en place
- ▶ Le résultat donne une vague idée du sujet du texte
- ▶ Problèmes avec l'ordre, le sens de mots, les expressions, ...

très mauvaise solution

# Traduction Automatique par transfert syntaxique

## Principes

1. Analyse (morpho-) syntaxique de la phrase source
2. Ré-arrangement des composantes
3. Traduction des mots
4. Génération de la phrase source

## Pros & Cons

- ▶ Aborde le réordonnement des mots
- ▶ Nécessite des règles de transfert pour chaque paire de langue
- ▶ Travail difficile, besoin d'experts (linguistes bi-lingues)

Peut être performant, mais lourd à développer et à maintenir,  
faible robustesse

# Traduction Automatique par transfert syntaxique

## Principes

1. Analyse (morpho-) syntaxique de la phrase source
2. Ré-arrangement des composantes
3. Traduction des mots
4. Génération de la phrase source

## Pros & Cons

- ▶ Aborde le réordonnement des mots
- ▶ Nécessite des règles de transfert pour chaque paire de langue
- ▶ Travail difficile, besoin d'experts (linguistes bi-lingues)

Peut être performant, mais lourd à développer et à maintenir,  
faible robustesse

# Utilisation d'une langue pivot

## Approche Interlingua

- ▶ Utiliser une “langue” intermédiaire pour le transfert par exemple la logique des prédicats
- ▶ Traduire: analyser l'entrée, construire la forme logique, générer en langue cible

Traiter plusieurs paires de langues est facilité

# Utilisation d'une langue pivot

## Approche Interlingua

- ▶ Utiliser une “langue” intermédiaire pour le transfert par exemple la logique des prédicats
- ▶ Traduire: analyser l'entrée, construire la forme logique, générer en langue cible

Traiter plusieurs paires de langues est facilité



# Traduction à base d'exemples (EBMT)

## Les grands principes

- ▶ systématiser l'utilisation de **mémoires de traduction** = phrases déjà traduites
- ▶ pour généraliser: décomposer les traductions mémorisées en plusieurs appariements
- ▶ pour traduire: décomposer la phrase cible, l'analyser, extraire des exemples exploitables, recombinaison des fragments extraits

## Défis

- ▶ Localiser des phrases/fragments similaires dans un corpus
- ▶ Alignement des fragments
- ▶ Choisir les bons fragments
- ▶ Recombinaison des fragments dans la phrase cible
- ▶ Sélection parmi plusieurs solutions

# Traduction à base d'exemples (EBMT)

## Les grands principes

- ▶ systématiser l'utilisation de **mémoires de traduction** = phrases déjà traduites
- ▶ pour généraliser: décomposer les traductions mémorisées en plusieurs appariements
- ▶ pour traduire: décomposer la phrase cible, l'analyser, extraire des exemples exploitables, recombinaison des fragments extraits

## Défis

- ▶ Localiser des phrases/fragments similaires dans un corpus
- ▶ Alignement des fragments
- ▶ Choisir les bons fragments
- ▶ Recombinaison des fragments dans la phrase cible
- ▶ Sélection parmi plusieurs solutions

# BLEU : une métrique pour la TA

- ▶ Évaluer les systèmes de traduction est une question difficile
- ▶ Il existe des métriques subjectives et objectives
- ▶ La mesure BLEU ...
- ▶ quantifie la ressemblance avec des traductions de références
- ▶ Formellement, une moyenne géométrique de la précision  $n$ -gram

## Calcul du score BLEU

I am feeling good      Ref1: I am happy

Ref2: I am feeling very good

# BLEU : une métrique pour la TA

- ▶ Évaluer les systèmes de traduction est une question difficile
- ▶ Il existe des métriques subjectives et objectives
- ▶ La mesure BLEU ...
- ▶ quantifie la ressemblance avec des traductions de références
- ▶ Formellement, une moyenne géométrique de la précision  $n$ -gram

## Calcul du score BLEU

I am feeling good      Ref1: I am happy

I am feel ing      I am

Ref2: I am feeling very good

I am feel ing very good

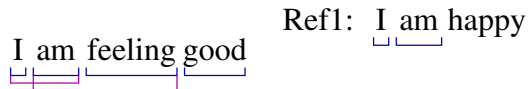
$$p_1 = 1$$

# BLEU : une métrique pour la TA

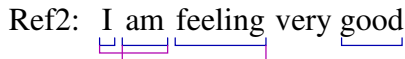
- ▶ Évaluer les systèmes de traduction est une question difficile
- ▶ Il existe des métriques subjectives et objectives
- ▶ La mesure BLEU ...
- ▶ quantifie la ressemblance avec des traductions de références
- ▶ Formellement, une moyenne géométrique de la précision  $n$ -gram

## Calcul du score BLEU

I am feeling good      Ref1: I am happy



Ref2: I am feeling very good



$$p_1 = 1 \quad p_2 = \frac{2}{3}$$

# BLEU : une métrique pour la TA

- ▶ Évaluer les systèmes de traduction est une question difficile
- ▶ Il existe des métriques subjectives et objectives
- ▶ La mesure BLEU ...
- ▶ quantifie la ressemblance avec des traductions de références
- ▶ Formellement, une moyenne géométrique de la précision  $n$ -gram

## Calcul du score BLEU

I am feeling good      Ref1: I am happy

Ref2: I am feeling very good

$$p_1 = 1 \quad p_2 = \frac{2}{3} \quad p_3 = \frac{1}{2} \quad p_4 = \frac{0}{1}$$

# Outline

Traduire

Anatomie d'un système de traduction statistique

Traduction statistique mot-à-mot et alignements

Les alignements mot-à-mot

Modèles de segments

Décodage et recherche

Traduire la parole

Perspectives

# Le modèle du canal bruité

## Canal Bruité

**f** une phrase *source* [Français], traduire **f**= résoudre:

$$\operatorname{argmax}_e P(\mathbf{e}|\mathbf{f}) = \operatorname{argmax}_e P(\mathbf{f}|\mathbf{e})P(\mathbf{e})$$

où le maximum est calculé sur l'ensemble de toutes les phrases **e** du langage *cible* [English],

## Les deux modèles

- ▶  $p(\mathbf{f}|\mathbf{e})$  définit le **modèle de transfert**:
  - ▶ mesure la qualité de l'appariement entre **e** et **f**
  - ▶ estimée sur des **corpus parallèles**
- ▶  $p(\mathbf{e})$  définit le **modèle de langue**
  - ▶ mesure la "qualité" (syntaxique, sémantique) de la traduction en cible
  - ▶ estimée sur des corpus monolingues
  - ▶ (cf. cours modélisation du langage)



# Le modèle du canal bruité

## Canal Bruité

**f** une phrase *source* [Français], traduire **f** = résoudre:

$$\operatorname{argmax}_e P(\mathbf{e}|\mathbf{f}) = \operatorname{argmax}_e P(\mathbf{f}|\mathbf{e})P(\mathbf{e})$$

où le maximum est calculé sur l'ensemble de toutes les phrases **e** du langage *cible* [English],

## Les deux modèles

- ▶  $p(\mathbf{f}|\mathbf{e})$  définit le **modèle de transfert**:
  - ▶ mesure la qualité de l'appariement entre **e** et **f**
  - ▶ estimée sur des **corpus parallèles**
- ▶  $p(\mathbf{e})$  définit le **modèle de langue**
  - ▶ mesure la "qualité" (syntaxique, sémantique) de la traduction en cible
  - ▶ estimée sur des corpus monolingues
  - ▶ (cf. cours modélisation du langage)

# Pourquoi le canal bruité ?

## Un modèle paradoxal

Pour produire une phrase en français, le locuteur choisit une phrase en anglais (avec probabilité  $P(\mathbf{e})$ ), puis la passe dans un canal bruité.

## L'intérêt du canal bruité

Décomposer:

$$\operatorname{argmax}_{\mathbf{e}} P(\mathbf{e}|\mathbf{f}) = \operatorname{argmax}_{\mathbf{e}} P(\mathbf{f}|\mathbf{e})P(\mathbf{e})$$

permet de séparer le problème:

- ▶  $P(\mathbf{f}|\mathbf{e})$  sélectionne des phrases cibles qui sont des appariements probables de  $\mathbf{f}$
- ▶  $P(\mathbf{e})$  sélectionne des phrases cibles qui sont probablement des phrases correctes

# Pourquoi le canal bruité ?

## Un modèle paradoxal

Pour produire une phrase en français, le locuteur choisit une phrase en anglais (avec probabilité  $P(\mathbf{e})$ ), puis la passe dans un canal bruité.

## L'intérêt du canal bruité

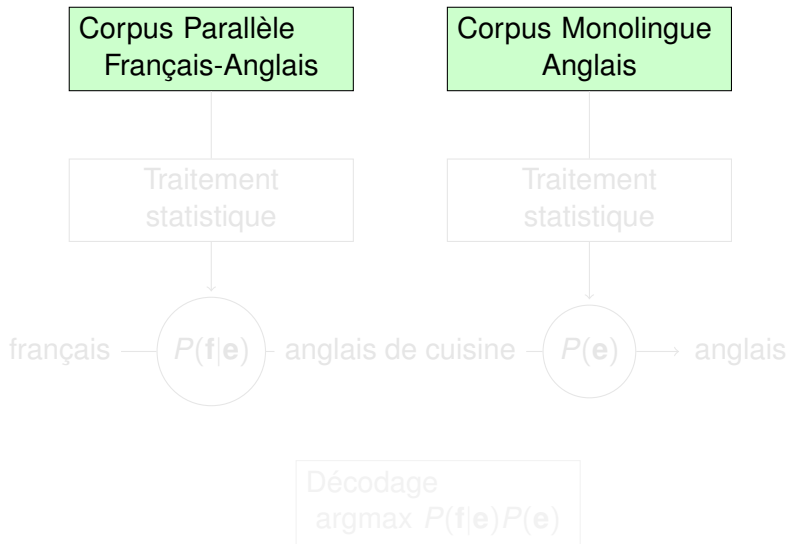
Décomposer:

$$\operatorname{argmax}_{\mathbf{e}} P(\mathbf{e}|\mathbf{f}) = \operatorname{argmax}_{\mathbf{e}} P(\mathbf{f}|\mathbf{e})P(\mathbf{e})$$

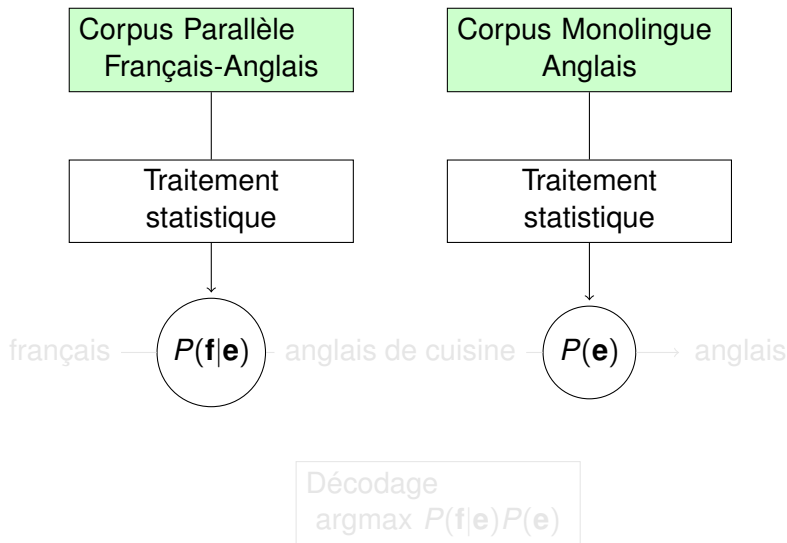
permet de séparer le problème:

- ▶  $P(\mathbf{f}|\mathbf{e})$  sélectionne des phrases cibles qui sont des appariements probables de  $\mathbf{f}$
- ▶  $P(\mathbf{e})$  sélectionne des phrases cibles qui sont probablement des phrases correctes

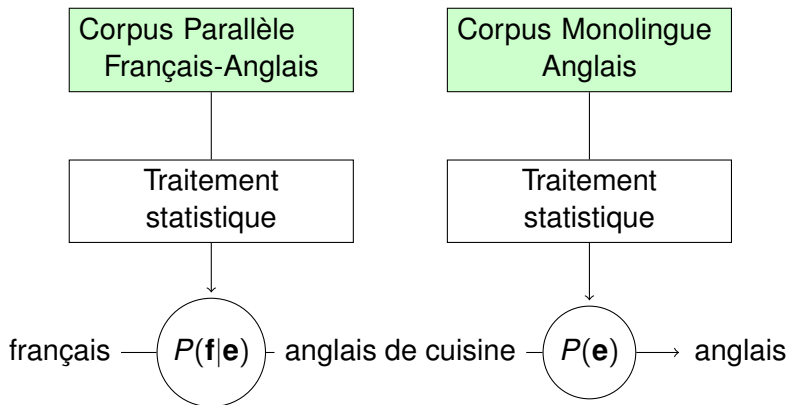
# Des ressources, des modèles, des algorithmes



# Des ressources, des modèles, des algorithmes

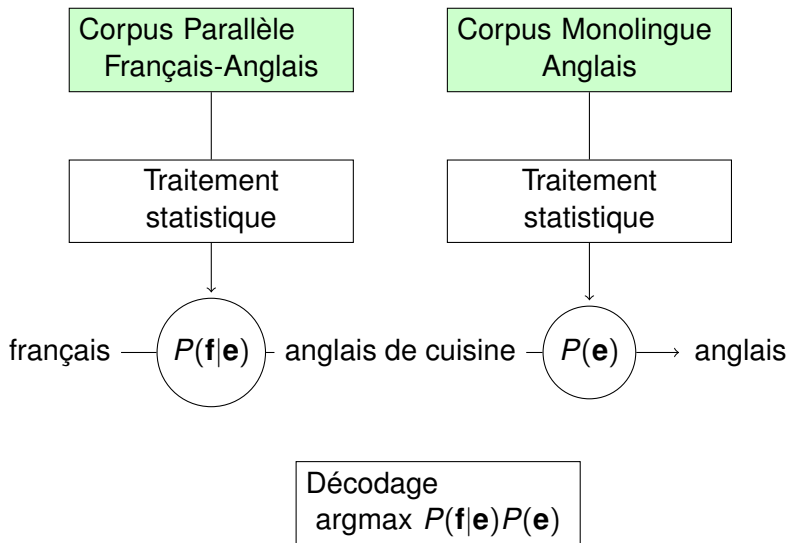


# Des ressources, des modèles, des algorithmes



Décodage  
 $\operatorname{argmax} P(\mathbf{f}|\mathbf{e})P(\mathbf{e})$

# Des ressources, des modèles, des algorithmes



# Les corpus alignés: le nerf de la guerre

Pourquoi donc les producteurs d'armes de l'UE devraient-ils s'enrichir sur le dos de personnes innocentes ?	So why should EU arms producers profit at the expense of innocent people?
---	---

## Sources principales:

- ▶ Institutions multi-lingues (Canada, Suisse, EU)
- ▶ Textes techniques (documentations, cours)
- ▶ "Classiques" de la littérature (Bible, etc)

Ces ressources sont rares, mais faciles à aligner phrase à phrase



# Les corpus alignés: le nerf de la guerre

Pourquoi donc les producteurs d'armes de l'UE devraient-ils s'enrichir sur le dos de personnes innocentes ?	So why should EU arms producers profit at the expense of innocent people?
---	---

## Sources principales:

- ▶ Institutions multi-lingues (Canada, Suisse, EU)
- ▶ Textes techniques (documentations, cours)
- ▶ “Classiques” de la littérature (Bible, etc)

Ces ressources sont rares, mais faciles à aligner phrase à phrase

# Les corpus alignés: le nerf de la guerre

Pourquoi donc les producteurs d'armes de l'UE devraient-ils s'enrichir sur le dos de personnes innocentes ?	So why should EU arms producers profit at the expense of innocent people?
---	---

## Sources principales:

- ▶ Institutions multi-lingues (Canada, Suisse, EU)
- ▶ Textes techniques (documentations, cours)
- ▶ “Classiques” de la littérature (Bible, etc)

Ces ressources sont rares, mais faciles à aligner phrase à phrase

# Outline

Traduire

Anatomie d'un système de traduction statistique

**Traduction statistique mot-à-mot et alignements**

Les alignements mot-à-mot

Modèles de segments

Décodage et recherche

Traduire la parole

Perspectives

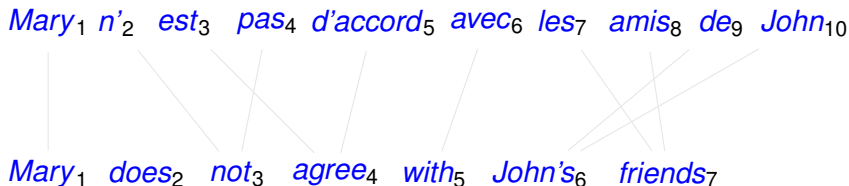
# Introduction des alignements

- ▶ l'estimation directe de  $P(\mathbf{f}|\mathbf{e})$  est impossible
- ▶ la décomposition  $P(\mathbf{f}|\mathbf{e}) = \prod_i P(f_i|e_i)$  est trop simpliste
- ▶  $\Rightarrow$  décomposition via des **alignements latents**:

$$P(\mathbf{f}|\mathbf{e}) = \sum_{\mathbf{a} \in \mathcal{A}} P(\mathbf{a}, \mathbf{f}|\mathbf{e})$$

où  $\mathcal{A}$  est l'ensemble des *alignements* entre  $\mathbf{e}$  et  $\mathbf{f}$

# Alignement de mots



- ▶ un alignement = relation sur  $I \times J$ .

$$a = \{(1, 1), (2, 3), (3, 4), (4, 3), (5, 4) \dots\}$$

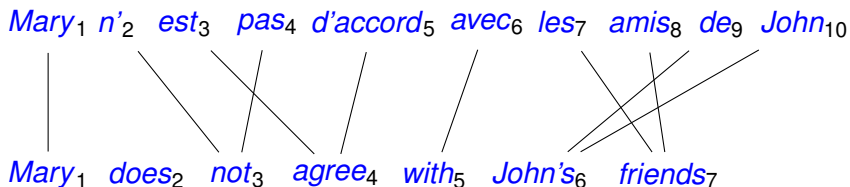
$2^{I \times J}$  relations possibles

- ▶ un alignement = application partielle de  $J$  vers  $I$ :

$$a = [1, 3, 4, 2, 4, 5, 7, 7, 6, 6]$$

“seulement”  $I^J$  applications possibles

# Alignement de mots



- un alignement = relation sur  $I \times J$ .

$$a = \{(1, 1), (2, 3), (3, 4), (4, 3), (5, 4) \dots\}$$

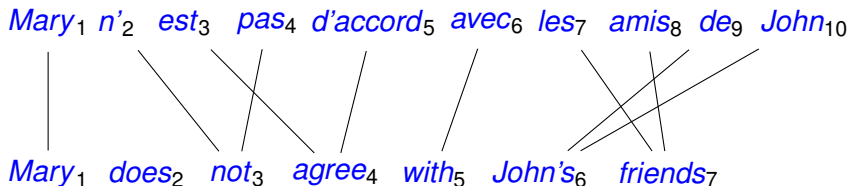
$2^{I \times J}$  relations possibles

- un alignement = application partielle de  $J$  vers  $I$ :

$$a = [1, 3, 4, 2, 4, 5, 7, 7, 6, 6]$$

“seulement”  $I^J$  applications possibles

# Alignement de mots



- un alignement = **relation** sur  $I \times J$ .

$$a = \{(1, 1), (2, 3), (3, 4), (4, 3), (5, 4) \dots\}$$

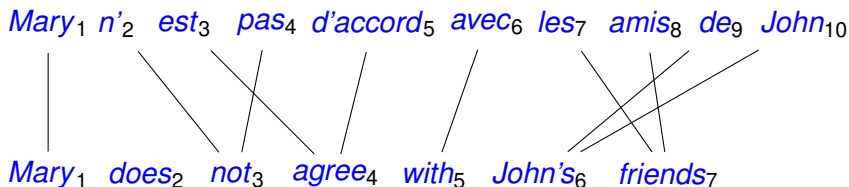
$2^{I \times J}$  relations possibles

- un alignement = **application partielle** de  $J$  vers  $I$ :

$$a = [1, 3, 4, 2, 4, 5, 7, 7, 6, 6]$$

“seulement”  $I^J$  applications possibles

# Alignement de mots



- un alignement = **relation** sur  $I \times J$ .

$$a = \{(1, 1), (2, 3), (3, 4), (4, 3), (5, 4) \dots\}$$

$2^{I \times J}$  relations possibles

- un alignement = **application partielle** de  $J$  vers  $I$ :

$$a = [1, 3, 4, 2, 4, 5, 7, 7, 6, 6]$$

“seulement”  $I^J$  applications possibles



# Apories des alignements mots à mots

- ▶ chaque mot source apparié à exactement un mot cible ?
  - ▶ e = *slap* / f = *donner une claque*
  - ▶ e = *thank you* / f = *merci*
  - ▶ e = *in accordance with* / f = *dans le respect de*
- ▶ et les mots non alignés (en cible) ?
- ▶ asymétrie des alignements

# Modélisation avec alignements cachés

## Notations

- ▶  $\mathbf{f} = f_1^J = f_1 \dots f_J$  la phrase source ( $J$  mots)
- ▶  $\mathbf{e} = e_1^I = e_1 \dots e_I$  la phrase cible ( $I$  mots)
- ▶ problème: décomposer  $P(\mathbf{a}, \mathbf{f} | \mathbf{e})$

## Structure du modèle génératif

- ▶ choisir  $J$  sachant  $e_1^I$
- ▶ pour chaque position  $j \in [1 : J]$ 
  - ▶ choisir  $a_j$  sachant  $J, a_1^{j-1}, f_1^{j-1}, e_1^I$
  - ▶ choisir  $f_j$  sachant  $J, a_1^j, f_1^{j-1}, e_1^I$

$$P(a_1^J, f_1^J | e_1^I) = P(J | e_1^I) \prod_j P(a_j | a_1^{j-1}, f_1^{j-1}, e_1^I) P(f_j | a_1^j, f_1^{j-1}, e_1^I)$$

# Modélisation avec alignements cachés

## Notations

- ▶  $\mathbf{f} = f_1^J = f_1 \dots f_J$  la phrase source ( $J$  mots)
- ▶  $\mathbf{e} = e_1^I = e_1 \dots e_I$  la phrase cible ( $I$  mots)
- ▶ problème: décomposer  $P(\mathbf{a}, \mathbf{f} | \mathbf{e})$

## Structure du modèle génératif

- ▶ choisir  $J$  sachant  $e_1^I$
- ▶ pour chaque position  $j \in [1 : J]$ 
  - ▶ choisir  $a_j$  sachant  $J, a_1^{j-1}, f_1^{j-1}, e_1^I$
  - ▶ choisir  $f_j$  sachant  $J, a_1^j, f_1^{j-1}, e_1^I$

$$P(a_1^J, f_1^J | e_1^I) = P(J | e_1^I) \prod_j P(a_j | a_1^{j-1}, f_1^{j-1}, e_1^I) P(f_j | a_1^j, f_1^{j-1}, e_1^I)$$

# Modéliser les alignements: avec des HMMs

## Hypothèses

- ▶  $J$  ne dépend que de  $I$
- ▶  $a_j$  ne dépend que de  $a_{j-1}$  et  $e_1^I$  (dépendance Markovienne)
- ▶  $f_j$  ne dépend que de  $e_{a_j}$  (le mot aligné avec  $f_j$ )

## Histoire générative de $f$ sachant $e$

- ▶ choisir  $J$  sachant  $I$
- ▶ pour chaque position  $j \in [1 : J]$ 
  - ▶ choisir  $a_j | I, a_{j-1}, e_1^I$ , "indice du mot qui engendre  $f_j$ "
  - ▶ choisir  $f_j | J, e_{a_j}$ : sachant le mot cible, choisir le mot source

$$P(a_1^J, f_1^J | e_1^I) = P(J | I) \prod_j P(a_j | a_{j-1}, e_1^I) P(f_j | e_{a_j})$$

# Modéliser les alignements: avec des HMMs

## Hypothèses

- ▶  $J$  ne dépend que de  $I$
- ▶  $a_j$  ne dépend que de  $a_{j-1}$  et  $e_1^I$  (dépendance Markovienne)
- ▶  $f_j$  ne dépend que de  $e_{a_j}$  (le mot aligné avec  $f_j$ )

## Histoire générative de $f$ sachant $e$

- ▶ choisir  $J$  sachant  $I$
- ▶ pour chaque position  $j \in [1 : J]$ 
  - ▶ choisir  $a_j | I, a_{j-1}, e_1^I$ , 'indice du mot qui engendre  $f_j$
  - ▶ choisir  $f_j | J, e_{a_j}$ : sachant le mot cible, choisir le mot source

$$P(a_1^J, f_1^J | e_1^I) = P(J | I) \prod_j P(a_j | a_{j-1}, e_1^I) P(f_j | e_{a_j})$$

# Engendrer des alignements avec un HMM

Target:  $e_1^9 =$

*this<sub>1</sub> seems<sub>2</sub> to<sub>3</sub> me<sub>4</sub> to<sub>5</sub> be<sub>6</sub> a<sub>7</sub> workable<sub>8</sub> solution<sub>9</sub>*

$J \sim P(J|9) \Rightarrow J = 9$

1								
2	$a_1$	$P(a_1 1)$	$a_2 = 2$	$t_1$	$P(t_1 seems)$	$t_2 = me$		
3								
4	$a_3$	$P(a_3 2)$	$a_4 = 3$	$t_3$	$P(t_3 to)$	$t_4 = que$		
5								
6	$a_5$	$P(a_5 5)$	$a_6 = 6$	$t_5$	$P(t_5 be)$	$t_6 = est$		
7								
8	$a_7$	$P(a_7 7)$	$a_8 = 8$	$t_7$	$P(t_7 workable)$	$t_8 = bonne$		
9								

Source:  $S =$

*the best way to find a good solution*

# Engendrer des alignements avec un HMM

Target:  $e_1^9 =$

*this<sub>1</sub> seems<sub>2</sub> to<sub>3</sub> me<sub>4</sub> to<sub>5</sub> be<sub>6</sub> a<sub>7</sub> workable<sub>8</sub> solution<sub>9</sub>*

$$J \sim P(J|9) \Rightarrow J = 9$$

1	$a_1 \sim P(a_1) : a_1 = 1$	$f_1 \sim P(f this) : f_1 = il$
2	$a_2 \sim P(a_2 1) : a_2 = 2$	$f_2 \sim P(f seems) : f_2 = me$
3	$a_3 \sim P(a_3 2) : a_3 = 2$	$f_3 \sim P(f seems) : f_3 = semble$
4	$a_4 \sim P(a_4 2) : a_4 = 3$	$f_4 \sim P(f to) : f_4 = que$
5	$a_5 \sim P(a_5 3) : a_5 = 5$	$f_5 \sim P(f to) : f_5 = c'$
6	$a_6 \sim P(a_6 5) : a_6 = 6$	$f_6 \sim P(f be) : f_6 = est$
7	$a_7 \sim P(a_7 6) : a_7 = 7$	$f_7 \sim P(f a) : f_7 = une$
8	$a_8 \sim P(a_8 7) : a_8 = 8$	$f_8 \sim P(f workable) : f_8 = bonne$
9	$a_9 \sim P(a_9 8) : a_9 = 9$	$f_9 \sim P(f solution) : f_9 = solution$

# Engendrer des alignements avec un HMM

Target:  $e_1^9 =$

*this<sub>1</sub> seems<sub>2</sub> to<sub>3</sub> me<sub>4</sub> to<sub>5</sub> be<sub>6</sub> a<sub>7</sub> workable<sub>8</sub> solution<sub>9</sub>*

$$J \sim P(J|9) \Rightarrow J = 9$$

1	$a_1 \sim P(a_1) : a_1 = 1$	$f_1 \sim P(f this) : f_1 = il$
2	$a_2 \sim P(a_2 1) : a_2 = 2$	$f_2 \sim P(f seems) : f_2 = me$
3	$a_3 \sim P(a_3 2) : a_3 = 2$	$f_3 \sim P(f seems) : f_3 = semble$
4	$a_4 \sim P(a_4 2) : a_4 = 3$	$f_4 \sim P(f to) : f_4 = que$
5	$a_5 \sim P(a_5 3) : a_5 = 5$	$f_5 \sim P(f to) : f_5 = c'$
6	$a_6 \sim P(a_6 5) : a_6 = 6$	$f_6 \sim P(f be) : f_6 = est$
7	$a_7 \sim P(a_7 6) : a_7 = 7$	$f_7 \sim P(f a) : f_7 = une$
8	$a_8 \sim P(a_8 7) : a_8 = 8$	$f_8 \sim P(f workable) : f_8 = bonne$
9	$a_9 \sim P(a_9 8) : a_9 = 9$	$f_9 \sim P(f solution) : f_9 = solution$

Source:  $f_1^9 =$

*il<sub>1</sub> me<sub>2</sub> semble<sub>3</sub> que<sub>4</sub> c'<sub>5</sub> est<sub>6</sub> une<sub>7</sub> bonne<sub>8</sub> solution<sub>9</sub>*



# Engendrer des alignements avec un HMM

Target:  $e_1^9 =$

*this<sub>1</sub> seems<sub>2</sub> to<sub>3</sub> me<sub>4</sub> to<sub>5</sub> be<sub>6</sub> a<sub>7</sub> workable<sub>8</sub> solution<sub>9</sub>*

$$J \sim P(J|9) \Rightarrow J = 9$$

1	$a_1 \sim P(a_1) : a_1 = 1$	$f_1 \sim P(f this) : f_1 = il$
2	$a_2 \sim P(a_2 1) : a_2 = 2$	$f_2 \sim P(f seems) : f_2 = me$
3	$a_3 \sim P(a_3 2) : a_3 = 2$	$f_3 \sim P(f seems) : f_3 = semble$
4	$a_4 \sim P(a_4 2) : a_4 = 3$	$f_4 \sim P(f to) : f_4 = que$
5	$a_5 \sim P(a_5 3) : a_5 = 5$	$f_5 \sim P(f to) : f_5 = c'$
6	$a_6 \sim P(a_6 5) : a_6 = 6$	$f_6 \sim P(f be) : f_6 = est$
7	$a_7 \sim P(a_7 6) : a_7 = 7$	$f_7 \sim P(f a) : f_7 = une$
8	$a_8 \sim P(a_8 7) : a_8 = 8$	$f_8 \sim P(f workable) : f_8 = bonne$
9	$a_9 \sim P(a_9 8) : a_9 = 9$	$f_9 \sim P(f solution) : f_9 = solution$

Source:  $f_1^9 =$

*il<sub>1</sub> me<sub>2</sub> semble<sub>3</sub> que<sub>4</sub> c'<sub>5</sub> est<sub>6</sub> une<sub>7</sub> bonne<sub>8</sub> solution<sub>9</sub>*

# Engendrer des alignements avec un HMM

Target:  $e_1^9 =$

*this<sub>1</sub> seems<sub>2</sub> to<sub>3</sub> me<sub>4</sub> to<sub>5</sub> be<sub>6</sub> a<sub>7</sub> workable<sub>8</sub> solution<sub>9</sub>*

$$J \sim P(J|9) \Rightarrow J = 9$$

1	$a_1 \sim P(a_1) : a_1 = 1$	$f_1 \sim P(f this) : f_1 = il$
2	$a_2 \sim P(a_2 1) : a_2 = 2$	$f_2 \sim P(f seems) : f_2 = me$
3	$a_3 \sim P(a_3 2) : a_3 = 2$	$f_3 \sim P(f seems) : f_3 = semble$
4	$a_4 \sim P(a_4 2) : a_4 = 3$	$f_4 \sim P(f to) : f_4 = que$
5	$a_5 \sim P(a_5 3) : a_5 = 5$	$f_5 \sim P(f to) : f_5 = c'$
6	$a_6 \sim P(a_6 5) : a_6 = 6$	$f_6 \sim P(f be) : f_6 = est$
7	$a_7 \sim P(a_7 6) : a_7 = 7$	$f_7 \sim P(f a) : f_7 = une$
8	$a_8 \sim P(a_8 7) : a_8 = 8$	$f_8 \sim P(f workable) : f_8 = bonne$
9	$a_9 \sim P(a_9 8) : a_9 = 9$	$f_9 \sim P(f solution) : f_9 = solution$

Source:  $f_1^9 =$

*il<sub>1</sub> me<sub>2</sub> semble<sub>3</sub> que<sub>4</sub> c'<sub>5</sub> est<sub>6</sub> une<sub>7</sub> bonne<sub>8</sub> solution<sub>9</sub>*

# Engendrer des alignements avec un HMM

Target:  $e_1^9 =$

*this<sub>1</sub> seems<sub>2</sub> to<sub>3</sub> me<sub>4</sub> to<sub>5</sub> be<sub>6</sub> a<sub>7</sub> workable<sub>8</sub> solution<sub>9</sub>*

$$J \sim P(J|9) \Rightarrow J = 9$$

1	$a_1 \sim P(a_1) : a_1 = 1$	$f_1 \sim P(f this) : f_1 = il$
2	$a_2 \sim P(a_2 1) : a_2 = 2$	$f_2 \sim P(f seems) : f_2 = me$
3	$a_3 \sim P(a_3 2) : a_3 = 2$	$f_3 \sim P(f seems) : f_3 = semble$
4	$a_4 \sim P(a_4 2) : a_4 = 3$	$f_4 \sim P(f to) : f_4 = que$
5	$a_5 \sim P(a_5 3) : a_5 = 5$	$f_5 \sim P(f to) : f_5 = c'$
6	$a_6 \sim P(a_6 5) : a_6 = 6$	$f_6 \sim P(f be) : f_6 = est$
7	$a_7 \sim P(a_7 6) : a_7 = 7$	$f_7 \sim P(f a) : f_7 = une$
8	$a_8 \sim P(a_8 7) : a_8 = 8$	$f_8 \sim P(f workable) : f_8 = bonne$
9	$a_9 \sim P(a_9 8) : a_9 = 9$	$f_9 \sim P(f solution) : f_9 = solution$

Source:  $f_1^9 =$

*il<sub>1</sub> me<sub>2</sub> semble<sub>3</sub> que<sub>4</sub> c'<sub>5</sub> est<sub>6</sub> une<sub>7</sub> bonne<sub>8</sub> solution<sub>9</sub>*

# Engendrer des alignements avec un HMM

Target:  $e_1^9 =$

*this<sub>1</sub> seems<sub>2</sub> to<sub>3</sub> me<sub>4</sub> to<sub>5</sub> be<sub>6</sub> a<sub>7</sub> workable<sub>8</sub> solution<sub>9</sub>*

$$J \sim P(J|9) \Rightarrow J = 9$$

1	$a_1 \sim P(a_1) : a_1 = 1$	$f_1 \sim P(f this) : f_1 = il$
2	$a_2 \sim P(a_2 1) : a_2 = 2$	$f_2 \sim P(f seems) : f_2 = me$
3	$a_3 \sim P(a_3 2) : a_3 = 2$	$f_3 \sim P(f seems) : f_3 = semble$
4	$a_4 \sim P(a_4 2) : a_4 = 3$	$f_4 \sim P(f to) : f_4 = que$
5	$a_5 \sim P(a_5 3) : a_5 = 5$	$f_5 \sim P(f to) : f_5 = c'$
6	$a_6 \sim P(a_6 5) : a_6 = 6$	$f_6 \sim P(f be) : f_6 = est$
7	$a_7 \sim P(a_7 6) : a_7 = 7$	$f_7 \sim P(f a) : f_7 = une$
8	$a_8 \sim P(a_8 7) : a_8 = 8$	$f_8 \sim P(f workable) : f_8 = bonne$
9	$a_9 \sim P(a_9 8) : a_9 = 9$	$f_9 \sim P(f solution) : f_9 = solution$

Source:  $f_1^9 =$

*il<sub>1</sub> me<sub>2</sub> semble<sub>3</sub> que<sub>4</sub> c'<sub>5</sub> est<sub>6</sub> une<sub>7</sub> bonne<sub>8</sub> solution<sub>9</sub>*

# Engendrer des alignements avec un HMM

Target:  $e_1^9 =$

*this<sub>1</sub> seems<sub>2</sub> to<sub>3</sub> me<sub>4</sub> to<sub>5</sub> be<sub>6</sub> a<sub>7</sub> workable<sub>8</sub> solution<sub>9</sub>*

$$J \sim P(J|9) \Rightarrow J = 9$$

1	$a_1 \sim P(a_1) : a_1 = 1$	$f_1 \sim P(f this) : f_1 = il$
2	$a_2 \sim P(a_2 1) : a_2 = 2$	$f_2 \sim P(f seems) : f_2 = me$
3	$a_3 \sim P(a_3 2) : a_3 = 2$	$f_3 \sim P(f seems) : f_3 = semble$
4	$a_4 \sim P(a_4 2) : a_4 = 3$	$f_4 \sim P(f to) : f_4 = que$
5	$a_5 \sim P(a_5 3) : a_5 = 5$	$f_5 \sim P(f to) : f_5 = c'$
6	$a_6 \sim P(a_6 5) : a_6 = 6$	$f_6 \sim P(f be) : f_6 = est$
7	$a_7 \sim P(a_7 6) : a_7 = 7$	$f_7 \sim P(f a) : f_7 = une$
8	$a_8 \sim P(a_8 7) : a_8 = 8$	$f_8 \sim P(f workable) : f_8 = bonne$
9	$a_9 \sim P(a_9 8) : a_9 = 9$	$f_9 \sim P(f solution) : f_9 = solution$

Source:  $f_1^9 =$

*il<sub>1</sub> me<sub>2</sub> semble<sub>3</sub> que<sub>4</sub> c'<sub>5</sub> est<sub>6</sub> une<sub>7</sub> bonne<sub>8</sub> solution<sub>9</sub>*

# Engendrer des alignements avec un HMM

Target:  $e_1^9 =$

*this<sub>1</sub> seems<sub>2</sub> to<sub>3</sub> me<sub>4</sub> to<sub>5</sub> be<sub>6</sub> a<sub>7</sub> workable<sub>8</sub> solution<sub>9</sub>*

$$J \sim P(J|9) \Rightarrow J = 9$$

1	$a_1 \sim P(a_1) : a_1 = 1$	$f_1 \sim P(f this) : f_1 = il$
2	$a_2 \sim P(a_2 1) : a_2 = 2$	$f_2 \sim P(f seems) : f_2 = me$
3	$a_3 \sim P(a_3 2) : a_3 = 2$	$f_3 \sim P(f seems) : f_3 = semble$
4	$a_4 \sim P(a_4 2) : a_4 = 3$	$f_4 \sim P(f to) : f_4 = que$
5	$a_5 \sim P(a_5 3) : a_5 = 5$	$f_5 \sim P(f to) : f_5 = c'$
6	$a_6 \sim P(a_6 5) : a_6 = 6$	$f_6 \sim P(f be) : f_6 = est$
7	$a_7 \sim P(a_7 6) : a_7 = 7$	$f_7 \sim P(f a) : f_7 = une$
8	$a_8 \sim P(a_8 7) : a_8 = 8$	$f_8 \sim P(f workable) : f_8 = bonne$
9	$a_9 \sim P(a_9 8) : a_9 = 9$	$f_9 \sim P(f solution) : f_9 = solution$

Source:  $f_1^9 =$

*il<sub>1</sub> me<sub>2</sub> semble<sub>3</sub> que<sub>4</sub> c'<sub>5</sub> est<sub>6</sub> une<sub>7</sub> bonne<sub>8</sub> solution<sub>9</sub>*

# Engendrer des alignements avec un HMM

Target:  $e_1^9 =$

*this<sub>1</sub> seems<sub>2</sub> to<sub>3</sub> me<sub>4</sub> to<sub>5</sub> be<sub>6</sub> a<sub>7</sub> workable<sub>8</sub> solution<sub>9</sub>*

$$J \sim P(J|9) \Rightarrow J = 9$$

1	$a_1 \sim P(a_1) : a_1 = 1$	$f_1 \sim P(f this) : f_1 = il$
2	$a_2 \sim P(a_2 1) : a_2 = 2$	$f_2 \sim P(f seems) : f_2 = me$
3	$a_3 \sim P(a_3 2) : a_3 = 2$	$f_3 \sim P(f seems) : f_3 = semble$
4	$a_4 \sim P(a_4 2) : a_4 = 3$	$f_4 \sim P(f to) : f_4 = que$
5	$a_5 \sim P(a_5 3) : a_5 = 5$	$f_5 \sim P(f to) : f_5 = c'$
6	$a_6 \sim P(a_6 5) : a_6 = 6$	$f_6 \sim P(f be) : f_6 = est$
7	$a_7 \sim P(a_7 6) : a_7 = 7$	$f_7 \sim P(f a) : f_7 = une$
8	$a_8 \sim P(a_8 7) : a_8 = 8$	$f_8 \sim P(f workable) : f_8 = bonne$
9	$a_9 \sim P(a_9 8) : a_9 = 9$	$f_9 \sim P(f solution) : f_9 = solution$

Source:  $f_1^9 =$

*il<sub>1</sub> me<sub>2</sub> semble<sub>3</sub> que<sub>4</sub> c'<sub>5</sub> est<sub>6</sub> une<sub>7</sub> bonne<sub>8</sub> solution<sub>9</sub>*

# Engendrer des alignements avec un HMM

Target:  $e_1^9 =$

*this<sub>1</sub> seems<sub>2</sub> to<sub>3</sub> me<sub>4</sub> to<sub>5</sub> be<sub>6</sub> a<sub>7</sub> workable<sub>8</sub> solution<sub>9</sub>*

$$J \sim P(J|9) \Rightarrow J = 9$$

1	$a_1 \sim P(a_1) : a_1 = 1$	$f_1 \sim P(f this) : f_1 = il$
2	$a_2 \sim P(a_2 1) : a_2 = 2$	$f_2 \sim P(f seems) : f_2 = me$
3	$a_3 \sim P(a_3 2) : a_3 = 2$	$f_3 \sim P(f seems) : f_3 = semble$
4	$a_4 \sim P(a_4 2) : a_4 = 3$	$f_4 \sim P(f to) : f_4 = que$
5	$a_5 \sim P(a_5 3) : a_5 = 5$	$f_5 \sim P(f to) : f_5 = c'$
6	$a_6 \sim P(a_6 5) : a_6 = 6$	$f_6 \sim P(f be) : f_6 = est$
7	$a_7 \sim P(a_7 6) : a_7 = 7$	$f_7 \sim P(f a) : f_7 = une$
8	$a_8 \sim P(a_8 7) : a_8 = 8$	$f_8 \sim P(f workable) : f_8 = bonne$
9	$a_9 \sim P(a_9 8) : a_9 = 9$	$f_9 \sim P(f solution) : f_9 = solution$

Source:  $f_1^9 =$

*il<sub>1</sub> me<sub>2</sub> semble<sub>3</sub> que<sub>4</sub> c'<sub>5</sub> est<sub>6</sub> une<sub>7</sub> bonne<sub>8</sub> solution<sub>9</sub>*



# Engendrer des alignements avec un HMM

Target:  $e_1^9 =$

*this<sub>1</sub> seems<sub>2</sub> to<sub>3</sub> me<sub>4</sub> to<sub>5</sub> be<sub>6</sub> a<sub>7</sub> workable<sub>8</sub> solution<sub>9</sub>*

$$J \sim P(J|9) \Rightarrow J = 9$$

1	$a_1 \sim P(a_1) : a_1 = 1$	$f_1 \sim P(f this) : f_1 = il$
2	$a_2 \sim P(a_2 1) : a_2 = 2$	$f_2 \sim P(f seems) : f_2 = me$
3	$a_3 \sim P(a_3 2) : a_3 = 2$	$f_3 \sim P(f seems) : f_3 = semble$
4	$a_4 \sim P(a_4 2) : a_4 = 3$	$f_4 \sim P(f to) : f_4 = que$
5	$a_5 \sim P(a_5 3) : a_5 = 5$	$f_5 \sim P(f to) : f_5 = c'$
6	$a_6 \sim P(a_6 5) : a_6 = 6$	$f_6 \sim P(f be) : f_6 = est$
7	$a_7 \sim P(a_7 6) : a_7 = 7$	$f_7 \sim P(f a) : f_7 = une$
8	$a_8 \sim P(a_8 7) : a_8 = 8$	$f_8 \sim P(f workable) : f_8 = bonne$
9	$a_9 \sim P(a_9 8) : a_9 = 9$	$f_9 \sim P(f solution) : f_9 = solution$

Source:  $f_1^9 =$

*il<sub>1</sub> me<sub>2</sub> semble<sub>3</sub> que<sub>4</sub> c'<sub>5</sub> est<sub>6</sub> une<sub>7</sub> bonne<sub>8</sub> solution<sub>9</sub>*

# Deux finesses

## Les mots “vides”

Traiter des mots source non alignables: *ai* et *l'* dans:  
*j' ai eu l' occasion / I had occasion*

- ▶ état fictif dans la cible (d'indice 0) atteint avec  
 $P_0 = P(a_i = 0 | a_{i-1}, J)$
- ▶ une distribution associée à cet état  $P = P(f | \epsilon)$

## Modéliser les sauts

Rendre le modèle d'alignement indépendant des indices absolus:  $\Rightarrow$  remplacer  $P(a_i | a_{i-1})$  par  $P(a_i - a_{i-1} | a_{i-1} - a_{i-2})$

# Deux finesses

## Les mots “vides”

Traiter des mots source non alignables: *ai* et *l'* dans:

*j' ai eu l' occasion / I had occasion*

- ▶ état fictif dans la cible (d'indice 0) atteint avec  
 $P_0 = P(a_i = 0 | a_{i-1}, J)$
- ▶ une distribution associée à cet état  $P = P(f | \epsilon)$

## Modéliser les sauts

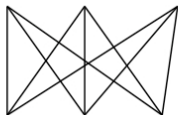
Rendre le modèle d'alignement indépendant des indices

absolus:  $\Rightarrow$  remplacer  $P(a_i | a_{i-1})$  par  $P(a_i - a_{i-1} | a_{i-1} - a_{i-2})$

# Émergence des alignements

La puissance de EM

... la maison ... la maison blue ... la fleur ...

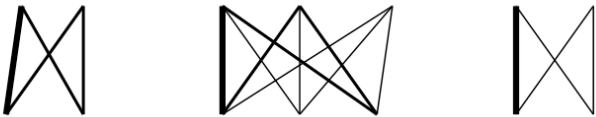


... the house ... the blue house ... the flower ...

Tous les alignements sont également probables

# Émergence des alignements

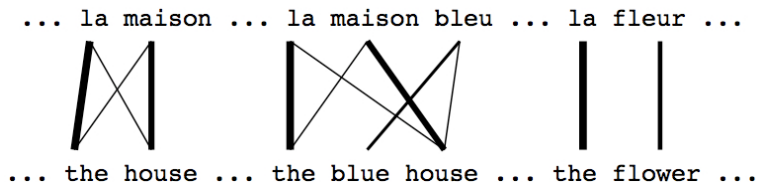
La puissance de EM

... la maison ... la maison blue ... la fleur ...  
  
... the house ... the blue house ... the flower ...

Tous les alignements sont également probables *la/the*,  
*maison/house* émergent...

# Émergence des alignements

La puissance de EM



Tous les alignements sont également probables *la/the*,  
*maison/house* se renforcent

# Émergence des alignements

La puissance de EM

...	la	maison	...	la	maison	bleu	...	la	fleur	...
	/				X					
...	the	house	...	the	blue	house	...	the	flower	...

Tous les alignements sont également probables *bleue/blue*,  
*fleur/flower* s'imposent (principe du *pigeonhole*)

# Estimation supervisée du modèle

- ▶ à alignements connus...
- ▶ ... les paramètres se déduisent par décompte:

$$\forall I \in [1 \dots I_{max}], J \in [1 \dots J_{max}], P(J|I) = \frac{n(I, J)}{n(I)}$$

$$\forall i, i' \in [1 \dots I_{max}], P(i'|i, J, I) = \frac{n(i, i')}{n(i')}$$

$$\forall e \in V_e, f \in V_f, P(f|e) = \frac{n(e, f)}{n(e)}$$



# Calculer les alignements (à modèle connu)

- ▶  $P(.|I)$  connu;  $P(.|a, I, J)$  connu ;  $P(f|e)$  connu
- ▶  $e_1^I$  et  $f_1^J$  sont observés
- ▶ trouver:

$$\begin{aligned} a^* &= \operatorname{argmax}_{a_1 \dots a_J} P(f_1^J, a_1^J | e_j) \\ &= \operatorname{argmax}_{a_1 \dots a_J} P(J|I) \prod_j P(a_j | a_{j-1}) P(f_j | e_{a_j}) \end{aligned}$$

- ▶ Résolution par programmation dynamique (Viterbi)

$$\begin{cases} \delta(i, 1) = P(a_1 = i), \forall i \in [1 \dots I] \\ \delta(i, j) = \max_{i' \in I} \delta(i', j-1) P(a_j = i | a_{j-1} = i') P(f_j | e_i) \forall i, j > 1 \end{cases}$$

# Estimation par EM

## Étape E(xpectation)

à paramètres connus (étape précédente):

$$P(a_1^J | e_1^J, f_1^J) = \frac{P(a_1^J, f_1^J | e_1^J)}{\sum_a P(a_1^J, f_1^J | e_1^J)}$$

Le dénominateur se calcule par l'algorithme *forward*.

## Étape M(aximisation)

$$\forall I \in [1 \dots I_{max}], J \in [1 \dots J_{max}], P(J|I) = \frac{n(I, J)}{n(I)}$$

$$\forall i, i' \in [1 \dots I], P(i'|i, J, I) = \frac{\sum_k P(a^{(k)} | e^{(k)}, f^{(k)}) n^{(k)}(i, i')}{\sum_i \sum_k P(a^{(k)} | e^{(k)}, f^{(k)}) n^{(k)}(i, i')}$$

$$\forall e, f, P(f|e) = \frac{\sum_{(k)} P(a^{(k)} | e^{(k)}, f^{(k)}) n^{(k)}(e, f)}{\sum_f \sum_{(k)} P(a^{(k)} | e^{(k)}, f^{(k)}) n^{(k)}(e, f)}$$

# Estimation par EM

## Étape E(xpectation)

à paramètres connus (étape précédente):

$$P(a_1^J | e_1^J, f_1^J) = \frac{P(a_1^J, f_1^J | e_1^J)}{\sum_a P(a_1^J, f_1^J | e_1^J)}$$

Le dénominateur se calcule par l'algorithme *forward*.

## Étape M(aximisation)

$$\forall I \in [1 \dots I_{max}], J \in [1 \dots J_{max}], P(J|I) = \frac{n(I, J)}{n(I)}$$

$$\forall i, i' \in [1 \dots I], P(i'|i, J, I) = \frac{\sum_k P(a^{(k)} | e^{(k)}, f^{(k)}) n^{(k)}(i, i')}{\sum_i \sum_k P(a^{(k)} | e^{(k)}, f^{(k)}) n^{(k)}(i, i')}$$

$$\forall e, f, P(f|e) = \frac{\sum_{(k)} P(a^{(k)} | e^{(k)}, f^{(k)}) n^{(k)}(e, f)}{\sum_f \sum_{(k)} P(a^{(k)} | e^{(k)}, f^{(k)}) n^{(k)}(e, f)}$$

# Initialiser avec des modèles simples: IBM1 et IBM2

Des modèles faciles à entraîner

## IBM1

Les probabilités des  $a_j$  sont uniformes:  $P(a_j|a_{j-1}, l, J) = \frac{1}{I+1}$

$$P(a_1^J, f_1^J | e_1^I) = \frac{P(J|I)}{(I+1)^J} \prod_j P(f_j | e_{a_j})$$

## IBM2

Les  $a_j$  ne dépendent que de  $j$ :  $P(a_j|a_{j-1}, l, J) = P(a_j|j, J)$

$$P(a_1^J, f_1^J | e_1^I) = P(J|I) \prod_j P(a_j|j, J) P(f_j | e_{a_j})$$

# Initialiser avec des modèles simples: IBM1 et IBM2

Des modèles faciles à entraîner

## IBM1

Les probabilités des  $a_j$  sont uniformes:  $P(a_j|a_{j-1}, l, J) = \frac{1}{I+1}$

$$P(a_1^J, f_1^J | e_1^I) = \frac{P(J|I)}{(I+1)^J} \prod_j P(f_j | e_{a_j})$$

## IBM2

Les  $a_j$  ne dépendent que de  $j$ :  $P(a_j|a_{j-1}, l, J) = P(a_j|j, J)$

$$P(a_1^J, f_1^J | e_1^I) = P(J|I) \prod_j P(a_j|j, J) P(f_j | e_{a_j})$$

# Au delà des HMMs: IBM3-IBM5

## Un décodeur... pas très bon

Après estimation du HMM, approximer:

$$P(\mathbf{f}|\mathbf{e}) = \sum_{\mathbf{a}} P(\mathbf{f}, \mathbf{a}|\mathbf{e}) \approx \max_{\mathbf{a}} P(\mathbf{f}, \mathbf{a}|\mathbf{e})$$

Traduire = résoudre  $\operatorname{argmax}_{\mathbf{e}} P(\mathbf{e}) \max_{\mathbf{a}} P(\mathbf{f}|\mathbf{e})$

Nouvelle Histoire générative de  $\mathbf{f}$  sachant  $\mathbf{e}$

- Pour chaque  $e_j$ , tirer le nombre de mots générés  $n_j$  (fertilité)
- Pour chaque  $e_j$ , tirer  $n_j$  mots cibles  $f_{j,1} \dots f_{j,n_j}$
- Pour chaque  $f_{j,i}$ , tirer les probabilités dans la table

# Au delà des HMMs: IBM3-IBM5

## Un décodeur... pas très bon

Après estimation du HMM, approximer:

$$P(\mathbf{f}|\mathbf{e}) = \sum_{\mathbf{a}} P(\mathbf{f}, \mathbf{a}|\mathbf{e}) \approx \max_{\mathbf{a}} P(\mathbf{f}, \mathbf{a}|\mathbf{e})$$

Traduire = résoudre  $\operatorname{argmax}_{\mathbf{e}} P(\mathbf{e}) \max_{\mathbf{a}} P(\mathbf{f}|\mathbf{e})$

## Nouvelle Histoire générative de $\mathbf{f}$ sachant $\mathbf{e}$

- ▶ Pour chaque  $e_j$ , tirer le nombre de mots générés  $n_j$  (fertilité)
- ▶ Pour chaque  $e_j$ , tirer  $n_j$  mots cibles  $f_{j,1} \dots f_{j,n_j}$
- ▶ Pour chaque  $f_k$ , tirer sa position dans la phrase

# Au delà des HMMs: IBM3-IBM5

## Un décodeur... pas très bon

Après estimation du HMM, approximer:

$$P(\mathbf{f}|\mathbf{e}) = \sum_{\mathbf{a}} P(\mathbf{f}, \mathbf{a}|\mathbf{e}) \approx \max_{\mathbf{a}} P(\mathbf{f}, \mathbf{a}|\mathbf{e})$$

Traduire = résoudre  $\operatorname{argmax}_{\mathbf{e}} P(\mathbf{e}) \max_{\mathbf{a}} P(\mathbf{f}|\mathbf{e})$

## Nouvelle Histoire générative de $\mathbf{f}$ sachant $\mathbf{e}$

- ▶ Pour chaque  $e_j$ , tirer le nombre de mots générés  $n_j$  (fertilité)
- ▶ Pour chaque  $e_j$ , tirer  $n_j$  mots cibles  $f_{j,1} \dots f_{j,n_j}$
- ▶ Pour chaque  $f_k$ , tirer sa position dans la phrase



# Au delà des HMMs: IBM3-IBM5

## Un décodeur... pas très bon

Après estimation du HMM, approximer:

$$P(\mathbf{f}|\mathbf{e}) = \sum_{\mathbf{a}} P(\mathbf{f}, \mathbf{a}|\mathbf{e}) \approx \max_{\mathbf{a}} P(\mathbf{f}, \mathbf{a}|\mathbf{e})$$

Traduire = résoudre  $\operatorname{argmax}_{\mathbf{e}} P(\mathbf{e}) \max_{\mathbf{a}} P(\mathbf{f}|\mathbf{e})$

## Nouvelle Histoire générative de $\mathbf{f}$ sachant $\mathbf{e}$

- ▶ Pour chaque  $e_j$ , tirer le nombre de mots générés  $n_j$  (fertilité)
- ▶ Pour chaque  $e_j$ , tirer  $n_j$  mots cibles  $f_{j,1} \dots f_{j,n_j}$
- ▶ Pour chaque  $f_k$ , tirer sa position dans la phrase

# Les ingrédients des modèles IBM

- ▶ Un modèle de fertilité:  $n(\phi|e)$  (nombre de mots produits)
- ▶ Un modèle lexical:  $t(f|e)$  (quels mots sont produits)
- ▶ Un modèle de distorsion:  $d(j|i, \dots)$  (où ces mots sont placés)
- ▶ Un paramètre  $p_0$  de *production spontanée de mots*
- ▶ Un modèle de langage  $P(\mathbf{e})$  (de la langue cible)

# D'IBM-3 à IBM5

- IBM-3 Les probabilités des alignements dépendent des positions dans  $f$  et  $e$  et des longueurs des deux phrases (comme pour IBM-2)
- IBM-4 Les probabilités des alignements dépendent en plus des mots français et anglais alignés et des autres mots qui sont alignés au même mot anglais
- IBM-5 Comme IBM-4, mais corrige des problèmes d'alignement (plusieurs mots assignés à la même position)

L'inférence exacte n'est plus possible  $\Rightarrow$  approches heuristiques, apprentissage très long.

# D'IBM-3 à IBM5

- IBM-3 Les probabilités des alignements dépendent des positions dans  $f$  et  $e$  et des longueurs des deux phrases (comme pour IBM-2)
- IBM-4 Les probabilités des alignements dépendent en plus des mots français et anglais alignés et des autres mots qui sont alignés au même mot anglais
- IBM-5 Comme IBM-4, mais corrige des problèmes d'alignement (plusieurs mots assignés à la même position)

L'inférence exacte n'est plus possible  $\Rightarrow$  approches heuristiques, apprentissage très long.

# D'IBM-3 à IBM5

- IBM-3 Les probabilités des alignements dépendent des positions dans  $f$  et  $e$  et des longueurs des deux phrases (comme pour IBM-2)
- IBM-4 Les probabilités des alignements dépendent en plus des mots français et anglais alignés et des autres mots qui sont alignés au même mot anglais
- IBM-5 Comme IBM-4, mais corrige des problèmes d'alignement (plusieurs mots assignés à la même position)

L'inférence exacte n'est plus possible  $\Rightarrow$  approches heuristiques, apprentissage très long.

# D'IBM-3 à IBM5

- IBM-3 Les probabilités des alignements dépendent des positions dans  $f$  et  $e$  et des longueurs des deux phrases (comme pour IBM-2)
- IBM-4 Les probabilités des alignements dépendent en plus des mots français et anglais alignés et des autres mots qui sont alignés au même mot anglais
- IBM-5 Comme IBM-4, mais corrige des problèmes d'alignement (plusieurs mots assignés à la même position)

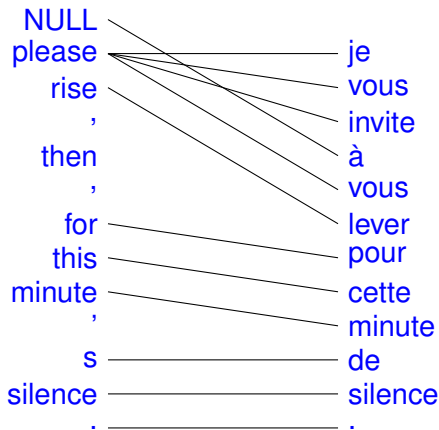
L'inférence exacte n'est plus possible  $\Rightarrow$  approches heuristiques, apprentissage très long.

# D'IBM-3 à IBM5

- IBM-3 Les probabilités des alignements dépendent des positions dans  $f$  et  $e$  et des longueurs des deux phrases (comme pour IBM-2)
- IBM-4 Les probabilités des alignements dépendent en plus des mots français et anglais alignés et des autres mots qui sont alignés au même mot anglais
- IBM-5 Comme IBM-4, mais corrige des problèmes d'alignement (plusieurs mots assignés à la même position)

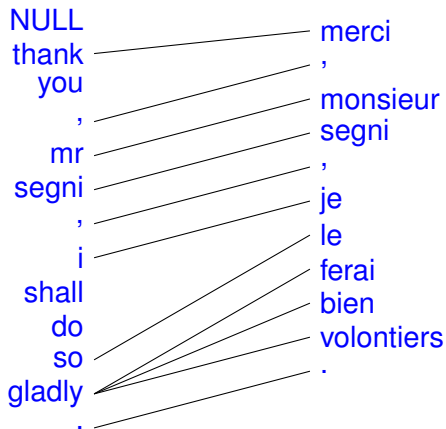
L'inférence exacte n'est plus possible  $\Rightarrow$  approches heuristiques, apprentissage très long.

# Des alignements... plus ou moins heureux

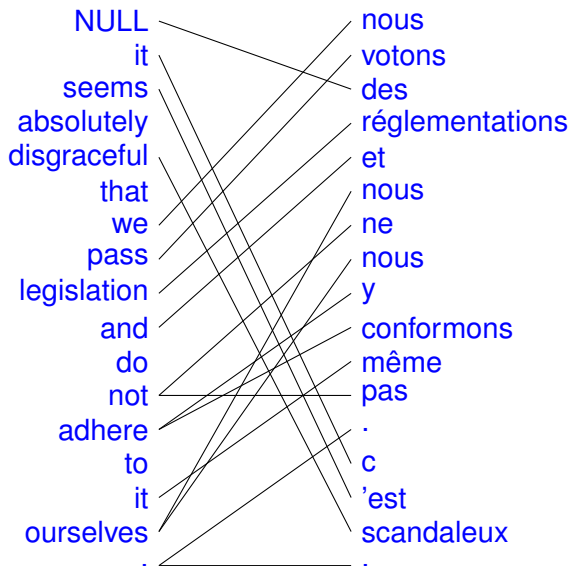




# Des alignements... plus ou moins heureux



# Des alignements... plus ou moins heureux



## Pour en savoir plus...

- ▶ The mathematics of statistical machine translation (Brown & al, 1993): publication de référence sur la traduction mot-à-mot et les modèles d'alignement
- ▶ A Statistical MT tutorial workbook (Knight, 1999): le même, en pédagogique
- ▶ Giza, Giza++, Giza-pp, MGiza: logiciels *open-source* pour la construction d'alignements

# Outline

Traduire

Anatomie d'un système de traduction statistique

Traduction statistique mot-à-mot et alignements

Les alignements mot-à-mot

**Modèles de segments**

Décodage et recherche

Traduire la parole

Perspectives

# Vers les modèles de segments

- ▶ Les alignements mot-à-mot sont problématiques
  - ▶ Le modèle lexical  $t(f|e)$  n'utilise pas de contexte: choix de traduction imprécis
  - ▶ Un modèle lexical  $t(f_j|e_{i-2}e_{i-1}e_i)$  est trop complexe à estimer
- ⇒ nouveau modèle de traduction, alignement de “blocs de mots” (**segments**).
- ▶ Apprentissage du modèle:
    - 1. la population des segments
    - 2. les modèles probabilistes de segments

# Vers les modèles de segments

- ▶ Les alignements mot-à-mot sont problématiques
  - ▶ Le modèle lexical  $t(f|e)$  n'utilise pas de contexte: choix de traduction imprécis
  - ▶ Un modèle lexical  $t(f_j|e_{i-2}e_{i-1}e_i)$  est trop complexe à estimer
- ⇒ nouveau modèle de traduction, alignement de “blocs de mots” (**segments**).
- ▶ Apprentissage du modèle:
    - 1. la segmentation des segments
    - 2. le modèle probabiliste des segments

# Vers les modèles de segments

- ▶ Les alignements mot-à-mot sont problématiques
  - ▶ Le modèle lexical  $t(f|e)$  n'utilise pas de contexte: choix de traduction imprécis
  - ▶ Un modèle lexical  $t(f_j|e_{i-2}e_{i-1}e_i)$  est trop complexe à estimer
- ⇒ nouveau modèle de traduction, alignement de “blocs de mots” (segments).
- ▶ Apprentissage du modèle:
- Exemple de phrase source et cible:
- Source: *Il y a un grand bâtiment à l'arrière de la maison.*
- Cible: *There is a big building behind the house.*

# Vers les modèles de segments

- ▶ Les alignements mot-à-mot sont problématiques
  - ▶ Le modèle lexical  $t(f|e)$  n'utilise pas de contexte: choix de traduction imprécis
  - ▶ Un modèle lexical  $t(f_j|e_{i-2}e_{i-1}e_i)$  est trop complexe à estimer
- ⇒ nouveau modèle de traduction, alignement de “blocs de mots” (**segments**).
- ▶ Apprentissage du modèle:
    - » acquisition des segments
    - » alignement des segments



# Vers les modèles de segments

- ▶ Les alignements mot-à-mot sont problématiques
  - ▶ Le modèle lexical  $t(f|e)$  n'utilise pas de contexte: choix de traduction imprécis
  - ▶ Un modèle lexical  $t(f_j|e_{i-2}e_{i-1}e_i)$  est trop complexe à estimer
- ⇒ nouveau modèle de traduction, alignement de “blocs de mots” (**segments**).
- ▶ Apprentissage du modèle:
    - ▶ acquisition des segments
    - ▶ modèles probabilistes de segments

# Vers les modèles de segments

- ▶ Les alignements mot-à-mot sont problématiques
  - ▶ Le modèle lexical  $t(f|e)$  n'utilise pas de contexte: choix de traduction imprécis
  - ▶ Un modèle lexical  $t(f_j|e_{i-2}e_{i-1}e_i)$  est trop complexe à estimer
- ⇒ nouveau modèle de traduction, alignement de “blocs de mots” (**segments**).
- ▶ Apprentissage du modèle:
    - ▶ acquisition des segments
    - ▶ modèles probabilistes de segments

# Vers les modèles de segments

- ▶ Les alignements mot-à-mot sont problématiques
  - ▶ Le modèle lexical  $t(f|e)$  n'utilise pas de contexte: choix de traduction imprécis
  - ▶ Un modèle lexical  $t(f_j|e_{i-2}e_{i-1}e_i)$  est trop complexe à estimer
- ⇒ nouveau modèle de traduction, alignement de “blocs de mots” (**segments**).
- ▶ Apprentissage du modèle:
    - ▶ acquisition des segments
    - ▶ modèles probabilistes de segments

# Des alignements de segments

## Phrases Anglaises :

- ▶ *I was reading reports from the united states*
- ▶ *To this first group of figures let us quickly add a second one*
- ▶ *One of my colleagues says that i have not had an answer yet*

## Phrases Françaises :

- ▶ *Je lisais des rapports qui nous parviennent des États-Unis*
- ▶ *À cette première vague de chiffres ajoutons rapidement une seconde*
- ▶ *L' un de mes collègues dit que je n' ai pas encore eu de réponse*

# Un algorithme d'extraction de segments

- ▶ Construction d'alignements de mots  $\mathbf{f} \rightarrow \mathbf{e}$  et  $\mathbf{e} \rightarrow \mathbf{f}$
- ▶ **Symétrisation** des alignements:
  - ▶ Création d'une matrice croisant les mots source et cible
  - ▶ Intersection des alignements
  - ▶ Etendus de façon heuristique au voisinage
- ▶ Extraction de segments  $(\tilde{f}_i, \tilde{e}_i)$
- ▶ Évaluation des segments

# Symétrisation des alignements

english to spanish

	bofetada				bruja			
	Maria	no	daba	una	a	la	verde	
Mary	■							
did					■			
not		■						
slap			■	■	■			
the						■		
green								■
witch							■	

spanish to english

	bofetada				bruja			
	Maria	no	daba	una	a	la	verde	
Mary	■							
did		■	■					
not								
slap						■		
the							■	
green								■
witch							■	

intersection

	bofetada				bruja			
	Maria	no	daba	una	a	la	verde	
Mary	■							
did								
not		■						
slap						■		
the							■	
green								■
witch							■	

# Symétrisation des alignements

				bofetada		bruja		
	Maria	no	daba	una	a	la	verde	
Mary								
did								
not								
slap								
the								
green								
witch								

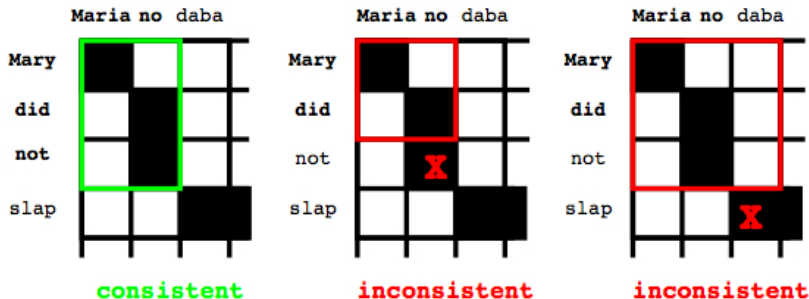
# Extraction des segments



Les alignements symétrisés



# Extraction des segments



Les contraintes de cohérence

$$\forall e_i \in \tilde{e}, (e_i, f_j) \in A \Rightarrow f_j \in \tilde{f}$$

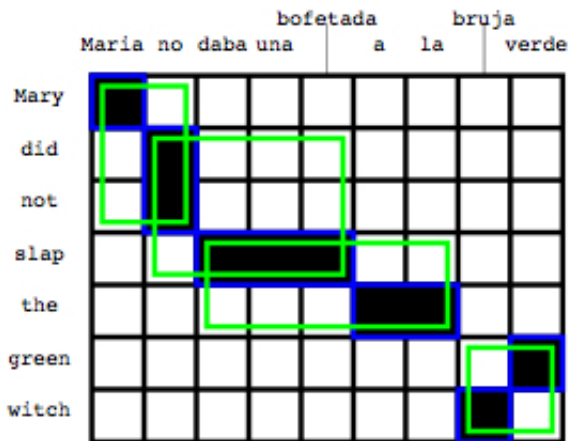
$$\forall f_j \in \tilde{f}, (e_i, f_j) \in A \Rightarrow e_i \in \tilde{e}$$

# Extraction des segments

	bofetada				bruja			
	Maria	no	daba	una	a	la	verde	
Mary	■							
did		■						
not		■						
slap			■	■	■			
the						■	■	
green								■
witch							■	

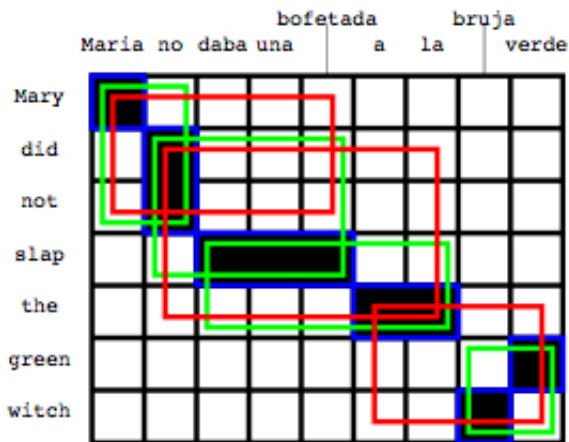
(Mary,Maria), (did not, no), (slap, daba una bofetada)...

# Extraction des segments



(Mary did not, Maria no), (did not slap, no daba una bofetada)...

# Extraction des segments



(Mary did not slap, Maria no daba una bofetada)...

# Les scores d'un fragment

- ▶ Au maximum de vraisemblance:

- ▶  $P(\tilde{f}|\tilde{e}) = \frac{c(\tilde{f},\tilde{e})}{c(\tilde{e})}$

- ▶  $P(\tilde{e}|\tilde{f}) = \frac{c(\tilde{e},\tilde{f})}{c(\tilde{f})}$

⇒ estimateurs très optimistes pour les longs segments

- ▶ Autres options:

- ▶  $P(\tilde{f}|\tilde{e}) = P_{IBM}(\tilde{f}|\tilde{e})$

- ▶  $P(\tilde{e}|\tilde{f}) = P_{IBM}(\tilde{e}|\tilde{f})$

- ▶ Pourquoi choisir ? ⇒ combinaison des scores

# Combinaison des scores et tuning

- Nouveau modèle de traduction (segments indépendants):

$$\begin{aligned} P(\mathbf{e}|\mathbf{f}) &= \sum_{a=s_1 \dots s_k} \prod_{i=1}^k P(\tilde{f}_i, s_i | \tilde{e}_{s_i}) \\ &\approx \max_{a=s_1 \dots s_k} \prod_{i=1}^k P(\tilde{f}_i, s_i | \tilde{e}_{s_i}) \end{aligned}$$

- Modèles probabilistes individuellement imprécis

⇒ Pondération de leur influence :

$$\begin{aligned} \mathbf{e}^* &= \operatorname{argmax}_{\mathbf{e}} \prod_k P_k(\mathbf{f}, \mathbf{e})^{\lambda_i} \\ \mathbf{e}^* &= \operatorname{argmax}_{\mathbf{e}} \sum_k \lambda_k \log P_k(\mathbf{f}, \mathbf{e}) \end{aligned}$$

- Comment déterminer les coefficients  $\lambda_i$  ?

# Optimisation du système: calcul des $\lambda$

- ▶ À la main ?

- ▶ Boucle exploratoire:

1. Choisir  $\lambda_k$  initiaux
2. Faire un découplage avec ces valeurs
3. Obtenir une solution et calculer son score EL/BU
4. Modifier les  $\lambda_k$  et recommencer à l'étape 2)
5. Terminer si le score EL/BU ne s'améliore plus

- ▶ Il n'est pas possible d'utiliser des algorithmes du type "descente de gradient"

⇒ Algorithmes itératifs de recherche

# Optimisation du système: calcul des $\lambda$

- ▶ À la main ?
  - ▶ Boucle exploratoire:
    1. Choisir  $\lambda_k$  initiaux
    2. Faire un décodage avec ces valeurs
    3. Obtenir une solution et calculer son score BLEU
    4. Modifier les  $\lambda_k$  et recommencer à l'étape 2)
    5. Terminer si le score BLEU ne s'améliore plus
  - ▶ Il n'est pas possible d'utiliser des algorithmes du type "descente de gradient"
- ⇒ Algorithmes itératifs de recherche



# Optimisation du système: calcul des $\lambda$

- ▶ À la main ?
  - ▶ Boucle exploratoire:
    1. Choisir  $\lambda_k$  initiaux
    2. Faire un décodage avec ces valeurs
    3. Obtenir une solution et calculer son score BLEU
    4. Modifier les  $\lambda_k$  et recommencer à l'étape 2)
    5. Terminer si le score BLEU ne s'améliore plus
  - ▶ Il n'est pas possible d'utiliser des algorithmes du type "descente de gradient"
- ⇒ Algorithmes itératifs de recherche

# Optimisation du système: calcul des $\lambda$

- ▶ À la main ?
  - ▶ Boucle exploratoire:
    1. Choisir  $\lambda_k$  initiaux
    2. Faire un décodage avec ces valeurs
    3. Obtenir une solution et calculer son score BLEU
    4. Modifier les  $\lambda_k$  et recommencer à l'étape 2)
    5. Terminer si le score BLEU ne s'améliore plus
  - ▶ Il n'est pas possible d'utiliser des algorithmes du type "descente de gradient"
- ⇒ Algorithmes itératifs de recherche

# Optimisation du système: calcul des $\lambda$

- ▶ À la main ?
  - ▶ Boucle exploratoire:
    1. Choisir  $\lambda_k$  initiaux
    2. Faire un décodage avec ces valeurs
    3. Obtenir une solution et calculer son score BLEU
    4. Modifier les  $\lambda_k$  et recommencer à l'étape 2)
    5. Terminer si le score BLEU ne s'améliore plus
  - ▶ Il n'est pas possible d'utiliser des algorithmes du type "descente de gradient"
- ⇒ Algorithmes itératifs de recherche

# Optimisation du système: calcul des $\lambda$

- ▶ À la main ?
  - ▶ Boucle exploratoire:
    1. Choisir  $\lambda_k$  initiaux
    2. Faire un décodage avec ces valeurs
    3. Obtenir une solution et calculer son score BLEU
    4. Modifier les  $\lambda_k$  et recommencer à l'étape 2)
    5. Terminer si le score BLEU ne s'améliore plus
  - ▶ Il n'est pas possible d'utiliser des algorithmes du type "descente de gradient"
- ⇒ Algorithmes itératifs de recherche

# Optimisation du système: calcul des $\lambda$

- ▶ À la main ?
  - ▶ Boucle exploratoire:
    1. Choisir  $\lambda_k$  initiaux
    2. Faire un décodage avec ces valeurs
    3. Obtenir une solution et calculer son score BLEU
    4. Modifier les  $\lambda_k$  et recommencer à l'étape 2)
    5. Terminer si le score BLEU ne s'améliore plus
  - ▶ Il n'est pas possible d'utiliser des algorithmes du type "descente de gradient"
- ⇒ Algorithmes itératifs de recherche

# Optimisation du système: calcul des $\lambda$

- ▶ À la main ?
- ▶ Boucle exploratoire:
  1. Choisir  $\lambda_k$  initiaux
  2. Faire un décodage avec ces valeurs
  3. Obtenir une solution et calculer son score BLEU
  4. Modifier les  $\lambda_k$  et recommencer à l'étape 2)
  5. Terminer si le score BLEU ne s'améliore plus
- ▶ Il n'est pas possible d'utiliser des algorithmes du type "descente de gradient"

⇒ Algorithmes itératifs de recherche

# Optimisation du système: calcul des $\lambda$

- ▶ À la main ?
  - ▶ Boucle exploratoire:
    1. Choisir  $\lambda_k$  initiaux
    2. Faire un décodage avec ces valeurs
    3. Obtenir une solution et calculer son score BLEU
    4. Modifier les  $\lambda_k$  et recommencer à l'étape 2)
    5. Terminer si le score BLEU ne s'améliore plus
  - ▶ Il n'est pas possible d'utiliser des algorithmes du type "descente de gradient"
- ⇒ Algorithmes itératifs de recherche

# Les ingrédients d'un modèle de segment

- Modèle de traduction :

$P(\tilde{f}|\tilde{e})$  traduction segments  $e \rightarrow f$

$P(f|e)$  traduction de mots  $e \rightarrow f$  (modèle lexical type IBM1)

$P(\tilde{e}|\tilde{f})$  traduction segments  $f \rightarrow e$

$P(e|f)$  traduction de mots  $e \rightarrow f$  (modèle lexical type IBM1)

- $e$  constante  $\rightarrow$  pénalité sur le nombre de segments

- + modèles de distortions (une autre fois)

- Modèle de langage :  $P(e)$

- constante 1  $\rightarrow$  pénalité de longueur

- ... et tout ce qui peut caractériser un segment...



# La table des segments

## quelques traductions de “A big”

```
A big ||| Le grand ||| 0.0106383 0.000152962 0.166667 0.00405915 2.718
A big ||| Un des principaux ||| 0.0434783 0.0005689 0.166667 1.56536e-05 2.718
A big ||| Un grand ||| 0.00961538 0.00957428 0.166667 0.0300893 2.718
A big ||| Une grande ||| 0.0108696 0.00360665 0.166667 0.0208976 2.718
A big ||| ont une grande ||| 0.0217391 1.12938e-05 0.166667 3.79597e-06 2.718
A big ||| une grande ||| 0.000256345 1.12938e-05 0.166667 0.00211983 2.718
```

Les scores :  $P(\tilde{f}|\tilde{e})$ ,  $P(e|f)$ ,  $P(\tilde{e}|\tilde{f})$ ,  $P(f|e)$  et  $\exp(1)$

# La table des segments (suite)

## 467 traductions de “European Commission”

```
European Commission ||| Commission européenne ||| 0.752696 0.812097 0.749849 0.455413 2.718
European Commission ||| Commission ||| 0.00265859 0.00194196 0.0511501 0.952132 2.718
European Commission ||| la Commission européenne ||| 0.0426116 0.812097 0.0352603 0.0174883 2
European Commission ||| Commission européenne , ||| 0.17041 0.812097 0.0195218 0.0364258 2.71
European Commission ||| de la Commission européenne ||| 0.0625 0.812097 0.0160412 0.00229579 :
```

## 38 traductions inverses de “Commission européenne”

```
European Commission ||| Commission européenne ||| 0.752696 0.812097 0.749849 0.455413 2.718
Commission ||| Commission européenne ||| 0.116208 0.490344 0.00548883 0.00587199 2.718
the European Commission ||| Commission européenne ||| 0.0095701 0.0437849 0.0119704 0.455413 :
Commission 's ||| Commission européenne ||| 0.00592435 0.00389219 0.0137227 0.00378834 2.718
Commission is ||| Commission européenne ||| 0.00303813 0.000335368 0.0036914 4.97013e-05 2.71
```

# La table des segments (suite et fin)

## 672 traductions de '!' !!!

```
! ||| ! ! ! ||| 0.375 0.588351 0.000338181 0.462852 2.718
! ||| ! ! ||| 0.153846 0.588351 0.000225454 0.598358 2.718
! ||| ! ||| 0.534388 0.588351 0.731372 0.773536 2.718
! ||| : non ! ||| 0.5 0.588351 0.000112727 2.60435e-07 2.718
...
! ||| , dit-on partout ! ||| 1 0.588351 0.000112727 4.76404e-12 2.718
! ||| , exigez que ||| 0.5 5.69e-05 0.000112727 1.92463e-10 2.718
! ||| , exigez ||| 0.333333 5.69e-05 0.000112727 1.20609e-08 2.718
! ||| , il est primordial que la ||| 0.333333 5.69e-05 0.000112727 3.20037e-15 2.718
! ||| , il est primordial que ||| 0.0277778 5.69e-05 0.000112727 8.33407e-14 2.718
...
! ||| Messieurs , il est primordial que la ||| 1 5.69e-05 0.000112727 4.92856e-19 2.718
! ||| Messieurs , il est primordial ||| 1 5.69e-05 0.000112727 8.04285e-16 2.718
...
```

# Outline

Traduire

Anatomie d'un système de traduction statistique

Traduction statistique mot-à-mot et alignements

Les alignements mot-à-mot

Modèles de segments

Décodage et recherche

Traduire la parole

Perspectives

# Recherche et décodage: un gros soucis

- ▶ Décodage monotone: ordre des mots identique en source et en cible
  - ▶ efficace
  - ▶ pas de réordonnancement
- ▶ Décodage avec distortion
  - ▶ résoudre  $\operatorname{argmax} P(\mathbf{e}|\mathbf{f})$  est NP-difficile (y compris avec IBM1 !)
  - ▶ méthodes heuristiques ( $A^*$ , recherche en faisceau, recherche locale, etc)
- ▶ L'espace de recherche est gigantesque
  - ▶ ne pas considérer toutes les traductions de phrase
  - ▶ ne pas considérer tous les réordonnements
  - ▶ élaguer pendant la recherche

This beautiful plant is unique

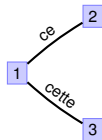
1

transfer table		
this	↔	ce
	↔	cette
beautiful	↔	belle
	↔	beau
plant	↔	plante
	↔	usine
is	↔	est
unique	↔	seule
	↔	unique
beautiful plant		
↕		
belle plante		
plante magnifique		

language model	
ce beau plante	:- (
cette belle usine	:-
belle usine est	:- )
...	



This beautiful plant is unique

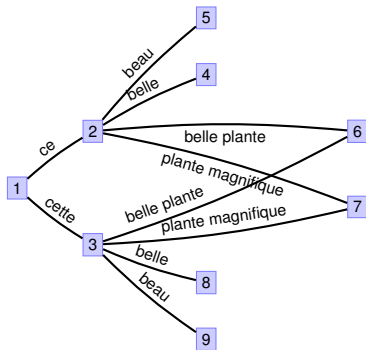


transfer table		
this	↔	ce
	↔	cette
beautiful	↔	belle
	↔	beau
plant	↔	plante
	↔	usine
is	↔	est
unique	↔	seule
	↔	unique
beautiful plant		
	↕	
belle plante		
plante magnifique		

language model	
ce beau plante	:- (
cette belle usine	:-
belle usine est	:- )
...	

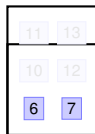


# This beautiful plant is unique



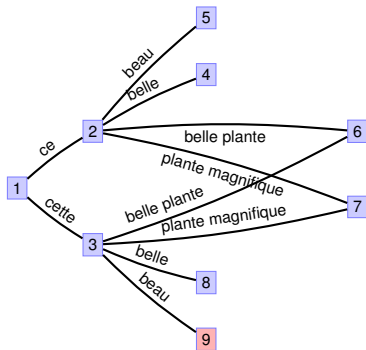
transfer table	
this	↔ ce
	↔ cette
beautiful	↔ belle
	↔ beau
plant	↔ plante
	↔ usine
is	↔ est
unique	↔ seule
	↔ unique
beautiful plant	
	↕
belle plante	
plante magnifique	

language model	
ce beau plante	:-
cette belle usine	:-
belle usine est	:-)
...	



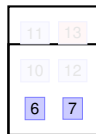


# This beautiful plant is unique

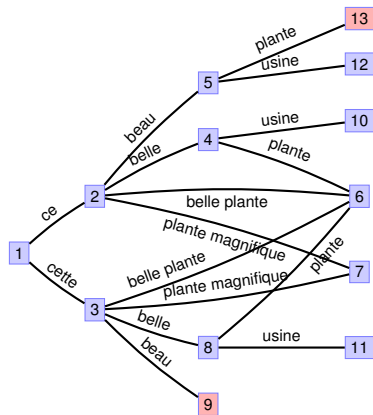


transfer table	
this	↔ ce
	↔ cette
beautiful	↔ belle
	↔ beau
plant	↔ plante
	↔ usine
is	↔ est
unique	↔ seule
	↔ unique
beautiful plant	
	↕
belle plante	
plante magnifique	

language model	
ce beau plante	:-
cette belle usine	:-
belle usine est	:-)
...	

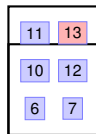


This beautiful plant is unique

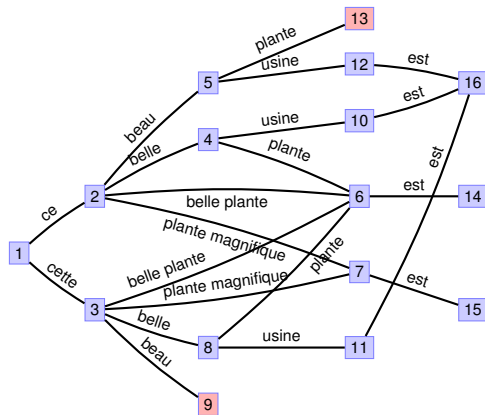


transfer table	
this	↔ ce
	↔ cette
beautiful	↔ belle
	↔ beau
plant	↔ plante
	↔ usine
is	↔ est
unique	↔ seule
	↔ unique
beautiful plant	
	↕
belle plante	
plante magnifique	

language model	
ce beau plante	:- (
cette belle usine	:-
belle usine est	:- )
...	



# This beautiful plant is unique

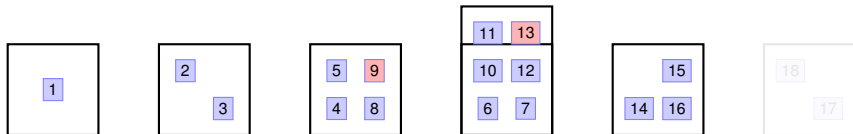


**transfer table**

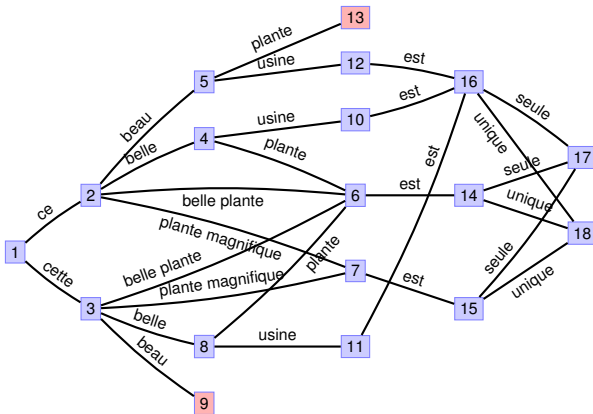
this	↔	ce
	↔	cette
beautiful	↔	belle
	↔	beau
plant	↔	plante
	↔	usine
is	↔	est
unique	↔	seule
	↔	unique
beautiful plant		
	↕	
belle plante		
plante magnifique		

**language model**

ce beau plante	:- (
cette belle usine	:-
belle usine est	:- )
...	

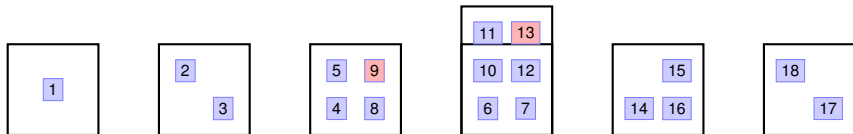


This beautiful plant is unique



transfer table		
this	↔	ce
	↔	cette
beautiful	↔	belle
	↔	beau
plant	↔	plante
	↔	usine
is	↔	est
unique	↔	seule
	↔	unique
beautiful plant ⇕ belle plante plante magnifique		

language model	
ce beau plante	:-)
cette belle usine	:-
belle usine est	:-)
...	



# Décodage en passes multiples

## Motivations

- ▶ Certains modèles sont difficiles à appliquer pendant la phase de décodage
  - ▶ Des modèles de langage avec un ordre élevé
  - ▶ Des modèles de phrases
  - ▶ Diverses analyses morpho-syntaxiques
  - ▶ ...
- ⇒ Utiliser une approche en deux passes :
  1. Décodage et création d'un ensemble de solutions plausibles
  2. Ajout de connaissances supplémentaires et sélection de la meilleure hypothèse

# Représentations intermédiaires

- ▶ Graphes de mots (*lattices*) a.k.a accepteurs pondérés

- ▶ Utilisés en reconnaissance de la parole (outils existants)
- ▶ Représentations compactes rès compacts
- ▶ Moins adaptés à la traduction (ordre de mots variables)

- ▶ Listes *n*-best

- ▶ Plus compactes et rapides qu'un arbre (mécanisme)
- ▶ Il est facile de gérer des listes *n*-best variables
- ▶ Permettent l'explication de décisions de phrases complètes (grâce à des règles)

# Représentations intermédiaires

- ▶ Graphes de mots (*lattices*) a.k.a accepteurs pondérés
  - ▶ Utilisés en reconnaissance de la parole (outils existants)
  - ▶ Représentations compactes rès compacts
  - ▶ Moins adaptés à la traduction (ordre de mots variables)
- ▶ Listes *n*-best
  - ▶ Liste compacte d'alternatives (un arbre n'est pas compact)
  - ▶ Permet de gérer des hypothèses multiples
  - ▶ Permettent l'exploration de l'espace des phrases possibles (au-delà des *n*-grammes)

# Représentations intermédiaires

- ▶ Graphes de mots (*lattices*) a.k.a accepteurs pondérés
  - ▶ Utilisés en reconnaissance de la parole (outils existants)
  - ▶ Représentations compactes rès compacts
    - ▶ Moins adaptés à la traduction (ordre de mots variables)
- ▶ Listes *n*-best



# Représentations intermédiaires

- ▶ Graphes de mots (*lattices*) a.k.a. accepteurs pondérés
  - ▶ Utilisés en reconnaissance de la parole (outils existants)
  - ▶ Représentations compactes rès compacts
  - ▶ Moins adaptés à la traduction (ordre de mots variables)
- ▶ Listes *n*-best
  - ▶ Moins compactes et variées qu'un *lattice* (redondance)

# Représentations intermédiaires

- ▶ Graphes de mots (*lattices*) a.k.a. accepteurs pondérés
  - ▶ Utilisés en reconnaissance de la parole (outils existants)
  - ▶ Représentations compactes rès compacts
  - ▶ Moins adaptés à la traduction (ordre de mots variables)
- ▶ Listes *n*-best
  - ▶ Moins compactes et variées qu'un lattice (redondance)
  - ▶ Il est facile de gérer des solutions très variables
  - ▶ Permettent l'application de modèle de phrases complexes (au-delà des n-grammes)

# Représentations intermédiaires

- ▶ Graphes de mots (*lattices*) a.k.a. accepteurs pondérés
  - ▶ Utilisés en reconnaissance de la parole (outils existants)
  - ▶ Représentations compactes rès compacts
  - ▶ Moins adaptés à la traduction (ordre de mots variables)
- ▶ Listes *n*-best
  - ▶ Moins compactes et variées qu'un lattice (redondance)
  - ▶ Il est facile de gérer des solutions très variables
  - ▶ Permettent l'application de modèle de phrases complexes (au-delà des n-grammes)

# Représentations intermédiaires

- ▶ Graphes de mots (*lattices*) a.k.a. accepteurs pondérés
  - ▶ Utilisés en reconnaissance de la parole (outils existants)
  - ▶ Représentations compactes vs compacts
  - ▶ Moins adaptés à la traduction (ordre de mots variables)
- ▶ Listes *n*-best
  - ▶ Moins compactes et variées qu'un lattice (redondance)
  - ▶ Il est facile de gérer des solutions très variables
  - ▶ Permettent l'application de modèle de phrases complexes (au-delà des *n*-grammes)

# Représentations intermédiaires

- ▶ Graphes de mots (*lattices*) a.k.a. accepteurs pondérés
  - ▶ Utilisés en reconnaissance de la parole (outils existants)
  - ▶ Représentations compactes vs compacts
  - ▶ Moins adaptés à la traduction (ordre de mots variables)
- ▶ Listes *n*-best
  - ▶ Moins compactes et variées qu'un lattice (redondance)
  - ▶ Il est facile de gérer des solutions très variables
  - ▶ Permettent l'application de modèle de phrases complexes (au-delà des *n*-grammes)

# Une liste de $n$ -best

```
0 ||| Notre déclaration des droits est la première de ce millénaire .  
||| lm: -53.1725 tm: -8.54868 -8.36703 -6.29597 -9.46295 8.99907 w: -11 ||| -2.01804  
0 ||| Notre déclaration des droits n ' est la première de ce millénaire .  
||| lm: -55.9546 tm: -4.29181 -8.36703 -5.70585 -16.96 7.99917 w: -13 ||| -2.10735  
0 ||| Notre déclaration des droits est le premier de ce millénaire .#  
||| lm: -52.6802 tm: -8.68783 -8.73413 -7.26683 -10.4078 8.99907 w: -11 ||| -2.11691  
0 ||| Notre déclaration des droits est la première de ce nouveau millénaire .  
||| lm: -53.4205 tm: -10.6899 -9.05756 -9.47818 -12.9981 8.99907 w: -12 ||| -2.21531  
0 ||| Notre déclaration des droits n ' est le premier de ce millénaire .  
||| lm: -55.7673 tm: -4.43096 -8.73413 -6.67671 -17.9048 7.99917 w: -13 ||| -2.22684  
0 ||| Notre déclaration de droits est la première de ce millénaire .  
||| lm: -59.4228 tm: -3.32516 -8.27583 -5.18753 -8.48052 7.99917 w: -11 ||| -2.2324  
0 ||| La déclaration des droits est la première de ce millénaire .  
||| lm: -47.5985 tm: -14.3938 -14.2576 -9.5678 -12.6795 8.99907 w: -11 ||| -2.29651  
0 ||| Notre déclaration des droits n ' est la première de ce nouveau millénaire .  
||| lm: -56.2027 tm: -6.43302 -9.05756 -8.88807 -20.4951 7.99917 w: -14 ||| -2.30462  
0 ||| Notre déclaration des droits , c ' est la première de ce millénaire .  
||| lm: -54.6978 tm: -9.60467 -8.81344 -10.844 -16.6753 8.99907 w: -13 ||| -2.30946  
0 ||| Notre déclaration des droits est la première de millénaire .  
||| lm: -53.9945 tm: -7.78823 -12.0139 -4.39036 -8.23918 6.99927 w: -10 ||| -2.31331
```

Remarque : les scores du modèle de distorsion ont été omis

# Outline

Traduire

Anatomie d'un système de traduction statistique

Traduction statistique mot-à-mot et alignements

Les alignements mot-à-mot

Modèles de segments

Décodage et recherche

**Traduire la parole**

Perspectives

# Traduire la Parole: plus ou moins facile ?

## De nouvelles difficultés

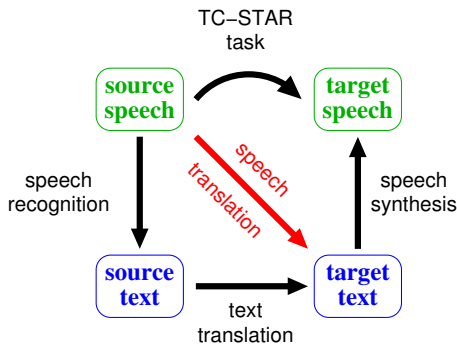
- ▶ des entrées “bruitées” (dysfluences)
- ▶ des données d’entraînement encore plus rares...
- ▶ ou encore moins bien adaptées

## Des facteur simplificateurs

- ▶ utilisation de la prosodie ?
- ▶ possibilité d’interactions avec l’utilisateur ?



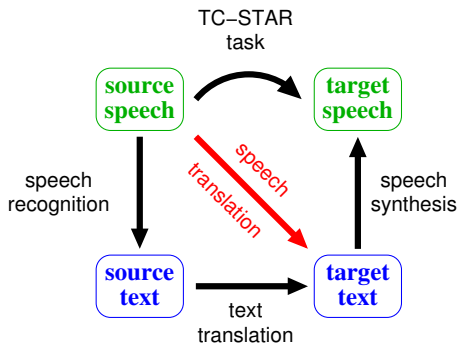
# Architectures pour la Traduction de la Parole



## Architectures

- ▶ Pipe line: reconnaissance | traduction | synthèse
- ▶ Peut-on faire mieux avec une interaction étroite ?
- ▶ Traduction assistée par ordinateur

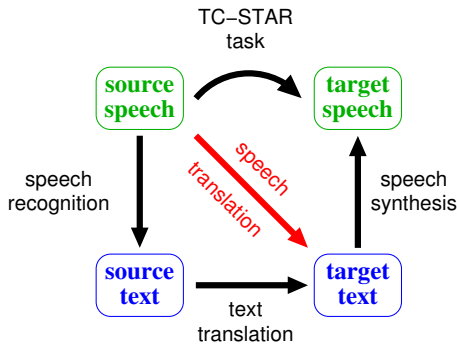
# Architectures pour la Traduction de la Parole



## Architectures

- ▶ Pipe line: reconnaissance | traduction | synthèse
- ▶ Peut-on faire mieux avec une interaction étroite ?
- ▶ Traduction assistée par ordinateur

# Architectures pour la Traduction de la Parole



## Architectures

- ▶ Pipe line: reconnaissance | traduction | synthèse
- ▶ Peut-on faire mieux avec une interaction étroite ?
- ▶ Traduction assistée par ordinateur

# La traduction de parole, décision probabiliste

## Analyse Théorique [Ney ICASSP'99]

$e$ : phrase cible,  $x$  signal source,  $f$  phrase source

$$\begin{aligned} e^* &= \operatorname{argmax}_e \Pr(e|x) = \operatorname{argmax}_e \Pr(e) \Pr(x|e) \\ &= \operatorname{argmax}_e \left( \Pr(e) \sum_f \Pr(x, f|e) \right) \\ &= \operatorname{argmax}_e \left( \Pr(e) \sum_f \Pr(f|e) \Pr(x|f, e) \right) \end{aligned}$$

Comment simplifier cette équation ?

# Sans modéliser les textes en source ??

Analyse Théorique (suite) :

$$\begin{aligned} e^* &= \operatorname{argmax}_e \left( \Pr(e) \sum_f \Pr(f|e) \Pr(x|f, e) \right) \\ &= \operatorname{argmax}_e \left( \Pr(e) \sum_f \Pr(f|e) \Pr(x|f) \right) \\ &\approx \operatorname{argmax}_e \left( \Pr(e) \max_f \Pr(f|e) \Pr(x|f) \right) \end{aligned}$$

## Questions

- ▶ Faut-il un modèle de langage source  $P(f)$  ?
- ▶ Comment obtenir une interaction étroite ?

# Sans modéliser les textes en source ??

Analyse Théorique (suite) :

$$\begin{aligned} e^* &= \operatorname{argmax}_e \left( \Pr(e) \sum_f \Pr(f|e) \Pr(x|f, e) \right) \\ &= \operatorname{argmax}_e \left( \Pr(e) \sum_f \Pr(f|e) \Pr(x|f) \right) \\ &\approx \operatorname{argmax}_e \left( \Pr(e) \max_f \Pr(f|e) \Pr(x|f) \right) \end{aligned}$$

## Questions

- ▶ Faut-il un modèle de langage source  $P(f)$  ?
- ▶ Comment obtenir une interaction étroite ?

# One step back

## Interprétation :

$$e^* \approx \arg \max_e \left( \Pr(e) \max_f \Pr(f|e) \Pr(x|f) \right)$$

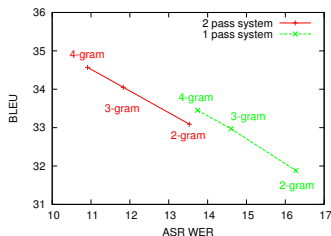
- ▶  $\Pr(f|e)$  probabilise **des phrases entières**
- ⇒ Induit une structure sur l'ensemble des phrases sources
- ▶ Mais les modèles de traduction actuels sont plutôt basés sur des mots

## Analyse Expérimentale :

- ▶ Analyser les performances de traduction pour différents modèles acoustiques et modèles de langage source
- ▶ Essayer d'utiliser les scores du module de reconnaissance afin d'améliorer la traduction

# Taux d'erreur et erreurs de traduction

## Score BLEU pour différents systèmes de reconnaissance



Traduction de texte :  
score bleu de 40.5

Erreur de la Reco :  
Système 1: 10.9%  
Système 2: 13.7%

## Conclusions

- ▶ Peu de perte par rapport à la traduction de texte
- ▶ Dépendance linéaire entre l'erreur de mots et score BLEU ?



# Une interface plus riche: les listes d'hypothèses

- ▶ Reconnaissance de la parole et génération de listes  $n$ -best
  - ▶ Traduction de toutes les hypothèses
  - ▶ Combinaison des scores de traduction et des scores de transcription
- ⇒ Reclassement des hypothèses traduites
- ▶ Généralisation: reclassement de treillis d'hypothèses, de réseaux de confusion

# Outline

Traduire

Anatomie d'un système de traduction statistique

Traduction statistique mot-à-mot et alignements

Les alignements mot-à-mot

Modèles de segments

Décodage et recherche

Traduire la parole

Perspectives

# Perspectives de recherche

- ▶ Plus de traitements linguistiques (en source, en cible)
- ▶ Modèles discriminants pour l'alignement
- ▶ Utilisation de tables de segments généralisées
- ▶ Meilleure évaluation/sélection des segments (utilisation de la source: segmentation, désambiguïsation)
- ▶ Meilleurs modèles de réordonnancement (traduction de paires de langues très différentes)
- ▶ Vers la traduction de documents (coréférence, discours...)
- ▶ Couplage Reconnaissance/traduction:
  - ▶ quelles interfaces ?
  - ▶ utilisation des informations prosodiques (segmentation, intonation)
- ▶ Construction semi-supervisée de modèles de traduction

# Incorporation de Connaissances Linguistiques

## Motivation

### Motivation

- ▶ Les méthodes statistiques permettent facilement de construire des systèmes performants
  - ▶ Mais la traduction ne correspond pas toujours à un texte grammaticalement correct, facile à comprendre ou bien structuré
- ⇒ De la place pour des améliorations

### Directions de Recherche

- ▶ Meilleurs modèles statistiques
- ▶ Incorporation de connaissances linguistiques

# Incorporation de Connaissances Linguistiques

## Motivation

### Motivation

- ▶ Les méthodes statistiques permettent facilement de construire des systèmes performants
  - ▶ Mais la traduction ne correspond pas toujours à un texte grammaticalement correct, facile à comprendre ou bien structuré
- ⇒ De la place pour des améliorations

### Directions de Recherche

- ▶ Meilleurs modèles statistiques
- ▶ Incorporation de connaissances linguistiques

# Incorporation de Connaissances Linguistiques

## Idées :

- ▶ Incorporation de connaissances dans le modèle de langage
- ▶ Incorporation de connaissances dans le modèle de traduction
- ▶ Utilisation d'autres ressources, p.ex. des dictionnaires bilingues

# Incorporation de Connaissances Linguistiques

## Idées :

- ▶ Incorporation de connaissances dans le modèle de langage
- ▶ Incorporation de connaissances dans le modèle de traduction
- ▶ Utilisation d'autres ressources, p.ex. des dictionnaires bilingues

# Incorporation de Connaissances Linguistiques

## Idées :

- ▶ Incorporation de connaissances dans le modèle de langage
- ▶ Incorporation de connaissances dans le modèle de traduction
- ▶ Utilisation d'autres ressources, p.ex. des dictionnaires bilingues



# Améliorer les modèles statistiques

## Modèles de Traduction

- ▶ Grammaires synchrones
- ▶ Modèles discriminants
- ▶ Désambiguïsation lexicale
- ▶ Modèles de traduction factorisés

## Modèles de Langue

- ▶ Traitement en deux passes (traduction avec  $n$ -gramme + rescoring)
- ▶ Modèles  $n$ -grammes de catégories morpho-syntaxiques
- ▶ Modèles fondés sur la syntaxe (SCFG, dépendances)
- ▶ Pour l'instant, peu d'améliorations observées (sauf en ajoutant du corpus)

# Segments généralisés (avec trous)

**f**= tu ne veux pas dormir

**e**= you don't want to sleep

	tu	ne	veux	pas	dormir
sleep					
to					
want					
don't					
you					

- ▶ (*want*; *veux*) a sub-phrase of (*don't want* ; *ne veux pas*)
- ▶ segment à trous  $N(\textit{don't } X ; \textit{ne } X \textit{ pas})++$
- ▶ meilleure généralisation

# Segments généralisés (avec trous)

f=je ne le comprends plus

e= I don't understand it anymore

	je	ne	le	comprend	plus
anymore					
it					
understand					
don't					
I					

- ▶ même idée, avec deux non-terminaux
- ▶  $N(\textit{don't } X_1 X_2 \textit{ anymore} ; \textit{ne } X_2 X_1 \textit{ plus})^{++}$
- ▶ définit un modèle de réordonnancement lexicalisé

# Un système hiérarchique

Quelques innovations dues à D. Chiang

- ▶ segments à trous = règles d'une grammaire hors-contexte synchrone
  - ▶ segment classique  $(e; f) \Leftrightarrow \text{règle terminale } X \rightarrow e; f$
  - ▶ segment à trou  $(\alpha; \beta) \Leftrightarrow X \rightarrow \alpha; \beta$
  - ▶ règle de concaténation  $S \rightarrow SX \mid X$
  - ▶ estimation ML (+ lissage)
- ▶ traduction pendant le décodage

$$\mathbf{e} = \operatorname{argmax}_{\mathbf{e} \in E} \lambda_1 \log P_{LM}(\mathbf{e}) + \lambda_2 \log P_G(\mathbf{f}; \mathbf{e}) + \dots$$

- ▶ Avantage
  - ▶ des segments plus généraux
  - ▶ un meilleur modèle de réordonnancement
  - ▶ les performances [?]
- ▶ Problèmes
  - ▶ taille de la grammaire
  - ▶ coût du décodage

# Désambiguïsation lexicale

An : You must make the first **move**.

---

An : You must first **move** the car.

Comment traduire le mot *move* ?

# Désambiguïsation lexicale

An : You must make the first **move**.

Fr : Tu dois faire le premier **pas**.

---

An : You must first **move** the car.

Fr : Tu dois d'abord **déplacer** la voiture.

Comment traduire le mot *move* ?  $\left\{ \begin{array}{l} \text{pas} \\ \text{déplacer} \end{array} \right.$

# Désambiguïisation lexicale

An : You must make the **first move**.

Fr : Tu dois faire le **premier pas**.

---

An : You must **first move** the car.

Fr : Tu dois **d'abord déplacer** la voiture.

Traduction de *first move* ?  $\left\langle \begin{array}{l} \text{premier pas} \\ \text{d'abord déplacer} \end{array} \right.$

# Désambiguïisation lexicale

An : You must make the first **move**.

PP MD VV DT JJ **NN**

Fr : Tu dois faire le premier pas.

---

An : You must first **move** the car.

PP MD RB **VV** DT NN

Fr : Tu dois d'abord déplacer la voiture.

Les catégories lexicales permettraient de désambiguïiser



# Désambiguïisation lexicale

An : You must make the first **move**.

PP MD VV DT JJ NN

Fr : Tu dois faire le premier **pas**.

---

An : You must first **move** the car.

PP MD RB VV DT NN

Fr : Tu dois d'abord **déplacer** la voiture.

Traductions :     $\text{move}_{NN} \rightarrow \text{pas}$   
                       $\text{move}_{VV} \rightarrow \text{déplacer}$

# Utilisation de morpho-syntaxe

## Principe :

- ▶ Étiqueter les textes parallèles avec des informations morpho-syntaxiques (les deux côtés)
- ▶ Enrichir les mots avec les catégories lexicales :

*You<sub>P</sub> must<sub>V</sub> make<sub>V</sub> the<sub>D</sub> first<sub>Adj</sub> move<sub>N</sub>.*

*Tu<sub>P</sub> dois<sub>V</sub> faire<sub>V</sub> le<sub>D</sub> premier<sub>Adj</sub> pas<sub>N</sub>.*

- ▶ Construire un **système statistique** complet sur ce vocabulaire enrichi
- ▶ En sortie :
  - ▶ Suppression des étiquettes
  - ▶ Réutilisation des étiquettes (ML morpho-syntaxique)

# Modèles de Traduction Factorisés

## Motivation

- ▶ Seuls sont disponibles les segments du corpus parallèle d'apprentissage
- ▶ Pas de généralisation lexicale

## Exemple

- ▶ *La voiture rouge est belle*  
→ *The red car is nice*
- ▶ *Les vélos rouges sont beaux*  
→ *The red bikes are nice*
- ▶ *Les voitures rouges sont belles*  
Traduction de cette phrase sachant les deux autres ?

# Modèles de Traduction Factorisés

## Motivation

- ▶ Seuls sont disponibles les segments du corpus parallèle d'apprentissage
- ▶ Pas de généralisation lexicale

## Exemple

- ▶ *La voiture rouge est belle*  
→ *The red car is nice*
- ▶ *Les vélos rouges sont beaux*  
→ *The red bikes are nice*
- ▶ *Les voitures rouges sont belles*  
Traduction de cette phrase sachant les deux autres ?

# Modèle de Traduction Factorisé

## Principe

- ▶ L'approche actuelle de traduction par syntagmes traite des formes
  - ▶ Savoir traduire une forme ne permet pas de traduire les formes apparentés (conjugaison, flexion ...)
- ⇒ Décomposer les mots en lemme, genre, nombre, ...
- ▶ Traduire ces **facteurs** séparément
  - ▶ Recomposer le mot dans la langue cible à partir de la traductions des facteurs
  - ▶ **Processus de génération**



# Conclusion

## État de l'art

- ▶ applications grand public: traduction approximative (aide à la traduction, RI cross-lingue, etc): Google Translate, Microsoft Translator... vers une navigation transparente
- ▶ applications professionnelles: modules statistiques dans Systran 7, partenariat Reverso/Language Weaver, etc
- ▶ applications militaires traduction de parole déployées

## Des verrous ?

- ▶ Robustesse aux énoncés bruités, aux changements de domaine
- ▶ Des modèles moins lourds
- ▶ Discours, aspects culturels
- ▶ Faire avec le manque de données
- ▶ Hybrider SMT/EBMT/RBMT