



HAL
open science

L'informaticien face au traitement du langage : que peut-on faire et que croit-on que nous faisons ?

Maxime Amblard

► **To cite this version:**

Maxime Amblard. L'informaticien face au traitement du langage : que peut-on faire et que croit-on que nous faisons ?. Séminaire 2015/2016 " Images du savoir pratique : les figures de l'informaticien et du médecin dans les récits de fiction populaire contemporains, Feb 2016, Strasbourg, France. hal-01280932

HAL Id: hal-01280932

<https://inria.hal.science/hal-01280932>

Submitted on 4 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Plan

- 1 Traitement Automatique des Langues
- 2 Que fait on ?
- 3 Que croit-on que nous faisons ?
- 4 Conclusion

Plan

- 1 **Traitement Automatique des Langues**
- 2 Que fait on ?
- 3 Que croit-on que nous faisons ?
- 4 Conclusion

Traitement Automatique des Langues

Manipulations numériques de données exprimées en langue naturelle

Traitement Automatique des Langues

Manipulations numériques de données exprimées en langue naturelle

- capacité d'exprimer une pensée au moyen de signes avec une sémantique. *ie* un moyen de communication

Traitement Automatique des Langues

Manipulations numériques de données exprimées en langue naturelle

- capacité d'exprimer une pensée au moyen de signes avec une sémantique. *ie* un moyen de communication
- langage naturel : langue parlée par un être humain
- langage non naturel :
 - $a^n b^n c^n$
 - définir des programmes (des outils concrets)
 - penser des algorithmes (des outils abstraits de modélisation)

Traitement Automatique des Langues

Manipulations numériques de données exprimées en langue naturelle

- capacité d'exprimer une pensée au moyen de signes avec une sémantique. *ie* un moyen de communication
- langage naturel : langue parlée par un être humain
- langage non naturel :
 - $a^n b^n c^n$
 - définir des programmes (des outils concrets)
 - penser des algorithmes (des outils abstraits de modélisation)
- si on comprend un langage, on pourrait le définir pour qu'une machine le comprenne (?)

Qu'est-ce que la langue naturelle ?

- difficile de dire précisément si une phrase est grammaticale/non-grammaticale
- pas d'ensemble officiel de règles grammaticales

Qu'est-ce que la langue naturelle ?

- difficile de dire précisément si une phrase est grammaticale/non-grammaticale
 - pas d'ensemble officiel de règles grammaticales
- ⇒ les modèles de la langue ne sont que des approximations cherchant à définir des processus cognitifs
- ⇒ on cherche à définir les propriétés formelles :
- suisse allemand $S_1 S_2 S_3 V_1 V_2 V_3 O_1 O_2 O_3$.

Plan

1 Traitement Automatique des Langues

2 **Que fait on ?**

3 Que croit-on que nous faisons ?

4 Conclusion

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie

- syntaxe

- sémantique

- pragmatique

Modéliser la langue = gérer les ambiguïtés

Ginette Garcin

- phonologie
- morphologie

- syntaxe

- sémantique

- pragmatique

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie

autossa = dans l'auto
autossaniko = dans mon auto

- syntaxe
- sémantique
- pragmatique

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie

la : pronom, nom ou déterminant ?

- syntaxe
- sémantique
- pragmatique

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie

- syntaxe

Jean regarde (la fille avec un télescope)
Jean regarde (la fille) (avec un télescope)

- sémantique

- pragmatique

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie

- syntaxe

Esther pense que Bertille croit que Hélène dit qu'elle conduira la voiture

- sémantique
- pragmatique

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie

- syntaxe

- sémantique

- pragmatique

La belle ferme la porte
étoile : célébrité ou astre ?

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie

- syntaxe

- sémantique

- pragmatique

La pièce est dans le porte-monnaie
Le porte-monnaie est dans la pièce

Où et pour faire quoi ?

- mails, SMS, Twitter, Facebook, commentaires, blogs, articles, pages web, ...

Où et pour faire quoi ?

- mails, SMS, Twitter, Facebook, commentaires, blogs, articles, pages web, ...

- interface homme-machine en langue naturelle
- dialogues automatiques (réservations d'hôtel, Siri)
- génération automatique de textes (résumé, bulletin météo)
- reconnaissance d'écriture, identification vocale

Où et pour faire quoi ?

- synthèse vocale
- traduction automatique, correction automatique
- recherche d'information, moteurs de recherche
- jeux en ligne (syntaxe, sémantique)
- détection d'opinion, de leader, suivi de marques sur internet
- génération de sous-titres
- avatar (LSF)

Des réalisations récentes - Fiction du réel 1/2

- Watson, qui a gagné Jeopardy
- Benjamin Gotesman et le Test de Turing.
- jeu de Go : victoire de AlphaGo de Google face au champion d'Europe. Mars 2016 : match contre Lee Sedol, champion du monde
- assistant virtuel



I'LL BE BACK!



EUGENE GOOSTMAN
THE WEIRDEST CREATURE IN THE WORLD



Des réalisations récentes - Fiction du réel 1/2

- **redaction automatique**

Did a Human or a Computer Write This?

A shocking amount of what we're reading is created not by humans, but by computer algorithms. Can you tell the difference? Take the quiz.

MARCH 7, 2015

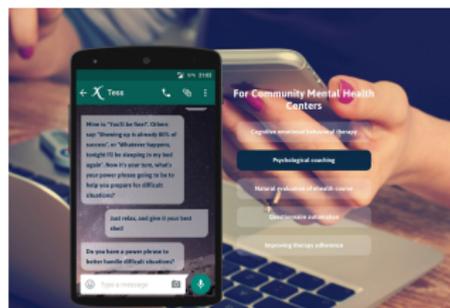
- **FaceBook** : identifier les mots manquants, cartographier la population mondiale
- **gestion de la santé** : identification de pathologies

```
Context : 1 do they had to fail (a tiny way)
2 do they not share their hair (a lower amount)
3 do they could n't get them out again .
4 that 's all .
5 thank you , (said Alice ) 's very interesting .
6 I have them on my (said Alice ) 's very interesting .
7 I can tell you more (said Alice ) 's very interesting .
8 do you know why it 's (said Alice ) 's very interesting .
9 I never thought about it (said Alice ) 's very interesting .
10 why ?
11 (said Alice ) 's very interesting .
12 (said Alice ) 's very interesting .
13 (said Alice ) 's very interesting .
14 (said Alice ) 's very interesting .
15 she repeated it (said Alice ) 's very interesting .
16 why , what (said Alice ) 's very interesting .
17 (said Alice ) 's very interesting .
18 (said Alice ) 's very interesting .
19 (said Alice ) 's very interesting .
20 (said Alice ) 's very interesting .

Query : "books and poems under the sea" (a tiny way) is a long
message , and some (a tiny way) is a long
message .

Candidate Answers : Alice , 8009 , Gryphon , 9090 , number , fall , month , tone , way , white .

MontiNo (window + self=up.)) Oryphon
```



réussites, mais sans modèle interprétatif

Plan

- 1 Traitement Automatique des Langues
- 2 Que fait on ?
- 3 Que croit-on que nous faisons ?**
- 4 Conclusion

Que croit-on que nous faisons ? Dans les séries américaines récentes

Évidence qu'un ordinateur utilise la LN

- l'informatique étant basée sur un langage particulier, l'informaticien parle ce langage
- il est donc bilingue entre sa langue maternelle et la langue de l'informatique
- or si un humain a métaphoriquement cette capacité, cela présuppose qu'il existe une traduction entre les deux langues
- ou dit autrement, tout ce qui peut être exprimé dans une langue peut être exprimé dans l'autre langue

Évidence qu'un ordinateur utilise la LN

Si les deux langues partagent des propriétés (existence d'une syntaxe, d'une sémantique,...),

Évidence qu'un ordinateur utilise la LN

Si les deux langues partagent des propriétés (existence d'une syntaxe, d'une sémantique,...), c'est évidemment absurde.

Évidence qu'un ordinateur utilise la LN

Si les deux langues partagent des propriétés (existence d'une syntaxe, d'une sémantique,...), c'est évidemment absurde.

- ❶ la classe des langages humains est supposée plus petite que les langages exprimables par une machine de Turing (cf hiérarchie de Chomsky)

Évidence qu'un ordinateur utilise la LN

Si les deux langues partagent des propriétés (existence d'une syntaxe, d'une sémantique,...), c'est évidemment absurde.

- 1 la classe des langages humains est supposée plus petite que les langages exprimables par une machine de Turing (cf hiérarchie de Chomsky)
- 2 les informaticiens n'ont pas encore réussi à définir des représentations conceptuelles de la langue naturelle. Le lien entre langue et logique est une autre question importante et connexe.

Évidence qu'un ordinateur utilise la LN

Les ordinateurs sont basés sur :

- des langages machines (explicitants les réactions des composants)
- des langages de programmation qui permettent de modéliser les chaînes de comportements
- des langages abstraits (algorithmique) qui permettent de penser les calculs

Évidence qu'un ordinateur utilise la LN

Les ordinateurs sont basés sur :

- des langages machines (explicitants les réactions des composants)
- des langages de programmation qui permettent de modéliser les chaînes de comportements
- des langages abstraits (algorithmique) qui permettent de penser les calculs

Les algorithmes sont généralement d'abord exprimés en langue naturelle et réinterprétés dans d'autres représentations dédiées.

Évidence qu'un ordinateur utilise la LN

Les ordinateurs sont basés sur :

- des langages machines (explicitants les réactions des composants)
- des langages de programmation qui permettent de modéliser les chaînes de comportements
- des langages abstraits (algorithmique) qui permettent de penser les calculs

Les algorithmes sont généralement d'abord exprimés en langue naturelle et réinterprétés dans d'autres représentations dédiées.

Et certains outils de l'informatique sont dédiés à l'usage des humains et donc passent par la langue naturelle.

l'informaticien serait omniscient

l'informatique permet d'accéder à une masse énorme de donnée, laissant croire que l'on peut tout savoir, tout faire. Donc l'informaticien est capable de :

l'informaticien serait omniscient

l'informatique permet d'accéder à une masse énorme de donnée, laissant croire que l'on peut tout savoir, tout faire. Donc l'informaticien est capable de :

- comprendre (instantanément) les systèmes d'information (complexes et interconnectés) qui contiennent les données ;

l'informaticien serait omniscient

l'informatique permet d'accéder à une masse énorme de donnée, laissant croire que l'on peut tout savoir, tout faire. Donc l'informaticien est capable de :

- comprendre (instantanément) les systèmes d'information (complexes et interconnectés) qui contiennent les données ;
- maîtriser les réseaux et la sécurité ;

l'informaticien serait omniscient

l'informatique permet d'accéder à une masse énorme de donnée, laissant croire que l'on peut tout savoir, tout faire. Donc l'informaticien est capable de :

- comprendre (instantanément) les systèmes d'information (complexes et interconnectés) qui contiennent les données ;
- maîtriser les réseaux et la sécurité ;
- extraire les informations pertinentes : manipulation de l'information : image, son, son, texte, document, ... ;

l'informaticien serait omniscient

l'informatique permet d'accéder à une masse énorme de donnée, laissant croire que l'on peut tout savoir, tout faire. Donc l'informaticien est capable de :

- comprendre (instantanément) les systèmes d'information (complexes et interconnectés) qui contiennent les données ;
- maîtriser les réseaux et la sécurité ;
- extraire les informations pertinentes : manipulation de l'information : image, son, son, texte, document, ... ;
- lire le code comme s'il s'agissait de littérature (ce qui est absurde parce que la lecture n'est pas linéaire).

Marginalisation de son comportement

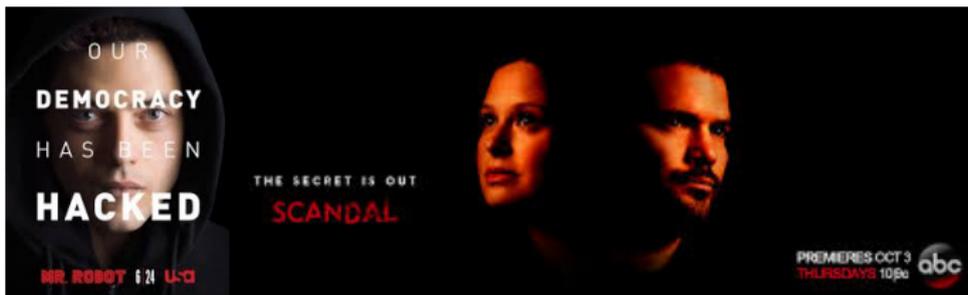
mais ça a un cout, l'informaticien n'a pas de comportement normal

Marginalisation de son comportement

mais ça a un cout, l'informaticien n'a pas de comportement normal

- Mr Robot
- Scandal

alors qu'il y a un énorme travail pour coller à la réalité



Marginalisation de son comportement

mais ça a un cout, l'informaticien n'a pas de comportement normal

- Mr Robot
- Scandal

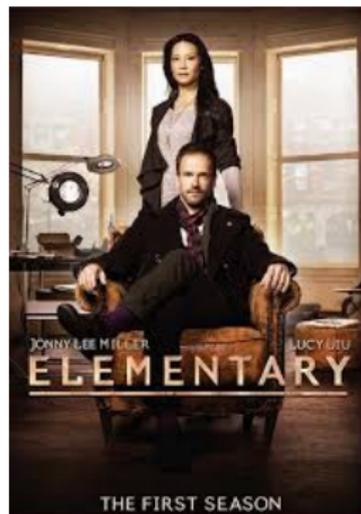
alors qu'il y a un énorme travail pour coller à la réalité



L'informaticien devient une entité multiple

Comme on ne comprend pas très bien ce que l'informatique permet de faire, on suppose parfois qu'un seul humain pour le faire n'est pas suffisant : communautés d'informaticiens ou organisations supposées omniscientes

- Everyone de Elementary



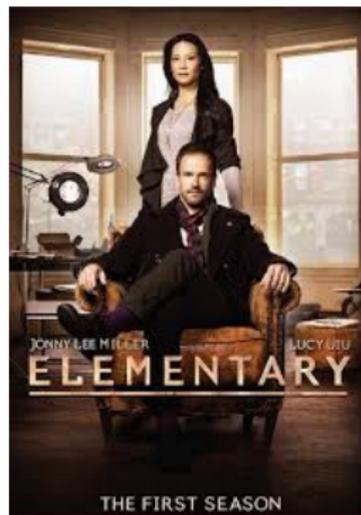
L'informaticien devient une entité multiple

Comme on ne comprend pas très bien ce que l'informatique permet de faire, on suppose parfois qu'un seul humain pour le faire n'est pas suffisant : communautés d'informaticiens ou organisations supposées omniscientes

- Everyone de Elementary

Et les organisation peuvent être gouvernementale ou institutionnalisée

- Person of Interest



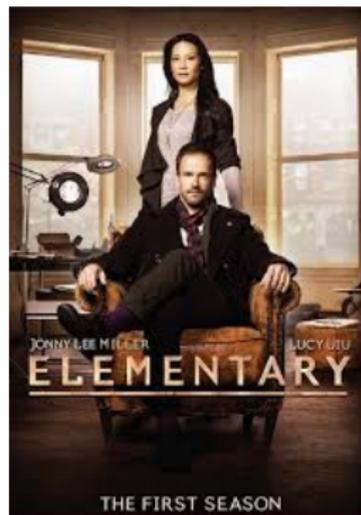
L'informaticien devient une entité multiple

Comme on ne comprend pas très bien ce que l'informatique permet de faire, on suppose parfois qu'un seul humain pour le faire n'est pas suffisant : communautés d'informaticiens ou organisations supposées omniscientes

- Everyone de Elementary

Et les organisation peuvent être gouvernementale ou institutionnalisée

- Person of Interest



ou une intelligence artificielle

jusqu'à produire des formes humanoïdes avancées d'ordinateur (et donc d'IA) qui passent parfaitement le test de Turing : Real Humans



ou une intelligence artificielle

jusqu'à produire des formes humanoïdes avancées d'ordinateur (et donc d'IA) qui passent parfaitement le test de Turing : Real Humans



Digne héritière de Hal



ou une intelligence artificielle

jusqu'à produire des formes humanoïdes avancées d'ordinateur (et donc d'IA) qui passent parfaitement le test de Turing : **Real Humans**



Digne héritière de Hal



Dans quelques rares fictions, l'informaticien est utilisé comme un autre élément narratif.

- The Code : l'informatique est l'objet de la recherche
- Halt and Catch fire : l'informatique est le contexte



Quoi que dans les deux cas, les informaticiens sont encore une fois marginaux, pour les besoins du scénario

Dans quelques rares fictions, l'informaticien est utilisé comme un autre élément narratif.

- The Code : l'informatique est l'objet de la recherche
- **Halt and Catch fire** : l'informatique est le contexte



Quoi que dans les deux cas, les informaticiens sont encore une fois marginaux, pour les besoins du scénario

Plan

1 Traitement Automatique des Langues

2 Que fait on ?

3 Que croit-on que nous faisons ?

4 Conclusion

Conclusion

l'informaticien, qu'en fait-on ?

l'informaticien :

- utilise un langage étrange, ce qui le rend difficile à comprendre
 - donc l'informaticien est marginal (syndrome d'Asperger léger dans le meilleur des cas)
- est capable d'analyser toutes les données en langue naturelle sans difficultés et en temps réel
 - ce qui est loin d'être le cas, comme nous l'avons évoqué dans la première partie
- peut produire une Intelligence Artificielle (au sens de Turing) parfaite.

Conclusion

Le TAL, que peut-on en faire ?



Conclusion

Le TAL, que peut-on en faire ?

- des mobiles qui donnent la météo après une requête à l'oral
- des textes courts résumants un ensemble de données numériques
- des systèmes qui analysent les échanges médicaux pour suivre les patients
- des outils permettant de rendre accessible la connaissance
- ...

Conclusion

Le TAL, que peut-on en faire ?

- des mobiles qui donnent la météo après une requête à l'oral
- des textes courts résumants un ensemble de données numériques
- des systèmes qui analysent les échanges médicaux pour suivre les patients
- des outils permettant de rendre accessible la connaissance
- ...

reste la question d'incarner ces aspects dans des outils narratifs

Conclusion

- la fiction est très décalée de la réalité sur l'informatique : la complexité des enjeux théoriques et techniques conduisent les auteurs à créer des représentations de l'informaticien
 - à la marge de la société, vivant dans un monde spécifique,
 - tout à fait fantasmagorique sur la réalité de ce qui peut être fait
- les auteurs n'utilisent pas les capacités technologiques actuelles de l'informatique
- dans un contexte complexe :
 - IA et singularité
 - loyauté des algorithmes