

Telecom ParisTech
vendredi 30 janvier 2015

Traitements Automatiques du Langage

Jean-Louis Dessalles

Shrdlu (T. Winograd 1971)

H- Peux-tu empiler les deux cubes rouges et soit un cube vert, soit une pyramide ?
M- **OK**
(pour réussir cette tâche, la machine commence par ôter la pyramide qui est posée sur l'un des deux cubes rouges).
H- Y a-t-il un gros cube derrière une pyramide ?
M- **Oui, il y en a trois. Un grand rouge, un grand cube vert, et le bleu.**
H- Mets-en un petit sur le cube vert qui porte la pyramide.
M- **OK**
H- Mets la plus petite pyramide dessus
M- **OK**
...
H- Pourquoi as-tu libéré le cube rouge ?
M- **Pour le mettre sur le grand cube vert.**
H- Pourquoi as-tu fait cela ?
M- **Pour empiler le cube rouge, un grand cube rouge, et un grand cube vert.**
H- Pourquoi as-tu fait cela ?
M- **Parce que vous me l'avez demandé.**

Traduction Exemple

La Cigale, ayant chanté
Tout l'été
A été grandement défaut
Lorsque le vent du nord est venu
Pas un seul petit morceau
Fly ou un ver .
Elle alla crier
À la Fourmi sa voisine ,
La priant de lui prêter
Certains grains à
Jusqu'à ce que la nouvelle saison .
" Je vous paierai , dit-elle ,
Avant Août , la foi des animaux
Intérêt et principal . "
Ant ne est pas un prêteur :
Ce est là son moindre défaut .
Que faisiez-vous chaud ?
Elle a dit à cette emprunteur .
- Nuit et jour à tout venant
Je ai chanté , je vous prie .
- Vous chantie ? Je suis très à l'aise .
Eh bien ! danser maintenant .

The Cicada , having sung
All summer
Was greatly lacking
When the north wind came
Not one little piece
Fly or worm .
She went begging
At her neighbor the Ant ,
Beggng her to lend him
Some grains to
Until the new season .
" I'll pay you , she said ,
Before August , animal faith
Interest and principal . "
Ant is not a lender :
That is its least defect .
What were you hot ?
She said to this borrower . -
Night and day to all comers
I sang , if you please . -
You were singing ?
I am very comfortable .
Well ! dance now .

Une petite grammaire

grammaire

Test

s
| _gn
| | _gn
| | | _det : la
| | | _n : fille
| | | _gnp
| | | _pp : de
| | | _gn
| | | _det : la
| | | _n : voisine
| _gv
| _v : agace
| _gn
| | _det : la
| | | _n : tante
| | | _gnp
| | | _pp : de
| | | _gn
| | | _det : sa
| | | _n : bru

la fille de la voisine agace la tante de sa bru

APIL
Association des Professionnels
des Industries de la Langue
<http://www.apil.asso.fr>

>>Conseil
>>Aide à la rédaction
>>Veille
>>CRM (customer Relationship Management)
>>KM (Knowledge Management)
>>Apprentissage des langues
>>Dictionnaire électronique
>>Recherche en langage naturel et fédération d'information
>>Gestion électronique de documents et de l'information
>>Reconnaissance et synthèse vocale
>>Traduction automatique
>>Reconnaissance de caractères
>>Text-Mining
>>Ontologies et web sémantique

APIL
Association des Professionnels
des Industries de la Langue
<http://www.apil.asso.fr>

* 4DConcept	* DIAGONAL	* NEURO-CONCEPT
* ABILDOC	* DIALOCA	* NEUROSOFT
* ACETIC	* DIGIMIND	* NORMATICS
* ALBERT	* ELAN Informatique	* NOMEN
* ALMAS	* EURASAM	* Nominio Technologies Inc
* INGENIERIE	* EURODOC	* Organic Information Solutions
* ALGOC	* SOFLOG	* PERTINENCE
* ANGLO-INSTITUTE	* EVER TEAM	* MINING Sarl
* APPLIED	* GBConcept	* PROLOGIA
* SEMANTICS	* GLOBALINK	* PROTEOR
* ARISEM	* HOLISTIQUE	* RESUMIX
* ASCOLA	* COMMUNICATION	* QWAM SYSTEM
* ATLANTIC	* S	* SAIL LABS
* INTELLIGENCE	* IBM France	* SCHLUMBER-GERSEMA
* ATOMZ	* INFOSTICS	* SEMANTIA
* AURALOG	* IPHASE	* SEMIOSYS
* AUTONOMY	* ISCOPE	* SINEQUA
* BABELING	* KNOWINGS	* SOFTISSIMO
* BEA Conseil	* LINGUATEC	* SYNAPSE
* BILINGUA	* LINGWAY	* SYSTRAN
* Ingénierie	* MACHINA	* TAGMATICA
* Linguistique	* SAPIENS	* TELISMA
* BOWNE GLOBAL	* MEDIACONCEPT	* TEMIS
* SOLUTIONS	* MEMODATA	* TERAGRAM
* CAP GEMINI	* MINDMAKER Inc	* TRIPLEHOP
* INNOVATION	* MONDECA	* VERITY FRANCE
* CONNEXOR	* MONDOSOFT	
* CONVERA	* MYSOFT	
* CRIL	* Multilingual	
* TECHNOLOGY	* Computing, Inc.	
* DATALING	* NAMING	
* DELPHES	* COMPANY LTD	
* TECHNOLOGIES	* NEMESIA	
* DEMONIAK		

Exemples d'applications du TALN

- la correction orthographe
- la traduction automatique
- Le résumé
- la phonétisation
- le traitement linguistique
- l'aide contextuelle
- Interaction humaine
- l'approche 'conos'
- Fouille de données
- la recherche d'informations textuelles et sur Internet
- ...



Natural Language Processing(NLP): Market Size, 2013 - 2018 (\$Million), Market Growth, 2013 - 2018 (Y-o-Y %)



Les applications du TALN dans l'entreprise

Marché

- Portails et moteurs de recherche
- Analyse de logs
- Réponse automatique
- Gestion de campagne marketing
- m-commerce (mobile-commerce)
- Fouille de textes
- Classification automatique
- Web sémantique

Environnement

- Intelligence économique
- Veille
 - Concurrentielle
 - Technologique
 - de notoriété
 - Logistique

Gestion de la relation clients

- e-commerce
- Support en ligne
- Réponse automatique
- Traitement des courriels
- Traduction automatique
- Agents conversationnels

Ressources internes

- Travail collaboratif
- e-learning
- Gestion du capital intellectuel
- Documentation
- Gestion de l'innovation
- Recherche d'experts
- Gestion des meilleures pratiques
- Gestion de CV

La complexité du langage

Kivunjo (langue bantoue) : **Nāikim' lyiīā**
 "He eats it for her"

- N- marqueur de focus
- ā- accord de classe avec le sujet. Ici: humain singulier
- i- temps présent
- kī- accord avec l'objet, classe animaux (classe 7)
- m'- marqueur indiquant que le bénéficiaire de l'action est de la classe 1
- lyi- verbe manger
- i- modification du verbe pour indiquer la présence d'un rôle de bénéficiaire
- ā- mode indicatif (vs subjonctif)

(16 classes : humain singulier, humain pluriel, objets fins ou étendus, objets en paquets ou par paires, paires ou paquets, instruments, animaux, parties du corps, diminutifs, qualités abstraites, localisations précises, localisations vagues).
 (temps : aujourd'hui, plus tôt aujourd'hui, hier, pas plus tôt qu'hier, hier ou avant, dans le passé lointain, habituellement, en train, consécutivement, hypothétique, futur, temps indéterminé, pas encore, parfois).

La complexité du langage

récurSION

J'espère que tu t'es rendu compte que ce que Marie a dit à propos de tes sentiments à son égard ne colle pas du tout avec ce que tu m'avais raconté de l'évolution de vos relations.

B. Victorri

aspect

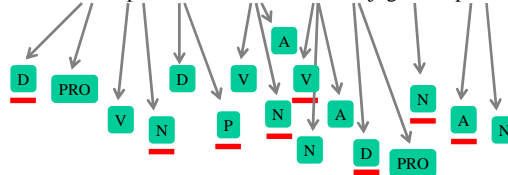
Tout en continuant à lire son journal, il prit un yaourt qu'il avait acheté plusieurs semaines auparavant et dont il allait regretter le soir même de ne pas avoir vérifié la date de péremption.

modalité

Normalement, Marie devrait participer au colloque à moins que Pierre ne décide au dernier moment d'y venir lui aussi, auquel cas je ne serais pas surpris qu'elle se désiste sous un prétexte quelconque si elle l'apprend à temps pour pouvoir annuler son voyage.

Ambiguïté

Le départ de l'abonné invalide le jugement pénal.



Ambiguïté

*Je n'ai pas bien vu les tableaux qui se trouvaient
dans la pièce du fond*

- *voir*: vision, compréhension, fréquentation, étude, constatation, passivation, etc.
- *tableau*: peinture, panneau mural, scène, t. de bord, t. noir, etc.
- *bien*: degré, assertion, jugement moral, concession, etc.
- *les*: générique ou spécifique
- *dans*: spatial, temporel, notionnel, approximation, etc.
- [*passé*]: accompli, inaccompli, irréel (une minute après, il était mort), futur dans le passé
- *qui*: descriptif ou restrictif
- [*sujet-verbe*]: agent, instrument, thème, etc.
- *du*: possessif, agent, objectif, etc.

ambiguïtés

www.dessalles.fr

dias

13

Niveaux de description du langage

Niveau	Contrexemple	Exemple
Phonologie	tparfch - aml - lch habe - paeuté	ramile - talifot - rostame
Lexique	dastimer	parodontose
Morphologie	ablemange - jetépré	mangement - rangeaison
Syntaxe	le hier beau princesse mangera	le jardin de la porte mange le ciel
Sémantique	la pomme va lire un garçon	ma cousine vient d'acheter un vélo
Pertinence	Hier, j'ai perdu une pièce de 20cts	Hier, j'ai perdu mon portefeuille
Sociologie	<i>faute de politesse, de tact, ...</i>	
...		

www.dessalles.fr

dias

14

Morphologie

<i>mot</i>	<i>Signification</i>	<i>âge</i>
enrailer	contraire de dérailier	3
senter	sentir	3
treindre	tenir	3
déprocher	éloigner	4
serrer	servir	4
écriture	texte écrit	4
explications	explications	5
habillage	acte de s'habiller	5
se dépyjamer	enlever son pyjama	5
levure	niveau de gonflage d'un matelas	5
poissonneur	poissonnier	5
senture	odeur	5
affaciler	faciliter	6
découder	découdre	6
détransformer	redonner la forme initiale	6
rallonguir	rallonger	6

<i>mot</i>	<i>Signification</i>	<i>âge</i>
recouer	recoudre	6
recouser	recoudre	6
se rejoindre	se rejoindre	6
sêcherie	sêcheresse	6
abrètemnt	abri	7
chantement	acte de chanter	7
décorages	décorations	7
jolissies	choses jolies	7
jupiersien	sur le modèle: martien	7
couratation	courant (rivière)	8
eruptioinaient	avaient une éruption	8
inapercevable	ne peut être aperçu	8
inoblîgé	non obligatoire	8
proposément	proposition	9
fortifige	fortification	9
interruption	interruption	9
menteries	mensonges	9
virager	virer	9

www.dessalles.fr

dias

15

Contraintes sur la structure

- Le fait qu'il dise que la personne à laquelle le frère de Jean a vendu sa voiture était un escroc n'a pas jeté le discrédit sur Jacques.



- Jean
- frère de Jean
- acheteur
- Jacques
- *I/ ?*

www.dessalles.fr

dias

16

Syntaxe

« Il pense que le frère de Jean est sourd »

[présent] → *penser* → *que* → *est* → *sourd*
 ↑
Il
 SD ↑
le → *frère* → *de* → *Jean*
 SD

« Le fait qu'il soit malade ennuie Jean »

[présent] → ennuyer → Jean

SD ↑ SN SC SI SA
le → fait → que → soit → malade
↑
il

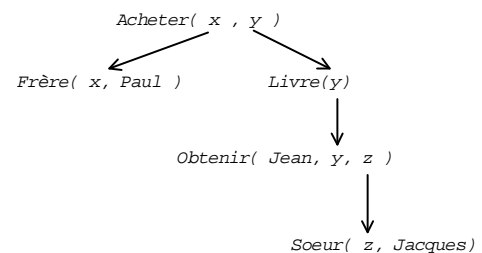
www.dessalles.fr

dias

17

Sémantique

"Le frère de Paul achète le livre que Jean a obtenu de la sœur de Jacques"



www.dessalles.fr

dias

18