



HAL
open science

Éthique et traitement automatique des langues

Karën Fort, Maxime Amblard

► **To cite this version:**

Karën Fort, Maxime Amblard. Éthique et traitement automatique des langues. Journée éthique et intelligence artificielle, Jul 2018, Nancy, France. <hal-01827579>

HAL Id: hal-01827579

<https://hal.science/hal-01827579v1>

Submitted on 2 Jul 2018

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.



HAL Authorization

Éthique et traitement automatique des langues

Karën Fort et Maxime Amblard

karen.fort@paris-sorbonne.fr / maxime.amblard@loria.fr

2 juillet 2018



Une thématique en éclosion dans un domaine "ancien"

Le traitement automatique des langues en bref

L'éthique dans le TAL

Éthiques et TAL : quelle approche ?

Conclusion et perspectives

Une thématique en éclosion dans un domaine "ancien"

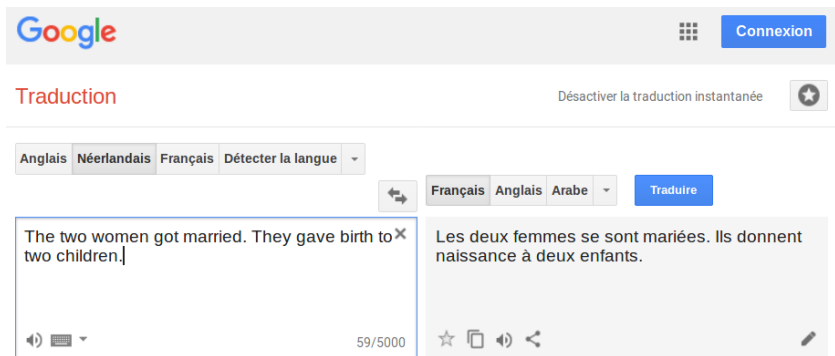
Le traitement automatique des langues en bref

L'éthique dans le TAL

Éthiques et TAL : quelle approche ?

Conclusion et perspectives

Des applications (et des problèmes) dans notre quotidien...



The screenshot shows the Google Translate web interface. At the top left is the Google logo. To the right is a grid icon and a blue button labeled "Connexion". Below this is the word "Traduction" in red, followed by the text "Désactiver la traduction instantanée" and a star icon in a grey box. The main interface has two rows of language selection buttons: "Anglais", "Néerlandais", "Français", and "Détecter la langue" (with a dropdown arrow). Below these is a double-headed arrow icon. The second row has "Français", "Anglais", and "Arabe" (with a dropdown arrow), followed by a blue button labeled "Traduire". The input area on the left contains the text "The two women got married. They gave birth to two children." with a close icon (X) at the end. Below the input is a speaker icon and a dropdown arrow, and the character count "59/5000". The output area on the right contains the French translation "Les deux femmes se sont mariées. Ils donnent naissance à deux enfants." Below the output are icons for a star, a copy icon, a speaker icon, and a share icon, along with an edit icon (pencil) on the far right.

Google Translate

... enfants de la guerre (froide)



Expérience Georgetown-IBM (janvier 1954)

Dan - Flickr : IBM 701 / CC BY-SA 2.0

Les 2 "révolutions" du TAL par l'exemple

et la multiplication des questions éthiques

Années 90 : apprentissage (statistique)

- ▶ traçabilité
- ▶ reconnaissance du travail des producteurs [Kenny, 2011]
- ▶ biais dans l'évaluation

Depuis quelques années : apprentissage (neuronal)

- ▶ traçabilité
- ▶ reconnaissance du travail des producteurs [Kenny, 2011]
- ▶ biais dans l'évaluation
- ▶ interprétabilité

À noter qu'en TAL les résultats des méthodes neuronales sont loin d'être aussi impressionnants que dans d'autres domaines

L'ogre a faim !

pour le nourrir, nous devenons des employeurs

Nécessité de **grandes** masses de données **annotées** pour

entraîner

et

évaluer les systèmes

Exemple :

→ 100 000 mots annotés pour entraîner un tagger :

Il/CLS est/V sain/ADJ et/CC sauf/P ./PONCT

⇒ besoin d'**annotateurs humains**, ce qui coûte cher (600 000 \$ pour le Prague Treebank)

Au final : des analyses plus ou moins maîtrisées

Environ 98 % d'exactitude pour l'analyse morphosyntaxique [Denis and Sagot, 2010]. . . c'est-à-dire **1 erreur toutes les 5 phrases** environ :

Il/CLS est/V sain/ADJ et/CC sauf/P ./PONCT

Entre 69 et 88 % d'exactitude pour la détection d'ironie [Karoui et al., 2015]

#Hollande est vraiment un bon diplomate #Algérie.

#Valls a appris la mise sur écoute de #Sarkozy en lisant le journal. Heureusement qu'il n'est pas Ministre de l'Intérieur.

95 % de F-mesure pour la désidentification (anonymisation) et une tâche par définition irréalisable :

mon père a fondé un le plus grand cabinet d'ophtalmologiste de la ville (Orléans, corpus ESLO)

Une thématique en éclosion dans un domaine "ancien"

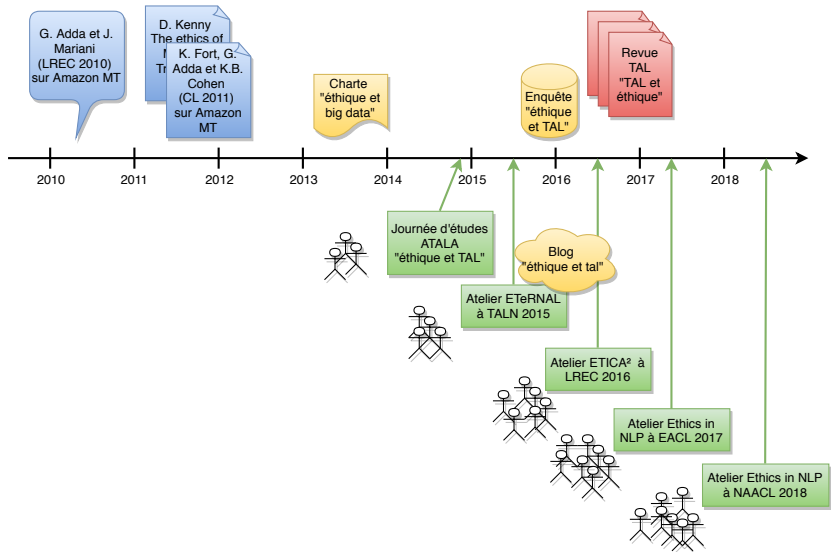
Le traitement automatique des langues en bref

L'éthique dans le TAL

Éthiques et TAL : quelle approche ?

Conclusion et perspectives

Une éclosion récente, des acteur·trice·s varié·e·s



Groupe éthique et TAL

Blog <http://www.ethique-et-tal.org/>

IR G. Adda (LIMSI-CNRS, Paris-Saclay)

MCF M. Amblard (LORIA / Université de Lorraine, Nancy)

PR J-Y. Antoine (LI / Université Rabelais, Tours)

Indus. A. Couillault (Apoliade, Paris)

MCF K. Fort (Sorbonne Univ., Paris)

Indus. H. de Mazancourt (Yseop, Paris)

CR A. Névéol (LIMSI-CNRS, Paris-Saclay)

Une thématique en éclosion dans un domaine "ancien"

Éthiques et TAL : quelle approche ?

Une vision souvent conséquentialiste

Les apports de l'éthique déontologique

Conclusion et perspectives

Une thématique en éclosion dans un domaine "ancien"

Éthiques et TAL : quelle approche ?

Une vision souvent conséquentialiste

Les apports de l'éthique déontologique

Conclusion et perspectives

Les apports du conséquentialisme



























Réflexion sur les applications développées [Lefeuvre et al., 2015] :

- ▶ qui représentent un risque concernant la vie privée des citoyens :
 - ▶ extraction automatique de « sentiments »,
 - ▶ identification de « leaders d'opinion », etc.
- ▶ dont les effets négatifs ne sont pas toujours évalués sur le temps long :
 - ▶ aide au handicap

Le temps long : l'exemple de l'aide au handicap

grâce à Sibylle, je peux communiquer avec mon père et ma

mots	picots	l	j	t	d	p	a	<	
e	c	s	m	r	i	f	u	q	
n	à	o	é	v	b	g	h	y	
ê	k	w	z	ù	x	ç	_	è	
,	.	-	?	!	.	@	;		

mère					
filie					
famille					
grand-mère					
soeur					
vie					
tante					

- Pathologie lourde avec perte de parole
- Clavier virtuel avec prédiction lexicale



**Vitesse
de
saisie**



**Maîtrise
de la
langue**



[Antoine and Lefevre, 2014]

Une thématique en éclosion dans un domaine "ancien"

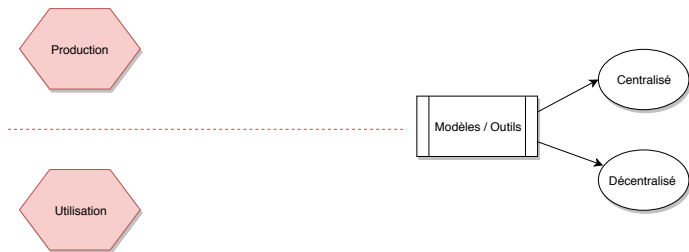
Éthiques et TAL : quelle approche ?

Une vision souvent conséquentialiste

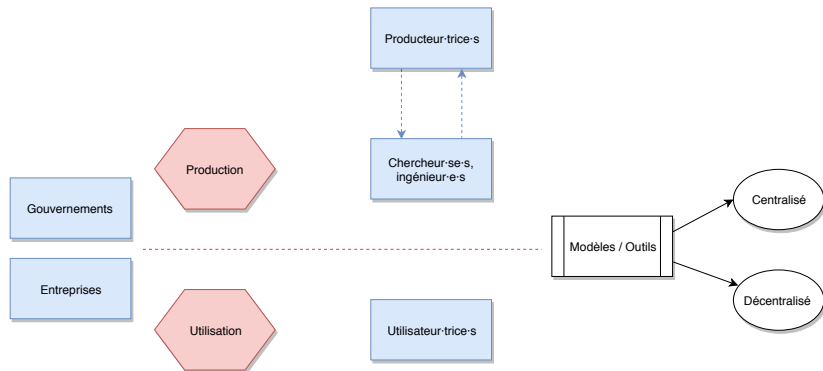
Les apports de l'éthique déontologique

Conclusion et perspectives

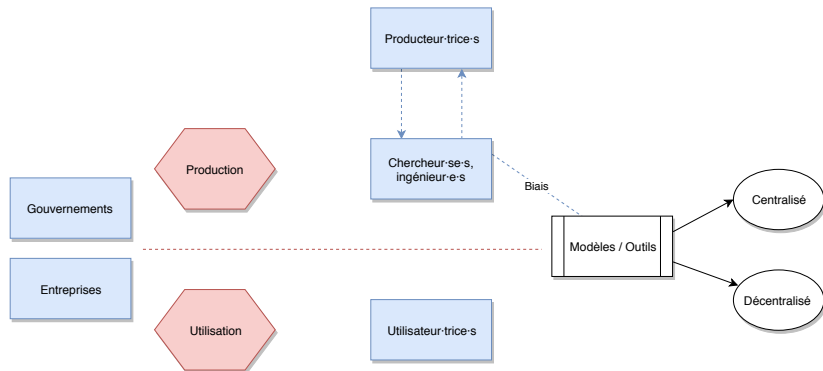
Environnement / Acteur·trice·s / Données



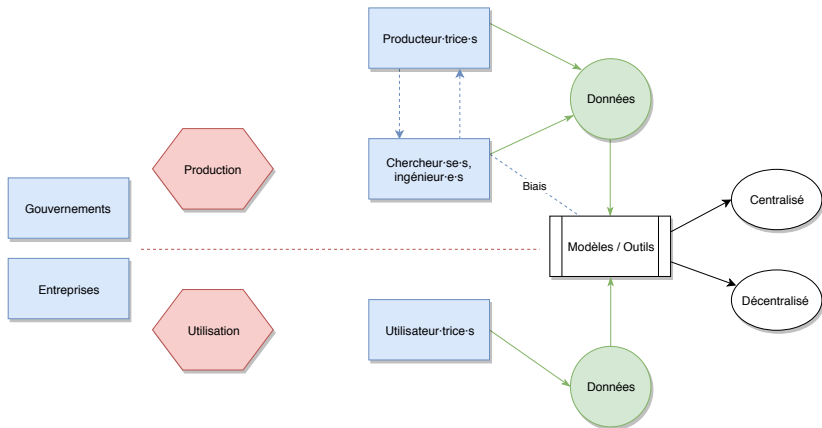
Environnement / Acteur·trice·s / Données



Environnement / Acteur·trice·s / Données



Environnement / Acteur-trice-s / Données



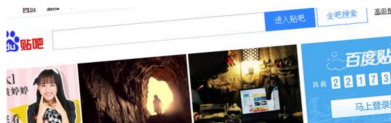
Nouveau monde. Baidu annonce pouvoir cloner une voix très facilement

L'entreprise chinoise Baidu (le Google chinois) a mis au point une intelligence artificielle capable de cloner une voix humaine très rapidement.



Mis à jour le 13/03/2018 | 18:23
publié le 13/03/2018 | 18:23

6
PARTAGES



S'ABONNER



Podcast via iTunes

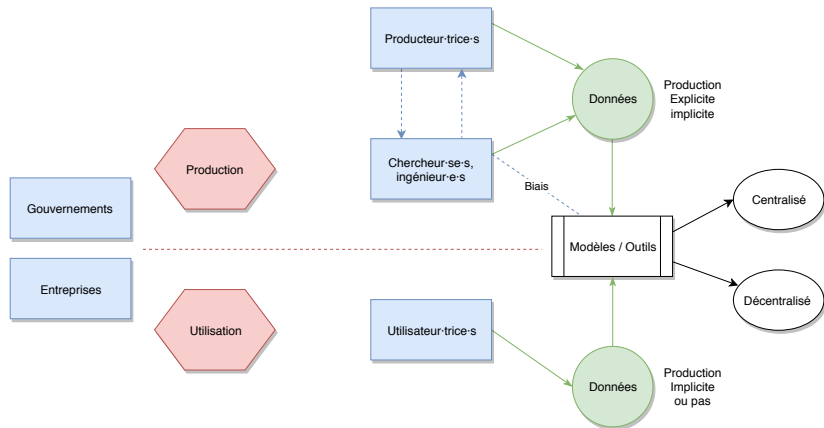


Podcast via RSS

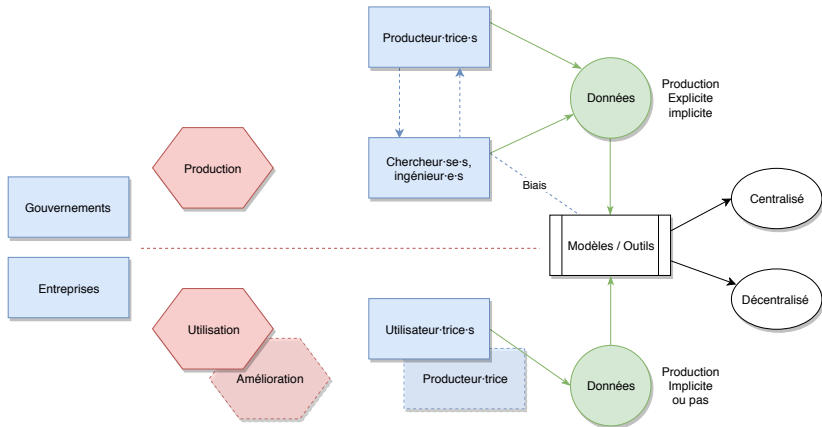


France TV Info

Type de création des données



⇒ La frontière entre production et utilisation s'atténue

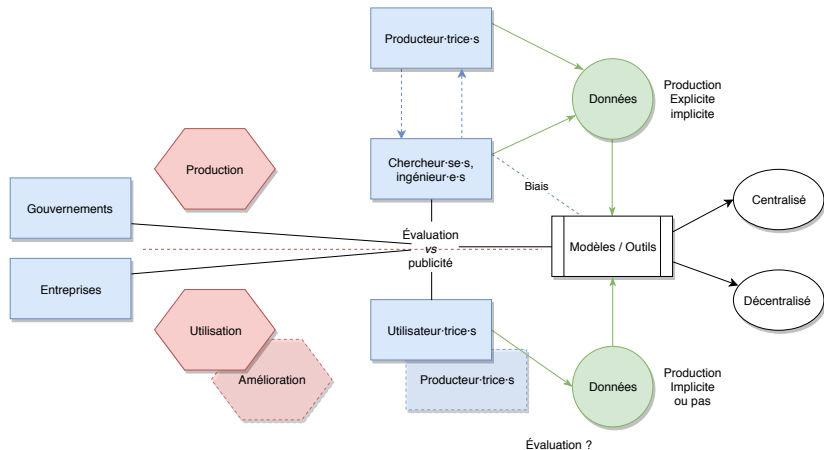


Intelligence artificielle : un robot renvoyé pour inefficacité

À Édimbourg, en Écosse, le premier robot salarié a été renvoyé car il n'était pas assez performant.

France TV Info

Évaluation



En Chine, les personnes qui ont une mauvaise "note sociale" pourraient être empêchées de prendre l'avion ou le train

Le Parti communiste chinois teste depuis plusieurs années un système de note sociale. A partir du 1er mai, il sera utilisé dans les trains et les avions. Les "mauvais élèves" interdits à bord seraient par exemple des personnes ayant utilisé des billets périmés ou ayant fumé dans un train.



Pouvoir sur les données : consentement éclairé (?)



[Utilisation de nos Services](#)

[Votre compte Google](#)

[Protection de la vie privée et des droits d'auteur](#)

[Vos contenus et nos Services](#)

[À propos des logiciels utilisés par ou présents dans nos Services](#)

[Modification et résiliation de nos Services](#)

[Garanties et clauses de non-responsabilité](#)

[Responsabilité pour nos Services](#)

[Utilisation de nos Services par une entreprise](#)

[À propos de ces Conditions d'Utilisation](#)

[Mises à jour](#)

Modification et résiliation de nos Services

Google n'a cessé de modifier et d'améliorer ses Services. Nous sommes donc susceptibles d'ajouter ou de supprimer des fonctionnalités ou des fonctions, et il peut également arriver que nous suspendions ou interrompions complètement un Service.

Vous pouvez cesser d'utiliser nos Services à tout moment. Nous espérons cependant que vous continuerez de les utiliser. Google est en droit de cesser de vous fournir tout ou partie des Services, ou d'ajouter ou de créer de nouvelles limites à l'utilisation des Services et ce, à tout moment.

Pour nous, vous restez propriétaire des données que vous nous confiez et nous pensons qu'il est important que vous puissiez y accéder. Si nous devons interrompre un Service, dans la mesure du possible, nous vous en avisons dans un délai raisonnable et vous donnons la possibilité de récupérer des informations de ce Service.

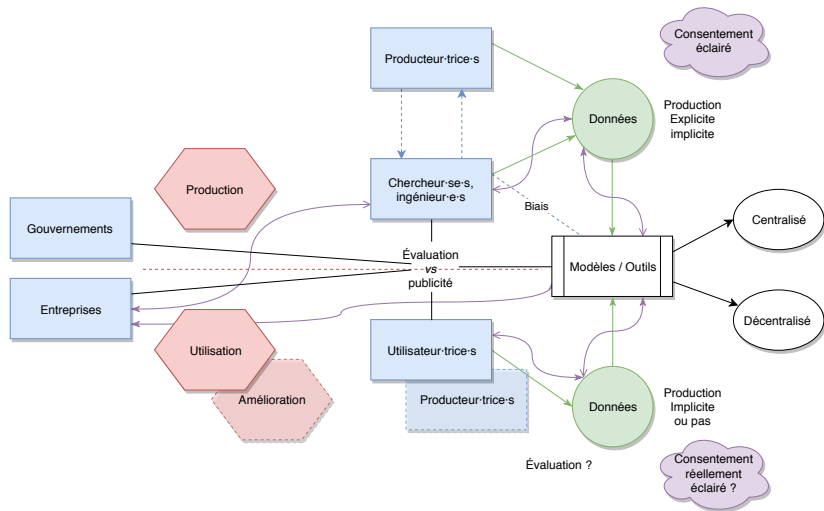
Garanties et clauses de non-responsabilité

Notre offre de Services est soumise à une obligation de moyens, dans les limites de ce qui est commercialement raisonnable. Nous espérons que vous trouverez plaisir à les utiliser. Nos Services font cependant l'objet d'une limitation de garantie.

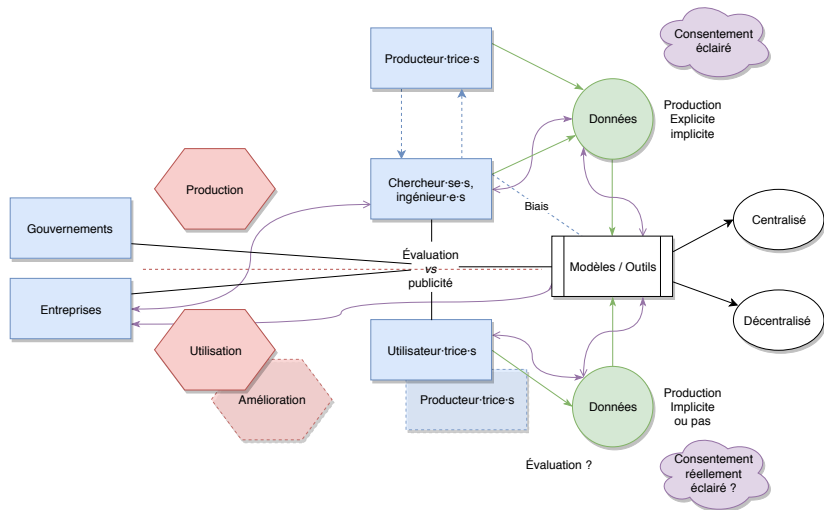
Sauf tel qu'expressément prévu par les présentes Conditions d'Utilisation ou des

[Conditions d'utilisation de Google](#)

Pouvoir sur les données : consentement éclairé (?)



Conclusion schéma 1/2 : traçabilité ?



Une thématique en éclosion dans un domaine "ancien"

Éthiques et TAL : quelle approche ?

Conclusion et perspectives

Un tournant ?

Actuellement :

- moins de soumissions aux ateliers sur le sujet (LREC 2018, NAACL 2018)
- + thématique reprise dans l'appel général de certaines conférences (TALN, LREC, ACL (dans Multidisciplinary))
- + certaines sous-thématiques prennent leur essor :
 - ▶ répliquabilité et reproductibilité (COLING 2018, LREC 2018)
 - ▶ biais de l'apprentissage (ACL 2018)

Dépasser les totems et les tabous

Totems :

- ▶ reproductibilité
- ▶ biais de l'apprentissage (en particulier genré)
- ▶ transparence des algorithmes

Tabous :

- ▶ traçabilité
- ▶ conditions de productions des ressources langagières (chercheur·se employeur·e)

▶ **déontologie**



En tant que chercheur·se·s

Difficultés :

- ▶ poser l'éthique comme un enjeu de recherche dans le domaine
- ▶ définir/circonscrire les problèmes

Besoins :

- ▶ collaborations (avec des philosophes, par exemple)
- ▶ déculpabiliser les chercheur·se·s

Merci

pour votre temps

et votre attention

Annexes

Bibliographie



Adda, G. and Mariani, J. (2010).

Language resources and amazon mechanical turk : legal, ethical and other issues.

In Legal Issues for Sharing Language Resources workshop in International Conference on Language Resources and Evaluation (LREC). European Language Resources Association (ELRA).



Antoine, J.-Y. and Lefevre, A. (2014).

Pour une réflexion éthique sur les conséquences de l'usage des ntic : le cas des aides techniques (à composante langagière ou non) aux personnes handicapées.

In Actes de la journée ATALA Éthique et TAL.



Denis, P. and Sagot, B. (2010).

Exploitation d'une ressource lexicale pour la construction d'un étiqueteur morphosyntaxique état-de-l'art du français.

In

Traitement Automatique des Langues Naturelles : TALN 2010,
Montréal, Canada.



Fort, K., Adda, G., and Cohen, K. B. (2011).
Amazon Mechanical Turk : Gold mine or coal mine?
Computational Linguistics (editorial), 37(2) :413–420.



Karoui, J., Benamara Zitoune, F., Moriceau, V.,
Aussenac-Gilles, N., and Hadrich Belguith, L. (2015).
Détection automatique de l'ironie dans les tweets en français.

In

22eme Conference sur le Traitement Automatique des Langues Naturelles
pages pp. 1–6, Caen, France.



Kenny, D. (2011).
The ethics of machine translation.
In New Zealand Society of Translators and Interpreters Annual
Conference 2011, Auckland, New Zealand.



Lefevre, A., Antoine, J.-Y., and Allegre, W. (2015).

Ethique conséquentialiste et traitement automatique des langues : une typologie de facteurs de risques adaptée aux technologies langagières.

In

Atelier Ethique et TRaitement Automatique des Langues (ETeRNAL
Actes de la 1e Ethique et TRaitement Automatique des
Langues (ETeRNAL'2015), Caen (France), pages 53–66, Caen,
France.