Traitement de la traduction des bigrammes et trigrammes nominaux par deux programmes de TA en ligne de l'anglais vers le français dans le domaine de la médecine nucléaire

François MANIEZ (francois.maniez@univ-lyon2.fr)

université lumière lyon 2

CeRLA
Centre de recherche en linguistique appliquée

1. Introduction
2. Corpus et méthode
3. GN binominaux (N1 N2)
4. GN trinominaux (N1 N2 N3)
5. Conclusion

1. Introduction

- prémodification nominale N1 N2 en anglais → complémentation dans les langues romanes
- développement des éléments prémodificateurs sous la forme d'un complément de nom :
  - pocket money → argent de poche

1. Introduction

- prémodification nominale N1 N2 en anglais → complémentation dans les langues romanes
- développement des éléments prémodificateurs sous la forme d'un complément de nom :
  - pocket money → argent de poche
- anglais : les deux formes (*packed/unpacked*) sont souvent possibles (Gledhill & Pecman 2018)
  - aspirin synthesis /synthesis of aspirin
  - cell proliferation / proliferation of cells

Variation de la préposition

- pocket money → argent **de** poche
- water pistol → pistolet **à** eau
- bank account → compte **en** banque
- wheel drive → entraînement **par** roue
- mound culture → culture **sur** buttes
- greenhouse cultivation → culture **sous/en** serre

- A **High Court** legal **challenge** against Brexit [...] →
  Une contestation judiciaire de la Haute Cour contre le Brexit (GT 2019)
  Une contestation judiciaire par la Haute Cour contre le Brexit (GT 2021)
- Un recours en justice devant la Haute Cour contre Brexit (DL 2019)

Variation de la détermination et du nombre

- horse race → course de **chevaux**,
  horse manure → crottin de **cheval**
- tumor growth →
  croissance [d'une tumeur / de la tumeur / des tumeurs]
- adjectif relationnel : croissance tumorale

TA des GN longs de la langue scientifique

- *tumor growth* (GS>2 000 000)

TA des GN longs de la langue scientifique

- *tumor growth* (GS>2 000 000)
- *tumor growth inhibition* (GS> 65 000)

TA des GN longs de la langue scientifique

- *tumor growth* (GS>2 000 000)
- *tumor growth inhibition* (GS> 65 000)
- *tumor growth inhibition effect* (GS >800)

TA des GN longs de la langue scientifique

- *tumor growth* (GS>2 000 000)
- *tumor growth inhibition* (GS> 65 000)
- *tumor growth inhibition effect* (GS >800)

→ effet d'inhibition de la croissance tumorale (DL/GT) **GS: 3**

TA des GN longs de la langue scientifique

- *tumor growth* (GS>2 000 000)
- *tumor growth inhibition* (GS> 65 000)
- *tumor growth inhibition effect* (GS >800)

→ effet d'inhibition de la croissance tumorale (DL/GT)

- effet inhibiteur de la croissance tumorale (GS : 10)

TA des GN longs de la langue scientifique

- *tumor growth* (GS>2 000 000)
- *tumor growth inhibition* (GS> 65 000)
- *tumor growth inhibition effect* (GS >800)

→ effet d'inhibition de la croissance tumorale (DL/GT)

- effet inhibiteur de la croissance tumorale (GS : 10)
- inhibition de la croissance tumorale (GS : 495)
- inhibiteur de la croissance tumorale (GS : 37/73)

TA des GN longs de la langue scientifique

- *tumor growth* (GS>2 000 000)
- *tumor growth inhibition* (GS> 65 000)
- *tumor growth inhibition effect* (GS >800)

→ effet d'inhibition de la croissance tumorale (DL/GT)

En allemand, regroupement en deux bigrammes :

- *Hemmwirkung auf das Tumorwachstum* (DL)
- *Tumorwachstum hemmende Wirkung* (DL)
- *Hemmwirkung des Tumorwachstums* (GT)

- Dans la prose scientifique, bon nombre de séquences binominales ou trinominales sont des termes dont la traduction varie peu, voire pas du tout.

- Dans la prose scientifique, bon nombre de séquences binominales ou trinominales sont des termes dont la traduction varie peu, voire pas du tout.
- Mais quand ces séquences ne sont pas lexicalisées, la variation de l’usage, logiquement reflétée par les mémoires de traduction, l’est également dans les résultats de la TA.

Questions

- A partir de quel degré de complexité des GN la TA devient-elle moins fiable ?
- L’un des deux logiciels (DL / GT) donne-t-il de meilleurs résultats ?

2. Corpus et méthode

- Corpus d'articles de la revue *Nuclear Medicine* (2007-2013, 4,5 M de mots)
- Étiquetage en PDD
- Extraction des GN de type NN et NNN
- Traduction des phrases contenant les GN les plus fréquents par DL et GT
- Les modèles de la TA traitent généralement les phrases de manière isolée (Tiedemann et Scherrer 2017) → phrases isolées de leur contexte.

- Comparaison des résultats obtenus à l'usage observé dans un corpus d'articles de la revue *Médecine Nucléaire* (2007-2012, 2M de mots), puis sur le moteur de recherche Google Scholar.
- https://perso.univ-lyon2.fr/~maniezf/Corpus/Corpus_Comparable_MedNucl_CRTT.htm

(Requête : maniezf corpus médecine nucléaire)

Fréquence des bigrammes et trigrammes nominaux dans le corpus Nuclear Medicine

Bigrammes		Trigrammes	
112	room temperature	77	positron emission tomography
84	dose rate	21	metal ion impurities
80	positron emission	20	dose rate patterns
49	reaction mixture	13	bone pain palliation
47	radionuclide therapy	10	silica gel column
47	tumor uptake	9	blood glucose levels
45	flow rate	9	serum glucose levels
37	molecular imaging	8	receptor imaging agents
35	biodistribution studies	8	plasma protein binding
28	residence time	7	rat brain slices

3. GN binominaux

- Très grande régularité de la traduction des termes :
  - **Complémentation** (positron emission, quality control, breast cancer)
  - **Adjectif dénominal/relationnel** (cell death, body weight, bone marrow)

Collocations

- brain regions (régions/zones cérébrales/du cerveau)
- tissue distribution (distribution tissulaire/dans les tissus)
- tumor uptake
- tracer uptake

Fréquence des ET potentiels de *tumor uptake* (GS)

ET (captation)	FRQ	ET (absorption)	FRQ
captation tumorale	131	absorption tumorale	7
captation de la tumeur	16	absorption de la tumeur	2
captation des tumeurs	5	absorption des tumeurs	5
captation de tumeurs	1	absorption de tumeurs	0
captation par la tumeur	6	absorption par la tumeur	4
captation par les tumeurs	10	absorption par les tumeurs	1
TOTAL	**169**	**TOTAL**	**19**
	90%		10%

- Traduction de 20 phrases contenant le GN *tumor uptake*
- GT et DL utilisent très majoritairement (90 et 85%) *absorption tumorale*.

absorption	Frq GS	captation	Frq GS
nucléaire **absorption**	42 200	nucléaire **captation**	16 800
traceurs **absorption**	12 100	traceurs **captation**	4 240
nucléaire traceurs **absorption**	4 970	nucléaire traceurs **captation**	2 090

Fréquence des ET potentiels de *tracer uptake* (GS)

ET (captation)	FRQ	ET (absorption)	FRQ
captation du traceur	100	absorption * du traceur	43
captation * du traceur	62	absorption du traceur	94
captation des traceurs	15	absorption des traceurs	19
captation * des traceurs	17	absorption * des traceurs	15
TOTAL	**194**	**TOTAL**	**171**
	53%		47%

Tracer uptake

- 20 phrases contenant le GN *tracer uptake*
- DL : captation (60%)
- GT : absorption (90%)

4. GN trinominaux

- positron emission tomography
- growth factor receptor
- plasma protein binding
- blood glucose level

Imbrication dans des GN de taille supérieure

- **silica gel column** chromatography → chromatographie sur (colonne de) gel de silice (GT & DL)

Imbrication dans des GN de taille supérieure

- **silica gel column** chromatography → chromatographie sur (colonne de) gel de silice (GT & DL)
- **peptide receptor radionuclide** therapy (PRRT) GS >11 000
  - thérapie radionucléide par récepteurs peptidiques
  - thérapie par radionucléides à récepteur(s) peptidique(s)
  - thérapie par radionucléide des récepteurs peptidiques
  - **radiothérapie de récepteurs peptidiques**

Collocations

- **cell** cycle arrest → arrêt du cycle **cellulaire** (GS 3630)
- **tumor** selectivity index → index de sélectivité **tumorale** (GS 0)

Structure ADJ1 ADJ2 N1 N2 N3 N4

Significant synergistic tumor growth inhibition effect was demonstrated […]	Un effet synergique significatif d'inhibition de la croissance tumorale a été démontré […] (DL&GT)

Structure ADJ N1 N2 N3

[…] to optimize skeletal tumor growth inhibition in vivo.	**[…] pour optimiser l'inhibition de la croissance des tumeurs squelettiques in vivo. (GT)**
	[…] pour optimiser l'inhibition de la **croissance tumorale du squelette** in vivo. (DL)

Structure ADJ1 ADJ2 N1 N2 N3

A novel transforming gene in a **human malignant melanoma cell line**	Un nouveau gène transformant dans une **lignée cellulaire de mélanome malin humain** (GT)	Un nouveau gène transformateur dans une **lignée cellulaire de mélanome malin humain** (DL)

A novel transforming gene in a human malignant melanoma cell line	Un nouveau gène transformant dans une lignée cellulaire de mélanome malin humain (GT)	Un nouveau gène transformateur dans une lignée cellulaire de mélanome malin humain (DL)

EN	GS	FR	GS
melanoma cell line	112 000	lignée cellulaire de/du mélanome	72
malignant melanoma cell line	3 200	lignée cellulaire de/du mélanome humain	33
human melanoma cell line	17 100	lignée cellulaire de/du mélanome malin	2
human malignant melanoma cell line	1 810	lignée cellulaire de mélanome humain malin (du : 1)	0

7. Conclusion

- Les difficultés de la TA en LS sont apparentes dès le niveau des bigrammes s'ils ne sont pas lexicalisés.
- Elles deviennent plus aiguës quand les bigrammes et trigrammes nominaux sont précédés par des adjectifs.
- Il n'y a pas d'avantage déterminant de l'un ou de l'autre des logiciels de TA dans ce type de discours spécialisé, mais DL semble meilleur dans les tests de « rétrotraduction » (*repacking*).

Lowering bone mineral affinity of bisphosphonates as a therapeutic strategy to optimize skeletal tumor growth inhibition in vivo.	**Réduction de l'affinité minérale osseuse des bisphosphonates comme stratégie thérapeutique pour optimiser l'inhibition de la croissance des tumeurs squelettiques in vivo.**	Reduction of the bone mineral affinity of bisphosphonates **as a therapeutic strategy to optimize** the inhibition of the growth of skeletal tumors **in vivo. (GT)**
		Bone mineral affinity reduction of bisphosphonates as a therapeutic strategy to optimize skeletal tumor growth inhibition in vivo. (DL)

Traitement de l'ambiguïté syntaxique générée par les structures coordonnées prémodifiées (ADJ N and N of N)

Repeated [endoscopic examinations and analyses of stools] did not reveal pathologic findings.	Les examens endoscopiques et les analyses répétés des tabourets n'ont pas indiqué des résultats pathologiques. (Systran 2008)
	Les examens endoscopiques répétés et les analyses des selles n'ont pas révélé de signes pathologiques. (GT 2021)
	Des examens endoscopiques répétés et des analyses de selles n'ont pas révélé de résultats pathologiques. (DL 2021)

Références

Gledhill, C., & Pecman, M. (2018). On alternating pre-modified and post-modified nominals such as aspirin synthesis vs. synthesis of aspirin: Rhetorical and cognitive packing in English science writing. *Fachsprache-Journal of Professional and Scientific Communication*, *40*(1+ 2), 24-46.

Maniez, F. (2008). Traduction automatique et ambiguïté syntaxique: le cas de la coordination dans les groupes nominaux complexes en anglais médical, in HEIDEN, Serge et Bénédicte PINCEMIN (dir.) *JADT 2008 : actes des 9es journées internationales d'analyse statistique des données textuelles, Lyon, 12-14 mars 2008,* pp. 765-776. Lyon: Presses universitaires de Lyon.

Maniez, F. (2017). An appraisal of recent breakthroughs in machine translation: the case of past participle-based compound adjectives in ESP. *ASp. la revue du GERAS*, (72), 29-48.

Tiedemann, J., & Scherrer, Y. (2017). Neural machine translation with extended context. In Proceedings of the Third Workshop on Discourse in Machine Translation, DISCOMT'17, pages 82–92, Copenhagen, Denmark. Association for Computational Linguistics.

Voita, E., Sennrich, R., & Titov, I. (2019). When a good translation is wrong in context: Context-aware machine translation improves on deixis, ellipsis, and lexical cohesion. *arXiv preprint arXiv:1905.05979*.