

Pièges méthodologiques des corpus parallèles

Olga Nádvorníková

Prague

olga.nadvornikova@ff.cuni.cz

- [FR] « *Moi, se dit le petit prince, si j'avais cinquante-trois minutes à dépenser, je marcherais tout doucement vers une fontaine ...* » (Antoine de Saint-Exupéry, *Le Petit prince*)
- [ALL] "*Wenn ich dreiundfünfzig Minuten übrig hätte*", sagte der kleine Prinz, "würde ich ganz gemächlich zu einem Brunnen laufen..." (trad. par Leitgeb, Grete; Leitgeb, Josef)
- [ANG] "*As for me,*" said the little prince to himself, "*if I had fifty-three minutes to spend as I liked, I should walk at my leisure toward a spring of fresh water.*" (trad. par Katherine Woods)
- [CS] *Kdybych já měl padesát tři minuty nazbyt, řekl si malý princ, šel bych docela pomaloučku ke studánce...* (trad. par Zdeňka Stavinohová)
- [HIN] "*अगर मेरे पास तिरपन मिनट बिताने को होते*" छोटे राजकुमार ने सोचा, '*तो मैं धीरे- धीरे एक जलाशय की ओर चल पड़ता ... ।*" (trad. par किशोर बलवीर, जगवंश)
- [IT] "*Io*", disse il piccolo principe, "*se avessi cinquantatré minuti da spendere, camminerei adagio adagio verso una fontana...*" (trad. par Nini Bompiani Bregoli)
- [RO] "*Eu, își spuse micul prinț, dacă aș avea de irosit cincizeci și trei de minute, aș porni în liniște spre o fântână.*"

Source : corpus parallèle InterCorp

Pièges

1. Problèmes techniques de l'exploitation de corpus parallèle
 2. Spécificités de la langue de traduction
 - a) Interférences
 - b) « universaux » de la traduction (simplification, explicitation, normalisation)
- ? Translationese
- ? Troisième code

Corpus parallèle InterCorp

- Institut du Corpus national tchèque
(www.korpus.cz, www.korpus.cz/intercorp)
- 39 langues (ar, be, bg, ca, da, de, el, en, es, et, fi, fr, he, hi, hr, hu, is, it, ja, lt, lv, mk, ms, mt, nl, no, pl, pt, ro, ru, sk, sl, sq, sr, sv, tr, uk, vi, **cs**)
- Accès gratuit (enregistrement - <https://www.korpus.cz/toolbar/signup.php>)
- Bibliographie : <http://biblio.korpus.cz/>

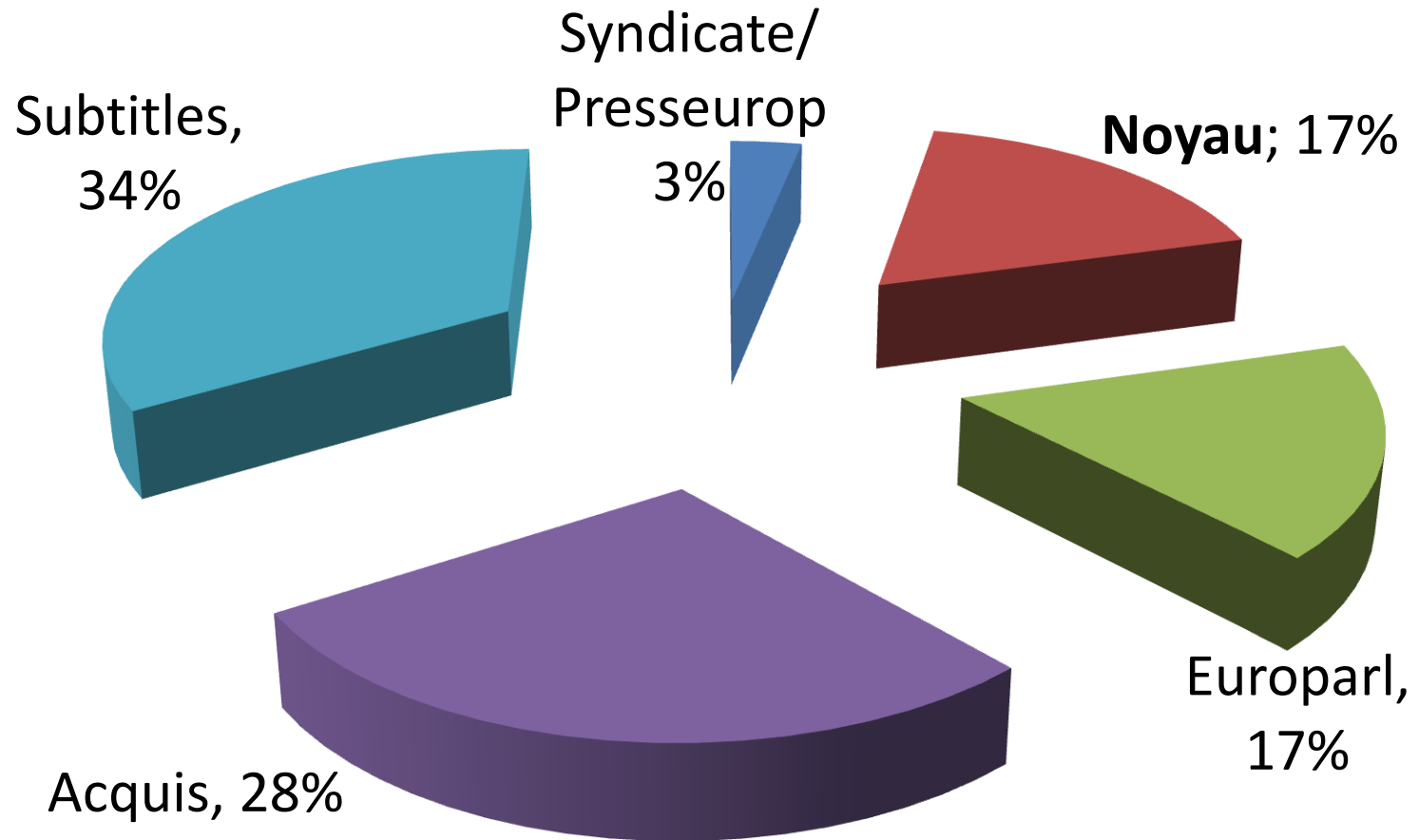
- new query
- word List
- ?
- ve
- ?
- ew
- Custom
- KWIC
- Sentence
- Alignment
- rt
- Custom
- Left | Right
- Node
- References
- Shuffle
- ter
- equency

Hits: 30 i.p.m.: 0.48 (related to the whole corpus) | ARF: 1

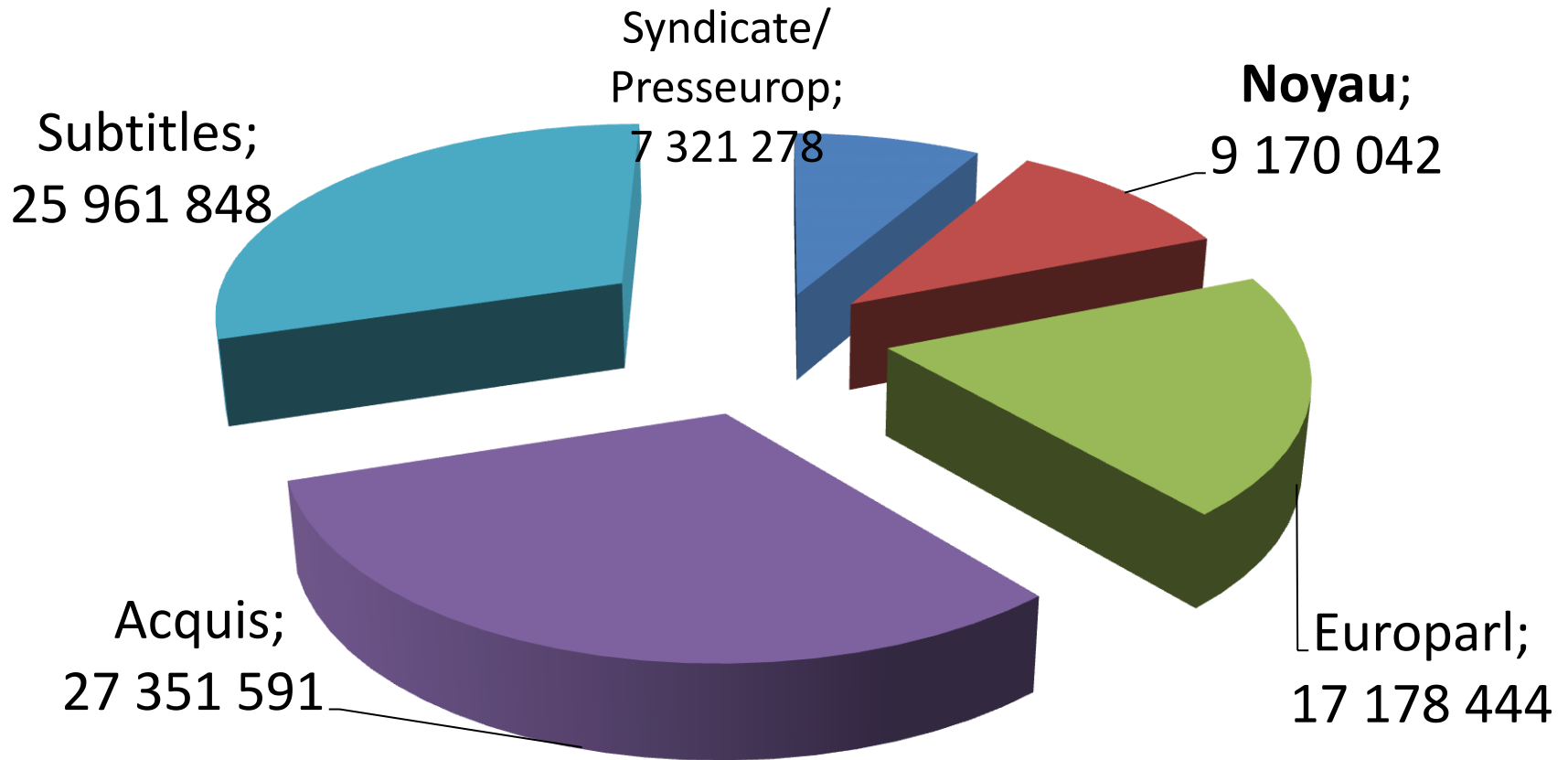
Page of 2 [Next](#) | [Last](#)

	<u>intercorp_fr</u>	<u>intercorp_en</u>
#2407897	<p><p> Ce fut au coin de la rue qu' il remarqua pour la première fois un détail insolite : un chat qui lisait une carte routière .</p>	#9581712 <p><p> It was on the corner of the street that he noticed the first sign of something peculiar -- a cat reading a map .</p>
#2407938	Il y avait bien un chat tigré , assis au coin de Privet Drive , mais pas la moindre trace de carte routière .	#9581760 <p>There was a tabby cat standing on the corner of Privet Drive , but there was n't a map in sight .</p>
#2407994	Mr Dursley cligna des yeux et regarda fixement le chat .	#9581800 <p>Mr. Dursley blinked and stared at the cat . It stared back .</p>
#2408024	Tandis qu' il tournait le coin de la rue et s' engageait sur la route , Mr Dursley continua d' observer le chat dans son rétroviseur .	#9581813 <p>As Mr. Dursley drove around the corner and up the road , he watched the cat in his mirror .</p>
#2408061	Les chats sont incapables de lire des cartes ou des écriteaux .	#9581833 <p>It was now reading the sign that said Privet Drive -- no , looking at the sign ; cats could n't read maps or signs .</p>
#2408079	Mr Dursley se ressaisit et chassa le chat tigré de son esprit .	#9581859 <p>Mr. Dursley gave himself a little shake and put the cat out of his mind . As he drove toward town he thought of nothing except a large order of drills he was hoping to get that day . </p></p>
#2408207	<p> Lorsqu' il s' engagea dans l' allée du numéro 4 de sa rue , la première chose qu' il vit - et qui n' améliora	#9582783 <p><p> As he pulled into the driveway of number four , the first thing he saw -- and it did n't improve his</p>

Composition du corpus InterCorp



Partie française du corpus InterCorp (nombre de mots)



Annotation morphologique

CS – Morče :

[1] Nepotřebuješ **svoje** /*P8F-4*----- oči, králi!

(Milan Kundera, *Žert*, 1967, 1991)

FR – TreeTagger :

[2] Tu n'as pas besoin de **tes** /**DET:POS** yeux,
mon roi !

(*La Plaisanterie*, Marcel Aymonin, 1991)

Alignement

- Au niveau de phrases
- Automatique : *hunalign*
- Contrôle semi-manuel de l'alignement – éditeur de textes parallèles ***InterText***
- Lemmatisation + annotation morphologique
 - *TreeTagger* (fr, esp, it...),
 - *Morče* (tchèque, slovaque), etc.

Segments manquants

Browser tabs: (4007) Doručené ... x Intercorp x ELRA Catalogue of L... x Konkordance x InterCorp: datab... x marcel aymonin ... x InterText x

Address bar: https://trnka.ff.cuni.cz/ucnk/intercorp/intertext/index.php?offset=1380

Navigation: EPSON E-Web Print | Print | Clip | Enable Flash printing

Search: Find: empty segment on both sides

1380	☆	» V šestnáctce setkal se Švejek s devatenácti muži bez kalhot.	» Au 16, Chvéík trouva une vingtaine d'hommes déshabillés de la même façon que lui.	✓
1381	☆	» To byli všichni, kteří měli na spisech napsanou poznámku "Streng behüten, beobachten!" a o které nyní pečovali velice starostlivě, aby jim neutekli.	» C'étaient tous des gens dont le dossier portait la fameuse note Streng behüten, beobachten, et qu'on gardait donc à vue avec une sollicitude particulière, pour les empêcher de prendre la fuite.	✓
1382	☆	» Kdyby ty podvlékačky byly čisté a nebyly v oknech mříže, na první pohled dělalo by to dojem, že jste v šatně nějakých lázní.		✓
1383	☆	» Švejka přijal od šikovatele Řepy cimrkomandant, zarostlý chlap v rozhalené košili. » Zapsal si jeho jméno na kousek papíru visícího na zdi a řekl k němu:	» Le sergent-major Riha remit Chvéík aux soins du « chef de chambrée », un gaillard poilu à la chemise baïllante. » Celui-ci inscrivit le nom de Chvéík sur un bout de papier épinglé au mur et lui dit :	✓
1384	☆	» "Zejtra máme u nás divadlo.	» « Demain, il y aura du bon chez nous.	✓
1385	☆	» Povedou nás do kaple na kázání.	» On nous conduira au sermon à la chapelle.	✓
1386	☆	» My všichni v podvlíkačkách stojíme zrovna pod kazatelnu	» Nous autres, tous en caleçon comme nous voilà, on nous fait mettre tout à fait près de la chaire	✓

Segments ajoutés

Explicitation :

[3]

Naopak!

(Václav Havel, *Largo desolato*)

Oh, grand Dieu, non! Au contraire.

(Abrams, Erika; Meldegg, Stephan)

Solution 1 – corpus comparable de traductions (Jerome)

JEROME	nbre de positions	point point	point d'interrog	point d'exclam	virgule	points de susp.	deux-points	point-virgule
Romans orig Fq abs	1768079	106 300	11263	7044	132760	4041	4977	715
Romans orig ipm		60 122	6370	3984	75087	2286	2815	404
Romans trad Fq abs	1765433	101549	11099	5741	134788	6501	6400	5266
Romans trad ipm		57521	6287	3252	76348	3682	3625	2983
Chi-carré		Signif. at p < .001	Not signif.	Signif. at p < .001	Signif. at p < .001	Signif. at p < .001	Signif. at p < .001	Signif. at p < .001
DIN		-2	-1	-10	1	23	13	76

Solution 2 – corpus unilingues

Unilingue FR



Parallèle FR trad.



Parallèle FR orig.

Unilingue CS



Parallèle CS trad.



Parallèle CS orig.

Tableau 3 – FRANTEXT vs InterCorp (fr)

Lettres	point	point d'interrog	point d'exclam	virgule	points de susp.	deux-points	point-virgule
FRANTEXT ipm (31 610 109 positions)	42835	4722	3941	70325	4403	4126	1933
InterCorp fr ORIG ipm (3 555 295 positions)	39890	5848	10720	62958	14005	3858	1199
InterCorp fr TRADUCTION ipm (3 016 305 positions)	48252	6783	12087	80449	16333	4283	1347

Points de suspension et d'exclamation

[4] je peux pas tendre la main !... même pour les A.S... même les A.M.C... je démordrai pas !... idiot d'orgueil !

(Céline, *D'un château à l'autre*)

→nedokážu natáhnout ruku!... ani na pokladnu... ani na bezplatnou lékařskou péči... budu si paličatě stát na svém!... pitomá pýcha!
(Kareninová)

Idiolecte de traducteurs

[5] Já bych vyrobil tatarskou omáčku – načali
bychom nějaké dobré víno – má přijít Lucy – tuhle
Olda by jistě taky zůstal – cítím, že by mi to udělalo dobře
– trochu se odreagovat – přijít na jiné myšlenky –
(Havel, *Largo desolato*)

→ **Je ferais de la sauce tartare... On ouvrirait une bonne
bouteille... Lucy doit venir... Je suis sûr qu'Olda
resterait aussi... J'ai vraiment besoin de me distraire
un peu, changer d'idées, parler du bon vieux temps...**
(Abrams; Meldegg)

Solution 3 – bi-directionnalité du corpus

Unilingue FR

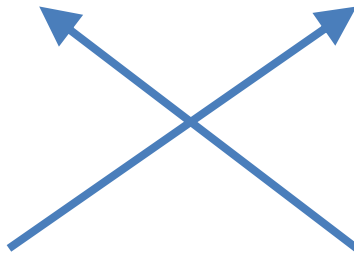
Unilingue CS

Parallèle FR trad.

Parallèle CS trad.

Parallèle FR orig.

Parallèle CS orig.



Interférence ? (ex. – gérondif)

[6] ... dit-il **en souriant**.

→ řekl **usmívaje se**. (transgressif, *vieilli*, 2%)

→ řekl **a usmál se**. (proposition, 67 %)

[il a dit et il a souri]

→ řekl **s úsměvem**. (SP, 16 %)

[dit-il avec sourire]

Fq relative du gérondif

Parallèle FR traductions 1 920 ipm

Parallèle FR originaux 1 748 ipm

Universaux de la traduction (ponctuation)

4 000 occurrences de non1:1 segments

[7] J'ai fait le chemin à pied. J'ai voulu voir
maman tout de suite.

(Camus, *L'Étranger*)

→ Došel jsem tam pěšky a hned jsem chtěl vidět
maminku.

[... à pied et tout de suite j'ai voulu...]

(Cizinec, Žilina)

Analyse bi-directionnelle

[8]

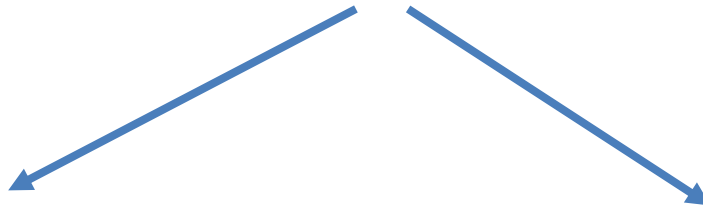
Ti soudruzí jezdcí. Kde se vzali? Odkud?
(Topol, *Noční práce*, 2001)

→ D'où avaient bien pu sortir ces camarades à cheval ?

(*Missions nocturnes*, Canavaggio, 2002)

CONCLUSION

RECHERCHE SUR CORPUS PARALLELES



Analyse **contrastive** + Analyse **traductologique**

- qualité de l'alignement/de la traduction
- corpus unilingues et corpus comparables de traductions
- bi-directionnalité de l'analyse

Références bibliographiques

- BAKER, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and Translation: Studies in language engineering, in honour of Juan C. Sager* (pp. 175–186). Amsterdam: John Benjamins.
- BAKER, M. (1998). Réexplorer la langue de la traduction: une approche par corpus. *Meta : journal des traducteurs / Meta: Translators' Journal* 43 (4), 1–10. [online]. <http://www.erudit.org/revue/meta/1998/v43/n4/001951ar.pdf>. [Cit.11.4.2015].
- ČERMÁK, F., Klégr, A., Corness, P. (eds). 2010. *InterCorp : Exploring a Multilingual Corpus*. Praha: Nakladatelství Lidové noviny.
- ČERMÁKOVÁ, A., Fárová, L. 2010. “Keywords in Harry Potter and their Czech and Finnish translation equivalents”. In F. Čermák, A. Klégr, P. Corness (eds). *InterCorp : Exploring a Multilingual Corpus*. Praha: Nakladatelství Lidové noviny, 177-188.
- CHLUMSKÁ, L. (2013). Korpus Jerome – comparable corpus of Czech translation and non-translation language. Praha: ÚČNK FF UK, www.korpus.cz.
- LAVIOSA, S. (2006). *Corpus-based Translation Studies. Theory, Findings, Applications*. Amsterdam/New York: Rodopi.
- NÁDVORNÍKOVÁ, O. 2010a. « Les corpus parallèles : L'Espace pour l'analyse contrastive ». *Études Romanes de Brno*, 2010, 31, 1, 7-27.
- NÁDVORNÍKOVÁ, O. 2010b. *The French gérondif and its Czech equivalents*. In Čermák, Fr. – Corness, P. – Klégr, A. (eds). *InterCorp: Exploring a Multilingual Corpus*. Praha: Nakladatelství Lidové noviny/Ústav Českého národního korpusu, 2010, 83-96.
- NÁDVORNÍKOVÁ, O. Les gérondifs antéposés : quelles relations avec les contextes de gauche et de droite ? *Verbum*, 35, 1-2, 161-174.
- NÁDVORNÍKOVÁ, O. Le corpus multilingue InterCorp et les possibilités de son exploitation. XXVII^e Congrès international de linguistique et de philologie romanes (CILPR), Nancy 15-20 juillet 2013.
- OLOHAN, M. (2004). *Introducing Corpora in Translation Studies*. London and New York: Routledge.
- ROSEN, A. 2012. « Mediating between incompatible tagsets ». In Ahrenberg et al. (eds), *Proceedings of the Workshop on Annotation and Exploitation of Parallel Corpora (AEPC), Volume 10 of NEALT Proceedings Series*. Tartu: Northern European Association for Language Technology, 53-62.

This presentation resulted from the implementation of the Czech National Corpus project (LM2011023) funded by the Ministry of Education, Youth and Sports of the Czech Republic within the framework of Large Research, Development and Innovation Infrastructures.