



**ÉPREUVES
DE L'ÉTRANGER**
Une expérience de traduction
et de rétrotraduction avec Gérard Macé



« Chenilles et papillons »: Les métamorphoses d'un texte dans les filets de l'alignement multitextuel

Olivier Kraif et Pascale Roux (Université Grenoble Alpes)

Les hypothèses de départ

- ⇒ La traduction est une co-crédation
- ⇒ Un texte qui est passéd par une langue distante est susceptible de « revenir » en français plus modifiéd qu'un texte qui est passéd par une langue proche
- ⇒ Le texte poétique est plus sujet à distorsion dans la traduction qu'un autre type de texte

Les nouvelles hypothèses

- ⇒ La **posture du traducteur** influe sur la distance entre les textes de départ et les rétrotraductions.
- ⇒ On peut établir une **typologie des séquences** textuelles en fonction du critère de la **distorsion / stabilité** traductive.
- ⇒ Les **virtualités interprétatives** réalisées dans les rétrotraductions sont en germe dans le **texte de départ**.
- ⇒ Les choix linguistiques et stylistiques dans le texte traduit peuvent être appréhendés à travers **leur relation à la langue standard** : distance (effet d'étrangeté linguistique)/ proximité (effet de transparence de la langue traduisante)

Stylistique et humanités numériques:
trois directions à explorer

1. L'alignement

1. L'alignement

En sortant de la caverne, nous verrons les mirages
qu'on rêve depuis toujours de traverser à pied sec :
des flaques d'eau qui s'évaporent quand on avance,
des nappes de lumière dans lesquelles on voit des lacs.

G.Macé

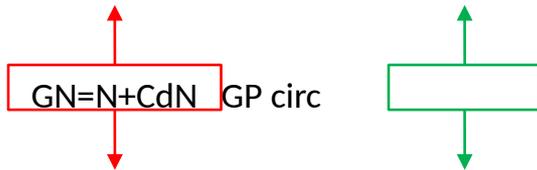
Hors de la caverne nous verrons ces images
Qu'en rêve chaque fois nous foulons à pied sec.
Des marais s'asséchant tandis que l'on avance
Et des semblants de lacs dans la lueur étale.

M.Capponi, GR(R) (F. Dell'Oro)

1. L'alignement

Correspondance syntaxique entre les vers:

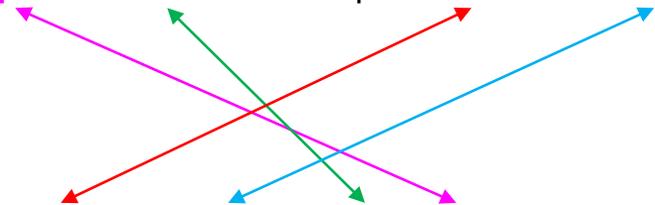
des nappes de lumière dans lesquelles on voit des lacs



Et des semblants de lacs dans la lueur étale

Correspondance sémantique entre les vers:

des nappes de lumière dans lesquelles on voit des lacs



Et des semblants de lacs dans la lueur étale

2. La distance entre versions

2. La distance entre versions

En sortant de la **caverne**, nous verrons les **mirages**
qu'on **rêve** depuis toujours de traverser à **pied** sec :
des flaques d'eau qui s'évaporent quand on avance,
des **nappes** de lumière dans lesquelles on voit des lacs.

G.Macé

2. La distance entre versions

Un monde où l'on ferrait les chevaux
dans l'odeur de la corne, des chevaux battus

qui regardaient droit devant eux, à cause des œillères
que n'osaient pas porter les hommes.

G.Macé

Dans le parfum des cornes,
Du temps où l'on ferrait les sabots,

Sous les œillères que les humains n'osaient porter,
Les chevaux vaincus regardaient droit devant.

Lee Taeyeon et A. Marchand, KR(R) (Lee Soyeon)

C'était au temps du ferrage quand

Dans l'odeur du sabot

Les chevaux fixaient droit devant eux

Grâce aux œillères que les hommes n'avaient pas le courage de porter.

J. Lévy, JAP(R) (Kohei Kuwada)

3. La relation à une langue standard

3. La relation à une langue standard

Les montagnes à contrejour ressemblent à des enclumes,
et le soleil éclaire cette forge abandonnée.

Les étincelles, c'est le marteau du philosophe
qui les fait jaillir en tapant sur le vieux monde.

Un monde où l'on ferrait les chevaux
dans l'odeur de la corne, des chevaux battus
qui regardaient droit devant eux, à cause des œillères
que n'osaient pas porter les hommes.

G.Macé

Adossée au jour, la montagne ressemble à une enclume,
le soleil illumine cette forge abandonnée.

L'éclair, ce qui fait jaillir l'éclair en cognant sur le vieux monde,
c'est le marteau du philosophe.

Ce bas monde où, respirant l'odeur de la corne de l'enclume,
on ferrait les chevaux, les chevaux qui, sous les coups de cravache,
ne regardaient que devant eux,
avec ces œillères que jamais l'homme n'osa porter.

Cha Hyangmi et T.Laplanche, KR(R) (Moon Soyoung)

L'alignement multi-textuel

L'alignement multi-textuel

L'alignement multi-textuel

- Objectifs
 - Identifier des correspondances entre vers puis entre mots
 - Observer les points d'ancrage et les espaces d'instabilité
 - en fonction des mots
 - en fonctions des parties du discours
 - en fonction des langues impliquées

L'alignement multi-textuel

- L'alignement automatique
 - Segmentation
 - Appariements des segments équivalents
- A différents niveaux (granularité)
 - Phrases (ou groupes de phrases)
 - Vers (ou groupes de vers)
 - Mots (ou groupes de mots)
- En général 2 à 2 → bitexte

L'alignement multi-textuel

- Première étape : structuration XML des poèmes

```
<text>
  <body>
    <lg>
      <l>En sortant de la caverne, nous verrons les mirages</l>
      <l>qu'on rêve depuis toujours de traverser à pied sec :</l>
      <l>des flaques d'eau qui s'évaporent quand on avance,</l>
      <l>des nappes de lumière dans lesquelles on voit des lacs.</l>
    </lg>
    <lg>
      <l>Aucun prophète, aucun pharaon pour marcher devant nous.</l>
      <l>Pas de mer rouge qui s'ouvrirait d'elle-même</l>
      <l>comme les grandes lèvres des femmes, mais un au-delà</l>
      <l>qui recule en même temps que l'horizon.</l>
    </lg>
  <p>*</p>
```

L'alignement multi-textuel

- Deuxième étape :
alignement
automatique 2 à 2
(ici *la*)

```
<tu tuid='33'>
  <tuv id='33' xml:lang='fr'>
    <seg>Un monde où l'on ferrait les chevaux</seg>
  </tuv>
  <tuv id='33' xml:lang='f2'>
    <seg>Alors, les sabots étaient recouverts</seg>
  </tuv>
</tu>
<tu tuid='34'>
  <tuv id='34' xml:lang='fr'>
    <seg>dans l'odeur de la corne, des chevaux battus</seg>
  </tuv>
  <tuv id='34' xml:lang='f2'>
    <seg>de corne odorante, pour les chevaux battus </seg>
  </tuv>
</tu>
<tu tuid='35'>
  <tuv id='35' xml:lang='fr'>
    <seg>qui regardaient droit devant eux, à cause des œillères</seg>
  </tuv>
  <tuv id='35 36' xml:lang='f2'>
    <seg>qui ne pouvaient regarder qu'en avant, </seg>
    <seg>parce que leurs yeux portaient des œillères</seg>
  </tuv>
</tu>
<tu tuid='36'>
  <tuv id='36' xml:lang='fr'>
    <seg>que n'osaient pas porter les hommes.</seg>
  </tuv>
  <tuv id='37' xml:lang='f2'>
    <seg>que n'osaient porter les hommes.</seg>
  </tuv>
</tu>
</body>
</tmx>
```

L'alignement multi-textuel

- Troisième étape :
multi-alignement
par clôture
transitive (ici *fr-la-ko-ja-it-grk-fa-de-ar*)

```
<tu tuid='22'>
  <tuv id='35 36' xml:lang='fr'>
    <seg>qui regardaient droit devant eux, à cause des œillères</seg>
    <seg>que n'osaient pas porter les hommes.</seg>
  </tuv>
  <tuv id='35 36 37' xml:lang='f2'>
    <seg>qui ne pouvaient regarder qu'en avant, </seg>
    <seg>parce que leurs yeux portaient des œillères</seg>
    <seg>que n'osaient porter les hommes.</seg>
  </tuv>
  <tuv id='43 44' xml:lang='f3'>
    <seg>ne regardaient que devant eux, </seg>
    <seg>avec ces œillères que jamais l'homme n'osa porter.</seg>
  </tuv>
  <tuv id='35 36' xml:lang='f4'>
    <seg>droit devant eux braquaient les yeux, les hommes ne songeant à porter</seg>
    <seg>les œillères qu'ils leur appliquaient</seg>
  </tuv>
  <tuv id='35 36' xml:lang='f5'>
    <seg>qui regardaient droit devant eux, contraints par les œillères</seg>
    <seg>que n'osaient porter les hommes.</seg>
  </tuv>
  <tuv id='19' xml:lang='f6'>
    <seg> Les chevaux, frappés, regardaient aussitôt droit devant eux, à cause des œillères
  </tuv>
  <tuv id='33 34' xml:lang='f7'>
    <seg>Qui avec les œillères ne voyaient que devant eux</seg>
    <seg>Et craignaient d'être la monture des hommes</seg>
  </tuv>
  <tuv id='35 36' xml:lang='f8'>
    <seg>qui regardaient devant eux à cause des œillères</seg>
    <seg>que les hommes n'osaient pas porter.</seg>
  </tuv>
  <tuv id='34 35' xml:lang='f9'>
    <seg>Regardent droit devant car ils portent des œillères </seg>
    <seg>Que les gens n'osaient point mettre.</seg>
  </tuv>
</tu>
```

L'alignement multi-textuel

- Pas de correspondance fine entre segments (correspondances sur des fragments de vers entraînant des regroupements en cascade).
- Solution trouvée : insérer manuellement des repères `<milestone/>` pour aligner avec les vers de l'original

L'alignement multi-textuel

- Ex. *fr - la2*

```
<lg>
  <milestone />
  <l>Le nôtre est un soleil d'hiver</l>
  <milestone />
  <l>qui éclaire le lieu du crime :</l>
  <milestone />
  <l>la goutte de sang, les plumes d'oiseau,</l>
  <milestone />
  <l>était-ce le festin des dieux</l>
  <milestone />
  <l>ou le repas du serpent ?</l>
</lg>
```

```
<lg>
  <milestone />
  <l>Sur nous, un soleil d'hiver</l>
  <milestone />
  <l>qui brille dans des lieux criminels :</l>
  <milestone />
  <l>une goutte de sang, des plumes d'oiseau</l>
  <milestone />
  <l>
    était-ce le festin des dieux
  <milestone />
    ou le repas du dragon ?
  </l>
</lg>
```

L'alignement multi-textuel

- Quatrième étape :
annotation des tokens

```
<lg>
  <milestone />
  <l>
    <t id="t1" f="En" l="En" c="ADJ" />
    <t id="t2" f="sortant" l="sortant" c="NOM" />
    <t id="t3" f="de" l="de" c="PRP" />
    <t id="t4" f="la" l="le" c="DET:ART" />
    <t id="t5" f="caverne" l="caverne" c="NOM" />
    <t id="t6" f="," l="," c="PUN" />
    <t id="t7" f="nous" l="nous" c="PRO:PER" />
    <t id="t8" f="verrons" l="voir" c="VER:futu" />
    <t id="t9" f="les" l="le" c="DET:ART" />
    <t id="t10" f="mirages" l="mirage" c="NOM" />
  </l>
  <milestone />
  <l>
    <t id="t11" f="qu'" l="que" c="KON" />
    <t id="t12" f="on" l="on" c="PRO:PER" />
    <t id="t13" f="rêve" l="rêver" c="VER:pres" />
    <t id="t14" f="depuis" l="depuis" c="PRP" />
    <t id="t15" f="toujours" l="toujours" c="ADV" />
    <t id="t16" f="de" l="de" c="PRP" />
    <t id="t17" f="traverser" l="traverser" c="VER:infi" />
    <t id="t18" f="à" l="à" c="PRP" />
    <t id="t19" f="pied" l="pied" c="NOM" />
    <t id="t20" f="sec" l="sec" c="ADJ" />
    <t id="t21" f=":" l=":" c="PUN" />
  </l>
```

L'alignement multi-textuel

- Statistiques
- (nb de tokens)
 - grande variabilité entre les traducteurs
 - difficile de tirer des conclusions par langue

<u>Rétrotraduction</u>	<u>Lignes</u>	<u>Mots</u>
original	32	274
persan1	30	280
persan2	33	264
grec ancien1	33	236
grec ancien2	9	308
italien1	32	264
italien2	33	265
latin1	41	297
latin2	36	293
japonais1	32	235
japonais2	32	273
coréen1	32	247
coréen2	32	297
allemand1	33	275
allemand2	32	280
arabe1	31	314
arabe2	31	284

L'alignement multi-textuel

- Cinquième étape : alignement des tokens
 - Alignement manuel (Kevin Woringer)

	A	B	C	D	E	F
1		allemand1R_DD (de KM)	allemand2R_NRFS (de MGM)	arabe1R_SA (de SG)	arabe2R_JEG (de HJ)	coréen1R_LTAM (de LS)
2	"En sortant de "	"Au sortir de"	id	id	"Une fois sortis de"	"à la sortie"
3	"la caverne"	id	id	id	id	"une grotte"
4	"nous verrons"	id	id	id	id	des mirages qui éblouissent
5	"les mirages"	id	id	"le mirage"	"le mirage"	"les mirages"
6	" on rêve"	id	"nous rêvions"	"nous avons rêvé"	"nous rêvons"	"un vieux rêve"
7	"depuis toujours"	id	id	"depuis l'éternité"	"depuis l'éternité"	"un vieux rêve"
8	"de traverser"	id	id	id	"de le franchir"	"nous traverserons"
9	"à pied sec"	id	id	ans nous mouiller les pieds	ans nous mouiller les pieds	id
10	"des flaques d'eau"	id	id	"Les mares d'eau"	id	"les flaques d'eau"
11	"s'évaporent"	id	id	id	id	"disparaissent"
12	"quand"	"dès"	id	"au fur et à mesure"	"à mesure que"	
13	"on avance"	"on s'en approche"	"on se met en route"	id	progressons dans notre m	"en avançant"
14	"des nappes de lumière"	"Des étendues lumineuses"	"des couche de lumières"	"Des sources de lumière"	"nuées de lumière"	"Les strates de lumière"

- Puis synchronisation avec les tokens de l'étape précédente

L'alignement multi-textuel

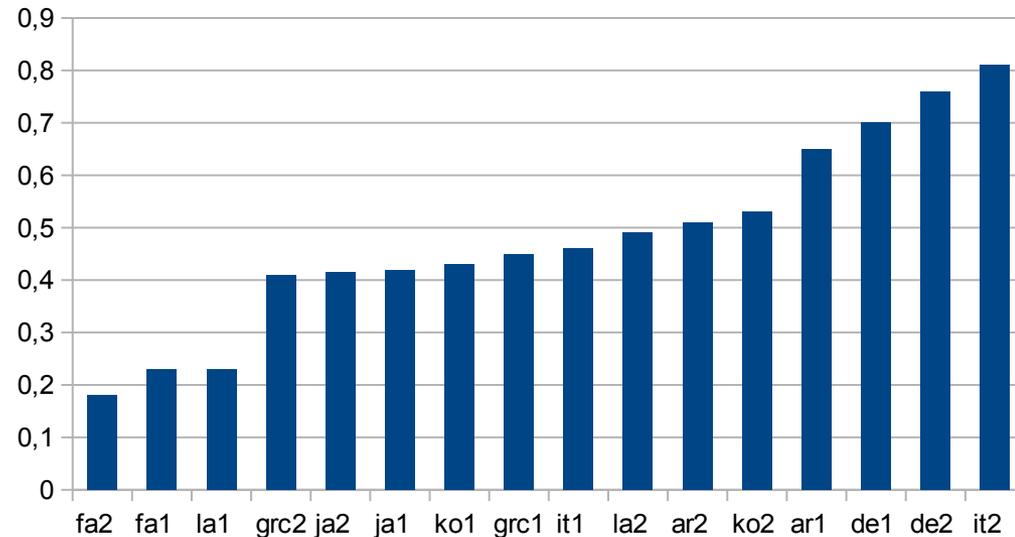
- A l'intérieur de chaque groupe de mots alignés, calcul de la **stabilité lexicale**
 - mot existe dans le groupe original : 1
 - mot absent du groupe original : 0

L'alignement multi-textuel

- **Stabilité lexicale par langue**

- → variations importantes
- liées aux traductions
- ex. *it* et *la*
- liées aux langues ?
 - diff. moy. entre deux langues ≠
 - 0,215
 - diff. moy. au sein de la même lang.
 - 0,1255
- → t-test ($p=0,05$) pas significative
- mais de justesse ($t=2,10 < 2,36$)

Stabilité moyenne des formes par langue



L'alignement multi-textuel

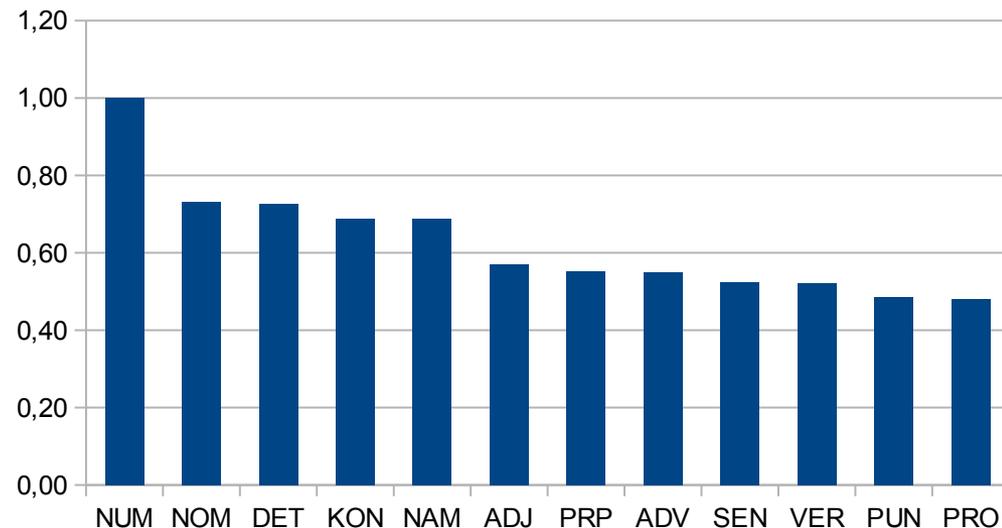
- **Stabilité lexicale par catégorie**

- Adj. numéraux (*deux, trois, quatre*)
- Noms et nom propres stables
- Déterminants (article et poss.)
- Conjonctions
- (*mais, et, ou, quand, que, ...*)

- Les moins stables : verbes, ponctuations et pronoms

- → pas d'opposition mots pleins - mots outils

Stabilité des lemmes en fonction des parties du discours



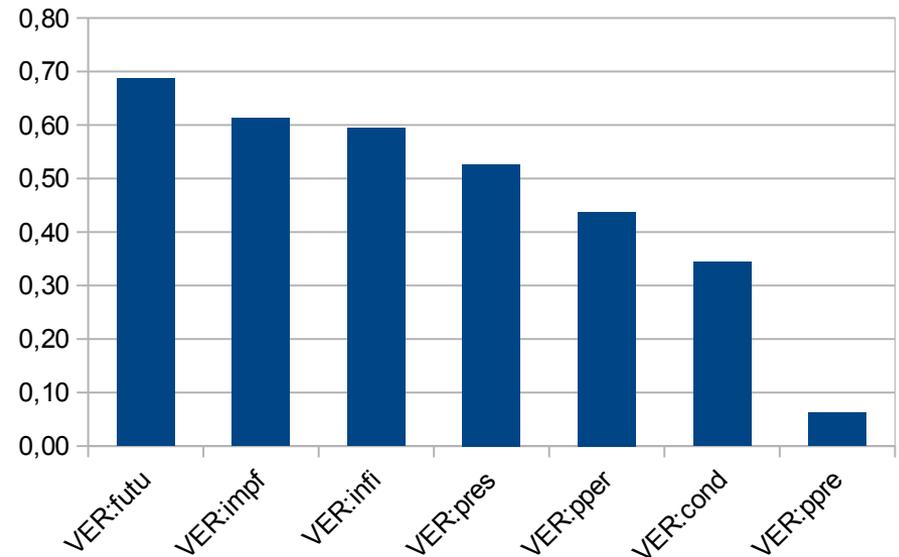
L'alignement multi-textuel

- **Stabilité lexicale pour les**

verbes

- Futur (1 occ. "nous verrons les mirages"), imparfait et infinitif plutôt stable
- Conditionnel (2 occ.) et participe présent (1 occ. "en tapant sur le vieux monde") très instable

Stabilité des lemmes pour les verbes



L'alignement multi-textuel

- Stabilité par lemme

Occurrences	Stables	Instables	Taux
prophète	16	0	100%
pharaon	16	0	100%
mer	16	0	100%
lèvre	16	0	100%
femme	16	0	100%
deux	16	0	100%
ou	32	0	100%
trois	16	0	100%
quatre	16	0	100%
sang	16	0	100%
œillère	16	0	100%

L'alignement multi-textuel

- Stabilité par lemme

- Beaucoup de noms concrets
- Grande intensité sensorielle
 - *rouge, sang, neige, briller*

- Constellations métaphoriques

- *enclume, forge, étincelle,*

marteau, cheval

- *horizon, soleil, ciel*
- *serpent, oiseau, dieux*
- *prophètes, pharaons*

- Seuls deux verbes :

- *attendre, briller*

neige	31	1	97%
rouge	15	1	94%
horizon	15	1	94%
soleil	45	3	94%
goutte	15	1	94%
oiseau	15	1	94%
nom	15	1	94%
attendre	15	1	94%
jour	15	1	94%
montagne	15	1	94%
enclume	15	1	94%
marteau	15	1	94%
pied	14	2	88%
ciel	28	4	88%
entier	14	2	88%
plume	14	2	88%
dieu	14	2	88%
serpent	14	2	88%
?	14	2	88%
briller	14	2	88%
forge	14	2	88%
étincelle	14	2	88%
philosophe	14	2	88%
cheval	28	4	88%

L'alignement multi-textuel

- Et à l'autre extrême ?

- beaucoup de mots outils
- beaucoup de verbes
- deux noms :
 - *nappe, campagne*

elle-même	6	10	38%
tant	6	10	38%
ressembler	6	10	38%
pas	6	10	38%
nappe	11	21	34%
en	5	11	31%
sortant	5	11	31%
avec	5	11	31%
sur	15	33	31%
nôtre	5	11	31%
vieux	5	11	31%
cause	5	11	31%
on	19	45	30%
ouvrir	4	12	25%
quand	3	13	19%
lequel	3	13	19%
autant	3	13	19%
campagne	3	13	19%
marcher	2	14	13%
hagard	2	14	13%
mettre	1	15	6%
taper	1	15	6%

L'alignement multi-textuel

- Ces deux "disparitions" sont plus révélatrices que les éléments stables :
 - *la campagne en fleurs*
 - un pays florissant, prairie fleurie, des champs de fleurs, le champ fleuri à l'infini, la plaine en fleurs, les champs en fleur, les champs couverts de fleurs, la nature en fleur, le champ fleuri, les fleurs des vallées finales, les fleurs éparses dans les champs, La ferme pleine de fleurs, les villages plein de fleurs
- l'image des fleurs phagocyte *campagne*, qui n'est au fond qu'un support

L'alignement multi-textuel

- *des nappes de lumière*
- Des étendues lumineuses, des couche de lumières, Des sources de lumière, nuées de lumière, Les strates de lumière, feuilles de lumière, la lueur étale, les étendues brillantes de l'eau stagnante, étendues de lumières, puits de lumière, L'étendue de lumière, les couches de lumière, les lumières qui s'ourdissent, plaines de lumière
- → idem ici, *lumière* prend le dessus sur *nappe*. Ici on est proche d'une relation base+collocatif.

L'alignement multi-textuel

- Une expérience de traduction automatique

Échantillon du texte original	Échantillon de la rétrotraduction automatique stabilisée en italien
Avec la neige la nappe est mise	Mettez la nappe sur la neige
sur les prés. Le ciel tout entier	sur les pelouses Le ciel entier
pourrait tenir dans une cuiller en argent,	pourrait tenir dans une cuillère en argent,
le ciel et deux ou trois soleils	le ciel et deux ou trois soleils
sur les quatre que comptaient les Aztèques.	des quatre qui ont compté les Aztèques.
Le nôtre est un soleil d'hiver	Notre soleil d'hiver
qui éclaire le lieu du crime :	cela illumine la scène du crime:
la goutte de sang, les plumes d'oiseau,	la goutte de sang, les plumes d'un oiseau,
était-ce le festin des dieux	c'était la fête des dieux
ou le repas du serpent ?	ou le repas du serpent?

Conclusion ?

- Pas de conclusion générale vu le faible volume des données (pas de représentativité)
- On constate de très fortes **variations liées aux traductions**, plus qu'aux langues
- Certaines **parties du discours sont plus stables** :
 - noms, déterminants, conjonctions
 - les verbes, pronoms sont très instables
 - les ponctuations sont globalement instables sauf '?'

Conclusion ?

- Les phénomènes impliqués dans la stabilité lexicale ne sont pas seulement ... lexicaux :
 - saillance des images
 - selon différents axes : saillance sensorielle, symbolique, référentielle
- L'étude des points de stabilité et des zones de variabilité peut fournir des clés d'interprétations

Conclusion ?

- La traduction / rétrotraduction comme *métamorphose* (déploiement des virtualités interprétatives du texte) et comme *crible* (identification des constantes interprétatives)