

Damien Hansen

Adapter la traduction automatique aux domaines culturels

Point de départ

- Recherches sur la perception et l'utilisation des technologies de la traduction dans les domaines culturels, notamment la littérature et les jeux vidéo.
- Position unique de la traduction automatique, en particulier depuis la mise en application des approches « neuronales », car elle donne parfois lieu à des discours tout à fait opposés, et parfois peu nuancés, dans la sphère publique.
- Nécessité de mettre les outils à l'épreuve, d'offrir un aperçu objectif de leurs performances et de leurs limites, d'aborder les enjeux sociétaux et les changements attendus pour ces secteurs.





BLOOMSBURY

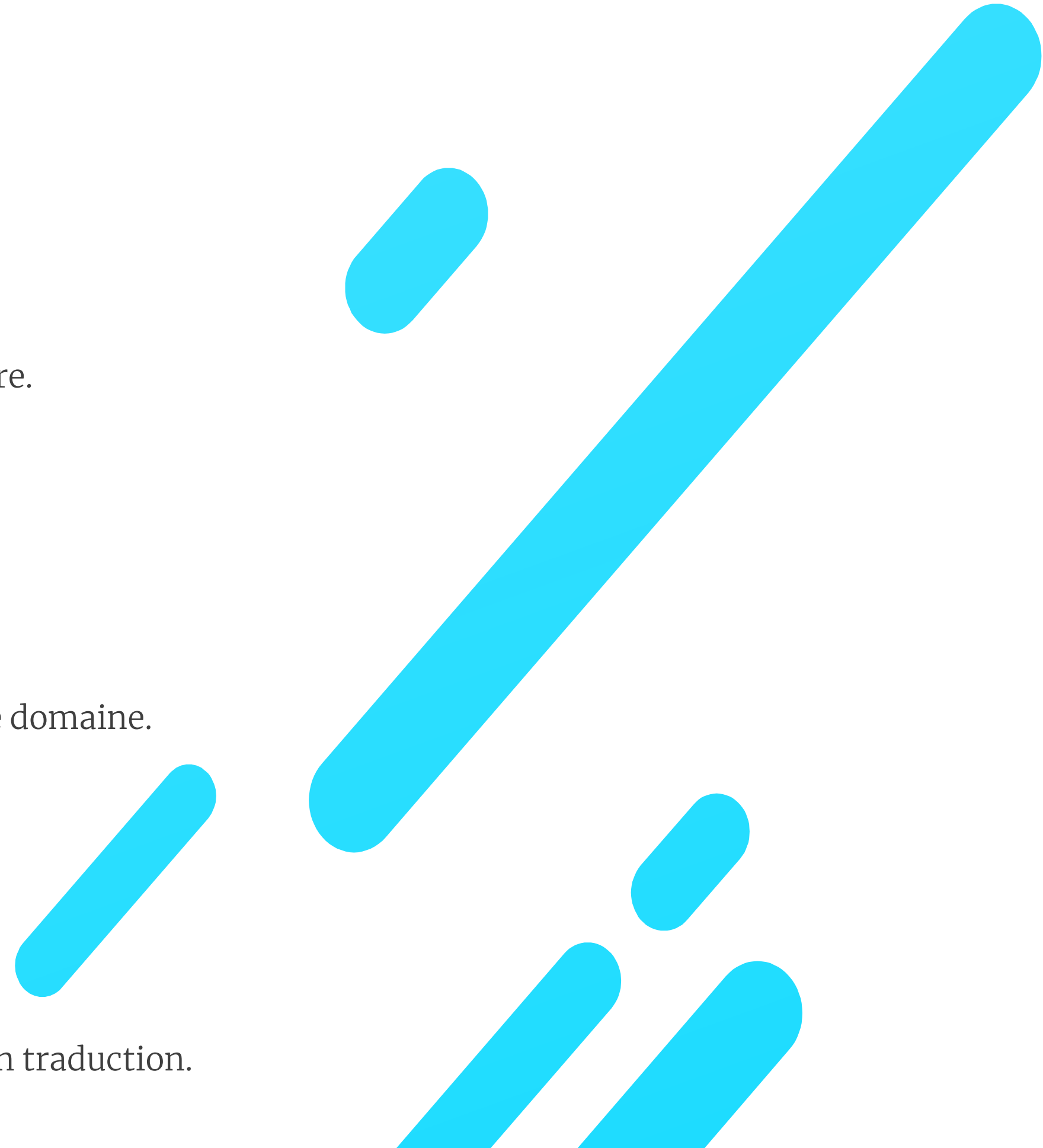
SEPTIMUS HEAP WIZARD APPRENTICE

Le cas de la littérature

La saga *Septimus Heap*

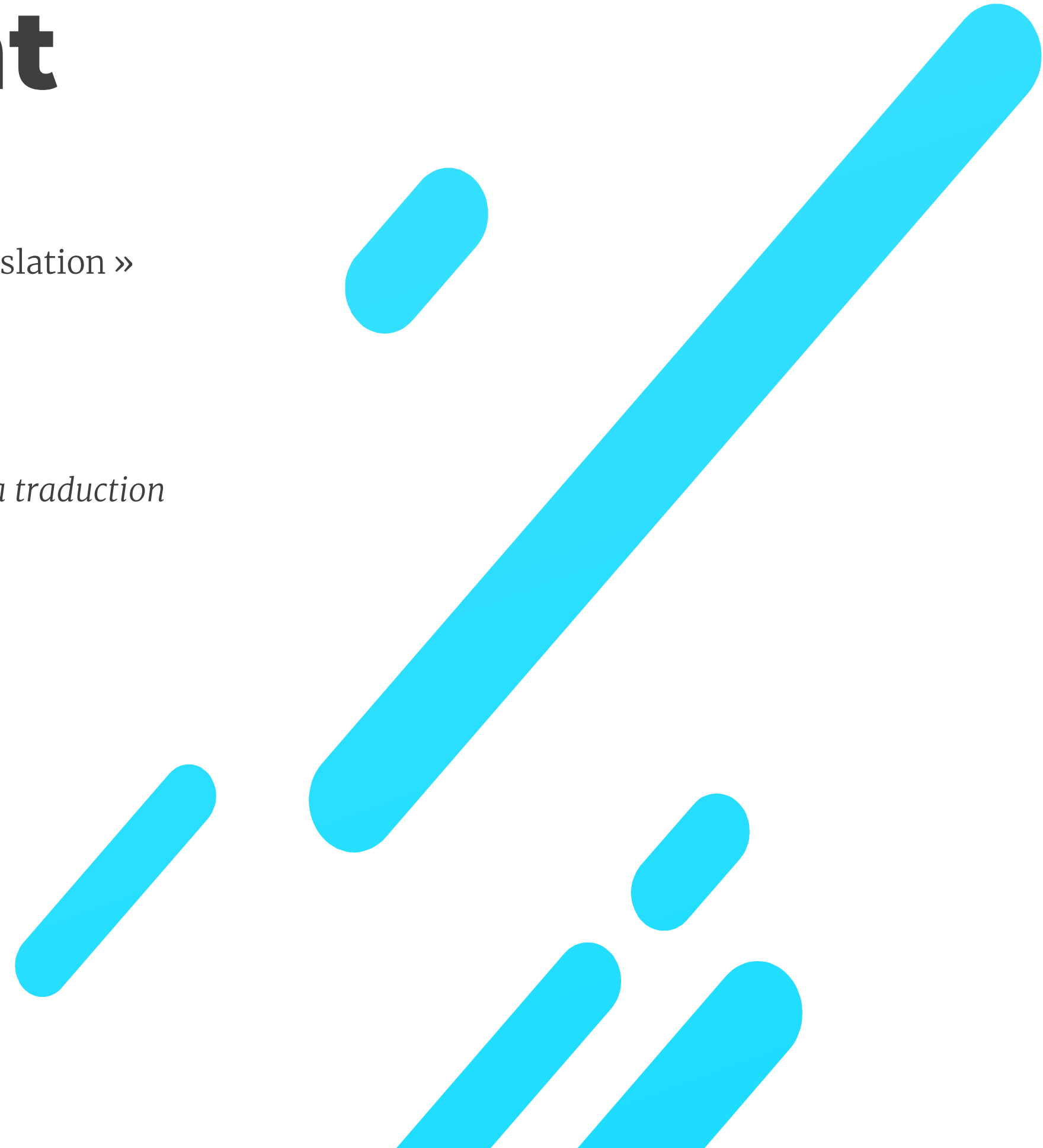
Pourquoi ?

- Nuancer le débat.
- Offrir un [aperçu concret et objectif](#) des capacités de la TA en littérature.
- Voir quelle [aide](#) elle peut apporter [au traducteur littéraire](#).
- Et quelle est la [plus-value](#) apportée par l'humain.
- Évaluer les principales [difficultés rencontrées par la machine](#) dans ce domaine.
- Nourrir la réflexion sur l'[évaluation de la TA](#).
- Aborder les [aspects sociétaux](#) attendus pour l'ensemble du secteur.
- Contribuer aux [\(in\)formations sur l'utilisation des nouveaux outils](#) en traduction.



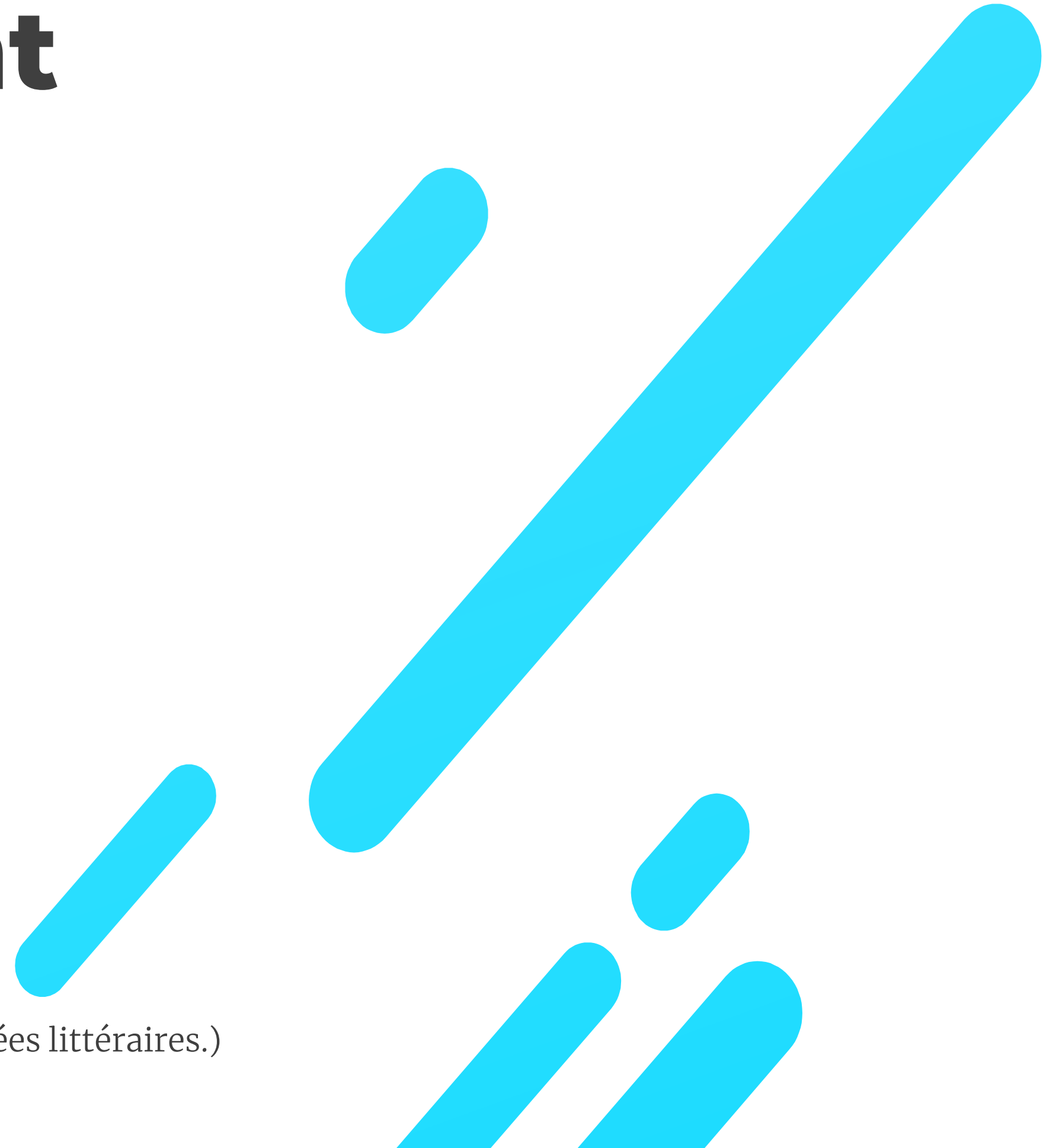
Un intérêt croissant

- 2019 : *MT Summit*, « Workshop on Qualities of Literary Machine Translation »
- 2019 : *EST*, « Translation technologies for creative-text translation »
- 2020 : *La traduction littéraire et SHS à la rencontre des technologies de la traduction*
- 2020 : *Creative Translation and Technologies Expert Meeting*
- 2021 : *BCBF*, « Machine Translation and Human Creativity »
- 2021 : *Conference on Computer-Assisted Literary Translation (CALT)*
- 2021-2022 : *TRACT*, « L'Impact de l'IA sur la traduction littéraire »



Un intérêt croissant

- Tezcan et coll. (2019–2020) pour la paire [EN-NL](#).
(Évaluation sur Google Traduction.)
- Ó Murchú (2019) pour la paire [GLA-GLE](#).
(Évaluation sur Intergaelic.)
- Kuzman et coll. (2019) pour la paire [EN-SLO](#).
(Évaluation d'un système maison affiné sur des données littéraires.)
- Matusov (2019) pour les paires [EN-RU](#) et [DE-EN](#).
(Évaluation de systèmes maison affinés sur des données littéraires.)
- Toral et coll. (2018–2021) pour la paire [EN-CAT](#).
(Évaluation de systèmes maison entraînés uniquement sur des données littéraires.)



Le corpus



Sage, Angie. *Septimus Heap*. HarperCollins, 2005–2013. 7 vol.

Serval, Nathalie, traductrice. *Magyk*. Par Angie Sage, Albin Michel, 2005–2013. 6 vol.

Affiner la TA ?

- Comment entraîner et « spécialiser » un système de TA ?
(⇒ Exemple plus simple avec le cas des jeux vidéo.)
- Cas plus complexe ici lié à la taille très limitée des **données** !
- Corpus littéraire personnalisé : **45 K phrases** (séparées en 3 corpus).
- Seul corpus littéraire publiquement accessible : **127 K phrases**
(bruitées et mal alignées).
- On estime qu'il faut au moins **2 M de phrases** pour entraîner des systèmes basiques.
↳ **6 M de phrases** (Blin 2021)

Affiner la TA ?

- Entraîner un système générique (3-4 M de phrases).
- Relancer un second entraînement en affinant sur les données littéraires.
- Conserver les données génériques pour éviter la dégradation du système.
- Donner un poids plus haut aux données spécialisées.
- (Autres questions de segmentation, vocabulaire, paramétrage...)

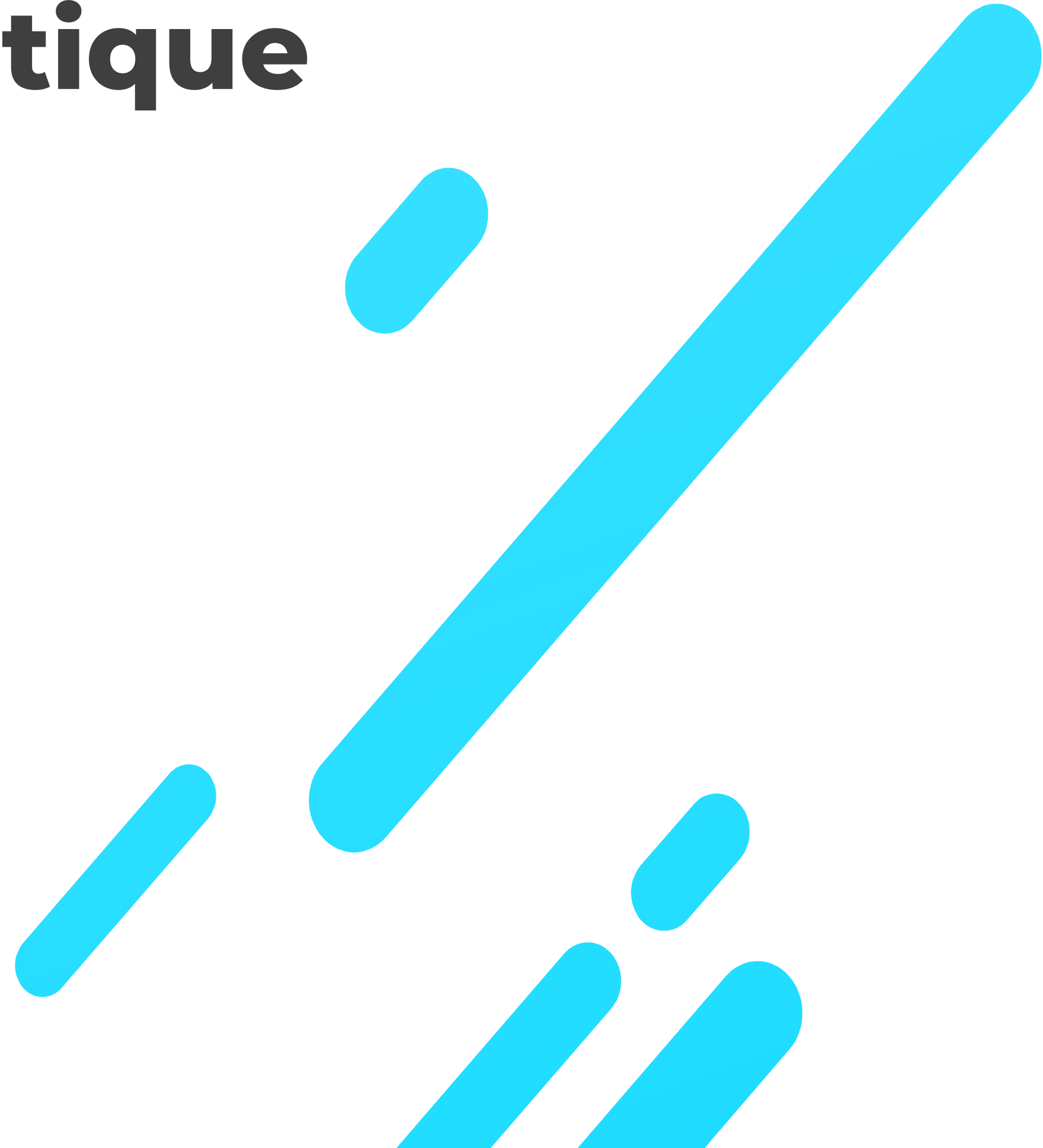
Évaluation automatique

- Mesure à quel point la traduction est proche de la référence produite par la traductrice officielle.

	BLEU ↑	chrF2++ ↑	TER ↓
Google Tr.	10.79	35.20	91.08
DeepL	10.04	34.88	92.81
Custom	18.07	40.29	76.09

Métriques d'évaluation sacreBLEU (Post 2018)

- BLEU calcule combien de *tokens* sont similaires à la référence, chrF fait de même, caractère par caractère, et TER mesure le nombre de changements nécessaires pour produire la référence.
- Un score BLEU de ~20 serait un score faible, mais acceptable, pour une sortie de TA « utile » (Toral & Way 2015).



D'autres résultats

- Comparaison avec d'autres essais d'adaptation à la littérature (L = LSTM ; T = Transformer)

Système (T)	BLEU
Générique	09.90
Affiné	18.07
Google Trad.	10.79

Système maison, EN-FR

Système (L)	BLEU
Générique	17.50
Affiné	20.75
Google Trad.	21.97

Kuzman et coll. (2019), EN-SLO

Système (L)	BLEU
Générique	14.20
Affiné	15.20
Google Trad.	13.90

Matusov (2019), EN-RU

Système (T)	BLEU
Générique	18.50
Affiné	16.20
Google Trad.	20.20

Matusov (2019), DE-EN

- Seule l'évolution générique < affiné est comparable, séparément pour chaque tableau.
- Affinage mitigé, voire moins performant, excepté chez Kuzman et coll. (2019) : entraînement sur un roman de l'auteure et de la traductrice !
- Explique aussi nos très bons résultats (+ 8 BLEU), avec pourtant très peu de données. Mais aussi avec un roman visiblement complexe (cf. résultats Google Trad. ; Hansen 2021).



Difficulté des comparaisons

- Différentes paires de langues.
- Différentes directions.
- Différentes approches.
- Différents corpus d'entraînement.
- Différent·e·s auteur·e·s.
- Différent·e·s traducteur·trice·s.
- Différents genres littéraires...



Annotation d'erreurs

Respect global de tous les termes/néologismes propres à la série

Quelques observations (étude en cours) :

De bêtes erreurs (souvent avec les déterminants et les prépositions)

Source	Two books— The Undoing of the Darkenesse and The Darke Index —lay open on Marcia's desk . In the shadow of the books a small silver box and a tiny silver disc rested on a piece of purple velvet.
TA	Deux livres, L'Art de vaincre la Ténèbre et L'Codex Tenebrae , étaient ouverts sur le bureau de Marcia tandis qu'une petite boîte d'argent et un petit disque d'argent reposaient sur un morceau de velours pourpre.
Référence	Deux livres – L'Art de vaincre la Ténèbre et le Codex Tenebrae – étaient ouverts sur la table de travail de Marcia, à côté d'une petite boîte en argent et d'un disque du même métal posé sur un carré de velours violette.

Des fusions
surprenantes de
certaines phrases

Des traductions relativement littérales

Source	Marcia had found The Undoing of the Darkenesse in Aunt Zelda's cottage a few years back and she knew that somewhere within its pages lay The Great UnDoing —the legendary Anti-Darke incantation that practitioners of the Darke feared above all else.
TA	Quelques années plus tôt, Marcia avait trouvé L'Art de vaincre la Ténèbre dans le cottage de tante Zelda et elle savait qu'à l'intérieur de ses pages se trouvait l'incantation anti-Ténèbre , une incantation légendaire que les praticiens de la Ténèbre craignaient par-dessus tout.
Référence	Puis Marcia avait découvert L'Art de vaincre la Ténèbre au cottage de tante Zelda, quelques années plus tôt. Elle savait qu'il contenait la formule du Grand Dénouement , le contresort mythique que les adeptes de la Ténèbre redoutaient par-dessus tout.

Mais parfois
malvenues

Des paraphrases intéressantes

Annotation d'erreurs

Quelques observations (étude en cours) :

À nouveau, une fusion de segments

Source	Jenna looked around the entrance hall, which was full of shadows and ghosts . She reached into her pocket for the Candlelight Charm Septimus had given her that afternoon.
TA	Jenna promena son regard autour de la salle d'entrée, remplie d'ombres et de fantômes , et plongea la main dans sa poche, cherchant le charme chocolaté que Septimus lui avait offert cet après-midi-là.
Référence	Jenna promena son regard autour du hall, peuplé d'ombres et de fantômes , puis elle sortit de sa poche l' allume-bougies magique que Septimus lui avait offert cet après-midi-là.

??? (hallucination)

Des traductions relativement littérales

Source	Aunt Zelda settled herself beside the fire . She looked at the duck in Jenna's arms.
TA	Tante Zelda s'installa près du feu et considéra la cane dans les bras de Jenna.
Référence	Tante Zelda s'assit près de la cheminée et son regard se posa sur la cane , blottie dans les bras de Jenna.

Pas de sexisme avec les animaux

Source	She turned to Septimus. "You really don't have to come," she said.
TA	— Tu n'es pas obligé de m'accompagner, dit-elle à Septimus.
Référence	Là, elle se tourna vers Septimus et lui dit : — Tu n'es pas obligé de venir.

Un respect surprenant des conventions typographiques littéraires, même sans aucun indice dans la source (p. ex. un « : » précédant le dialogue)

Conclusions préliminaires

- Est-il possible d'adapter un système de TA à la littérature ?
 - ➔ Oui, et même l'adapter à un couple auteure/traducteur·trice en particulier !
- Mais il reste encore du chemin à parcourir.
 - ➔ Pour l'heure, on peut surtout l'envisager comme une aide à la lecture ou une aide à l'apprentissage des langues étrangères (Oliver et coll. 2019 ; Matusov 2019) ou pour équilibrer les chances de publication (Castro 2020).
- À terme, peut-être, comme un véritable outil pour les professionnel·les.
 - ➔ Mais cela pose de nouvelles questions (éthiques, sociétales, pédagogiques...)





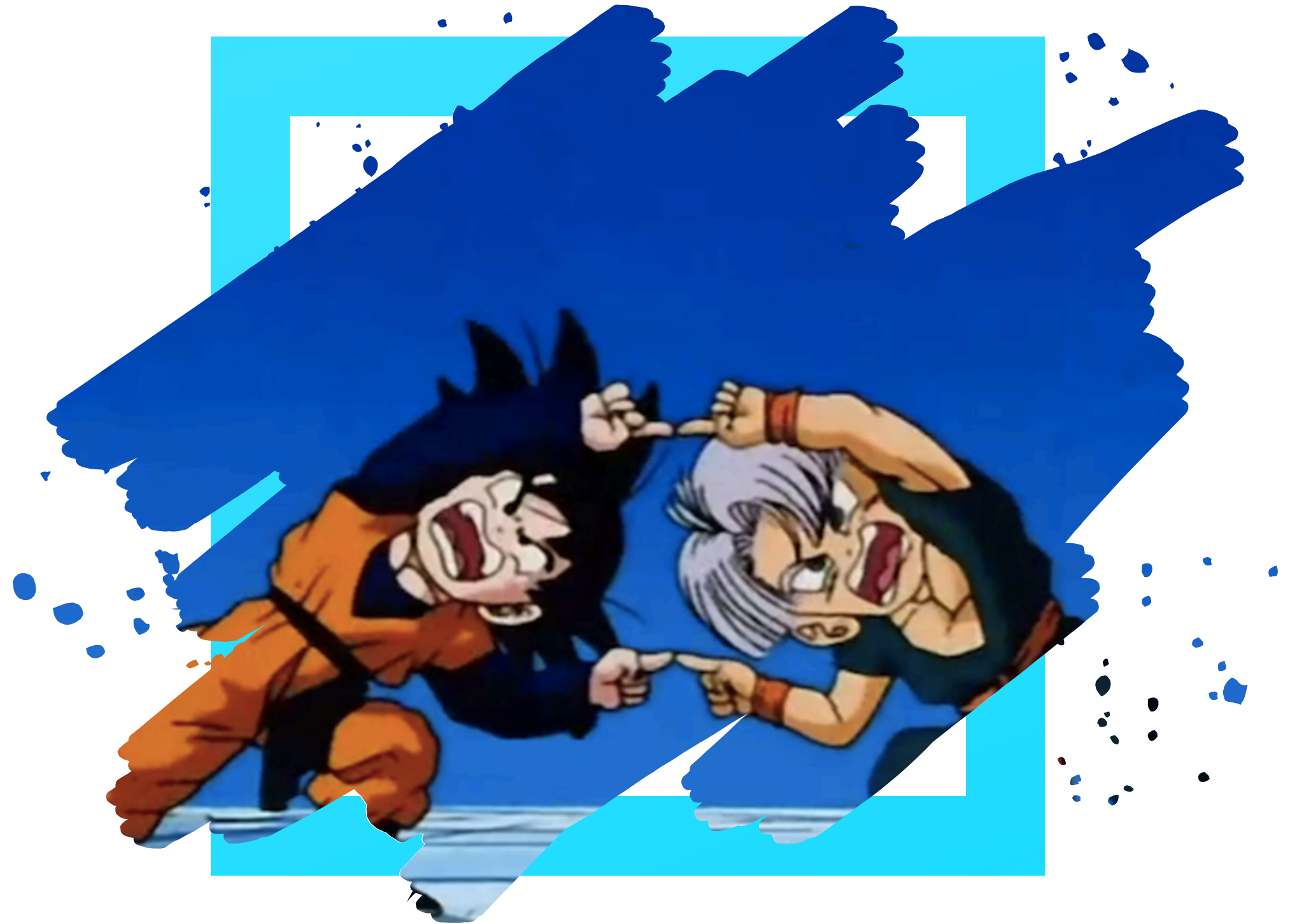
Le cas des jeux vidéo

Les séries *Elder Scrolls* et *Fallout* (Definitive Legendary Special Anniversary Edition HD Remastered II)

Pourquoi ?

(La raison scientifique)

- ★ Thèse n° 1
La traduction automatique
- ★ Thèse n° 2
La traduction de jeux vidéo
- ➔ La traduction automatique de jeux vidéo



Pourquoi ?

(La véritable raison)

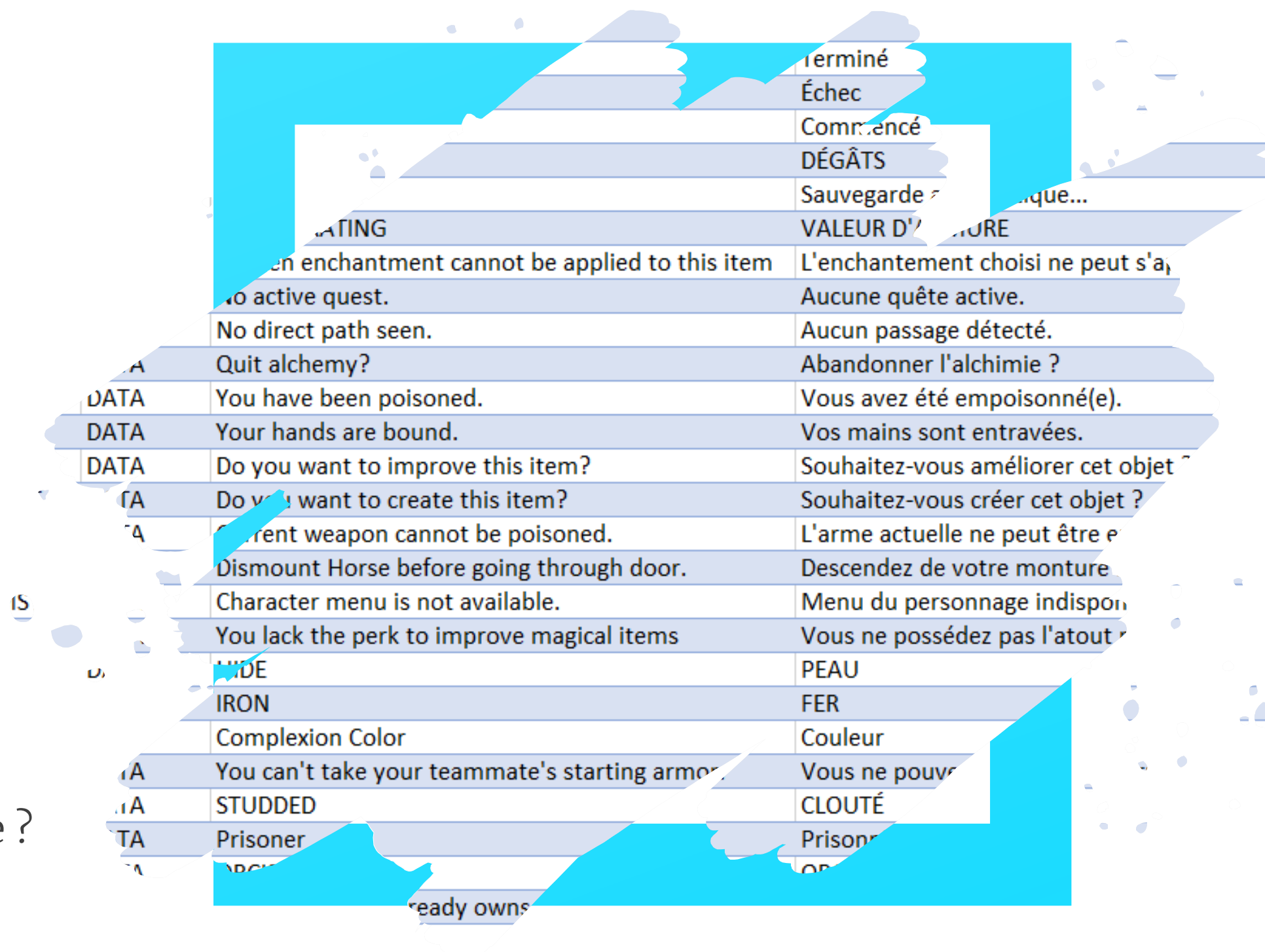
★ Raison n° 1

Des corpus

★ Raison n° 2

Une envie de collaborer sur un projet commun

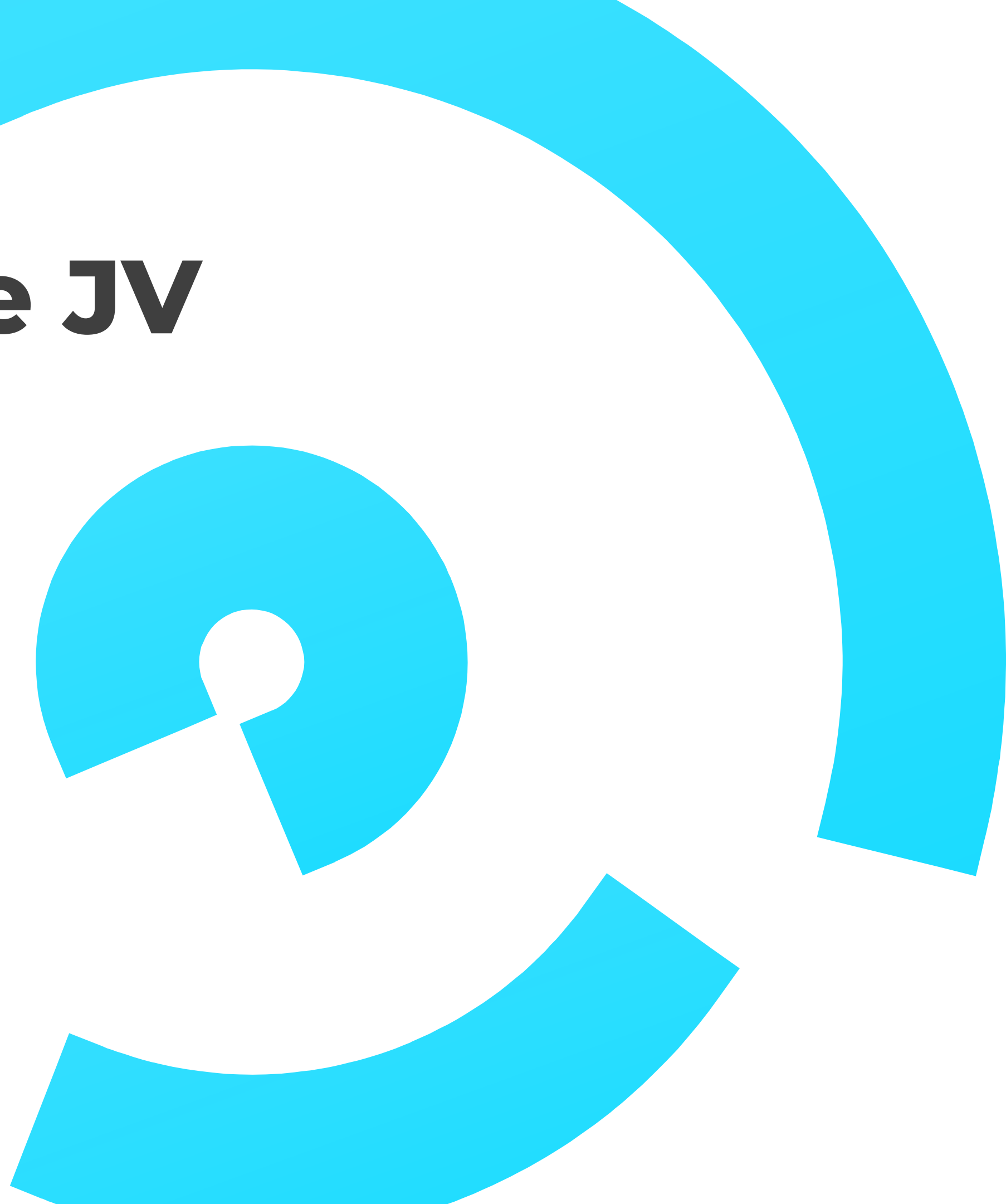
➡ Hé, et si on passait tous ces corpus à la machine ?



Terminé	Échec	Commencé
DÉGÂTS	Sauvegarde	VALEUR D'
en enchantment cannot be applied to this item	L'enchantement choisi ne peut s'a	Aucune quête active.
No direct path seen.	Aucun passage détecté.	Abandonner l'alchimie ?
Quit alchemy?	Abandonner l'alchimie ?	Vous avez été empoisonné(e).
DATA You have been poisoned.	DATA Your hands are bound.	DATA Do you want to improve this item?
DATA Do you want to create this item?	DATA Current weapon cannot be poisoned.	DATA Dismount Horse before going through door.
Character menu is not available.	Menu du personnage indispon	You lack the perk to improve magical items
Vous ne possédez pas l'atout	PEAU	FER
Complexion Color	Couleur	You can't take your teammate's starting armor
STUDDER	CLOUTÉ	Prisoner
ready owns		

Les spécificités de la traduction de JV

- [Voir la présentation de Pierre-Yves Houlmont]
- Traditionnellement :
 - la fragmentation des textes ;
 - la présence de variable ;
 - le manque de contexte ;
 - l'utilisation de sous-titres (Bernal-Merino 2007, 2015).
- Plus particulièrement :
 - la multimodalité ;
 - l'association de plusieurs sémiotiques ;
 - les (dis)jonctions narratives ;
 - les différends culturels ;
 - les stratégies de traduction (Houlmont 2021).



Le corpus

- Il existe des amateur·e·s. Il existe des créateurs et créatrices amateur·e·s. Il existe aussi des [traducteurs et traductrices amateur·e·s](#).
- Ces amateur·e·s [corrigent des traductions](#) (officielles) jugées de mauvaise qualité. Ils/elles [proposent des traductions](#) pour les jeux et le contenu qui ne sera jamais traduit (mods). Ils/elles [créent des logiciels](#) pour aider celles et ceux qui se lancent dans ces projets (plusieurs milliers à centaines de milliers de phrases ; 260 000 pour le mod *Beyond Skyrim*).
- Sans le savoir, ils/elles créent et utilisent des outils de TAO, en s'aidant de mémoires de traduction.
- ➔ Et si on faisait pareil et qu'on dépouillait les traductions de quelques jeux pour [entraîner un système uniquement sur des données JV](#) ?

Le corpus

	Game	Developer	Size
1.	<i>Baldur's Gate</i>	BioWare, 1998	19 K phrases
2.	<i>Baldur's Gate II: Shadows of Amn</i>	BioWare, 2000	62 K phrases
3.	<i>Darkest Dungeon</i>	Red Hook Studios, 2016	8 K phrases
4.	<i>Divinity: Original Sin II</i>	Larian Studios, 2017	85 K phrases
5.	<i>Fallout</i>	Black Isle Studios, 1997	26 K phrases
6.	<i>Fallout 2</i>	Black Isles Studios, 1998	56 K phrases
7.	<i>Fallout 3</i>	Bethesda Game Studios, 2008	49 K phrases
8.	<i>Fallout: New Vegas</i>	Obsidian Entertainment, 2010	64 K phrases
9.	<i>Fallout 4</i>	Bethesda Game Studios, 2015	126 K phrases
10.	<i>Planescape: Torment</i>	Black Isle Studios, 1999	39 K phrases
11.	<i>Pillars of Eternity</i>	Obsidian Entertainment, 2015	48 K phrases
12.	<i>Star Wars: Battlefront II</i>	Pandemic Studios, 2005	4 K phrases
13.	<i>The Elder Scrolls III: Morrowind</i>	Bethesda Softworks, 2002	37 K phrases
14.	<i>The Elder Scrolls IV: Oblivion</i>	Bethesda Game Studios, 2006	40 K phrases
15.	<i>The Elder Scrolls V: Skyrim</i>	Bethesda Game Studios, 2011	70 K phrases
16.	<i>The Witcher 2: Assassins of Kings</i>	CD Projekt, 2011	32 K phrases
17.	<i>The Witcher 3: Wild Hunt</i>	CD Projekt RED, 2015	78 K phrases
18.	<i>Torment: Tides of Numenéra</i>	inXile Entertainment, 2017	49 K phrases
19.	<i>Ultima VII: The Black Gate</i>	Origin Systems, 1992	12 K phrases
20.	<i>Ultima VIII: Pagan</i>	Origin Systems, 1994	6 K phrases
21.	<i>Ultima IX: Ascension</i>	Origin Systems, 1999	9 K phrases
22.	<i>Wasteland 2</i>	inXile Entertainment, 2014	37 K phrases
Total			956 K phrases

Le corpus

- Sortir les fichiers EN et FR de chaque jeu.
- Aligner les traductions dans un tableur.
- Premier traitement dans *Excel*.
(éliminer les données inutiles, harmoniser la typo, corriger l'encodage...)
- Transformer les documents en mémoires de traduction.
- Deuxième traitement dans un logiciel de nettoyage de bitextes.
(segments vides, segments non traduits, doublons...)
- Séparer les textes EN et FR (toujours alignés) dans des fichiers .txt.
- Troisième traitement avec des scripts de nettoyage de corpus.
(traductions dans une autre langue que le FR, segments mal alignés...)

Pourquoi la TA ?

- La TA fait clairement ressortir les **spécificités** et **difficultés** dans ce domaine particulier.
- Elle permet également d'aborder les problématiques de **traduction amateur**, **d'ergonomie** et de facilitation du travail des traducteurs et traductrices pros, mais aussi de **valorisation de la traduction** chez les studios qui ont tendance à voir cela comme quelque chose de dispensable (« *All your base are belong to us* »).
- Il y aurait également des retombées directes, soit pour **faciliter la traduction de contenu amateur**, soit pour les **professionnel·les** qui sont en permanence confronté·es au manque de contexte et aux problèmes de terminologie dans les textes qui leur sont soumis.

Comment la TA ?

Prétraitement :

- *The Greybeards have summoned the Dragonborn to High Hrothgar.*
- *There's a dragon. Here. In Dragonsreach.*
- *In the ancient tongue, you are Dovahkiin - Dragonborn!*

- Tokenisation

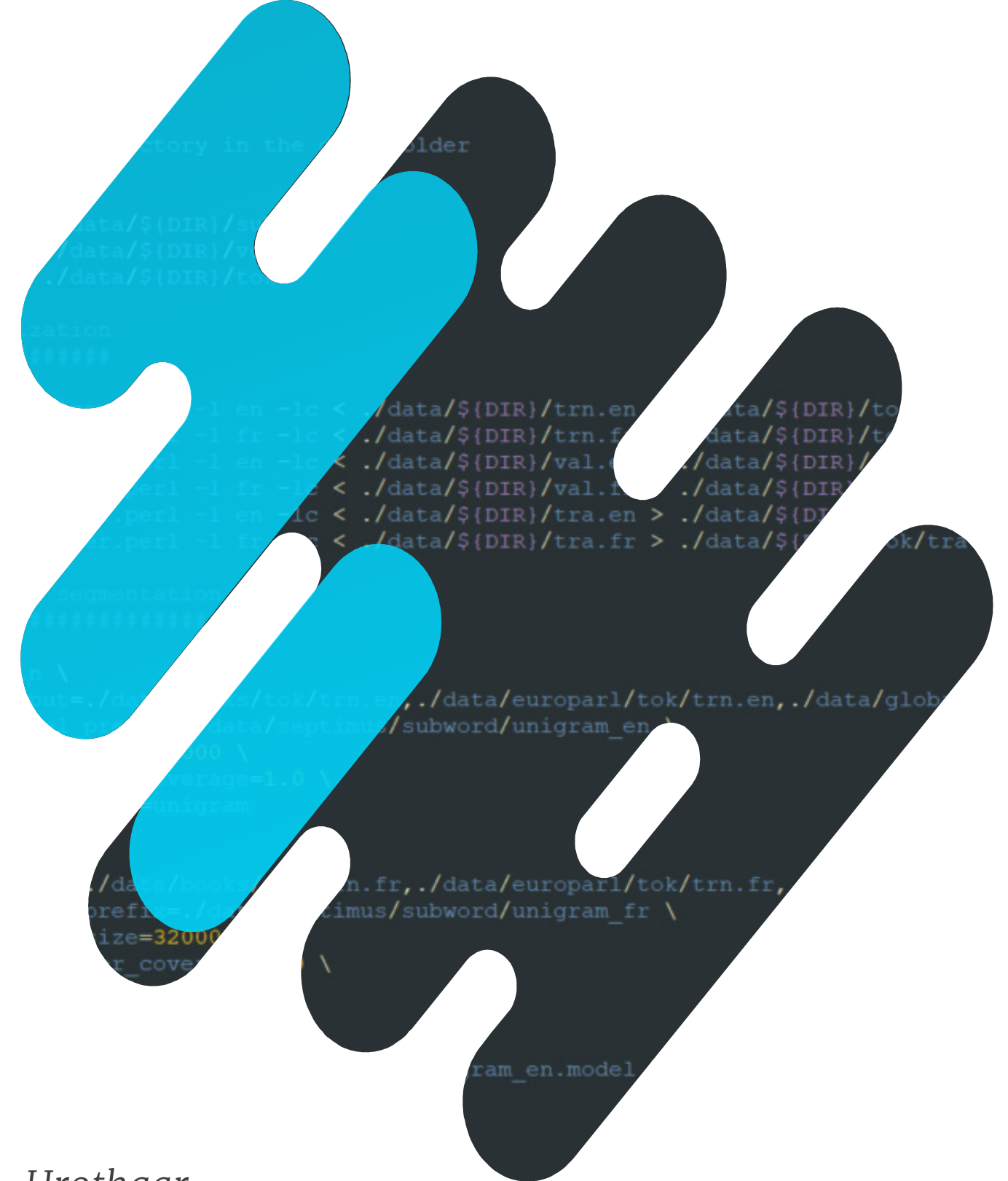
- *The Greybeards have summoned the Dragonborn to High Hrothgar .*
- *There 's a dragon . Here . In Dragonsreach .*
- *In the ancient tongue , you are Dovahkiin - Dragonborn !*

- Segmentation en sous-mots

- *_The _Greybeard s _have _summon ed _the _Dragon born _to _High _Hrothgar _.*
- *_There _'s _a _dragon _ . _Here _ . _In _Dragons reach _.*
- *_In _the _ancient _tongue _ , _you _are _Dovah kiin _- _Dragon born _ !*

- Création d'un vocabulaire utilisable par la machine

Segmentation dite « *language-agnostic* », fonctionne même pour les langues inventées !




```
train:
  path_src: ./data/skyrim/tok/trn.e
  path_tgt: ./data/skyrim/tok/trn.fr
  transforms: [filtertoolong, sentencepiece]
  weight: 1
valid:
  path_src: ./data/skyrim/tok/val.en
  path_tgt: ./data/skyrim/tok/val.fr
  transforms: [filtertoolong, sentencepiece]
src_seq_length: 200
tgt_seq_length: 200
src_subword_model: ./data/skyrim/subword/ur_en.model
tgt_subword_model: ./data/skyrim/subword/ur_fr.model
src_subword_vocab: ./data/skyrim/subword/ur_en.vocab
tgt_subword_vocab: ./data/skyrim/subword/ur_fr.vocab
src_subword_nbest: 20
tgt_subword_nbest: 20
src_subword_alpha: 0.1
tgt_subword_alpha: 0.1

# Training parameters:
batch_type: "tokens"
batch_size: 4096
valid_batch_size: 16
batch_size_multiple: 1
max_generator_batches: 0
accum_count: [3]
accum_steps: [0]
train_steps: 200000
valid_steps: 5000
report_every: 100
save_checkpoint_steps: 10000
queue_size: 10000
bucket_size: 32768

# Optimization
model_dtype: "fp32"
optim: "adam"
```

Comment la TA ?

Entraînement :



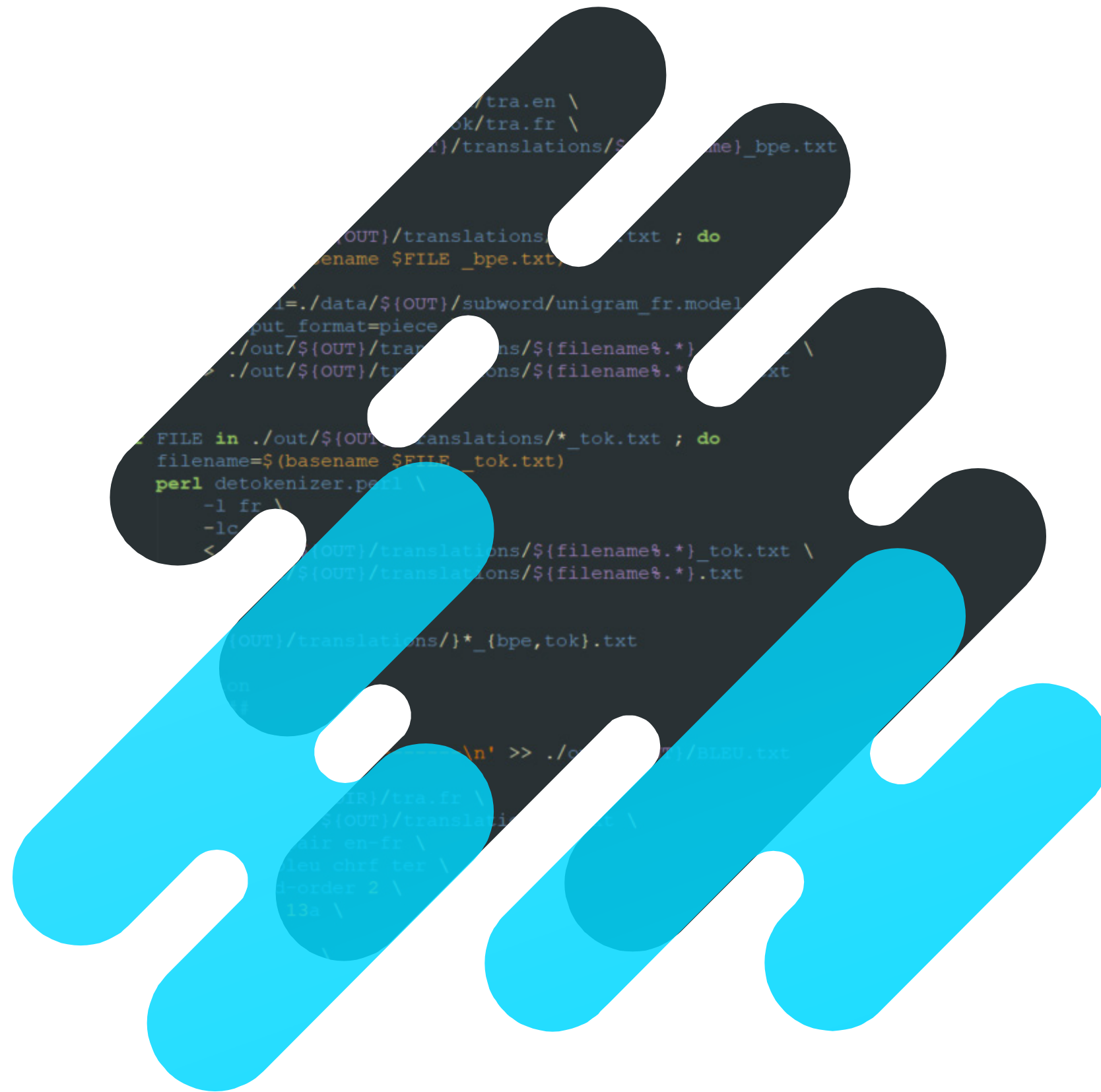
Klein et coll.
(2020)

- OpenNMT (open source neural machine translation system)
- Séparer le corpus en trois : entraînement, validation, test
- Choisir une architecture : LSTM / Transformer
- Définir les paramètres d'entraînement (durée, taille, opti, suivi)
- Lancer et attendre (longtemps)

Comment la TA ?

Traduction et évaluation :

- Traduire chaque set de test :
 - *Fallout 4*
 - *Skyrim*
- Comparer avec la référence
- Calculer un score automatique pour évaluer la distance avec la référence



Résultats

- Évaluation sacreBLEU sur trois métriques : BLEU, chrF et TER (testé en ligne en décembre 2020).

	BLEU ↑	chrF2++ ↑	TER ↓
Google Tr.	27.75	48.25	66.75
DeepL	29.27	50.04	61.26
Custom	37.14	55.80	53.32

The Elder Scrolls V: Skyrim (Bethesda Game Studios, 2011)

	BLEU ↑	chrF2++ ↑	TER ↓
Google Tr.	26.05	45.35	72.39
DeepL	27.60	47.04	67.94
Custom	31.18	48.80	62.96

Fallout 4 (Bethesda Game Studios, 2015)

- Le système maison affiné sur les jeux vidéo obtient (sans surprise) les meilleurs résultats.
- Mais surtout, il obtient de très bons résultats, avec pourtant très peu de données (1 M) !
- Un peu moins avec *Fallout*, dans lequel on trouve un registre très cru et familier.

Exemples



Traduction des termes spécifiques à la saga

Source

The *Greybeards* have arranged a truce between the *Stormcloaks* and *Imperials*. How long can such a thing last?

Thalmor in the *Ratway*?
What's next, *Spriggans* in the *Bee and Barb*?

DeepL

Les *Barbes grises* ont conclu une trêve entre les *Manteaux d'orage* et les *Impériaux*.
Combien de temps une telle chose peut-elle durer ?

Thalmor dans le *Ratway* ? Et ensuite, *Spriggans* dans le *Bee and Barb* ?

TA maison

Les *Grisés-barbes* ont conclu une trêve entre les *Sombrages* et les *Impériaux*.
Combien de temps cela peut-il durer ?

Les *Thalmor* dans la *Souricière* ? Et maintenant, qu'est-ce qu'il y a, les *spriggans* dans le *Dard de l'abeille* ?

Référence

Les *Grisés-barbes* ont arrangé une trêve entre les *Sombrages* et les *Impériaux*, mais combien de temps pourra-t-elle durer ?

Des *Thalmor* à la *Souricière* ? Et puis quoi, encore ? Des *spriggans* au *Dard de l'abeille* ?

Exemples



Traduction des termes spécifiques à la saga

Source

Hmm... There's a lot of groups in the Commonwealth that take people. *Raiders*, *Super Mutants*, the *Gunners*, and of course, there's the *Institute*.

It's *synth* detective, jackass.

DeepL

Hmm... Il y a beaucoup de groupes dans le Commonwealth qui prennent les gens. Les *Raiders*, les *Super Mutants*, les *Gunners*, et bien sûr, il y a *l'Institut*.

C'est un détective *de synthèse*, crétin.

TA maison

Hmm... Il y a beaucoup de groupes dans le Commonwealth qui prennent les gens. Des *pillards*, des *super mutants*, des *Artilleurs* et bien sûr, il y a *l'Institut*.

C'est un détective *synthétique*, imbécile.

Référence

Hmm... C'est pas les kidnappeurs qui manquent dans le Commonwealth. Les *pillards*, les *super mutants*, les *Artilleurs*... Sans parler de *l'Institut*.

C'est détective *synthétique*, andouille.

Exemples



Tendance à traduire littéralement

Source

That old woman who runs the orphanage has been murdered. Say, didn't I see you coming out of there...?

That is one... big hammer.

DeepL

La vieille femme qui dirige l'orphelinat a été assassinée. Dites, je ne vous ai pas vu sortir de là...

C'est un... gros marteau.

TA maison

La vieille femme qui dirige l'orphelinat a été assassinée. Dites, je ne vous ai pas vu sortir de là...

C'est un... grand marteau.

Référence

La vieille femme qui dirigeait l'orphelinat a été assassinée. Mais au fait, ce n'est pas de là que vous venez ?

En voilà un gros marteau !

Exemples



Tendance à traduire littéralement

Source	DeepL	TA maison	Référence
<p>Maybe some kind of improvised armor? Lot of hired guns do that <i>to look tough</i>.</p> <hr/> <p>The hazard suit is interesting. <i>Not many mercs can afford</i> something that fancy.</p>	<p>Peut-être une sorte d'armure improvisée ? Beaucoup de mercenaires font cela <i>pour avoir l'air dur</i>.</p> <hr/> <p>La combinaison de survie est intéressante. <i>Peu de mercenaires peuvent se permettre</i> une telle fantaisie.</p>	<p>Une sorte d'armure improvisée, peut-être ? Beaucoup d'armes à la main font ça <i>pour avoir l'air solide</i>.</p> <hr/> <p>La combinaison de risques est intéressante. <i>Peu de mercenaires peuvent se permettre</i> une chose pareille.</p>	<p>Une protection improvisée, peut-être ? Beaucoup de porte-flingues en ont, <i>ils pensent que ça leur donne l'air coriace</i>.</p> <hr/> <p>Intéressant, la combinaison. <i>Pas le genre de matos à la portée du premier mercenaire venu</i>.</p>

Exemples



Registre

Source

*In the ancient tongue, **you** are Dovahkiin - Dragonborn!*

*You... **you** were talking to that murderous savage. And then he escaped. What are **you** hiding?*

DeepL

*Dans l'ancienne langue, **tu** es Dovahkiin - Dragonborn !*

*Tu... **tu** parlais à ce sauvage meurtrier. Et puis il s'est échappé. Qu'est-ce que **tu** caches ?*

TA maison

*Dans la langue ancienne, **vous** êtes Dovahkiin, Enfant de dragon !*

*Vous... **vous** étiez en train de parler à ce sauvage assassin. Et puis, il s'est échappé. Que cachez-**vous** ?*

Référence

*Dans la langue ancienne, **vous** êtes Dovahkiin... Enfant de dragon.*

*Vous... **vous** avez parlé à ce meurtrier, et il s'est échappé. Qu'est-ce que **vous** cachez ?*

Exemples



Registre

Source

There ain't no test. That's the problem. You know what I say? No mayor. We let the city council run things from now on.

DeepL

Il n'y a pas de test. C'est ça le problème. Tu sais ce que je dis ? Pas de maire. Nous laissons le conseil municipal diriger les choses à partir de maintenant.

TA maison

Il n'y a pas de test. C'est ça le problème. Vous savez ce que je dis ? Pas de maire. On laisse le conseil municipal s'occuper de tout.

Référence

Y'a pas de test, c'est ça le problème. La seule solution, c'est de pas avoir de maire du tout. On laisse le conseil municipal gérer la ville.

Exemples



Locutions / tournures familières

Source	DeepL	TA maison	Référence
<p><i>Whoa, whoa, whoa!</i> <i>Watch the magic!</i></p> <hr/> <p><i>Some are saying the Psijic Monks have been seen in Skyrim. They're an ancient order. I had 'em figured for dead.</i></p>	<p><i>Whoa, whoa, whoa !</i> <i>Regardez la magie !</i></p> <hr/> <p><i>Certains disent que les Moines Psijic ont été vus à Skyrim. C'est un ordre ancien. Je les ai fait croire qu'ils étaient morts.</i></p>	<p><i>Woh, woh, woh ! Attention à la magie !</i></p> <hr/> <p><i>Certains disent que les moines de l'Ordre des Psijiques ont été vus en Bordeciel. C'est un ordre ancien. Je les avais crus morts.</i></p>	<p><i>Holà, holà ! Attention à la magie !</i></p> <hr/> <p><i>Certains disent que des moines de l'Ordre des Psijiques auraient été aperçus en Bordeciel. Je croyais qu'ils étaient tous morts.</i></p>

Exemples



Locutions / tournures familières

Source	DeepL	TA maison	Référence
<p><i>Ah, the mayor's okay in my book. He's kept the Ghoul's out, hasn't he?</i></p> <hr/>	<p><i>Ah, le maire est bien dans mon livre. Il a gardé les Ghoul's à l'écart, n'est-ce pas ?</i></p> <hr/>	<p><i>Ah, le maire est bien dans mon livre. Il a protégé les goules, pas vrai ?</i></p> <hr/>	<p><i>Moi, j'ai rien à lui reprocher, au maire. Il nous a protégés des goules, non ?</i></p> <hr/>
<p><i>Don't know. Whoever it was, they're no one now.</i></p> <hr/>	<p><i>Je ne sais pas. Qui que ce soit, il n'y a plus personne maintenant.</i></p> <hr/>	<p><i>Je ne sais pas. Qui que ce soit, ils ne sont plus personne.</i></p> <hr/>	<p><i>Je sais pas, personne que je connaissais en tout cas.</i></p> <hr/>
<p><i>Hold your horses...</i></p>	<p><i>Tenez vos chevaux...</i></p>	<p><i>Tenez vos chevaux...</i></p>	<p><i>Un instant...</i></p>

Exemples



Discours oralisé / raccourcis

Source

You know, I could have gone into that Sanctuary. Killed all those Dark Brotherhood types. But I was... sick that day.

For a dragon to attack Kynesgrove... Why? Why there? Is no place safe?

DeepL

Vous savez, j'aurais pu aller dans ce Sanctuaire. Tuer tous ces types de la Fraternité des Ténèbres. Mais j'étais... malade ce jour-là.

Pour qu'un dragon attaque Kynesgrove... Pourquoi ? Pourquoi là-bas ? Aucun endroit n'est sûr ?

TA maison

Vous savez, j'aurais pu entrer dans ce sanctuaire. J'ai tué tous ces types de la Confrérie noire. Mais j'étais... malade ce jour-là.

Pour qu'un dragon attaque le Bosquet de Kyne... Pourquoi ? Pourquoi là-bas ? Il n'y a pas d'endroit sûr ?

Référence

Vous savez, j'aurais pu aller à ce sanctuaire et tuer tous les agents de la Confrérie noire, mais j'étais... malade, ce jour-là.

Un dragon a attaqué le Bosquet de Kyne... Pourquoi ? Aucun lieu n'est-il donc sûr ?

Exemples



Ambiguïtés

	Source	DeepL	TA maison	Référence
Choix de dialogue	<i>Describe</i> Vault 111 <i>Bribe</i> (250 Caps)	<i>Décrire</i> le coffre-fort 111 <i>Corruption</i> (250 Caps)	<i>Décrire</i> l'Abri 111 <i>Corruption</i> (250 capsules)	<i>Décrire</i> l'abri 111 <i>Soudoyer</i> (250 capsules)
Ordre	<i>Take a seat.</i> <i>Give me a 'sec...</i>	<i>Prenez place.</i> <i>Donnez-moi une seconde...</i>	<i>Asseyez-vous.</i> <i>Donne-moi une seconde...</i>	<i>Asseyez-vous.</i> <i>Donnez-moi une seconde...</i>
Objectif	<i>Go to</i> Diamond City <i>Go to</i> Valentine's Detective Agency	<i>Aller</i> à Diamond City <i>Aller</i> à l'agence de détectives Valentine	<i>Aller</i> à Diamond City <i>Allez</i> à l'agence de détectives Valentine.	<i>Allez</i> à Diamond City. <i>Allez</i> à l'agence de détectives Valentine.

Exemples



Contamination des données (registre : *Fallout* vs *Skyrim*)

Source	DeepL	TA maison	Référence
<p><i>We were in a Vault when it happened. Vault 111. It was some kind of cryo facility.</i></p> <hr/> <p><i>Maybe we can reach an agreement? I just happen to have a pile of caps here...</i></p>	<p><i>Nous étions dans une chambre forte quand c'est arrivé. Voûte 111. C'était une sorte de cryo-station.</i></p> <hr/> <p><i>Peut-être pouvons-nous parvenir à un accord ? Il se trouve que j'ai une pile de casquettes ici...</i></p>	<p><i>Nous étions dans un abri quand c'est arrivé. L'abri 111. C'était une sorte de complexe cryogénique.</i></p> <hr/> <p><i>Peut-être pourrions-nous parvenir à un accord ? Il se trouve que j'ai justement un tas de capsules sur moi...</i></p>	<p><i>On était dans un abri quand c'est arrivé. L'abri 111, un genre de complexe cryogénique.</i></p> <hr/> <p><i>On pourrait peut-être s'arranger ? J'ai un joli tas de capsules qui me brûle les poches...</i></p>

Exemples



Contamination des données (registre : *Fallout vs Skyrim*)

Source

Hey! You're the one who killed Alain Dufont, up in Raldbthar. *Oh yeah*, word about that got around. Man was a *piece of bandit trash*, is what I hear.

Filthy she-hound Elisif got what she deserved. Let her serve as "High Queen" of Oblivion, for all I care.

DeepL

Hé ! C'est vous qui avez tué Alain Dufont, à Raldbthar. *Oh oui*, ça s'est su. L'homme était *une ordure de bandit*, c'est ce que j'ai entendu dire.

La *sale chienne* Elisif a eu ce qu'elle méritait. Qu'elle serve de "Haute Reine" d'Oblivion, pour ce que j'en ai à faire.

TA maison

Hé ! C'est vous qui avez tué Alain Dufont, à Raldbthar. *Ouais*, j'ai entendu parler de ça. L'homme était *une ordure de bandit*, à ce qu'on m'a dit.

Cette *pourriture* d'Élisif a eu ce qu'elle méritait. Qu'elle serve de "haute-reine d'Oblivion", je m'en moque.

Référence

Hé ! C'est vous qui avez tué Alain Dufont, à Raldbthar. *Oh, oui*, ce genre de nouvelle circule vite. Ce bandit était *vraiment un sale type*, à ce que j'en sais.

Cette *garce* d'Élisif n'a eu que ce qu'elle méritait. Elle n'a qu'à être la haute-reine d'Oblivion, maintenant. Ça lui fera les pieds.

Exemples



Lissage des insultes ? Non, au contraire!!

Source	DeepL	TA maison	Référence
<p><i>Kellogg is here. He's been taunting me over the loudspeakers. Time to confront this asshole, face-to-face.</i></p> <hr/> <p><i>Take me to that son of a bitch...</i></p> <hr/> <p><i>A Commonwealth mutt like him can track a man's scent for miles...</i></p>	<p><i>Kellogg est là. Il m'a nargué par les haut-parleurs. Il est temps d'affronter ce connard, face à face.</i></p> <hr/> <p><i>Amène-moi à ce fils de pute...</i></p> <hr/> <p><i>Un clébard de la République comme lui peut suivre l'odeur d'un homme sur des kilomètres...</i></p>	<p><i>Kellogg est ici. Il m'a provoqué par les haut-parleurs. Il est temps d'affronter cet enfoiré, face à face.</i></p> <hr/> <p><i>Conduisez-moi à ce fils de pute...</i></p> <hr/> <p><i>Un bâtard du Commonwealth comme lui peut pister l'odeur d'un homme à des lieues à la ronde...</i></p>	<p><i>Kellogg est ici. Il me provoque par les haut-parleurs de la base. Il est temps d'affronter cet enfoiré.</i></p> <hr/> <p><i>Retrouve cette ordure...</i></p> <hr/> <p><i>Les braves toutous du Commonwealth peuvent suivre une odeur sur des kilomètres...</i></p>

Exemples



Neutralisation du genre

Source

Stay out of trouble, [Argonian](#).

Stay out of trouble, [Nord](#).

Watch the skies, [traveler](#).

There was a dragon held captive in Whiterun, and you... you released it? By the gods, [woman](#), why?

DeepL

Reste en dehors des ennuis, [Argonien](#).

Restez en dehors des ennuis, [Nord](#).

Regarde le ciel, [voyageur](#).

Il y avait un dragon prisonnier à Whiterun, et vous... vous l'avez relâché ? Par les dieux, [femme](#), pourquoi ?

TA maison

Évitez les ennuis.

Évitez les ennuis, [Nordique](#).

Surveillez le ciel.

Il y avait un dragon emprisonné à Blancherive et vous... vous l'avez libéré ? Par les dieux, pourquoi ?

Référence

Évitez de vous attirer des ennuis.

Évitez de vous attirer des ennuis, [Nordique](#).

Pensez à surveiller le ciel.

Qu'il y avait un dragon captif à Blancherive et que vous l'avez délivré ? Mais pourquoi, par les dieux ?

Exemples



Contresens

Source

*He's kept the Ghouls out,
hasn't he?*

DeepL

*Il a gardé les Ghouls à
l'écart, n'est-ce pas ?*

TA maison

*Il a protégé les goules,
pas vrai ?*

Référence

*Il nous a protégés des
goules, non ?*

Exemples



Glissement de sens

Source

A mercantile group that has established ports all over Tamriel. They pretty much dominate the whole shipping industry.

*Keep your eyes on him, he's quite **crafty**.*

DeepL

Un groupe mercantile qui a établi des ports dans tout Tamriel. Ils dominant à peu près toute l'industrie du transport maritime.

*Gardez les yeux sur lui, il est assez **rusé**.*

TA maison

Un groupe de marchands qui a établi des ports dans tout Tamriel. Ils dominant toute l'industrie maritime.

*Gardez l'oeil sur lui, il est assez **habile**.*

Référence

Un groupe mercantile qui a établi des ports dans tout Tamriel. Ils dominant la quasi-totalité du marché de l'import-export.

*Gardez-le à l'oeil, c'est une **vipère**.*

Exemples



Juste pour le plaisir

Source

Follow Dogmeat

DeepL

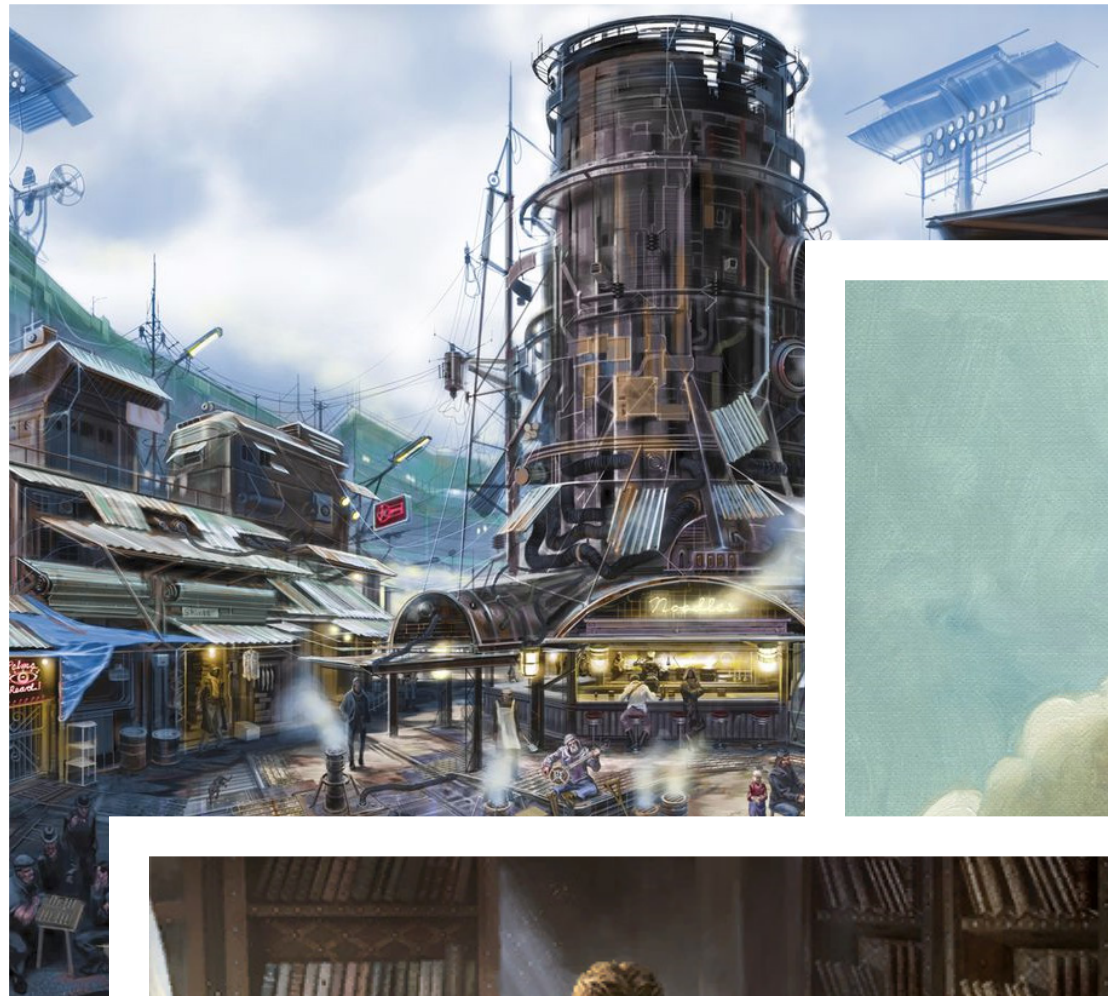
Suivre la viande de chien

TA maison

Suivez Canigou.

Référence

Suivez Canigou.



Conclusion

- Aide précieuse pour la termino, la cohérence entre les titres, la masse de travail...
- Révèle aussi des choix de traduction et des tendances très difficiles à voir pour l'œil humain.
- Possibilité de rendre le système plus robuste ;
- mais il faut tout de même rester très attentif aux glissements et aux contresens !
- Possibilité d'ajuster le poids de chaque sous-corpus pour régler les problèmes de registre, ou de prévoir l'étape de traduction pendant le développement, par exemple en ajoutant des balises qui seraient tout aussi utiles à l'humain (désambiguïsation).
- Nécessité de s'assurer que le travail de l'humain n'en sera pas desservi (qualité, ergonomie... > rapidité)

Discussion

D'autres utilités pour ces corpus ? Quelles implications si un tel outil venait à être « mis sur le marché » ?

Comment s'assurer que la TA ne nuise pas à la créativité ? Statut de la traduction ?

Traduction amateur vs professionnelle ?