

RÉSUMÉ

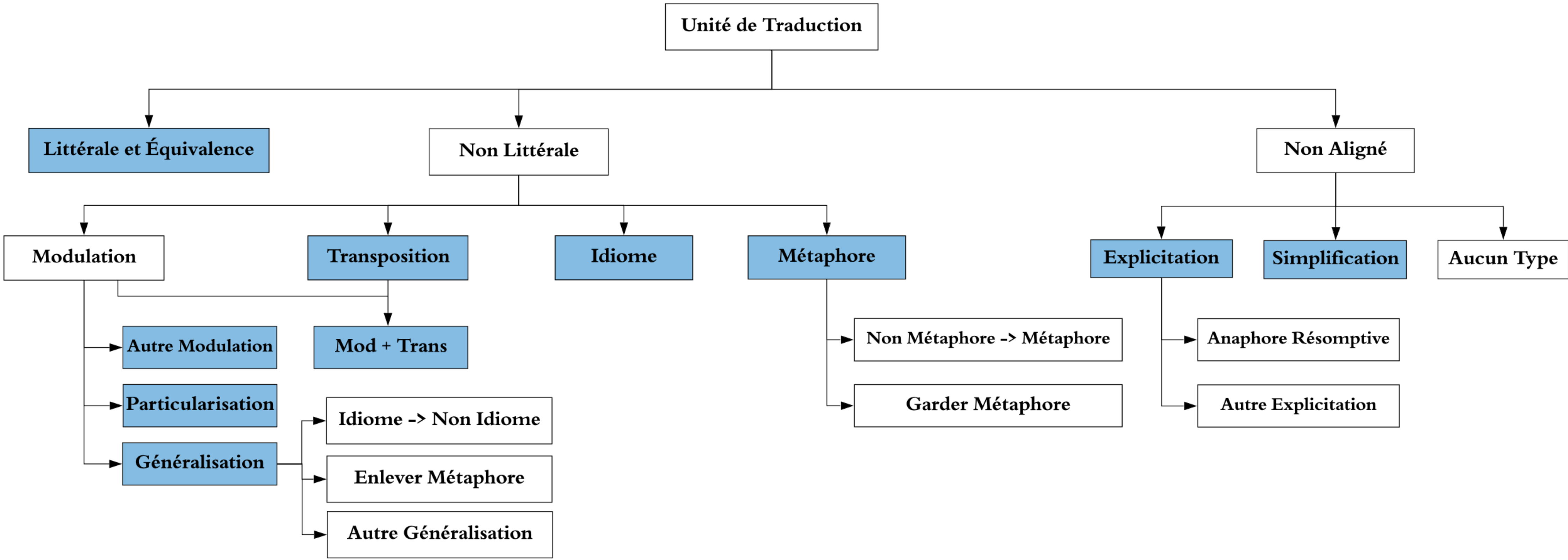
- Catégoriser les relations de traduction (littérale *vs* non littérale)
- Annotation manuelle dans un corpus parallèle trilingue (anglais, français, chinois) de présentations orales (*TED Talks*)

ANNOTATION

- Accord inter-annotateurs (Kappa de Cohen entre deux annotateurs)
 - EN-FR : 0,672, EN-ZH : 0,610
- Annotation à trois passes afin de garantir une bonne qualité

HIÉRARCHIE DES RELATIONS

- Fondée sur les théories explicitées dans (Chuquet & Paillard, 1989) et sur les phénomènes rencontrés pendant notre étude de corpus.
- Les relations sur fond bleu sont celles qui ont été annotées.



RELATIONS ET EXEMPLES

Littérale Équivalence	<i>What time is it?</i> → <i>Quelle heure est-il ?</i> <i>Birds of a feather flock together.</i> → <i>Qui se ressemble s'assemble.</i> <i>magic trick</i> → <i>tour de magie</i>
Modulation	<i>this is a completely unsustainable pattern</i> → <i>il est absolument impossible de continuer sur cette tendance</i>
Particularisation	<i>the director [said]</i> → <i>le directeur [déclara]</i>
Généralisation	<i>language loss</i> → <i>l'extinction du langage</i> <i>look carefully at</i> → <i>regardez</i> <i>trial and error</i> → <i>procéder par tâtonnements</i>
Transposition	<i>astorishingly inquisitive</i> → <i>dotée d'une curiosité stupéfiante</i>
Mod. + Trans.	<i>this is a people [who cognitively do not distinguish]</i> → <i>c'est un peuple [dont l'état des connaissances ne permet pas de faire la distinction]</i>
Idiome	<i>at any given moment</i> → <i>à un instant "t"</i>
Métaphore	<i>the Sun begins to [bathe] the slopes of the landscape</i> → <i>le soleil qui [inonde] les flancs de ce paysage</i>
Non aligné	
Explicitation	<i>feel their past in the wind</i> → <i>ressentent leur passé [souffler] dans le vent</i>
Simplification	<i>and you'll [suddenly] discover what it would be like</i> → <i>et vous découvrirez ce que ce serait</i>
Aucun type	<i>minus 271 degrees, colder than</i> → <i>moins 271 degrés, [ce qui] est plus froid</i>

ANALYSE CONTRASTIVE

	anglais	français	%EN tokens	anglais	chinois	%EN tokens
Littérale	2 796	2 896	68,63%	2 141	3 414	52,55%
Équivalence	247	310	6,06%	326	478	8,00%
Modulation	343	357	8,42%	386	549	9,47%
Transposition	106	150	2,60%	112	171	2,75%
Mod+Trans	128	126	3,14%	29	47	0,71%
Généralisation	58	37	1,42%	158	154	3,88%
Particularisation	132	180	3,24%	210	505	5,15%
Idiome	0	0	0,00%	0	0	0,00%
Métaphore	10	15	0,25%	6	10	0,15%
Simplification	74	-	1,82%	178	-	4,37%
Explicitation	-	46	-	-	459	-
Incertain	35	31	0,86%	128	238	3,14%
Tous les types	3 929	4 148	96,44%	3 674	6 025	90,18%
Aucun Type	145	171	3,56%	400	214	9,82%
Nb tokens total	4 074	4 319		4 074	6 239	

TRAVAUX EN COURS ET FUTURS

- Finaliser l'annotation (2 436 lignes de phrases parallèles trilingues)
- Annoter plus finement des grands blocs pour apprendre des patrons
- Détecter automatiquement les relations de traduction et intégrer ces informations pour améliorer la recherche de paraphrases strictes ou de paires de relation d'implication

MATRICE DE CONFUSION (CORPUS DE CONTRÔLE ANGLAIS-FRANÇAIS, NOMBRE D'INSTANCES)

	Équivalence	Littérale	Modulation	Transposit.	Mod+Trans	Généralisat.	Particulari.	Explicitation	Simplificat.	Idiome	Métaphore	Incertain
Équivalence	21	4	0	5	0	0	1	0	1	0	0	0
Littérale	27	1857	26	6	0	0	10	0	0	0	0	7
Modulation	4	8	37	7	1	1	3	0	2	0	0	0
Transposition	6	7	10	30	0	1	0	0	0	0	0	0
Mod+Trans	0	1	6	2	2	2	0	0	0	0	0	0
Généralisation	0	1	0	0	0	17	0	0	0	0	0	1
Particularisation	4	13	6	2	0	1	29	0	0	0	0	2
Explicitation	0	0	0	0	0	0	0	10	0	0	0	0
Simplification	0	0	0	0	0	0	0	0	40	0	0	0
Idiome	0	0	0	0	0	0	0	0	0	0	0	0
Métaphore	1	0	0	0	0	1	1	0	0	0	1	0
Incertain	1	8	2	2	0	4	0	0	1	0	0	4

REMERCIEMENT

Merci à Yaqiu Liu pour son annotation du corpus de contrôle anglais-chinois et pour son aide à l'analyse du corpus.