

Analyse sémantique AMR pour le français par transfert translingue

Jeongwoo Kang^{1,2} Maximin Coavoux¹ Cédric Lopez² Didier Schwab¹

(1) Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

(2) Emvista, 10 Rue Louis Breguet Bâtiment D, 34000, Montpellier, France

¹{prénom}. {nom}@univ-grenoble-alpes.fr

²{prénom}. {nom}@emvista.com

RÉSUMÉ

Abstract Meaning Representation (AMR) est un formalisme permettant de représenter la sémantique d'une phrase sous la forme d'un graphe, dont les nœuds sont des concepts sémantiques et les arcs des relations typées. Dans ce travail, nous construisons un analyseur AMR pour le français en étendant une méthode translingue zéro-ressource proposée par [Procopio et al. \(2021\)](#). Nous comparons l'utilisation d'un transfert bilingue à un transfert multi-cibles pour l'analyse sémantique AMR translingue. Nous construisons également des données d'évaluation pour l'AMR français. Nous présentons enfin les premiers résultats d'analyse AMR automatique pour le français. Selon le jeu de test utilisé, notre parseur AMR entraîné de manière zéro-ressource, c'est-à-dire sans donnée d'entraînement, obtient des scores Smatch qui se situent entre 54,2 et 66,0.

ABSTRACT

French AMR parsing with cross-lingual transfer learning

Abstract Meaning Representation (AMR) is a formalism for representing the semantics of a sentence with a graph of concepts connected with typed edges. In this work, we build a French AMR parser by extending a *zero-shot* cross-lingual method proposed by [Procopio et al. \(2021\)](#). We investigate the benefits of using bilingual data over multi-lingual data in cross-lingual AMR parsing. We also create evaluation datasets for French AMR parsing. With two evaluation datasets, we obtained respectively 54.2 and 66.0 Smatch score for the French AMR. Our French AMR parser is trained in a zero-shot setting, i.e. without French AMR training data.

MOTS-CLÉS : Analyse sémantique automatique, transfert translingue, seq2seq, AMR.

KEYWORDS: Semantic parsing, cross-lingual transfert, seq2seq, AMR.

1 Introduction

Abstract Meaning Representation ([Banarescu et al., 2013](#), AMR) est un formalisme permettant de représenter la sémantique d'une phrase sous la forme d'un graphe, dont les nœuds sont des concepts et les arcs des relations typées (figure 1). AMR a été initialement conçu pour analyser des phrases en anglais. Pourtant, [Damonte & Cohen \(2018\)](#) ont montré qu'AMR peut être utilisé également pour l'analyse sémantique multilingue, notamment pour l'espagnol, l'italien, l'allemand et le chinois. De plus, [Damonte & Cohen \(2018\)](#) ont construit des données d'évaluation AMR pour ces quatre langues par traduction du jeu d'évaluation anglais, et les ont diffusées. Ces données d'évaluation ont mené au développement d'analyseurs AMR translingues ([Blloshmi et al., 2020](#); [Uhrig et al., 2021](#); [Procopio et al., 2021](#)) exploitant

les données d’entraînement anglaises (seules données d’entraînement existantes). Les méthodes translingues consistent à utiliser la projection d’annotations sur des phrases traduites automatiquement en langue cible (Damonte & Cohen, 2018; Biloshmi *et al.*, 2020), l’apprentissage par transfert multitâche (Procopio *et al.*, 2021; Xu *et al.*, 2021) ou encore la distillation (Cai *et al.*, 2021).

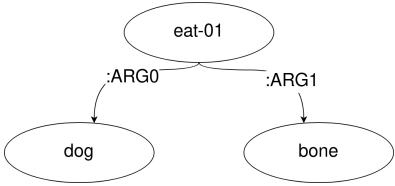


FIGURE 1 – Graphe AMR pour la phrase “The dog eats a bone.”

Notre objectif est de construire un parseur AMR pour le français, langue pour laquelle il n’existe pas de données AMR ni de travaux antérieurs, à l’exception de ceux de Vanderwende *et al.* (2015)¹. Nous reprenons l’approche de Procopio *et al.* (2021), basée sur un modèle séquence-à-séquence (seq2seq) multilingue pré-entraîné et affiné sur deux tâches : la prédiction de graphes sémantiques pour l’anglais, et la traduction automatique multilingue. Nous construisons par ailleurs notre propre jeu de test par projection d’annotations et traduction automatique suivie d’une vérification manuelle. En résumé nos contributions sont les suivantes :

- Nous reproduisons le travail de Procopio *et al.* (2021) et publions notre implémentation², qui est *de facto* la première implémentation disponible, Procopio *et al.* (2021) n’ayant pas publié leur code à ce jour.
- Nous construisons et publions le premier jeu d’évaluation AMR en français, obtenu par traduction automatique partiellement post-édité et projection d’annotations.
- Nous décrivons un système d’analyse AMR pour le français, par extension de Procopio *et al.* (2021) vers cette langue, et présentons les résultats de nos expériences d’apprentissage par transfert translingue. Nous obtenons des résultats intéressants (Smatch 54 et 66) sans donnée d’entraînement en français. À notre connaissance, notre travail est le premier à évaluer un modèle AMR sur le français.

2 Analyse AMR translingue

Speaking the Graph Languages (SGL) Nous utilisons l’approche SGL (Procopio *et al.*, 2021) fondée sur un modèle de séquence à séquence dont l’entrée est une phrase et la sortie est la forme linéarisée d’un graphe AMR (tableau 1). Procopio *et al.* (2021) adoptent un modèle de langue multilingue pré-entraîné : *mBART* (Liu *et al.*, 2020). Ils reformulent l’analyse AMR multilingue comme une tâche de traduction multilingue. Pendant l’entraînement, le modèle de traduction de *mBART* apprend à traduire des textes anglais en des graphes sémantiques AMR et UCCA (Abend & Rappoport, 2013, *Universal Conceptual Cognitive Annotation*).³ En plus de ces deux formalismes de graphes sémantiques, le modèle apprend la

1. Vanderwende *et al.* (2015) utilisent un analyseur sémantique pour extraire la forme logique d’une phrase et appliquent un ensemble de règles de conversion vers AMR. Ils ne proposent pas d’évaluation empirique de leur méthode pour le français. À l’inverse, nous générons un graphe AMR directement à partir d’une phrase d’entrée et évaluons notre système.

2. Le code est accessible sur <https://github.com/Emvista/French-Amr-Parser.git> et également archivé sur <https://doi.org/10.5281/zenodo.7944999>.

3. UCCA est un formalisme sémantique inspiré par la Théorie Linguistique de Base (Dixon, 2009, 2010, *Basic Linguistic Theory*). Contrairement à l’AMR pour lequel il n’y a pas d’alignement entre les concepts et les tokens, UCCA donne des représentations ancrées, c’est-à-dire que chaque token d’une phrase est une feuille du graphe sémantique.

traduction multilingue avec quatre corpus parallèles : EN↔ES, EN↔DE, EN↔IT, EN↔ZH (tableau 1). Notons que, contrairement aux données de traduction multilingue, pour l’analyse sémantique, la direction de la traduction est unidirectionnelle, depuis texte source anglais vers les graphes sémantiques. Puisque SGL fonctionne avec deux formalismes d’analyse sémantique, AMR et UCCA, leur approche n’est pas seulement multilingue mais aussi multi-formalismes.

entrée : <en>I have a dog.	sortie : <es>Tengo un perro.
entrée : <it>sono andato a letto.	sortie : <en>I went to bed.
entrée : <en>The boy wants to go.	sortie : <amr>(w / want-01 :ARG0 (b/boy) :ARG1 (g/go-02 :ARG0 b))
entrée : <en>John kicked his ball.	sortie : <ucca> [<root_0> H [<H_0> A [<A_0> T [John]] P [<P_0> T [kicked]] A [<A_0> E [<E_0> T [his]] C [<C_0> T [ball]]]]]

TABLEAU 1 – Exemples de données d’entraînement.

Adaptation de l’approche SGL au français Le choix de cette méthode a deux motivations. Tout d’abord, il s’agit d’une approche zéro-ressource qui ne nécessite pas de données annotées en AMR pour la langue cible et permet donc de pallier leur absence pour le français. À l’entraînement, le modèle apprend l’analyse en AMR, uniquement à partir de textes anglais. Au moment de l’inférence, il parvient à générer des graphes AMR à partir de textes dans des langues pour lesquelles le modèle n’a pas vu de données AMR correspondantes. Ceci est rendu possible par la capacité multilingue de *mBART* ainsi que par la tâche de traduction multidirectionnelle, pour laquelle le modèle est affiné. Deuxièmement, SGL est facile à étendre à une nouvelle langue. Pour permettre au modèle de générer des AMR pour des textes en français, il suffit d’ajouter un corpus parallèle FR ↔ EN aux données d’entraînement. Nous utilisons notre propre implémentation de SGL pour l’adapter au français.

Multilingue vs. Bilingue Des travaux (Zhang *et al.*, 2020; Conneau *et al.*, 2020) montrent l’avantage d’utiliser des données multilingues pour entraîner un modèle de traduction. Tang *et al.* (2021) montrent que le modèle de traduction multilingue est plus performant que le modèle de traduction bilingue quand il s’agit d’une traduction *many-to-one* où le modèle apprend à encoder en plusieurs langues et à décoder en une seule langue, notamment l’anglais. D’autre part, le transfert négatif (Wang *et al.*, 2019) est également un problème connu dans les modèles multilingues (Wang *et al.*, 2020; Conneau *et al.*, 2020) : les tâches multilingues entravent les performances individuelles pour chaque tâche. La méthode SGL est potentiellement sujette à ce problème, étant formalisée comme une tâche de traduction multilingue.

Nous souhaitons comparer l’usage d’un seul modèle multilingue à celui de multiples modèles bilingues (un pour chaque langue cible). Pour répondre à cette question, nous entraînons un parseur AMR français dans deux contextes : **Bilingue_FR** où le modèle est entraîné avec des données cibles uniquement (corpus AMR/UCCA en anglais et corpus parallèle EN↔FR) et **Multilingue** où le modèle est entraîné avec des données cibles mais aussi avec des données multilingues (corpus AMR/UCCA en anglais et corpus parallèles EN↔ES, DE, IT, ZH, FR). Concernant le modèle bilingue, nous entraînons également un modèle espagnol (**Bilingue_ES**) et un modèle allemand (**Bilingue_DE**) de la même manière avec un corpus parallèle pour chaque langue cible. Cela nous permettra de comparer plusieurs modèles bilingues de manière plus complète (section 4). Une comparaison similaire entre modèles bilingues et multilingues est décrite par Biloshmi *et al.* (2020). En revanche, notre perspective est différente : Biloshmi *et al.* (2020) utilisent des données cibles synthétiques, alors que nous sommes en contexte zéro-ressource.

3 Données pour l'évaluation de l'analyse AMR en français

Pour évaluer notre système, nous construisons des données d'évaluation pour le français par deux méthodes : la projection d'annotations et la traduction automatique.

Le petit prince Le corpus du *Petit Prince* d'Antoine de Saint-Exupéry est annoté pour l'analyse AMR en anglais et est disponible sur le site d'AMR⁴. Le livre étant initialement écrit en français, nous pouvons obtenir des paires [phrase française, graphe AMR] en alignant chaque phrase anglaise du corpus AMR sur son équivalent français. Notons que les textes français et anglais doivent être alignés au niveau de la phrase ou au-delà du niveau de la phrase puisqu'un graphe AMR correspond à une ou plusieurs phrases dans le corpus *Le Petit Prince*. À cette fin, nous utilisons un outil d'alignement de textes pré-entraîné, *LF Aligner* d'Andras Farkas⁵ et nous post-éditons les erreurs d'alignement manuellement. À la suite de la projection des annotations, nous avons obtenu 1 562 graphes AMR pour le français. Nous appelons ce jeu de données FR_LPP_GOLD.

AMR-2.0 Nous obtenons également un deuxième jeu de données d'évaluation en traduisant le texte source anglais du jeu de test AMR-2.0 (LDC2017T10)⁶ vers le français. Comme nous utilisons la traduction automatique DeepL⁷ sans la post-éditer manuellement, nous appelons ce jeu de données FR_SILVER. Nous évaluons empiriquement la fiabilité des résultats sur ces données dans le paragraphe suivant.

Examen de fiabilité des données de test : FR_SILVER Nous avons examiné les 200 premiers⁸ exemples de FR_SILVER et post-édité la traduction manuellement. Ainsi, 56 phrases ont été corrigées sur 200 lors de la post-édition (Correspondance Exacte : 0.72). Nous avons considéré ces 200 exemples post-édités comme la traduction de référence et évalué les traductions correspondantes de FR_SILVER par rapport à ces données avec le score BLEU (Papineni *et al.*, 2001). Le score BLEU de cet échantillon de traduction est égal à 0,89. Ensuite, nous avons examiné l'effet de l'utilisation de FR_SILVER sur l'évaluation d'analyses AMR. Pour cela, nous avons pris deux modèles différents, **Multilingue** et **Bilingue_FR** et avons évalué chaque modèle avec deux jeux de données : 200 exemples de FR_SILVER ainsi que ses correspondants post-édités. L'objectif de cette expérience est de comparer les deux scores Smatch et de considérer la différence comme un indicateur de la fiabilité de FR_SILVER en tant que jeu de test. Nous avons reproduit l'expérience en espagnol et en allemand en traduisant les données d'évaluation à l'aide de DeepL (ES_SILVER, DE_SILVER). Puis, nous avons évalué les modèles entraînés deux fois, sur les données d'évaluation originelles et les données automatiquement traduites (tableau 2). La différence entre les deux scores Smatch en allemand est égale à 4,5 % en moyenne pour les deux modèles et pour l'espagnol à 2,4 %. Pour le français, la différence est de $\pm 1\%$, ce qui est dans la marge d'erreur. En raison de cet écart insignifiant pour le français, nous n'incluons pas ces résultats dans la discussion approfondie dans 4 et considérons FR_SILVER comme des données de test qui donnent une approximation relativement fiable pour évaluer nos modèles en français.

4. <https://amr.isi.edu/download.html>

5. Nous utilisons une version en ligne : <http://phraseotext.univ-grenoble-alpes.fr/webAlignToolkit/>.

6. Nous avons choisi la version 2.0 pour être cohérent avec les données d'évaluation AMR multilingues (Damonte & Cohen, 2018).

7. DeepL API sur <https://www.deepl.com/>

8. Les exemples redondants ou les phrases trop courtes telles que "SOB" et "EMR" ainsi que des interjections telles que "Haha", "Braawwk!" sont exclues.

	Bilingue_FR	Multilingue
DE	61,1	63,3
DE_SILVER	63,3	64,6
ES	66,4	66,8
ES_SILVER	67,7	64,9
FR_SILVER_POSTEDIT	58,2	55,6
FR_SILVER	58,2	55,0

TABLEAU 2 – Score Smatch avec les données de référence et les données silver.

4 Expériences

Pour l’expérimentation, nous suivons principalement les spécifications du modèle mBART_{ft} de [Procopio et al. \(2021\)](#). Nous renvoyons les lectrices et lecteurs à l’article original pour obtenir des informations générales et détaillées sur la reproduction des expériences. Sauf indication contraire, nous utilisons la même configuration que le modèle mBART_{ft} .

Données et pré-traitement [Procopio et al. \(2021\)](#) reproduisent l’environnement de la campagne d’évaluation CoNLL2019 ([Koller et al., 2019](#)) et utilisent les données d’analyse sémantique UCCA qui sont distribuées aux participant-es de l’atelier. Ces données n’étant pas directement accessibles, nous avons téléchargé les données d’entraînement UCCA correspondantes directement sur le site officiel⁹. Ce corpus compte 8 775 paires phrase-graphe alors que pour mBART_{ft} , le corpus correspondant compte 6 572 paires. Au cours de la linéarisation des graphes UCCA, les auteurs utilisent l’étiquette T à la fois pour le nœud *Time* et le nœud *Terminal*. Cependant, pour distinguer ces deux concepts différents, nous utilisons T pour Time et Z pour Terminal. Nous employons l’API UCCA¹⁰ pour la linéarisation et la délinéarisation des graphes UCCA. À la différence de mBART_{ft} , nous incluons le corpus parallèle français-anglais dans les données d’entraînement, plus précisément, *ParaCrawl* ([Esplà-Gomis et al., 2019](#)). Nous appliquons le même filtrage de corpus parallèle que celui de mBART_{ft} , par exemple, le filtrage selon la longueur de la phrase ainsi que le ratio relatif du nombre de caractères entre la phrase d’entrée et la phrase cible.

Détails de l’entraînement et de l’évaluation Dans la version originale de mBART_{ft} , [Procopio et al. \(2021\)](#) utilisent 8 000 tokens comme taille effective de batch pendant l’entraînement. Pour nos expériences, nous utilisons des batches de 128 exemples. Pour l’échantillonnage des données, mBART_{ft} suréchantillonne les graphes sémantiques par rapport aux données de traduction multilingues. Nous appliquons la même probabilité que celle utilisée dans mBART_{ft} , soit 0,8 pour les graphes sémantiques et 0,2 uniformément répartie entre les corpus de traduction multilingues. Pour être plus précis, dans un cadre bilingue, nous attribuons 0,8 pour les graphes sémantiques et 0,2 pour le corpus parallèle $\text{EN} \leftrightarrow \text{FR}$. Dans le cadre multilingue, 0,8 pour l’analyse sémantique et 0,2 est uniformément répartie entre cinq corpus parallèles : $\text{EN} \leftrightarrow \text{ES}$, $\text{EN} \leftrightarrow \text{DE}$, $\text{EN} \leftrightarrow \text{IT}$, $\text{EN} \leftrightarrow \text{ZH}$, $\text{EN} \leftrightarrow \text{FR}$.

Résultats Nous présentons nos résultats dans le tableau 3 et incluons des résultats pour l’anglais, pour les 4 langues évaluées par [Damonte & Cohen \(2018\)](#), ainsi que pour le français. Nous incluons dans le

9. <https://universalconceptualcognitiveannotation.github.io/>

10. <https://ucca.readthedocs.io/en/latest/api.html>

tableau 3 les résultats d’autres modèles d’analyse sémantique multilingue : USeA (Orlando *et al.*, 2022), XL-AMR (Biloshmi *et al.*, 2020) et XLPT-AMR (Xu *et al.*, 2021). Nous utilisons Smatch (Cai & Knight, 2013) en tant que mesure d’évaluation pour l’analyse AMR. Un graphe AMR peut être exprimé comme un ensemble de triplets, dont chacun contient deux nœuds et un arc qui relie les deux. L’algorithme SMATCH calcule le chevauchement entre deux graphes en comptant les triplets en communs (F-score). Le tableau 3 montre l’avantage d’utiliser des données bilingues pour une langue cible par rapport à l’utilisation de données multilingues dans les cas majoritaires. Pour les deux jeux de test en français, FR_LPP_GOLD et FR_SILVER, **Bilingue_FR** surpasse **Multilingue** de respectivement 1,8 et 1,9 score Smatch. Pour l’allemand, **Bilingue_DE** obtient un score Smatch supérieur de 1,5 par rapport à **Multilingue**. Pour l’espagnol, cependant, **Multilingue** surpasse **Bilingue_ES**. Une hypothèse pour expliquer ce résultat est la particularité du corpus parallèle utilisé pour entraîner ce modèle. MultiUN (Windsor *et al.*, 2010), le corpus parallèle utilisé pour l’espagnol, a un style d’écriture plus formel que Paracrawl qui est utilisé pour le français et l’allemand. La majorité des données AMR proviennent de forums ou de discussions en ligne dont le style d’écriture est relativement familier et moins formel. Par conséquent, **Multilingue** a probablement bénéficié des corpus français et allemand qui sont plus similaires aux données AMR, tandis que **Bilingue_ES** a probablement eu davantage de données moins représentatives.

	EN*	DE	ES	IT	ZH	FR_LPP_GOLD	FR_SILVER
Multilingue _{zero}	80,3	63,3	66,8	65,8	52,9	52,4	64,1
Bilingue_FR _{zero}	80,0	61,1	66,4	65,9	48,4	54,2	66,0
Bilingue_ES _{zero}	79,9	61,5	66,1	65,2	48,2	51,9	63,3
Bilingue_DE _{zero}	80,2	64,8	65,0	64,2	48,9	50,0	61,8
USeA _{zero} (Orlando <i>et al.</i> , 2022)	78,6	58,8	62,1	60,2	38,3	46,5	61,1
SGL _{zero} (Procopio <i>et al.</i> , 2021)	81,2	65,8	69,2	69,6	54,8	-	-
SGL (Procopio <i>et al.</i> , 2021)	-	69,8	72,4	72,3	58,0	-	-
XL-AMR _{zero} (Biloshmi <i>et al.</i> , 2020)	-	32,7	39,1	37,1	25,9	-	-
XL-AMR (Biloshmi <i>et al.</i> , 2020)	-	53,0	58,0	58,1	41,5	-	-
XLPT-AMR (Xu <i>et al.</i> , 2021)	-	70,4	71,7	70,8	-	-	-

TABLEAU 3 – Score Smatch pour l’analyse AMR multilingue. Les scores sont tels que rapportés dans les articles originaux à l’exception des scores français pour USeA qui sont nouvellement ajoutés dans notre travail. Nous avons exploité l’API fournie d’USeA pour effectuer le test sur le français. ZERO indique les modèles zéro-ressource (aucun graphe AMR de la langue cible utilisé lors de l’apprentissage). L’évaluation sur l’anglais n’est pas zéro-ressource dans aucun cas et est ainsi marqué avec *.

Limites Alors que les données FR_SILVER donnent une bonne approximation pour évaluer le modèle français en l’absence de données annotées, elles peuvent conduire à un biais pour évaluer le modèle translingue. La traduction automatique a souvent tendance à être similaire à l’anglais en matière d’ordre de mots et de choix de termes. Par conséquent, utiliser FR_SILVER comme données d’évaluation peut biaiser en faveur du modèle translingue qui est entraîné principalement sur des données en anglais. Il convient également de noter que nous n’avons pas pu reproduire fidèlement les résultats rapportés par Procopio *et al.* (2021). En reproduisant le résultat d’évaluation de mBART_{fr}, c’est-à-dire sans les données du français, notre modèle a montré un écart allant de 1,1% pour l’analyse AMR anglaise à 5,3% pour l’analyse AMR italienne. En l’absence du code original, il est difficile d’enquêter sur l’origine exacte de cette lacune. L’une des expli-

cations probables réside dans notre propre implémentation ainsi que la configuration de l'expérimentation mentionnée dans la section 4, qui est légèrement différente de celle de [Procopio et al. \(2021\)](#). Notre objectif, cependant, est de comparer l'effet des données multilingues à celui des données bilingues dans l'analyse AMR translingue. Nous estimons que les différences dans les détails de mise en œuvre et les résultats de l'entraînement par rapport au travail original n'affectent pas de manière importante notre enquête comparative.

5 Conclusion

Dans cet article, nous présentons un parseur AMR français zéro-ressource entraîné par transfert translingue, ainsi qu'un premier jeu de données d'évaluation pour l'analyse AMR en français. Nos résultats montrent que l'utilisation de données bilingues est plus bénéfique que l'utilisation de données multilingues pour construire un parseur AMR français avec l'approche zéro-ressource. Dans la mesure où les données cibles sont soigneusement choisies, cette observation pourrait être généralisée aux analyses AMR dans d'autres langues.

Remerciements

Nous tenons à remercier les relecteurices anonymes ainsi que Adrien Pupier, Bastien Giordano, Kevin Cousot et Sylvain Verdy pour leurs suggestions constructives qui nous ont permis d'enrichir l'article. Nous tenons à remercier également le premier auteur de [Procopio et al. \(2021\)](#) pour des échanges informatifs qui nous ont permis de reproduire leur travail avec le plus de détails possible. Ces travaux ont bénéficié d'un accès aux moyens de calcul de l'IDRIS au travers de l'allocation de ressources 2023-AD011012853R1 attribuée par GENCI.

Références

- ABEND O. & RAPPOPORT A. (2013). Universal conceptual cognitive annotation (ucca). p. 228–238.
- BANARESCU L., BONIAL C., CAI S., GEORGESCU M., GRIFFITT K., HERMIJAKOB U., KNIGHT K., KOEHN P., PALMER M. & SCHNEIDER N. (2013). Abstract meaning representation for sembanking. p. 178–186.
- BLOSHMI R., TRIPODI R. & NAVIGLI R. (2020). Xl-amr : Enabling cross-lingual amr parsing with transfer learning techniques. p. 2487–2500 : Association for Computational Linguistics (ACL). DOI : [10.18653/V1/2020.EMNLP-MAIN.195](#).
- CAI D., LI X., HO J. C.-S., BING L. & LAM W. (2021). Multilingual amr parsing with noisy knowledge distillation. p. 2778–2789 : Association for Computational Linguistics. DOI : [10.18653/V1/2021.FINDINGS-EMNLP.237](#).
- CAI S. & KNIGHT K. (2013). Smatch : an evaluation metric for semantic feature structures.
- CONNEAU A., KHANDELWAL K., GOYAL N., CHAUDHARY V., WENZKE G., GUZMÁN F., GRAVE E., OTT M., ZETTMAYER L. & STOYANOV V. (2020). Unsupervised cross-lingual representation learning at scale. p. 8440–8451.
- DAMONTE M. & COHEN S. B. (2018). Cross-lingual abstract meaning representation parsing. volume 1, p. 1146–1155 : Association for Computational Linguistics (ACL). DOI : [10.18653/V1/N18-1104](#).

- DIXON R. (2009). *Basic linguistic theory volume 1 : Methodology*.
- DIXON R. (2010). *Basic linguistic theory volume 2 : Grammatical topics*.
- ESPLÀ-GOMIS M., FORCADA M. L., RAMÍREZ-SÁNCHEZ G. & HOANG H. (2019). Paracrawl : Web-scale parallel corpora for the languages of the eu. p. 118–119.
- KOLLER A., OEPEN S. & SUN W. (2019). Graph-based meaning representations : Design and processing. p. 6–11 : Association for Computational Linguistics (ACL). DOI : [10.18653/V1/P19-4002](https://doi.org/10.18653/V1/P19-4002).
- LIU Y., GU J., GOYAL N., LI X., EDUNOV S., GHAZVININEJAD M., LEWIS M. & ZETTLEMOYER L. (2020). Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, **8**, 726–742. DOI : [10.1162/tacl_a_00343](https://doi.org/10.1162/tacl_a_00343).
- ORLANDO R., CONIA S., FARALLI S. & NAVIGLI R. (2022). Universal semantic annotator : the first unified api for wsd, srl and semantic parsing. p. 20–25.
- PAPINENI K., ROUKOS S., WARD T. & ZHU W.-J. (2001). Bleu : a method for automatic evaluation of machine translation. p. 311 : Association for Computational Linguistics. DOI : [10.3115/1073083.1073135](https://doi.org/10.3115/1073083.1073135).
- PROCOPIO L., TRIPODI R. & NAVIGLI R. (2021). Sgl : Speaking the graph languages of semantic parsing via multilingual translation. p. 325–337 : Association for Computational Linguistics (ACL). DOI : [10.18653/V1/2021.NAACL-MAIN.30](https://doi.org/10.18653/V1/2021.NAACL-MAIN.30).
- TANG Y., TRAN C., LI X., CHEN P.-J., GOYAL N., CHAUDHARY V., GU J., FAN A., AI F. & AI A. (2021). Multilingual translation from denoising pre-training. p. 3450–3466. Multilingual finetuning » Bilingual finetuning.
- UHRIG S., GARCIA Y., OPITZ J. & FRANK A. (2021). Translate, then parse ! a strong baseline for cross-lingual AMR parsing. In *Proceedings of the 17th International Conference on Parsing Technologies and the IWPT 2021 Shared Task on Parsing into Enhanced Universal Dependencies (IWPT 2021)*, p. 58–64, Online : Association for Computational Linguistics. DOI : [10.18653/v1/2021.iwpt-1.6](https://doi.org/10.18653/v1/2021.iwpt-1.6).
- VANDERWENDE L., MENEZES A. & QUIRK C. (2015). An amr parser for english, french, german, spanish and japanese and a new amr-annotated corpus. p. 26–30 : Association for Computational Linguistics (ACL). DOI : [10.3115/V1/N15-3006](https://doi.org/10.3115/V1/N15-3006).
- WANG Z., DAI Z., POZOS B. & CARBONELL J. (2019). Characterizing and avoiding negative transfer. volume 2019-June, p. 11285–11294 : IEEE Computer Society. Introduction of the word "negative transfer / inference", DOI : [10.1109/CVPR.2019.01155](https://doi.org/10.1109/CVPR.2019.01155).
- WANG Z., LIPTON Z. C. & TSVETKOV Y. (2020). On negative interference in multilingual models : Findings and a meta-learning treatment. p. 4438–4450 : Association for Computational Linguistics (ACL). Negative inference, DOI : [10.18653/V1/2020.EMNLP-MAIN.359](https://doi.org/10.18653/V1/2020.EMNLP-MAIN.359).
- WINDSOR L. C., CUPIT J. G. & WINDSOR A. J. (2010). Automated content analysis across six languages. volume 14 : Public Library of Science. DOI : [10.1371/JOURNAL.PONE.0224425](https://doi.org/10.1371/JOURNAL.PONE.0224425).
- XU D., LI J., ZHU M., ZHANG M. & ZHOU G. (2021). Xlpt-amr : Cross-lingual pre-training via multi-task learning for zero-shot amr parsing and text generation. p. 896–907 : Association for Computational Linguistics (ACL). DOI : [10.18653/V1/2021.ACL-LONG.73](https://doi.org/10.18653/V1/2021.ACL-LONG.73).
- ZHANG B., WILLIAMS P., TITOV I. & SENNRICH R. (2020). Improving massively multilingual neural machine translation and zero-shot translation. p. 1628–1639. DOI : [10.18653/V1/2020.ACL-MAIN.148](https://doi.org/10.18653/V1/2020.ACL-MAIN.148).