Annotation d'expressions polylexicales verbales en arabe : validation d'une procédure d'annotation multilingue

Najet Hadj Mohamed, Cherifa Ben Khelil, Agata Savary, Iskander Keskes, Jean Yves Antoine, Lamia Hadrich Belguith

Annotation d'expressions polylexicales verbales en arabe : validation d'une procédure d'annotation multilingue

Najet Hadj Mohamed[1, 2] Cherifa Ben Khelil [1] Agata Savary[3] Iskandar Keskes[2] Jean-Yves Antoine[1] Lamia Belguith Hadrich[1]
(1) LIFAT, 64 avenue Jean Portalis, 37200 Tours, France
(2) MIRACL, Pôle technologique de Sfax, 3021 Ville Sfax, Tunisie
(3) LISN, Campus universitaire bât 507 rue du Belvedère, 91400 Orsay, France
najat.hadjmohamed@etu.univ-tours.fr, cherifa.benkhelil@univ-tours.fr, agata.savary@universite-paris-saclay.fr, iskandarkeskes@gmail.com, jean-yves.antoine@univ-tours.fr, lamia.belguith@fsegs.usf.tn

RÉSUMÉ
Cet article décrit nos efforts pour étendre le projet PARSEME à l'arabe standard moderne. L'applicabilité du guide d'annotation de PARSEME a été testée en mesurant l'accord inter-annotateurs dès la première phase d'annotation. Un sous-ensemble de 1062 phrases du Prague Arabic Dependency Treebank (PADT) a été sélectionné et annoté indépendamment par deux locutrices natives arabes. Suite à leurs annotations, un nouveau corpus arabe avec plus de 1250 expressions polylexicales verbales (EPV) annotées a été construit.

ABSTRACT
Annotating Verbal Multiword Expressions in Arabic : Assessing the Validity of a Multilingual Annotation Procedure.

This paper describes our efforts to extend the PARSEME framework to Modern Standard Arabic. The applicability of the PARSEME guidelines was tested by measuring the inter-annotator agreement in the early annotation stage. A subset of 1062 sentences from the Prague Arabic Dependency Treebank PADT was selected and annotated by two Arabic native speakers independently. Following their annotations, a new Arabic corpus with over 1250 annotated VMWEs has been built.

MOTS-CLÉS : Expressions polylexicales arabes, PARSEME, guide d'annotation .

KEYWORDS : Arabic multiword expressions, PARSEME, annotation guidelines.

1 Introduction

L'importance des expressions polylexicales (EP), telles que ***bien que***, ***coup bas*** 'attaque mesquine' ou ***mordre la poussière*** 'être humilié', est reconnue depuis longtemps dans le langage naturel. Leur comportement idiosyncrasique (c'est-à-dire spécial et particulier) nécessite des ressources linguistiques dans lesquelles elles sont explicitement identifiées et décrites. De plus, afin de permettre des études inter-langues de l'idiosyncrasie, la modélisation des EP devrait idéalement être effectuée dans un cadre unifié. La communauté PARSEME a entrepris cet effort de mise en place de guide d'annotation unifié pour les EP verbales (EPV) dans de nombreuses langues (Savary *et al.*, 2018 ; Ramisch *et al.*, 2018, 2020). Le principe de ce cadre est de représenter, de manière unifiée, uniquement les phénomènes qui sont vraiment similaires mettant ainsi l'accent sur ceux qui sont spécifiques à des

langues particulières. Jusqu'à présent, vingt-cinq équipes nationales ont préparé des corpus des EPV dans leurs langues. Ces corpus, publiés sous licences ouvertes, ont été annotés manuellement selon le guide de PARSEME. Chaque fois qu'une nouvelle langue rejoint PARSEME, l'applicabilité du guide d'annotation est testée. Le guide est par la suite modifié ou étendu si nécessaire. Cet article décrit nos efforts pour étendre le cadre PARSEME à l'arabe standard moderne (ASM), que nous appellerons désormais arabe en abrégé.

2 Travaux antérieurs

De nombreuses recherches ont été menées sur les ressources des EP dans plusieurs langues. Néanmoins concernant l'arabe, il existe relativement moins de ressources lexicales et surtout de corpus incluant des EP. Attia *et al.* (2010) ont présenté une méthode linguistique semi-automatique basée sur des expressions régulières pour extraire des EP dans des textes arabes. Hawwari *et al.* (2012) ont créé une liste de 5 000 EPs extraites manuellement de dictionnaires arabes et regroupées en fonction de leur type syntaxique. Al-Badrashiny *et al.* (2016) ont construit une ressource de 1 884 EP arabes en utilisant un détecteur de paradigmes. Ghoneim & Diab (2013) ont utilisé le corpus parallèle arabe-anglais LDC GALE (Grimes *et al.*, 2010) pour représenter les EP dans une chaîne de traitement de traduction automatique statistique (SMT). Une liste d'EP extraite de la base de données anglaise WordNet 3.0 [1] est également utilisée et les entités nommées y sont considérées comme un type d'EP. Ces travaux antérieurs sur les EP arabes concernaient principalement la construction de ressources lexicales et grammaticales à travers des méthodes d'extractions. Nous nous intéressons, à la construction d'un corpus arabe annoté en EP. Nous avons choisi de modéliser les EP arabes dans le cadre multilingue unifié PARSEME. Ainsi, nous nous concentrons non seulement sur les idiomes, mais aussi sur d'autres types de EPV et nous testons la pertinence de la typologie PARSEME pour l'arabe. Des efforts ont déjà été entrepris pour créer un corpus PARSEME arabe (Ramisch *et al.*, 2018). Cependant, ces efforts n'ont pas entièrement suivi la méthodologie PARSEME et le corpus n'a pas été publié ouvertement et n'est plus disponible. En raison de ces contraintes de disponibilité du corpus, l'arabe n'a jamais été couvert par les systèmes développés dans le cadre des tâches partagées PARSEME sur l'identification automatique des EPV. Afin de combler cette lacune, nous avons entrepris la construction d'un corpus arabe PARSEME utilisant de nouvelles sources. Cet article décrit les premières mesures prises pour atteindre cet objectif.

3 Les EPV arabes et PARSEME

Comme dans d'autres langues, les EP en arabe sont composées de pas moins de deux composants (éventuellement agglutinés et/ou discontinus) et se produisent dans un large éventail de configurations lexicales et syntaxiques : en tant que collocations, إبتسامة عريضة (ibtisama arida | lit. 'large sourire') 'grand sourire' ; idiomes ضرب به عرض الحائط (drb bh ard al-ha'et | lit. 'a frappé avec ça la largeur du mur') 'ne pas se soucier de lui' ; noms composés جلسة عامة (glst 'amma | lit. 'session générale') 'audience plénière' ; entités nommées البحر الميت (albhr al mit) 'la Mer Morte', etc. Nous avons décidé de nous concentrer sur les EPV et d'adopter la typologie de PARSEME couvrant les 4 catégories suivantes (Savary *et al.*, 2018) :

- Light Verb Construction (LVC), ou construction à verbe support, elle est formée d'un verbe (support) et d'un nom (prédicatif). Sa particularité réside dans le fait que ce n'est pas le verbe qui remplit la fonction de prédicat de la phrase, mais plutôt le nom prédicatif. Considérons les deux exemples suivants : أعطى نصيحة (A'ata nasiha | lit. 'il a donné un conseil') et أعطى كتاب (Aata kitab | lit. 'il a

1. https://wordnet.princeton.edu/

donné un livre') . L'action, dans la première expression, est exprimée par le verbe أعطى 'a donné', tandis que dans la seconde, elle est exprimée par le nom نصيحة 'conseil'. PARSEME définit deux sous-catégories pour les LVC : ***LVC.full*** (le sujet syntaxique du verbe est l'argument sémantique du nom) et ***LVC.cause*** (le sujet du verbe est la cause ou la source de l'événement ou de l'état exprimé par le nom).

- Verbal Idiom (VID), pour idiome verbal, est toute expression idiomatique qui a au moins deux composants lexicalisés (dont un verbe principal et au moins un de ses dépendants) et qui ne satisfait pas les tests linguistiques des autres catégories de PARSEME. Le sens des VID n'est pas compositionnel et ils possèdent souvent une double lecture littérale / idiomatique par ex. أدار له ضهره (adar lh ahrh | lit. 'tourner pour lui son dos') 'l'abandonner'.

- Inherently Adpositional Verb (IAV), pour verbe intrinsèquement adpositionnel, est une catégorie expérimentale. Un IAV se compose d'un verbe et d'une adposition idiomatiquement sélectionnée. Par exemple, lorsque le verbe اشاد (asad | lit. 'élever (qqn ou une construction)') est associé à la préposition ب (bi) 'de', il change de sens vers *'exalter'*.

- Multi-verb construction (MVC) est la construction multi-verbes composée de deux verbes adjacents. Ils ont généralement le même sujet et dénotent des actions qui sont liées et considérées comme faisant partie du même événement. Par exemple, le proverbe arabe اصبر تنل (asbr tnl | lit. 'Soyez patient vous obtiendrez') 'Soyez patient et vous obtiendrez ce que vous voulez'.

4 Corpus et résultats

Une fois que le guide d'annotation a été testé sur des exemples de textes arabes dans l'annotation pilote, nous avons procédé à l'annotation systématique d'un corpus préexistant annoté syntaxiquement. Le format PARSEME nommé CUPT s'appuie sur l'annotation morphosyntaxique au format CoNLL-U [2], qui est une norme de facto pour l'annotation en dépendances définie par Universal Dependencies (UD). Comme, en outre, notre corpus arabe doit être publié sous licence libre, nous avons choisi le seul corpus arabe UD dont les données sont entièrement disponibles en libre accès, à savoir le Prague Arabic Dependency Treebank (PADT) (Hajic *et al.*, 2004). Ce corpus offre une description de la phrase à plusieurs niveaux : la morphologie fonctionnelle, la syntaxe de dépendance analytique et la représentation tectogrammatique. Il contient actuellement 7 664 phrases annotées (282 384 tokens) provenant des articles de presse. Sur la base de ces directives, nous avons annoté manuellement les occurrences d'EPV dans PADT.

L'annotation d'une expression se fait en plusieurs étapes. Premièrement, nous identifions un candidat composé d'un verbe avec au moins un autre mot qui pourrait former un EPV. Deuxièmement, le candidat est transformé en sa forme canonique et nous déterminons les composants lexicalisés (figés). Troisièmement, nous appliquons les arbres de décision pour tester si le candidat est effectivement idiomatique et quelle est sa catégorie.

4.1 Accord inter-annotateurs

A_1	A_2	F_{span}	κ_{span}	κ_{cat}
763	704	0.699	0.626	0.864

TABLE 1 : Accord inter-annotateurs sur un échantillon de 1062 phrases. A_1 et A_2 sont les nombres d'EPV annotées par chacune des annotatrices, respectivement. F_{span} est la F-mesure entre les annotatrices, κ_{span} est l'accord sur l'empan d'annotation et κ_{cat} est l'accord sur la catégorie d'EPV.

Le tableau 1 montre l'IAA calculé avec les outils PARSEME [3]. Les deux annotatrices ont annoté respectivement 763 et 704 occurrences d'EPV. Leur accord est mesuré séparément pour l'identifi-

2. https://universaldependencies.org/format.html
3. https://gitlab.com/parseme/utilities

cation des empans de texte correspondant à des EPV, et pour leur catégorisation. Comme indiqué par Ramisch *et al.* (2018), F_{span} est la F-mesure des annotations d'une annotatrice par rapport à l'autre, et vice versa. Avec cette mesure, une annotation est considérée comme correcte si les deux annotatrices ont identifié précisément les mêmes tokens comme appartenant à une EPV (ainsi, les chevauchements partiels sont considérés comme entièrement erronés). Ensuite, κ_{span} et κ_{cat} évaluent l'accord inter-annotateurs respectivement sur l'identification et la catégorisation en utilisant la mesure Kappa (Cohen, 1960). Suivant l'échelle de fiabilité proposée par Carletta (1996), l'IAA que nous observons doit être considéré comme *bonne* sur la tâche de catégorisation, et *satisfaisante* sur la tâche d'identification. Nous avons comparé ces scores IAA initiaux pour l'arabe à ceux de la suite PARSEME (éditions 1.1 et 1.2) [4]. Parmi les 26 estimations de l'IAA [5], l'arabe a maintenant :

— Le deuxième plus grand nombre (après le chinois) d'annotations d'EPV utilisées pour estimer l'IAA.
— La 12e, 14e et 12e meilleure valeur de F_{span}, κ_{span} et κ_{cat}, respectivement.

Il est à noter que, pour les autres langues, le corpus IAA a souvent été doublement annoté à l'étape finale de la campagne d'annotation, lorsque l'expertise des annotateurs a atteint son optimum. L'IAA pour l'arabe, à l'inverse, a été estimé au début de l'étape d'annotation, de manière à contrôler la préparation suffisante des annotateurs et la justesse de la méthodologie dès le démarrage de la campagne d'annotation.

4.2 Résultats

Le tableau 2 donne les statistiques du corpus arabe dans son état actuel. Sa taille, avec plus de 1250 EPV annotées, dépasse déjà les plus petits corpus de la suite PARSEME, et permet de premières observations. La densité d'EPV est d'environ 0,68 EPV par phrase. L'universalité (c'est-à-dire la présence dans toutes les langues étudiées) des catégories VID et LVC est confirmée, les LVC.full étant presque deux fois plus fréquents que les VID, et les LVC.cause étant sporadiques. Les catégories quasi-universelles des verbes intrinsèquement réfléchis (IRV) (fréquentes en langues romanes et slaves, ainsi qu'en allemand) et des constructions verbe-particule (VPC) (fréquentes en langues germaniques et le hongrois) ne se retrouvent pas en arabe. [6]

Phrases	**Tokens**	**Occurrences d'EPV**								
		VID	IRV	LVC.full	LVC.cause	VPC.full	VPC.semi	IAV	MVC	All
1,847	70,498	345	0	673	68	0	0	165	1	1252

TABLE 2 : Statistiques du corpus arabe d'EPV dans son état actuel, en termes du nombre de phrases et des insertions, ainsi que le nombre d'EPV annotées par catégorie et au total.

Comme Savary *et al.* (2018), nous avons également analysé le corpus en termes de longueur (c'est-à-dire de nombres des composants lexicalisés) des EPV annotées et de leurs discontinuités (c'est-à-dire de nombre d'éléments externes survenant entre le premier et le dernier composant d'une EPV). Les discontinuités sont un défi majeur pour la tâche d'identification d'EPs (Constant *et al.*, 2017), par conséquent leur distribution est une caractéristique importante de la langue et du corpus à l'étude. Le tableau 3 montre les résultats de cette analyse. En particulier, plus de 73% de toutes les EPV contiennent 2 composants (colonne 3), plus de 42% sont continues (colonne 7) mais plus de 17% (dernière colonne) ont plus de 3 insertions.

4. L'édition 1.0 n'est pas prise en compte ici car elle était basée sur une version différente du guide d'annotation.

5. Pour 3 langues de la suite PARSEME, l'IAA a été estimée deux fois : une fois dans l'édition 1.1 et une fois dans la 1.2. Nous négligeons le précédent corpus PARSEME arabe publiquement inaccessible.

6. Ceci contraste avec les statistiques du précédent corpus PARSEME de l'arabe, que nous n'avons pas pu recalculer en raison de l'indisponibilité de ce corpus. Là, 4 219 EPV ont été signalées dans 3 137 phrases (avec une densité de 1,35) réparties en : 1 769 LVC.full, 1 320 VID, 1 080 VPC, 17 IRV, 33 MVC et 0 LVC.cause. À noter notamment l'absence des VPC dans nos statistiques. Nous affirmons que les particules, telles que définies dans PARSEME, sont inexistantes ou très rares en arabe. Les VPC du corpus d'Hawwari pourraient probablement être des IAV.

Nous comparons ces résultats aux 18 langues de la suite PARSEME dans l'édition 1.0. [7] La longueur moyenne des EPV en arabe (2,26 dans la colonne 1 du tableau 3) n'est pas une valeur aberrante, puisque dans 17 (sur les 18) langues sa valeur est comprise entre 2,02 et 2,71. L'inexistence ou la rareté des EPV à élément unique [8] (2.00 dans la colonne 2 du tableau 3) est également une caractéristique de 14 langues (le hongrois, l'allemand et le portugais ayant des valeurs aberrantes dans cette catégorie). En termes de discontinuités, l'arabe est la deuxième langue la plus remarquable (après l'allemand). Il a 1,97 insertions en moyenne (l'allemand en a 2,96, le slovène 1,47, le tchèque 1,35, le hongrois 1,01 et toutes les autres langues en ont moins de 1). Il a également le 2ème pourcentage le plus bas d'EPV continues (42,17%) et le 2ème pourcentage le plus élevé de EPV avec le nombre d'insertions supérieur à 3 (17,33%), après l'allemand (35,7% et 30,5%, respectivement). Le corpus dans son état actuel est déjà disponible dans le dépôt PARSEME [9] sous la licence CC-BY v4 [10], y compris l'échantillon à double annotation utilisé pour le calcul de l'IAA. Ainsi, les résultats présentés ici sont entièrement reproductibles, en utilisant les outils PARSEME [11].

Longueurs d' EPV					Longueurs de discontinuités					
Avg	%1	%2	%3	%>3	Avg	%0	%1	%2	%3	%>3
2,26	2,00	73,72	21,09	3,19	1,97	42,17	23,48	8,95	8,07	17,33

TABLE 3 : Longueurs et discontinuités des occurrences EPV arabes : nombre moyen de tokens (col. 1) ; pourcentage d'EPV avec 1, 2, 3 et plus de 3 tokens (col. 2–5) ; longueur moyenne des discontinuités (col. 6) ; pourcentage d'EPV avec des discontinuités de longueur 0, 1, 2, 3 et plus de 3 (col. 7–11).

5 Analyse et défis

La tâche d'annotation manuelle a apporté son propre ensemble de défis liés au corpus source, au processus d'annotation et aux spécificités de la langue arabe. Nous listons ici les principales difficultés rencontrées, qui nous renseignent sur l'adaptation du cadre d'annotation PARSEME à l'arabe :

- Ambiguïté grammaticale des corpus : Les textes du corpus source proviennent de journaux en ligne qui comportent des fautes d'orthographe et de grammaire. Des erreurs d'annotation (de morphologie et de syntaxe) peuvent également s'être produites. Cela rend l'identification de l'EPV plus difficile. De plus, la non voyellisation peut créer une ambiguïté de la catégorie grammaticale du mot. Ainsi, le mot طرق, que l'on trouve dans l'EP طرق الباب (lit. 'frapper la porte') 'chercher une solution', peut correspondre à un nom طرق (torq) 'chemins' ou un verbe طرق (trq) 'frapper' lorsqu'il n'est pas voyellé.

- Discontinuités des éléments lexicalisés : Les phrases en arabe sont caractérisées par la position libre des mots. Cette variabilité dans l'ordre des mots provoque des ambiguïtés syntaxiques qui rendent les expressions manipulées difficiles à comprendre lorsque des éléments interviennent entre les composants d'un candidat, comme l'exemple suivant : لعب الامين العام لمنظمة الوحدة الافريقية سليم احمد سليم ممثل الامم المتحدة الخاص للكونغو كامل مرجان دور (l'b al-amin al-'am lmnzmt al-uhda al-afriqit slim ahmed slim mmthl al-amm al-mthdt al-has llkungu kaml mrgan dur) 'Joue le secrétaire général de

7. Les statistiques des éditions suivantes n'ont pas été publiées.
8. Un token peut contenir plusieurs mots agglutinés, il peut donc s'agir d'une EP.
9. https://gitlab.com/parseme/parseme_corpus_ar
10. https://creativecommons.org/licenses/by/4.0/
11. https://gitlab.com/parseme/utilities/-/blob/master/st-organizers/corpus-statistics/mwe-stats-simple.py et https://gitlab.com/parseme/utilities/-/blob/master/st-organizers/corpus-statistics/mwe-stats.py

l'Organisation de l'unité africaine, Salim Ahmed Salim, le représentant spécial des Nations unies pour le Congo, Kamel Morjane, un rôle'

- Agglutination : Comme une forme agglutinée peut avoir plusieurs segmentations possibles, nous avons dû choisir avec soin les bons composants lexicalisés de l'expression. Considérons par exemple وضعه على الرف (wad'ah 'ala raf | lit. 'le mettre sur l'étagère') 'l'ignorer'. Ici, l'agglutination du verbe وضع (wad'a) 'mettre' à l'enclitique ه (ho) 'lui' est requise. Avec l'omission de l'enclitique on perd en effet le sens idiomatique 'mettre quelque chose sur l'étagère'.

- Masdar : est une partie du discours spécifique à l'arabe et définie comme un nom verbal, qui exprime le même événement que la racine verbale correspondante mais ne porte aucune indication de temps, modalité, aspect etc. Une EPV peut ainsi intégrer un verbe représenté par son *masdar* comme par exemple :على المجتمع تحقيق نهضته (alai al-mugtama tahqiqa nahdatihi | lit. 'sur la société réalisation sa renaissance') 'la société doit réussir sa renaissance'. Le sens est porté par le nom نهضته (nahda) 'renaissance', tandis que le *masdar* تحقيق (thqiq) 'réalisation' dérivé du verbe حقق (haqqaqa) 'réaliser' se comporte comme un verbe support. Dans ce cas, l'occurrence EPV candidate est تحقيق نهضه (réalisation renaissance | lit. 'réalisation de renaissance') , et la forme canonique à laquelle les tests linguistiques sont appliqués est حقق نهضه (haqaq nahda) 'réaliser la renaissance'.
Une spécificité intéressante de l'arabe est précisément le caractère très productif du schéma syntaxique (verbe support + *masdar*) où le verbe et le *masdar* sont issus de la même racine verbale, ce qui conduit à une duplication sémantique, comme dans l'exemple : خرج خروجا (k-rg k'rug | lit. 'il a quitté exit') 'il est sorti'. De telles combinaisons verbe/*masdar* passent les tests LVC.

- Difficultés de catégorisation : Outre ces spécificités, certaines EPV arabes sont difficiles à catégoriser. Une source importante de désaccords entre annotateurs a ainsi concerné la distinction entre VID et expression littérale. Par example قطع الطريق مسرعا (qata' al-triq mosr' | lit. 'couper la route en courant') est une expression littérale faisant référence à l'action de traverser la rue. Inversement, قطع الطريق عليه (qta' al-triq 'lh | couper+la+route sur+lui | lit. 'couper sa route') est idiomatique, signifiant 'empêcher quelqu'un de faire ce qu'il veut faire'. Dans de tels cas, la catégorisation requiert une analyse très fine du contexte d'occurrence de l'expression. Comme dans d'autres langues, certaines catégorisations difficiles demandent aux annotateurs d'imaginer l'expression dans plusieurs contextes afin de se décider par exemple entre LVC et VID. Ainsi, شكل جزءا (s'kl gz | lit. 'former partie') 'faire partie de' est un VID et non un LVC car il ne permet pas d'omettre le verbe, bien que le nom soit prédicatif et garde son sens habituel. Inversement, dans le LVC.full أسدى النصيحة (asda al-nasiha | lit. 'il tisse un conseil') 'il donne un conseil', le verbe أسدى 'tisser' est sémantiquement riche dans d'autres contextes mais dans cette expression il agit comme un verbe support.

6 Conclusion et perspectives

Nous avons annoté manuellement le corpus PADT en utilisant le guide d'annotation de PARSEME, qui a montré son efficacité sur l'arabe moderne. Les types ont été annotés sur un échantillon de 1 062 phrases du PADT par 2 annotatrices et nous obtenons un accord inter-annotateurs raisonnable. Nous avons annoté 1 252 occurrences EPV avec un taux élevé d'expressions discontinues (58%). Le guide d'annotation PARSEME s'étant avérée applicable à l'arabe sans adaptation, nous considérons l'initiative initiale de la tâche d'annotation comme validée. Cependant, il est possible que nous ayons omis certains types de variation non représentés dans notre corpus. Les travaux futurs consisteront à annoter plus des textes, ce qui pourrait nous conduire à proposer des ajouts spécifiques à l'arabe dans le guide d'annotation PARSEME.

Références

AL-BADRASHINY M., HAWWARI A., GHONEIM M. & DIAB M. (2016). SAMER : a semi-

automatically created lexical resource for Arabic verbal multiword expressions tokens paradigm and their morphosyntactic features. In *Proceedings of the 12th Workshop on Asian Language Resources (ALR12)*, p. 113–122.

ATTIA M., TORAL A., TOUNSI L., PECINA P. & VAN GENABITH J. (2010). Automatic extraction of Arabic multiword expressions. In *Proceedings of the 2010 Workshop on Multiword Expressions : from Theory to Applications*, p. 19–27.

CARLETTA J. (1996). Assessing agreement on classification tasks : The kappa statistic. *Computational Linguistics*, **22**(2), 249–254.

COHEN J. (1960). Kappa : Coefficient of concordance. *Educ Psych Measurement*, **20**(37), 37–46.

CONSTANT M., ERYIĞIT G., MONTI J., VAN DER PLAS L., RAMISCH C., ROSNER M. & TODIRASCU A. (2017). Survey : Multiword expression processing : A Survey. *Computational Linguistics*, **43**(4), 837–892. DOI : 10.1162/COLI_a_00302.

GHONEIM M. & DIAB M. (2013). Multiword expressions in the context of statistical machine translation. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, p. 1181–1187, Nagoya, Japan : Asian Federation of Natural Language Processing.

GRIMES S., LI X., BIES A., KULICK S., MA X. & STRASSEL S. (2010). Creating arabic-english parallel word-aligned treebank corpora at ldc.

HAJIC J., SMRZ O., ZEMÁNEK P., ŠNAIDAUF J. & BEŠKA E. (2004). Prague arabic dependency treebank : Development in data and tools. In *Proc. of the NEMLAR Intern. Conf. on Arabic Language Resources and Tools*, p. 110–117.

HAWWARI A., BAR K. & DIAB M. (2012). Building an Arabic multiword expressions repository. In *Proceedings of the ACL 2012 Joint Workshop on Statistical Parsing and Semantic Processing of Morphologically Rich Languages*, p. 24–29.

RAMISCH C., CORDEIRO S. R., SAVARY A., VINCZE V., BARBU MITITELU V., BHATIA A., BULJAN M., CANDITO M., GANTAR P., GIOULI V., GÜNGÖR T., HAWWARI A., IÑURRIETA U., KOVALEVSKAITĖ J., KREK S., LICHTE T., LIEBESKIND C., MONTI J., PARRA ESCARTÍN C., QASEMIZADEH B., RAMISCH R., SCHNEIDER N., STOYANOVA I., VAIDYA A. & WALSH A. (2018). Edition 1.1 of the PARSEME shared task on automatic identification of verbal multiword expressions. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, p. 222–240, Santa Fe, New Mexico, USA : Association for Computational Linguistics.

RAMISCH C., SAVARY A., GUILLAUME B., WASZCZUK J., CANDITO M., VAIDYA A., BARBU MITITELU V., BHATIA A., IÑURRIETA U., GIOULI V., GÜNGÖR T., JIANG M., LICHTE T., LIEBESKIND C., MONTI J., RAMISCH R., STYMNE S., WALSH A. & XU H. (2020). Edition 1.2 of the PARSEME shared task on semi-supervised identification of verbal multiword expressions. In *Proceedings of the Joint Workshop on Multiword Expressions and Electronic Lexicons*, p. 107–118, online : Association for Computational Linguistics.

SAVARY A., CANDITO M., MITITELU V. B., BEJEK E., CAP F., ÉPLÖ S., CORDEIRO S. R., ERYIIT G., GIOULI V., VAN GOMPEL M., HACOHEN-KERNER Y., KOVALEVSKAIT J., KREK S., LIEBESKIND C., MONTI J., ESCARTÍN C. P., VAN DER PLAS L., QASEMIZADEH B., RAMISCH C., SANGATI F., STOYANOVA I. & VINCZE V. (2018). PARSEME multilingual corpus of verbal multiword expressions. In S. MARKANTONATOU, C. RAMISCH, A. SAVARY & V. VINCZE, Éds., *Multiword expressions at length and in depth : Extended papers from the MWE 2017 workshop*, p. 87–147. Berlin : Language Science Press. DOI : 10.5281/zenodo.1469555.