Constructing an Annotated Corpus of Verbal MWEs for English

Abigail Walsh
ADAPT Centre
Dublin City University
abigail.walsh@adaptcentre.ie

Claire Bonial
U.S. Army Research Laboratory
claire.n.bonial.civ@mail.mil

Kristina Geeraert
University of Alberta
geeraert@ualberta.ca

John P. McCrae
Insight Centre for Data Analytics
National University of Ireland Galway
john.mccrae@insight-centre.org

Nathan Schneider **Clarissa Somers**
Georgetown University
nathan.schneider@georgetown.edu

Abstract

This paper describes the construction and annotation of a corpus of verbal MWEs for English as part of the PARSEME Shared Task 1.1 on automatic identification of verbal MWEs. The criteria for corpus selection, the categories of MWEs used, and the training process are discussed, along with the particular issues that led to revisions in edition 1.1 of the annotation guidelines. Finally, an overview of the characteristics of the final annotated corpus is presented, as well as some discussion on inter-annotator agreement.

1 Introduction

Multiword expressions (MWE) present a challenge in Natural Language Processing (NLP) due to their idiosyncrasy, which necessitates an interpretation of these expressions that is distinct from that of compositional phrases (literal, non-idiosyncratic word combinations) that may be similar or identical in their surface forms (Sag et al., 2002). MWEs are extremely prevalent in our lexicon; up to half of our lexicon is composed of MWEs (Ramisch, 2015). Understanding MWEs can aid in a variety of NLP tasks, ranging from syntactic disambiguation, to conceptual understanding, to semantic tagging, to word alignment (Baldwin and Kim, 2010). Given the critical nature of identifying and interpreting MWEs correctly, MWEs have been the main item on the agenda of several working groups, including PARSEME, which aims to improve cross-linguistic understanding of MWEs through the development of manually annotated data and shared tasks focused on automatic MWE identification (Constant et al., 2017).

This paper describes the construction and annotation of a corpus of English verbal MWEs (VMWEs) for the second edition of the PARSEME Shared Task (Shared Task 1.1). The term MWE frequently encompasses a wide variety of linguistic phenomena such as idioms, compound nouns, verb particle constructions, institutionalized phrases, etc. Although the precise definition can differ depending upon the community of interest (Constant et al., 2017), the annotation guidelines for the PARSEME shared task[1] define MWEs as continuous or discontinuous sequences of words with the following properties:

- Some degree of idiosyncrasy in respect to grammar (statistical idiosyncrasy[2] is not considered here)
- Their component words include a head and at least one other syntactically dependent word
- At least two of the components are lexicalised

PARSEME is focused on VMWEs, which are MWEs whose syntactic head is a verb in its prototypical form.

After providing some background on the past PARSEME shared task, we will describe the data and annotation procedures, including some of the annotation challenges, of the 2018 English corpus. We provide a quantitative overview of the characteristics of the corpus, as well as inter-annotator agreement figures.

[1] http://parsemefr.lif.univ-mrs.fr/parseme-st-guidelines/1.1/?page=010_Definitions_and_scope/020_Verbal_multiword_expressions

[2] (Farahmand and Nivre, 2015)

1.1 Background: The PARSEME Shared Task 1.0

The Shared Task 1.1 (Ramisch et al., 2018) is based on a similar Shared Task 1.0 that took place in 2017 (Savary et al., 2017). The aim of the shared task is to identify VMWEs in running texts across a variety of languages, and to establish a consistent set of guidelines for the annotation of these VMWEs. The 1.0 task and data encompassed 18 languages. English, the focus of this paper, is one of the five languages added to this new edition of the shared task, along with Arabic, Basque, Croatian, and Hindi. The guidelines have evolved from version 1.0 to accommodate these new languages. Specifically, two of the categories (VPC and LVC) have now been extended to allow for more fine-grained categorisation (see section 3.1 for more information on categories).

2 Data

There were several considerations when selecting appropriate text for inclusion in this corpus. This section describes the selection criteria, followed by a description of the annotation tool used. The suggestions for selecting an appropriate source of data were provided by PARSEME in the language leader guidelines, and were informed by version 1.0 of the shared task. Of those suggestions, the following criteria were deemed to be of the highest priority:

1. The corpus should be available under an open licence
2. The text must be originally written in English
3. The text should be annotated for morphosyntactic information
4. The size of the corpus should allow for at least 3,500 MWE annotations
5. The language must be of sufficiently high quality

There were several corpora considered for selection, including the DiMSUM corpus (Schneider et al., 2016), the UP/TAP corpus,[3] Wikidata parallel text (Vrandečić and Krötzsch, 2014) and the Universal Dependencies (UD) treebanks.[4] Three corpora from the UD treebanks for English were ultimately selected as a source of data, as they alone fulfilled the criteria mentioned above: text was selected from the English-EWT corpus (Silveira et al., 2014),[5] the LinES parallel corpus (Ahrenberg, 2007) and the Parallel Universal Dependencies (PUD) treebank (Zeman et al., 2017).[6] The files were extracted in CoNLL-U format and converted to FoLiA XML format (see section 3) for annotating. The training, development and testing datasets for each treebank were concatenated, and then split into files of 201 sentences for annotation.

3 Annotation

During the data preparation period, annotators were trained in the use of the FoLiA Linguistic Annotation Tool (FLAT). FLAT is an open-source web-based environment,[7] using the XML-based FoLiA format. In order to aid annotators in annotating only verbal MWEs, FLAT highlights verbs using POS information taken from the CoNLL-U file. Figure 1 shows a screenshot of the FLAT platform, demonstrating how the selecting and annotating of lexicalised components works. Annotators were also trained to recognize and categorise VMWEs of different types, detailed in the sections to follow.

The annotation team was comprised of volunteers who had experience with or interest in annotating multiword expressions, and were all native speakers of English. Four dialects of English were represented: Irish English, British English, American English and Canadian English.

[3] Documentation for UP/TAP: `https://www.l2f.inesc-id.pt/~thomas/metashare/report-UP-TAP.pdf`

[4] Documentation for UD: `http://universaldependencies.org`

[5] Originally sourced from the English Web Treebank (Bies et al., 2012)

[6] Though not a part of the task dataset, we have also fully annotated the Reviews portion of the UD English-EWT corpus by adding VMWE types to the existing VMWEs in STREUSLE (Schneider et al., 2014; Schneider and Smith, 2015, `https://github.com/nert-gu/streusle/`); they were previously uncategorized. STREUSLE as of version 4.1 comprises 3812 sentences and 871 VMWE instances (121 IAV, 12 LVC.cause, 123 LVC.full, 310 VID, 206 VPC.full, 99 VPC.semi).

[7] `http://flat.readthedocs.io/en/latest/`

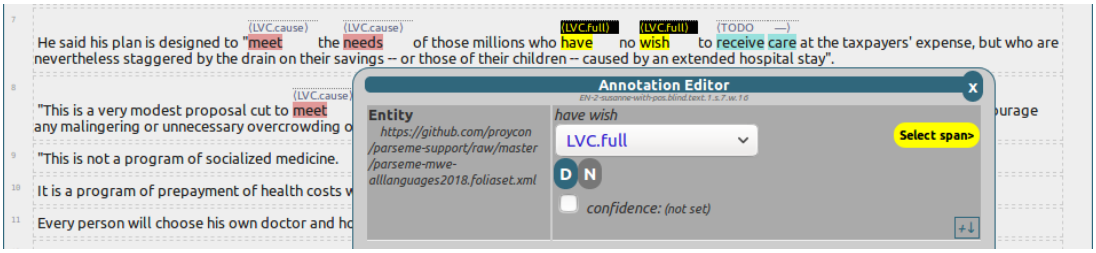

Figure 1: Screenshot of the FLAT Platform

3.1 Categories of VMWE

Seven categories of VMWE were used in the English annotation task: Verbal Idioms (VID), Verb-Particle Constructions (VPC.full and VPC.semi),[8] Light-Verb Constructions (LVC.full and LVC.cause),[8] Multi-Verb Constructions (MVC) and Inherently Adpositional Verbs (IAV). The categories are divided into universal categories (valid for all participating languages), quasi-universal categories (valid for some language groups or languages), and an experimental category (which may be optionally considered for some languages).

Verbal idioms (VIDs) and the **Light-Verb Constructions** (LVCs) constitute universal categories. VIDs have at least two lexicalised components, including a head and at least one dependent. Dependents can be of different grammatical roles and parts of speech, meaning VIDs may be confused with other categories of VMWEs, such as LVCs. VIDs also include sentential expressions with no open slots, such as proverbs.

VID: *to **take** something **with a pinch of salt***: VMWE with an adverbial complement

LVCs are formed by a verb and a single or compound dependent noun. The noun must be abstract and predicative. The verb can be of two types: a 'light' verb, which arguably contributes no extra semantics to the expression beyond the semantics denoted by the predicative noun (annotated as LVC.full), and a 'causative' verb, which contributes only the semantics of causation, as the subject of the verb is the cause or source of the event or state expressed by the dependent noun (annotated as LVC.cause).

LVC.full: *to **make** a **decision***: verb adds nothing substantive to the semantics of 'decision'
LVC.cause: *to **give** a **headache***: the subject of 'give' is the cause of the headache

Verb-Particle Constructions (VPCs) and **Multi-Verb Constructions** (MVCs) are quasi-universal categories that are applicable to English. VPCs are formed by a verb and a dependent particle. The verb can be either fully non-compositional, where the addition of the particle changes the meaning of the verb significantly (annotated as VPC.full), or semi-non-compositional where the particle adds a partially predictable but non-spatial meaning to the verb (annotated as VPC.semi).

VPC.full: *to **check in** upon arrival*: omitting 'in' leads to very different meaning
VPC.semi: *to **eat** the cookies **up***: 'up' adds a sense of completion, but not a spatial meaning

MVCs are composed of two adjacent verbs, one of which is a governing verb and the other a dependent verb; together they function as a single predicate. The test for this category in English involves replacing the dependent verb with another verb from the same semantic class. If this leads to ungrammaticality or an unexpected change in meaning, the expression is categorised as MVC.

MVC: *to **let go*** : replacing 'go' with 'depart', 'move', etc. changes the meaning significantly

Inherently Adpositional Verbs (IAVs) constitute an experimental category that has been included in the English annotation. IAVs consist of a verb and an adposition that is integral to the meaning of the expression. The guidelines include a test to differentiate between adpositions and particles, the former of which are exclusively used in IAVs.[9]

[8] New categories added to edition 1.1 of the shared task
[9] Adpositions are fixed in occurring exclusively before a noun phrase, unlike particles, which either modify an intransitive

IAV: *to **come across** something*: omitting adposition 'across' leads to very different meaning

IAVs may also contain particles: e.g. *to **put up with** something* (verb+particle+preposition) means to endure it, and cannot have this meaning absent *up* or *with*.

3.2 Pilot Annotation Tasks

Three pilot annotation tasks were held to allow annotators to familiarize themselves with FLAT and the guidelines, as well as raise any potential issues and disagreements concerning the categorisation of VMWEs in English. Two small corpora were used for the first two pilot annotations, consisting of 200 sentences taken from the Brown corpus.[10]

Pilot annotation 1 was held in the beginning of June 2017, using version 1.0 of the guidelines. Following a discussion of this task, many disagreements seemed to stem from LVC tests that were difficult to apply and did not cover all cases. For example:

*The grand jury **took** a **swipe** at the State Welfare Department...*: While annotators felt this should be categorised as LVC, the original tests for LVC state that the noun must be used in one of its original senses, i.e. non-idiomatic use of the word, which would cause this expression to fail as an LVC. In response, the noun requirements within LVCs were generalized such that the noun must only be predicative, but need not retain one of its senses used outside of LVCs.

Annotators came across cases of LVCs that were fairly straightforward because the verb quite clearly adds little semantics beyond that of the predicative noun (e.g., *She has a terrible headache*). However, variant expressions with a different light verb were not clearly LVCs, given a minimal amount of causative semantics contributed by the verb (e.g., *The buzzing radio gave him a headache*). Such cases were another source of disagreement in LVC annotations. To accommodate both types of LVCs while maintaining an acknowledgment of the causative semantics, it was decided after discussion to provide a distinction in the guideline tests between a fully light verb (LVC.full) and a causative light verb construction (LVC.cause). Causative light verbs, unlike fully light verbs, contribute the semantics of causation to the expression by licensing an outside causer or agent semantic role assigned to the verb's subject.

Other disagreements centered around unclear tests for particles, particularly particles which contribute aspectual or other subtle information, but do not significantly alter the meaning of the verb, leading to inconsistencies in VPC annotation. For example:

*...the Senate **passed** the bill **on** to the House*: Here the verb keeps its meaning but the particle contributes non-compositionally. Like LVCs, it was decided to subdivide the VPC category into the VPC.full and VPC.semi categories described above; thus improving agreement on borderline VPC.semi examples like this instance.

Pilot annotation 2 took place towards the end of November 2017, following the rewriting of the guidelines into version 1.1. In discussing the new annotation guidelines, some issues were raised. Many of these again centered around LVCs, including the productive nature of candidates in the new LVC.cause category (such productivity runs somewhat counter to the expectations of idiosyncrasy and lexicalisation for all VMWEs), as well as disagreements surrounding nouns categorized as either concrete or abstract (nouns within LVCs must be abstract and predicative). For example:

*A certain **vagueness** may also be **caused** by tactical appreciation of the fact...*: Here, and in all cases of *cause* in combination with an abstract/predicative noun, it was debated as to whether these should be considered LVC.cause. *Cause* expressions seemed to defy the normative expectations of idiosyncracy and lexicalization put forth for all VMWEs given that, unlike other light verbs, *cause* seems to combine productively with any predicative noun and the resulting expression is felicitous while maintaining its purely compositional semantics. After discussion, it was decided that such cases should be included as LVC.cause; however, a note that these cases do not exhibit some of the hallmarks of other LVCs and MWEs was added to the guidelines.

*The scholarship plan would **provide** federal **contributions** to each medical and dental school equal to $1500...*: It is unclear here if the noun *contributions* should be understood as abstract and

verb (***check in***) or are mobile with respect to full noun phrase complements (***eat** the cookies **up**/**eat up** the cookies*).

[10] Access the Brown Corpus Manual here: `http://clu.uni.no/icame/manuals/BROWN/INDEX.HTM`

predicative, or if it refers to the concrete contribution of the specific sum of money mentioned later in the sentence. If the noun is understood as abstract and predicative, then the expression could be considered a case of LVC.cause, given that presumably the *scholarship plan* is an outsider causer of the contribution, while *federal* likely refers to the actual contributer. Additional guidance on distinguishing abstract and concrete nouns was added in response.

Related to distinguishing IAVs from VPCs, annotators also expressed confusion regarding the difference between particles and adpositions (and the recently added test to differentiate). For example:

...to ***set aside*** *the privilege resolution*: The categorisation of this expression was controversial because of uncertainty as to whether 'aside' could be considered a particle, and thus, belonging to a VPC. Following this confusion, tests for differentiating between adpositions and particles were featured more prominently as part of the decision tree for categorising VPCs.

After clarifying some of the intended interpretations and tests in the guidelines, it was decided to hold a third round of pilot annotations for English, reusing the corpus from the second pilot task, during the month of December 2017. **Pilot annotation 3** led to a more informed, robust discussion of the previous issues, and concluded with amendments to the guidelines, including notes regarding the productive characteristic of many LVC.cause VMWEs and additional pointers for distinguishing IAVs with adpositions from VPCs with particles.

Table 1 in Section 4 shows the number of VMWEs that were annotated during each pilot task, and the breakdown of categories that were annotated. Note that the categories VPC.full and LVC.full represent VPC and LVC respectively for Pilot 1, as the fine-grained labels did not exist in version 1.0 of the guidelines. Similarly, the optional category IAV was not considered for the first pilot task.

After iteration throughout piloting, the 1.1 edition of the guidelines were finalized for all languages. Several of the changes to the guidelines came about due to challenges with annotation of English VMWEs during the pilot annotation task, namely the subdivision of the LVC and VPC categories.

4 Corpus Annotation and Results

The annotation of the final corpus took place between the start of January 2018 and the end of February 2018. During this period, a total of 7437 sentences (124,202 tokens) were annotated. 4221 of these sentences were from the English Web Treebank, 3015 were from the LinES parallel corpus, and the remaining 201 sentences were from the PUD treebank. Out of a total of 14,121 verbs, 832 were annotated as VMWEs. Table 1 displays the categories of VMWE that were annotated. The most commonly annotated category of English VMWE is full Verb-Particle Constructions, followed by full Light-Verb Constructions.

Following the end of the annotation period, the corpus was prepared for release. The annotated files were downloaded from FLAT in FoLiA XML format and aligned with the original CoNLL-U files. The annotated data from each annotator was consolidated, and a consistency check was performed to ensure that VMWEs were consistently annotated across all the data. Following this stage, the FoLiA files were then merged with the aligned CoNLL-U files to be converted into PARSEME TSV format, which is the format of the released data.[11]

Following the release of the annotated corpus, a portion of the corpus (804 sentences) was selected for annotation by all four annotators, in order to measure the quality of the corpus. The categorisation of VMWE types is shown in table 2. The table shows the greatest level of disagreement in the categorisation of LVCs, particularly the LVC.cause category. Despite having provided additional guidance on the subject in the guidelines, the general VMWE definitional requirement of idiosyncrasy may have affected the categorisation of LVC.cause, as many instances of LVC.cause appear regular, and thus annotators may find it counter-intuitive to label these candidates as VMWEs.

The IAA scores between all the pairs of annotators are given in Table 3. The agreement between annotators is fair, showing moderate agreement when calculating the span of annotation (*F-score* and *Kappa*), and substantial agreement when calculating the agreement of categorisation only (*Kappa-cat*).

[11] The full PARSEME shared task data can be found at: `https://gitlab.com/parseme/sharedtask-data/tree/master/1.1`

Category	Pilot 1	Pilot 2	Pilot 3	Final
VPC.full	40	33	49	297
VPC.semi	0	25	25	45
LVC.full	37	43	82	244
LVC.cause	0	21	44	43
VID	38	19	30	139
MVC	0	2	1	4
IAV	0	15	34	60
Total	**115**	**158**	**265**	**832**

Table 1: Number of annotations per category.

Category	A1	A2	A3	A4
VPC.full	27	41	62	41
VPC.semi	17	3	9	23
LVC.full	77	32	43	42
LVC.cause	28	2	5	11
VID	13	14	25	41
MVC	4	0	0	1
IAV	22	9	9	17
Total	**188**	**101**	**153**	**176**

Table 2: VMWEs in doubly annotated corpus.

We see from the table that the agreement between the two annotators who completed all three pilot tasks (A3 and A4) is higher than the agreement between the two annotators who did not participate in the three pilot tasks (A1 and A2), in as far as annotating the span of the VMWEs (*F-score* and *Kappa*). This is not the case when only the category of the VMWE is considered (*Kappa-cat*).

Pair	#X	#Y	*F-score*	*Kappa*	*Kappa-cat*
1x2	188	101	0.436	0.396	0.661
1x3	188	153	0.452	0.402	0.647
1x4	188	176	0.478	0.427	0.635
2x3	101	153	0.480	0.446	0.773
2x4	101	176	0.513	0.479	0.636
3x4	153	176	0.529	0.487	0.625

Table 3: IAA scores between annotator pairs (X and Y) for a subset (804 sentences) of the corpus. *F-score* is the F-measure between annotators, and is an optimistic measure that ignores agreement due to chance. The kappa scores used for *Kappa* and *Kappa-cat* are variants of 2-raters Cohen's kappa. *Kappa* is a calculation of the rate of agreement of annotation for all verbs in the corpus, while *Kappa-cat* takes into account only those VMWEs where both annotators agreed on the span, and measures the agreement of categorisation for these VMWEs.

4.1 Conclusions & Future Work

The development of manually annotated training data and support of shared tasks facilitating automatic identification of MWEs is critical for enabling the idiosyncratic interpretation required of these prevalent, but often ignored, elements of natural language. The challenging nature of this task for automatic systems is made evident by how challenging consistent VMWE identification and categorization can be, even for trained human annotators. The results of the pilot tasks in Section 3.2 demonstrate that MWE categorisation is a task requiring specialised training and clear guidelines. The IAA scores in Section 4 indicate that even with specialised training, disagreements in labelling VMWEs is to be expected. The process of annotating VMWEs in English led to some interesting discussion surrounding the categorisation of light-verb constructions and verb-particle constructions in particular. We look forward to the results of the second edition of the PARSEME shared-task, which may highlight English annotation gaps and inconsistencies to be addressed in the future.

5 Acknowledgements

The first author's work is funded by the Irish Government Department of Culture, Heritage and the Gaeltacht under the GaelTech Project, and is also supported by Science Foundation Ireland in the ADAPT Centre (Grant 13/RC/2106) (`http://www.adaptcentre.ie`) at Dublin City University.

References

Lars Ahrenberg. 2007. LinES: An English-Swedish parallel treebank. In *Proc. of NODALIDA*, pages 270–273, Tartu, Estonia, May.

Timothy Baldwin and Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkhya and Fred J. Damerau, editors, *Handbook of Natural Language Processing, Second Edition*, pages 267–292. CRC Press, Taylor and Francis Group, Boca Raton, FL.

Ann Bies, Justin Mott, Colin Warner, and Seth Kulick. 2012. English Web Treebank. Technical Report LDC2012T13, Linguistic Data Consortium, Philadelphia, PA.

Mathieu Constant, Gülşen Eryiğit, Johanna Monti, Lonneke van der Plas, Carlos Ramisch, Michael Rosner, and Amalia Todirascu. 2017. Multiword expression processing: a survey. *Computational Linguistics*, 43(4):837–892, December.

Meghdad Farahmand and Joakim Nivre. 2015. Modeling the statistical idiosyncrasy of multiword expressions. In *Proc. of the 11th Workshop on Multiword Expressions*, pages 34–38, Denver, Colorado, June.

Carlos Ramisch, Silvio Ricardo Cordeiro, Agata Savary, Veronika Vincze, Verginica Barbu Mititelu, Archna Bhatia, Maja Buljan, Marie Candito, Polona Gantar, Voula Giouli, Tunga Güngör, Abdelati Hawwari, Uxoa Iñurrieta, Jolanta Kovalevskaitė, Simon Krek, Timm Lichte, Chaya Liebeskind, Johanna Monti, Carla Parra Escartín, Behrang QasemiZadeh, Renata Ramisch, Nathan Schneider, Ivelina Stoyanova, Ashwini Vaidya, and Abigail Walsh. 2018. Edition 1.1 of the PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions. In *Proc. of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, Santa Fe, New Mexico, USA, August.

Carlos Ramisch. 2015. *Multiword Expressions Acquisition: A Generic and Open Framework*. Theory and Applications of Natural Language Processing. Springer.

Ivan Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multiword expressions: a pain in the neck for NLP. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, volume 2276 of *Lecture Notes in Computer Science*, pages 189–206. Springer, Berlin.

Agata Savary, Carlos Ramisch, Silvio Ricardo Cordeiro, Federico Sangati, Veronika Vincze, Behrang QasemiZadeh, Marie Candito, Fabienne Cap, Voula Giouli, Ivelina Stoyanova, and Antoine Doucet. 2017. The PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions. In *Proc. of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 31–47, Valencia, Spain, April.

Nathan Schneider and Noah A. Smith. 2015. A corpus and model integrating multiword expressions and supersenses. In *Proc. of NAACL-HLT*, pages 1537–1547, Denver, Colorado, June.

Nathan Schneider, Spencer Onuffer, Nora Kazour, Emily Danchik, Michael T. Mordowanec, Henrietta Conrad, and Noah A. Smith. 2014. Comprehensive annotation of multiword expressions in a social web corpus. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proc. of LREC*, pages 455–461, Reykjavík, Iceland, May.

Nathan Schneider, Dirk Hovy, Anders Johannsen, and Marine Carpuat. 2016. SemEval-2016 Task 10: Detecting Minimal Semantic Units and their Meanings (DiMSUM). In *Proc. of SemEval*, pages 546–559, San Diego, California, USA, June.

Natalia Silveira, Timothy Dozat, Marie-Catherine De Marneffe, Samuel R. Bowman, Miriam Connor, John Bauer, and Christopher D. Manning. 2014. A gold standard dependency corpus for English. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proc. of LREC*, pages 2897–2904, Reykjavík, Iceland, May.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: A free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, September.

Daniel Zeman, Martin Popel, Milan Straka, Jan Hajič, Joakim Nivre, Filip Ginter, Juhani Luotolahti, Sampo Pyysalo, Slav Petrov, Martin Potthast, Francis Tyers, Elena Badmaeva, Memduh Gökirmak, Anna Nedoluzhko, Silvie Cinková, Jan Hajič Jr., Jaroslava Hlaváčová, Václava Kettnerová, Zdeňka Urešová, Jenna Kanerva, Stina Ojala, Anna Missilä, Christopher D. Manning, Sebastian Schuster, Siva Reddy, Dima Taji, Nizar Habash, Herman Leung, Marie-Catherine de Marneffe, Manuela Sanguinetti, Maria Simi, Hiroshi Kanayama, Valeria de Paiva, Kira Droganova, Héctor Martínez Alonso, Çağrı Çöltekin, Umut Sulubacak, Hans Uszkoreit, Vivien Macketanz, Aljoscha Burchardt, Kim Harris, Katrin Marheinecke, Georg Rehm, Tolga Kayadelen, Mohammed

Attia, Ali Elkahky, Zhuoran Yu, Emily Pitler, Saran Lertpradit, Michael Mandl, Jesse Kirchner, Hector Fernandez Alcalde, Jana Strnadová, Esha Banerjee, Ruli Manurung, Antonio Stella, Atsuko Shimada, Sookyoung Kwak, Gustavo Mendonça, Tatiana Lando, Rattima Nitisaroj, and Josie Li. 2017. CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. In *Proc. of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–19, Vancouver, Canada, August.