

Annotating Verbal MWEs in Irish for the PARSEME Shared Task 1.2

Abigail Walsh Teresa Lynn Jennifer Foster

ADAPT Centre

School of Computing

Dublin City University

{[abigail.walsh](mailto:abigail.walsh@adaptcentre.ie), [teresa.lynn](mailto:teresa.lynn@adaptcentre.ie)}@adaptcentre.ie
jennifer.foster@dcu.ie

Abstract

This paper describes the creation of two Irish corpora (labelled and unlabelled) for verbal MWEs for inclusion in the PARSEME Shared Task 1.2 on automatic identification of verbal MWEs, and the process of developing verbal MWE categories for Irish. A qualitative analysis on the two corpora is presented, along with discussion of Irish verbal MWEs.

1 Introduction

Multiword expressions (MWEs) present a well-documented challenge in the field of NLP, given that they appear in a variety of forms, are idiosyncratic in nature, and prevalent in our lexicon (Jackendoff, 1997; Sag et al., 2002; Baldwin and Kim, 2010). That said, their correct handling can aid in a number of NLP tasks, including word-sense disambiguation, parsing, and machine translation (Constant et al., 2017). This has given rise to a number of working groups dedicated to identifying and interpreting MWEs. PARSEME is one such group, with the aim of improving cross-lingual processing of MWEs. Their shared task on the automatic identification of verbal MWEs (Savary et al., 2017; Ramisch et al., 2018) is now in its third iteration, and their guidelines have expanded to include 27 languages. This year saw the addition of Irish, as the first of the Celtic languages to participate. Two corpora of Irish text were created for this shared task: a small corpus consisting of manually labelled verbal MWEs (VMWEs), and a much larger corpus for use in unsupervised VMWE identification.

Research on MWEs in Irish is still sparse, and much work remains to define the types of MWEs that exist. Most of the literature on Irish linguistics and syntax focuses on a theoretical analysis of the language, and any discussion of idiomatic constructions, which are frequently exceptional cases, tends to be brief. (Stenson, 1981; Christian Brothers, 1999; Uí Dhonnchadha, 2009). Some studies offer more in-depth analysis on particular types of MWEs, such as light-verb constructions (Bloch-Trojnar, 2009; Bayda, 2015), the idiomatic use of prepositions with verbs (Ó Domhnalláin and Ó Baoill, 1975) and idioms (Ní Loingsigh, 2016). Others have offered a preliminary categorisation of Irish MWEs (Veselinović, 2006; Walsh et al., 2019). The categorisation carried out in our previous work (Walsh et al., 2019) is largely based on the annotation guidelines developed for the PARSEME shared tasks¹, and as such can be used as a starting point for the development of a comprehensive set of VMWE categories for Irish.

2 Verbal MWE Categories in Irish

Given that the focus of PARSEME is on the identification of verbal MWEs, some categories of MWEs considered in our previous work, such as nominal compounds or fixed expressions, are excluded. The categories examined here include two universal categories (verbal idioms and light verb constructions) that are found in all participating languages of the PARSEME shared task: two quasi-universal categories

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

¹<https://parseme.fr/lis-lab.fr/parseme-st-guidelines/1.1/?page=home>

(verb-particle constructions and inherently reflexive verbs) that are valid in many but not all participating languages, and one experimental category (inherently adpositional verbs) which can be optionally annotated.

Light verb constructions (LVCs) are described in the PARSEME guidelines as formed by a verb, *v*, and a (single or compound) noun, *n*, which either directly depends on *v* or is introduced by a preposition. Constructions where *v*'s syntactic subject is *n*'s semantic argument are full LVCs and annotated as `LVC.full`, while constructions where the subject of *v* is the cause or source of the event or state expressed by *n* are annotated as `LVC.cause`. Examples include the `LVC.full` *déan dearmad* (do negligence) ‘forget’ and the `LVC.cause` *cuir áthas* (put joy) ‘make happy’.

Verb particle constructions (VPCs) – sometimes called phrasal verbs – consist of a verb, and a dependent intransitive particle (usually a directional adverb in Irish), where the particle causes a significant shift in meaning in the verb. This change in meaning can be either fully non-compositional (annotated as `VPC.full`, e.g. *tabhair amach* (give out) ‘scold’) or semi-compositional (annotated as `VPC.semi`, e.g. *glan suas* (clean up) ‘clean up’).

Inherently adpositional verbs (IAVs) are considered an experimental category in the PARSEME guidelines and language teams may optionally annotate for this category as a final step in the annotation process. The construction consists of a verb and a dependent prepositional phrase, where the preposition is considered integral to the construction, i.e. “it cannot be omitted without markedly altering the meaning of the verb”. This construction occurs frequently in Irish (e.g. *buail le* (hit with) ‘meet’), and as such it was decided to annotate this category in the current edition, to determine whether future versions of the corpus should contain this category. VMWEs can themselves form part of the IAV construction, as in the IAV *cuir suas le* (put up with) ‘put up with’, which contains the VPC *cuir suas* (put up) ‘put up’, which is why this category must be annotated last.

Verbal idioms (VIDs) are idiomatic constructions with at least two lexicalised parts, including a head and at least one dependent. These dependents can be of several different categories (e.g. *tar an crú ar an tairne* (come the shoe on the nail) ‘come to the test’, *ag siúl le chéile* (at walking with each-other) ‘courting’). Also included in VIDs are sentential expressions with no open slots, such as proverbs (e.g. *Ní neart go cur le chéile* (is-not strength without put with each-other) ‘There’s strength in unity’).

2.1 Difficult Decisions

Annotating LVCs with IAV Many LVCs select for a specific preposition, and the construction never occurs without that preposition (e.g. *déan iarracht ar* (make attempt on) ‘make an attempt at’, and *bain triail as* (take test from) ‘try’). In analysis of the LVC, Irish scholars often include the preposition as an integral part of the construction (Stenson, 1981; Bloch-Trojnar, 2009; Bayda, 2015). There was some debate on whether to additionally annotate these LVC constructions with a selected preposition as IAV, as it was difficult to determine if the preposition was integral to the semantics of the construction, and whether omitting it caused a marked change in the meaning of the verb. It was decided not to extend these LVCs with the IAV label unless the preposition clearly caused a shift in meaning to the verb taken alone. This decision may be revisited in future versions of the corpus.

Terminology: VPC versus IAV The term *verb particle construction* is rarely used in Irish linguistic discourse, however *phrasal verbs* are discussed by various authors (Veselinović, 2006; Uí Dhonnchadha, 2009), although there seems to be a difference in the usage of the term. In the PARSEME guidelines, as with many other authors, the term *phrasal verb* is used synonymously with verb particle constructions. In English, particles are often homonymous with prepositions (though not always: e.g. *back*, *through*), although their behaviour is markedly different (Jackendoff, 2002). Uí Dhonnchadha (2009) uses the term *phrasal verb* to refer specifically to verbs that can combine with prepositions to give rise to idiomatic readings, as in *éirigh le* (rise with) ‘succeed’, whereas there does not appear to be any discussion of verb + adverb constructions such as *éirigh amach* (rise out) ‘revolt’. Furthermore, the preposition *le* ‘with’ in *éirigh le* does not appear to follow the specifications for a particle according to the PARSEME

guidelines (i.e. it should not govern a complement), given that it forms a constituent with the noun phrase rather than the verb, as in *d'éirigh léi* (succeeded with-she) ‘she succeeded’. In order to align with the categorisation of *VPCs* outlined by the PARSEME annotation guidelines, it was decided to annotate *éirigh le*, and similar constructions as *I*AV. To avoid confusion in the future, language-specific tests for identifying particles in Irish will be added to the guidelines.

Idiomatic constructions with the verb “be” There are two verbs for “be” in Irish: the substantive verb *bí* conjugates as a normal verb (past tense: *bhí*, present tense: *tá*) and is used to express state, including feelings and emotions, possession, location and existence; and the copula *is* that is used in many other constructions, such as classification, equivalency constructions, or comparisons (Christian Brothers, 1997).

The substantive *bí* can be combined with certain prepositions to express things like possession (*bí + ag/ar* (be at/on) ‘have’, *tá hata agam* (is hat at-me) ‘I have a hat’), desire (*bí + ó* (be from) ‘want’, *tá cáca uaim* (is cake from-me) ‘I want cake’), intention (*bí + faoi* (be under) ‘intend to’, *tá fúim é a dhéanamh* (is under-me it *<part>* doing) ‘I intend to do it’) and membership of a class (*bí + i + <possessive_pronoun>* (be in *<possessive_pronoun>*) ‘be of the class’, *tá mé i mo chócaire* (is I in my chef) ‘I am a chef’), among others. The latter construction was annotated as *VID* as it has two lexicalised dependents (the preposition *i* and the possessive pronoun²). The question of whether the prepositions were integral to the meaning of the other constructions was complicated by the fact that these prepositions could be applied to other verbs to give rise to a similar meaning (e.g. *teastaigh + ó* (be wanted from) = ‘wanting from’), making it unclear whether the prepositions were truly causing a shift in the meaning of the verb. Ultimately, such constructions were not annotated.

The copula appears in certain idiomatic constructions such as copula + preposition combinations (e.g. *is + le* (be with) (possession), *an leatsa an cupán?* (interrogative-be with-you the cup) ‘is the cup yours?’; *is + as* (be from) (origin), *is as Chiarraí mé* (is from Kerry me) ‘I am from Kerry’), copula + adjective combinations (*is + maith + le* (is good with) ‘like’, *is maith liom tae* (is good with-me tea) ‘I like tea’) and other unique idiomatic constructions (Josie + *is + ainm + di* (Josie be name to-her) ‘Josie is her name’). These cannot be categorised as *VMWEs*, given that the syntactic head of copular constructions is not a verb.

Inherently reflexive verbs (IRVs) are a quasi-universal category that occur rarely if at all in Irish. An *IRV* consists of a verb *v* and a reflexive clitic *RCLI* where either *v* never occurs without *RCLI*, or the meaning changes significantly. In Irish, the reflexive pronoun is formed through the combination of *féin* + personal pronoun. Very few constructions appear to require the reflexive pronoun to give a different meaning (possibly: *iompair mé* ‘I carry’ vs. *iompair mé féin* (carry I self) ‘I behave myself’). However, certain verb + inflected preposition constructions can imply reflexivity (e.g. *bailigh + leis* (gather with-him) ‘remove himself/be off’). It was decided to annotate such constructions with *IRV* in this version of the corpus, but this decision may be changed in the future, due to their scarcity and lack of an explicit *RCLI*.

3 Creation of Corpora

Previous editions of the shared task were focused on supervised training of *MWE* identification, through a manually annotated corpus of *VMWEs* that was also annotated for POS information, morphological tags, and dependency trees. This edition, however, included a corpus for unsupervised training, which contained no *VMWE* information, but was automatically tokenised, lemmatised and parsed using UD-Pipe (Straka and Straková, 2017).³

²Possessive pronouns in *VIDs* have special lexicalization status and can be realised by different lexemes depending on number and person.

³Note that the unlabelled corpus include *MWEs* of the kind annotated by UD (i.e. *flat*, *fixed*, *compound* and *compound:prt*)

3.1 Labelled Corpus

The 1,700 sentences in the labelled corpus were taken from version 2.5 of the Irish Universal Dependency Treebank (Zeman et al., 2019). The sentences contain gold-standard annotations at the following levels: POS-information, morphological features and dependency syntax.

Three annotators helped with the manual VMWE annotation. Annotator A had prior experience with the annotation of Irish MWEs and verbal MWEs for other languages according to the PARSEME guidelines, while Annotator B and Annotator C were practised experts in Irish linguistics and syntactic annotation. 100 sentences were annotated by Annotator A as a pilot annotation task, during which the categories `LVC.full`, `LVC.cause`, `VPC.full`, `VPC.semi`, `VID`, `IAV`, `IRV` were fixed upon. 600 sentences were then selected and used by Annotators B and C to test the categorisation guidelines through annotation. Annotator A annotated the other 1000 sentences, and then performed a review on all 1700 sentences, including the 100 pilot sentences and the 600 test sentences.

A portion of the corpus (800 sentences) was doubly annotated at the beginning and the end of the annotation period by Annotator A in order to measure intra-annotator agreement. The first pass of annotation found 312 VMWEs, while the second pass found 270. The F_{measure} was 0.71, the κ score was 0.66 (i.e. substantial agreement), and the κ_{cat} score was 0.84 (i.e. almost perfect agreement) (Landis and Koch, 1977). F_{measure} is an optimistic measure that ignores agreement due to chance, κ is an estimated Cohen’s κ that measures the rate of agreement of annotation for all verbs in the corpus, whereas κ_{cat} takes into account only those VMWEs where both passes agreed on the span.

In total, 662 MWEs were annotated. The most frequent category of VMWE was `LVC.full`, closely followed by `IAV`, while the least frequent category was `IRV`. When compared with the English corpus for edition 1.1 of the shared task (Walsh et al., 2018),⁴ it is clear that the density of VMWEs is much higher for Irish (1 per 2.6 sentences, or 1 out of every 8 verb phrases) versus English (1 per 8.9 sentences, or 1 out of every 47.8 verb phrases).⁵ Given that over a quarter of the VMWEs annotated were `IAV`, there is a strong argument for consistently annotating this category – in Irish if not for other languages.

Category	#Annotations	Category	#Annotations
LVC.full	201	VPC.full	28
IAV	183	VPC.semi	20
LVC.cause	119	IRV	6
VID	105	Total	662

Table 1: Number of annotations per category.

3.2 Unlabelled Corpus

The unlabelled corpus consists of 1,379,824 sentences compiled from the sources listed in Table 2.⁶ UD-Pipe trained on v2.5 of the Irish UD treebank was used to perform the following steps automatically: tokenisation, POS-tagging, lemmatisation, morphological analysis, and dependency parsing. To aid correct splitting of sentences, a pre-processing step was included where a period was added at the end of each line where it did not already exist. Based on a manual inspection of a subsection of the data (100 sentences from each source), some issues were noticed with the lemmatisation (e.g. *dtagraíonn* lemmatised to *tagraigh* when it should be *tagair*; lemma *n-oibrítí* has both initial mutation and is in its plural form), tokenisation (*d’imir* should be tokenised into *d’* and *imir*) and POS-tagging (*is* tagged as `AUX Cop` when it should be `CCONJ Coord`), which we assume affect parsing.⁷

3.3 Performance of the Shared Task Systems

The task of identification incorporates two subtasks: identifying the span of candidate VMWEs, and labelling these candidates. This edition of the shared task focused on the handling of *unseen* VMWEs,

⁴As Irish is the only Celtic language in the PARSEME shared task, English can be considered the closest language neighbour.

⁵# verb phrases estimated using POS information from released cupt files

⁶Text from Vicipéid Irish Wikipedia accessed 1/11/2019 and text from OPUS accessed at <http://opus.nlpl.eu/>

⁷To give an upper bound on parsing accuracy, UDPipe achieves UAS 0.85 and LAS 0.78 on the v2.5 test set.

Source	Size	License
Citizen’s Information website	10,297	CC BY 4.0
EU Bookshop (OPUS)	113,363	open-source
Paracrawl (OPUS)	782,769	Creative Commons CC0 Licence
Tatoeba (OPUS)	1,894	CC-BY 2.0 FR
Vicipéid	302,838	GNU Free Documentation License (GFDL)

Table 2: Sources of unlabelled data, size in # sentences, and licence of the source

i.e. identifying VMWEs that were not annotated in the training and development datasets. To that end, the annotated data was split so as to include at least 300 unseen VMWEs in the test set.

Of the 9 systems participating in the shared task, 6 were submitted for Irish, with 5 of them achieving $F1$ scores above 0. The highest achieved $F1$ score for unseen MWEs in Irish was 19.54, while the cross-lingual macro-average $F1$ score (based on unseen MWEs) for the same system was 38.53. The categories `IAV` and `VPC.full` appear the easiest to identify, while `VID` proved difficult. `VPC.semi` and `IRV` were not identified at all, possibly as the number of examples of each was too few (20 and 6 respectively).

In general, the systems performed more poorly on Irish when compared with other languages, particularly compared to Hindi, which had a similarly sized corpus, and the best unseen MWE-based $F1$ score was 53.11. The language that performed most similarly was Hebrew, where the best unseen MWE-based $F1$ score was 19.59. There are a number of reasons that could explain the poor performance on the Irish dataset. The dataset contained a relatively small number of VMWEs in the corpus (662), when compared to the second smallest number, which was 1034 in Hindi. In addition, as a result of including 301 unseen VMWEs in the Irish test set, the rate of unseen VMWEs with regards to the training and development set was 0.69, the highest for any language. Another possible reason for the high rate of unseen VMWEs occurring is the source of the annotated data; the sentences in the Irish UD treebank (Lynn and Foster, 2016) come from a balanced corpus with a mixture of domains and genres. This can result in MWEs of varying types occurring throughout the data. Given the proportionally higher rate of unseen VMWEs, coupled with the smaller amount of data overall, it is unsurprising that systems did not perform as well on the Irish data as on other languages.

4 Conclusion

This paper describes an initial attempt at the manual annotation of Irish verbal MWEs, including developing a categorisation scheme that aligns with the PARSEME annotation guidelines. It was found that seven of the categories were applicable to Irish language, and the experimental category of `IAV` occurred frequently.

The results of this annotation are explored, along with results from participating systems in the shared task. It appears that the submitted systems found the task of automatic identification particularly difficult for Irish; this is likely due to the small size of the corpus and number of VMWEs annotated, the high rate of unseen VMWEs in the test data when compared to other languages, and the relatively large number of potential categories that increases the complexity of the task.

In the future we plan to continue the work of manual annotation of these VMWEs, particularly in defining the categories more precisely, refining the Irish-specific guidelines and adding language-specific tests for certain categories such as `VPCs`, and expanding the size of the corpus.

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

- Timothy Baldwin and Su Nam Kim. 2010. Multiword expressions. *Handbook of Natural Language Processing, Second Edition*, pages 267–292, 01.
- Victor Bayda. 2015. Irish constructions with bain. *Yn llawen iawn, yn llawn iaith: Proceedings of the 6th International Colloquium of Societas Celto-Slavica. Vol. 7 of Studia Celto-Slavica. Johnston, D., Parina, E. and Fomin, M. (eds)*, 7:213–228, 01.
- Maria Bloch-Trojnar. 2009. On the Nominal Status of VNs in Light Verb Constructions in Modern Irish. In *PASE Papers 2008. Vol. 1: Studies in Language and Methodology of Teaching Foreign Languages*, page 25–33, Wrocław: Oficyna Wydawnicza ATUT.
- The Christian Brothers. 1997. *New Irish Grammar*. Dublin: Fallon.
- The Christian Brothers. 1999. *Graiméar Gaeilge na mBráithre Críostaí*. An Gúm, Baile Átha Cliath.
- 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.
- Ray Jackendoff. 1997. *The Architecture of the Language Faculty*. Linguistic Inquiry Monographs volume 28. MIT Press.
- Ray Jackendoff. 2002. English particle constructions, the lexicon, and the autonomy of syntax. *Verb-Particle Explorations*, pages 67–94.
- J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174.
- Teresa Lynn and Jennifer Foster. 2016. Universal Dependencies for Irish. In *Proceedings of the Second Celtic Language Technology Workshop*, Paris, France.
- Katie Ní Loingsigh. 2016. *Tiomsú agus Rangú i mBunachar Sonraí ar Chnuasach Nathanna Gaeilge as Saothar Pheadair Uí Laoghaire*. Ph.D. thesis, Dublin City University.
- Carlos Ramisch, Silvio Ricardo Cordeiro, Agata Savary, Veronika Vincze, Verginica Barbu Mititelu, Archana Bhatia, Maja Buljan, Marie Candito, Polona Gantar, Voula Giouli, Tunga Güngör, Abdelati Hawwari, Uxoia 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 *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 222–240. ACL.
- Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann A. Copestake, and Dan Flickinger. 2002. Multiword Expressions: A Pain in the Neck for NLP. In *Proceedings of Computational Linguistics and Intelligent Text Processing, Third International Conference*, pages 1–15, Mexico City, Mexico, 02.
- Agata Savary, Carlos Ramisch, Silvio 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 *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 31–47, Valencia, Spain, April. Association for Computational Linguistics.
- Nancy Stenson. 1981. *Studies in Irish syntax*. Ars linguistica. Tübingen: Gunter Narr Verlag.
- Milan Straka and Jana Straková. 2017. Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99, Vancouver, Canada, August. Association for Computational Linguistics.
- Elaine Uí Dhonnchadha. 2009. *Part-of-Speech Tagging and Partial Parsing for Irish using Finite-State Transducers and Constraint Grammar*. Ph.D. thesis, Dublin City University.
- Elvira Veselinović. 2006. How to put up with cur suas le rud and the bidirectionality of contact. *The Celtic Englishes IV*, page 173–190.

Abigail Walsh, Claire Bonial, Kristina Geeraert, John P. McCrae, Nathan Schneider, and Clarissa Somers. 2018. Constructing an annotated corpus of verbal MWEs for English. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 193–200, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

Abigail Walsh, Teresa Lynn, and Jennifer Foster. 2019. IIfhocail: A Lexicon of Irish MWEs. In *Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*, Florence, Italy, 08.

Daniel Zeman, Joakim Nivre, Mitchell Abrams, Noëmi Aeppli, Željko Agić, Lars Ahrenberg, Gabrielé Aleksandravičiūtė, Lene Antonsen, Katya Aplonova, Maria Jesus Aranzabe, Gashaw Arutie, Masayuki Asahara, Luma Ateyah, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, Colin Batchelor, John Bauer, Sandra Bellato, Kepa Bengoetxea, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Savas Cetin, Fabricio Chalub, Jinho Choi, Yongseok Cho, Jayeol Chun, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Elvis de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Hanne Eckhoff, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomáš Erjavec, Aline Etienne, Wograin Evelyn, Richárd Farkas, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Grioni, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Nizar Habash, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỹ, Na-Rae Han, Kim Harris, Dag Haug, Johannes Heinecke, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Jena Hwang, Takumi Ikeda, Radu Ion, Elena Irimia, Olájúké Ishola, Tomáš Jelínek, Anders Johannsen, Fredrik Jørgensen, Markus Juutinen, Hüner Kaşıkara, Andre Kaasen, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Elena Klementieva, Arne Köhn, Kamil Kopaciewicz, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Sookyoung Kwak, Veronika Laippala, Lorenzo Lambertino, Lucia Lam, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phng Lê H'ông, Alessandro Lenci, Saran Lertpradit, Herman Leung, Cheuk Ying Li, Josie Li, Keying Li, KyungTae Lim, Maria Liovina, Yuan Li, Nikola Ljubešić, Olga Loginova, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Cătălina Măranduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Yuji Matsumoto, Ryan McDonald, Sarah McGuinness, Gustavo Mendonça, Niko Miekka, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Tomohiko Morioka, Shinsuke Mori, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskiy, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Lng Nguy~ên Thị, Huy~ên Nguy~ên Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Adédayo Olúokun, Mai Omura, Petya Osenova, Robert Östling, Lilja Øvrelid, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Angelika Peljak-Łapińska, Siyao Peng, Cenel-Augusto Perez, Guy Perrier, Daria Petrova, Slav Petrov, Jason Phelan, Jussi Piitulainen, Tommi A Pirinen, Emily Pitler, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Peng Qi, Andriela Rääbis, Alexandre Rademaker, Loganathan Ramasamy, Taraka Rama, Carlos Ramisch, Vinit Ravishankar, Livy Real, Siva Reddy, Georg Rehm, Ivan Riabov, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Mykhailo Romanenko, Rudolf Rosa, Davide Rovati, Valentin Roşca, Olga Rudina, Jack Rueter, Shoal Sadde, Benoît Sagot, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Dage Särög, Baiba Saulīte, Yanin Sawanakunanon, Nathan Schneider, Sebastian Schuster, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Muh Shohibussirri, Dmitry Sichinava, Aline Silveira, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Antonio Stella, Milan Straka, Jana Strnadová, Alane Suhr, Umüt Sulubacak, Shingo Suzuki, Zsolt Szántó, Dima Taji, Yuta Takahashi, Fabio Tamburini, Takaaki Tanaka, Isabelle Tellier, Guillaume Thomas, Liisi Torga, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Francis Tyers, Sumire Uematsu, Zdeňka Uřešová, Larraitz Uribe, Hans Uszkoreit, Andrius Utka, Sowmya Vajjala, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Lars Wallin, Abigail Walsh, Jing Xian Wang, Jonathan North Washington, Maximilian Wendt, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Mary Yako, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Zhuoran Yu, Zdeněk Žabokrtský, Amir Zeldes, Manying Zhang, and Hanzhi Zhu. 2019. Universal dependen-

cies 2.5. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Tomás Ó Domhnalláin and Dónall Ó Baoill. 1975. *Réamhfhocail le briathra na Gaeilge*. Tuarascáil taighde. Institiúid Teangeolaíochta Éireann.