PARSEME corpora: Annotating verbal multiword expressions in a multilingual framework

Agata Savary

University of Tours, Blois, France

JeRTeh, 4 Feb 2021

Multiword expressions (MWEs)

What is so special about the highlighted expressions?

*The **prime time** speech by **first lady** **Michelle Obama** **set** the house **on fire**. She made **crystal clear** which issues she **took to heart** but she was **preaching to the choir**.*

Multiword expressions (MWEs)

What is so special about the highlighted expressions?

*The **prime time** speech by **first lady** **Michelle Obama** **set** the house **on fire**. She made **crystal clear** which issues she **took to heart** but she was **preaching to the choir**.*

Definition [1]

Combination of at least **two words** which exhibits lexical, morphological, syntactic, semantic and /or statistical **idiosyncrasies**.

Sample idiosyncrasies in MWEs

- **Non-compositional semantics**: the meaning of a MWE is surprising, given the meanings of its component words
  - EN *to **pull** one's **leg*** 'to tease someone playfully'
  - IT ***lasciar perdere*** 'to let lose'⇒'to give up'
- Morphosyntactic **irregularity** (token[a]-specific):
  - FR ***grand-mères*** 'grand$_{sing.masc}$-mothers$_{pl.fem}$' (defective agreement)
  - EN ***by and large*** 'mostly' (Prep Conj Adj is an irregular syntactic structure)
  - EN *to **go nuts*** 'to get crazy' (*go* alone is intransitive)
- Morphosyntactic **inflexibility** (type[b]-specific):
  - EN ***the die is cast*** 'a point of no-retreat has been passed' vs. *#someone cast the die*

[a] Token = individual occurrence

[b] Type = sets of surface realizations of the same expression

Major idiosyncrasy: Semantic non-compositionality

Semantic compositionality [3]

An expression E is semantically compositional if a **compositional semantic calculus** applies to it: given the meanings of E's components and E's **syntactic structure**, a grammar rule allows us to deduce the meaning of E.

Semantic non-compositionality – 3 cases

- A component has no individual meaning, it functions only within MWEs (*cranberry/fossil word*)
  - *to* ***go astray*** 'to become lost'
  - *to* ***let bygones be bygones*** 'to ignore a past offense'
- The syntactic structure is irregular
  - ***by and large*** 'mostly'
  - ***long live*** *the queen!* 'may she live for a long time'
  - *to* ***pretty-print*** 'use beautifying conventions for texts printing'
- The meaning is not deduced regularly
  - *a* ***hot dog*** 'a hot sausage served in a long bread roll' or 'a person showing off dangerous acts'
  - *to* ***pay*** *a* ***visit*** 'to visit'
  - *the* ***Black Sea*** 'a lake in Asia'

Defining idiosyncrasy

One usually tries to distinguish MWEs from "regular" or "free" constructions of the **same syntactic structure**.

Synt. structure	Regular construction	MWE
Adj N	*a hot soup*	*a **hot dog*** 'a hot sausage served in a long bread roll'
V Det N	*to pay a bill*, *to discuss a visit*	*to **pay** a **visit*** 'to visit'
V NP Prep Det N	*to throw fish to the dolphins*	*to **throw** Harry **to the lions*** 'to sacrifice or ruin Harry'
V Part NP	*to put up a flag*	*to **put up** a great performance* 'to show a great level of skill'
V Refl PP	*to wash oneself in the bath*	*to **find oneself** in times of trouble* 'to discover that one is in trouble'

Inflexibility of MWEs = a proxy for semantic non-compositionality

A MWE is (much) **less flexible** (variable) than a regular construction of the same syntactic structure.

Regular construction	**MWE**	**MWE property**
warm soup ≈[1] *hot soup* ≈ *warm stew*	***hot dog*** vs. *#warm dog* vs. *#hot terrier*	Lexical inflexibility
to throw meat to the lions ≈ *to throw meat to the lion*	*to* ***throw*** *someone* ***to the lions*** vs. *#to throw someone to the lion*	Morphological inflexibility
she held her elbow ≈ *she held his elbow*	*she* ***held*** *her* ***tongue*** 'she refrained from expressing her view' vs. *#she held his tongue*	Morpho-syntactic inflexibility

[1] '≈' means that the meaning shift is predictable from the formal change

Inflexibility of MWEs

Regular construction	MWE	MWE property
to throw meat to the lions ≈ *to throw meat to the hungry lions*	*to* ***throw*** *someone* ***to the lions*** vs. *#to throw someone to the hungry lions*	Syntactic inflexibility
he made it for her ≈ *It was made for her by him*	*he* ***made it*** *to the station well in advance* 'he managed to get to the station . . . ' vs. *#it was made by him to the station . . .*	
the die is stolen ≈ *someone stole the die*	***the die is cast*** 'a point of no-retreat has been passed' vs. *#someone cast the die*	
a text in red and blue ≈ *a text in blue and red*	*a photo in* ***black and white*** 'a photo in shades of gray' vs. *#a photo in white and black*	

Partial (in)flexibility of MWEs

Property	MWE respecting the property	MWE violating the property
free subject	*John **held** his **tongue*** ≈ *Adam **held** his **tongue***	***fear lends wings*** 'fear gives you unusual capacities' vs. *#Panic lends wings*
free object	***a little bird told** Suzy* 'Suzy received the information from a secret source' ≈ ***a little bird told** Mary*	*Suzy **crossed** her **fingers** for Tim* 'Suzy wishes good luck to Tim' vs. *#Suzy crossed her thumbs*
verb inflection	*Suzy **crossed** her **fingers*** ≈ *Suzy will **cross** her **fingers***	***a little bird told** Suzy* ≈ *#a little bird will tell Suzy*
object inflection	*Luke **held** his **tongue*** ≈ *Luke and Sue **held** their **tongues***	*Suzy **crossed** her **fingers*** vs. *Suzy crossed her finger*
object modification	*John **broke** my **fall*** 'John made my fall less forceful' ≈ *John **broke** my sudden **fall***	*Suzy **crossed** her **fingers*** vs. *Suzy **crossed** her long **fingers***
free poss. det.	*John **broke** my **fall*** ≈ *John **broke** his/her/our **fall***	*Suzy **crossed** her **fingers*** vs. *#Suzy crossed our fingers*
passive	*John **broke** my **fall*** ≈ *My **fall** was **broken** by John*	***fear lends wings*** vs. *#wings are lent by fear*

(In)flexibility as a matter of scale

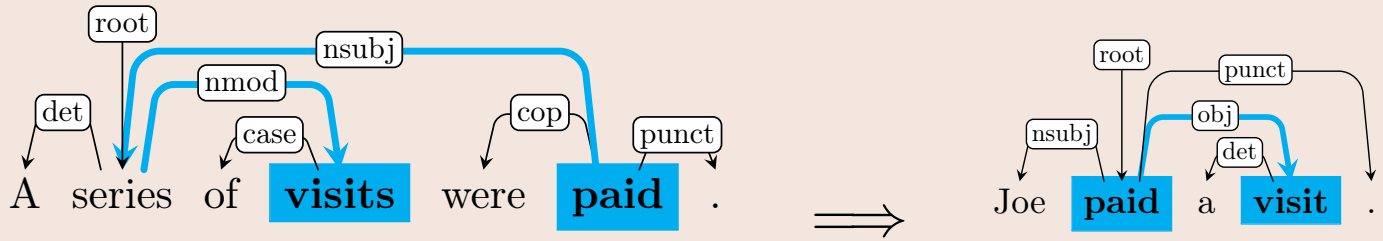
A MWE is **less flexible** than a regular construction of the same syntactic structure but it is often **not totally inflexible**.

Expression	Property						
	Free subject	Free object	Verb inflection	Object inflection	Object modif.	Free poss. det.	Passive
fear lends wings							
*Suzy **held** her **tongue***	✓		✓	✓			
*Suzy **crossed** her **fingers***	✓		✓				✓
***a little bird told** Suzy*		✓		✓	✓	✓	
*Suzy **broke** my **fall***	✓		✓	✓	✓	✓	✓
Suzy lends her books	✓	✓	✓	✓	✓	✓	✓
Suzy held her book	✓	✓	✓	✓	✓	✓	✓
Suzy crossed the road	✓	✓	✓	✓	✓	✓	✓
a little girl told Suzy	✓	✓	✓	✓	✓	✓	✓
Suzy broke my car	✓	✓	✓	✓	✓	✓	✓

Neutralizing flexibility

Canonical form

Least syntactically marked syntactic variant which preserves the idiomatic reading (active voice is less marked then passive, etc.)



Canonical forms are useful for **formalizing** the morpho-syntactic properties of MWEs. This is useful e.g. for **annotation guidelines**.

Lexicalization

MWE components

- **Lexicalized components** – mandatory components, always realized by the same lexemes; without them the MWE cannot occur. They are marked **in bold**.
- **Open slots** – mandatory components which can be realized (relatively) freely
- Example: *she set the house on fire* 'she made the people very excited'
  - *Michelle put the house on fire*, *His wife put the house on fire* → *she* is not lexicalized
  - *#she put the house on fire*[a], *#she set the house in fire*, *#she set the house in blaze* → *set*, *on* and *fire* are lexicalized
  - *she set the assembly/many lobbies on fire* → *the house* is not lexicalized
  - **she put on fire* → the direct object of *put* is an open slot
  - ⟹ *NP* ***set*** *NP* ***on fire***

[a] '#' and '*' signal the loss of idiomatic meaning and ungrammaticality, respectively.

Challenges for NLP

Pervasiveness

Up to 40% of words in a text belong to MWEs. [2, 7]

> *The **prime time** speech by **first lady Michelle Obama set** the house **on fire**. She made **crystal clear** which issues she **took to heart** but she was **preaching to the choir**.*

Here: 18 MWE components for 31 words of the text → 58%

Non-compositionality

Computational methods are mostly **compositional**. Complex phenomena are decomposed into simpler subproblems. Subproblems receive independent solutions, which are then composed to provide global solutions.
MWEs are **semantically non-compositional**. They are challenging for **semantically-oriented NLP applications**.

Machine translation

Word-to-word translations do not capture the idiomatic meaning.

Information retrieval

- The task: for a given query (one or more words), automatically find the relevant documents
- Bag-of-words approach:
  - Eliminate stop words, lemmatize the text, create an **index** (list of words contained in the text with their frequencies)
  - Example: *He **took the bull by the horns*** → {bull – 1, horn – 1, take – 1}
  - Each query word is looked up in the index. The documents containing the query words are weighted and returned.
- Challenges from MWEs:
  - A document contains *He **took the bull by the horns*** 'He dealt decisively with a difficult situation'
  - The query contains *horns of a bull*
  - The document is irrelevant but it will likely be returned

Opinion mining (= sentiment analysis)

- The task: automatically predict the valency (positive, neutral ou negative) of an opinion expressed by a text
- Examples:
  - *Huge respect to the French people for believing in better lives.*
  - *Nothing justifies violence or intimidation against an elected representative of the Republic.*
- Simple bag-of-word technique:
  - Single words are annotated with elementary valency: *respect* → 1, *violence* → -2, *justify* → 1, ...
  - Local rules modify elementary valency:
    - *huge*, *extreme* multiply the valency; *huge respect* → 2*1 = 2; *extreme violence* → 2*(-2) = -4
    - negation inverses valency: *nothing justifies* → -1*1=-1

Opinion mining – challenges from MWEs

Text	Comp. valency	True valency
kick$_0$ *the bucket*$_O$ 'die'	0	-2
go nuts$_O$ 'get crazy'		
make a mountain$_0$ *out of a molehill*$_0$ 'exaggerate'		
it's in the bag$_0$ 'success will obviously be achieved'		
kill$_{-2}$ *two birds*$_O$ *with one stone*$_O$ 'solve two problems with one single action'		
the sky's the limit$_{-1}$ 'there is no limit'		
beyond one's wildest$_{*(-1)}$ *dreams*$_1$ 'much better than expected'		
dark$_{-1}$ *horse* 'a person with a surprising ability'		

Opinion mining – challenges from MWEs

Text	Comp. valency	True valency
kick$_0$ *the bucket*$_O$ 'die'	0	-2
go nuts$_O$ 'get crazy'	0	-2
make a mountain$_0$ *out of a molehill*$_0$ 'exaggerate'	0	-1
it's in the bag$_0$ 'success will obviously be achieved'	0	2
kill$_{-2}$ *two birds*$_O$ *with one stone*$_O$ 'solve two problems with one single action'	-2	1
the sky's the limit$_{-1}$ 'there is no limit'	-1	2
beyond one's wildest$_{*(-1)}$ *dreams*$_1$ 'much better than expected'	-1	2
dark$_{-1}$ *horse* 'a person with a surprising ability'	-1	2

Solutions

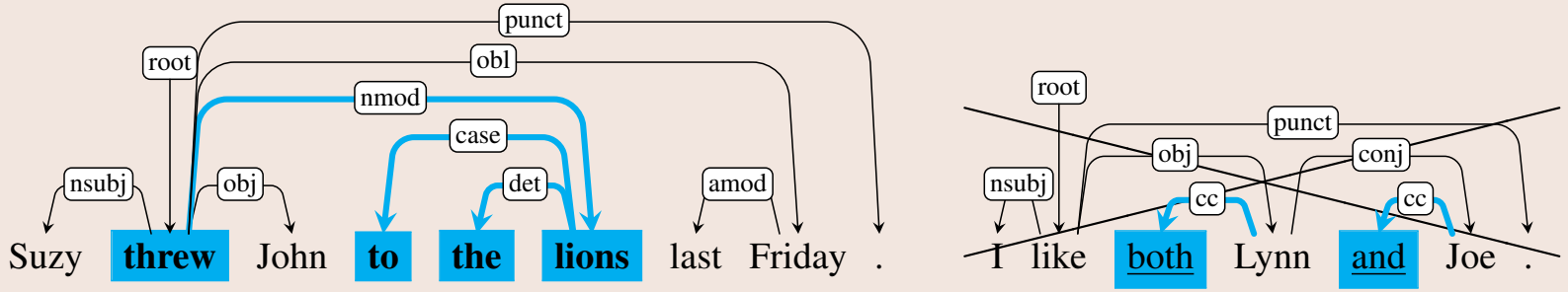
- Automatically identify the MWEs in the text, apply dedicated treatment
- Machine translation
  - rephrase the MWE prior to translation
  - *he **spilled** the **beans*** → *he revealed the secret* → *ha rivelato il segreto*
- Information retrieval
  - don't add the MWE components to the index
  - add the expression as a whole
  - *the re-election was **in the bag*** → {re-election – 1, in the bag – 1}
- Opinion mining
  - assing a valency to the whole expression
  - ***[kill two birds with one stone]***$_2$

Focus on **verbal** MWEs

Verbal MWEs (VMWEs)

Verbal MWEs – MWEs whose **canonical form** is such that:

- its syntactic head is a verb V
- its other lexicalized components form phrases directly dependent on V, i.e. the **dependency subgraph** of the lexicalized components is weakly **connected**



Challenges from **verbal** MWEs

- Discontinuity:
  - EN *Trying hard to **bear** all these more or less important indications **in mind***
  - DE *Klaus Kinkel (FDP) **ging** in seiner Würdigung des Mauerfalls zumindest auf den 9. November 1938 **ein**.*
- Variability: morphological, syntactic, lexical
  - EN *he **broke** my **fall*** vs. *both of my **falls** were hard to **break***
- Ambiguity: idiomatic vs. literal readings
  - EN *she **takes the cake*** 'she is the most outstanding' vs. *she takes the cake*
- Overlaps:
  - EN ***take** a **walk** and then a long **shower*** (coordination)
  - EN ***take** the fact that I **gave up** into account* (interleaving)
  - EN ***let** the cat **out** of the bag* (nesting)
- Multiword tokens
  - ES ***abstener/se*** 'abstain oneself'⇒'abstain' vs. *me abstengo*
  - DE ***auf/machen*** 'out|make'⇒'open' vs. *macht auf*
- Different languages ⇒ different behavior, linguistic traditions...

VMWEs: state of the art in NLP

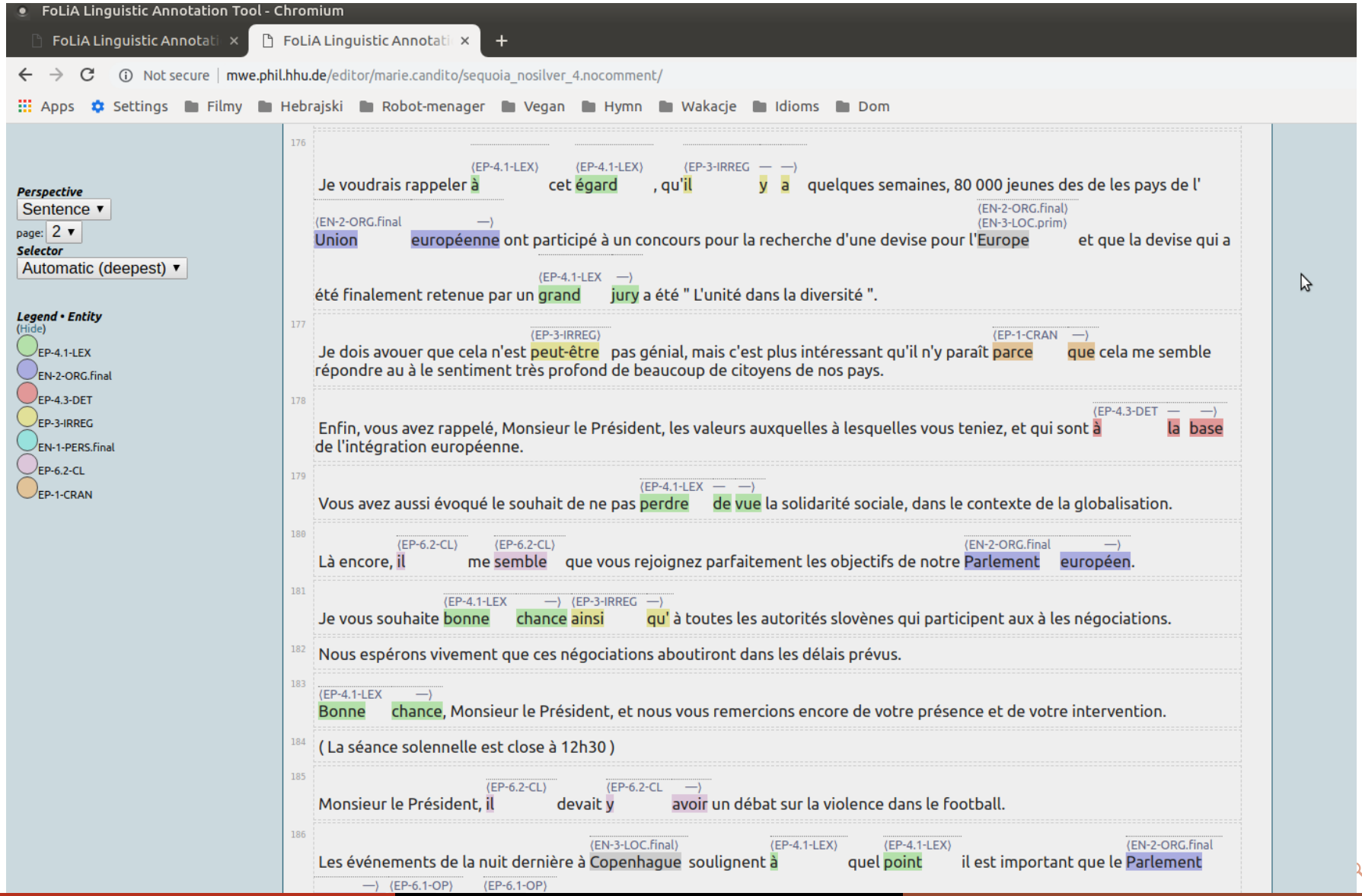
VMWE modeling via corpus annotation

- PARSEME corpus of verbal MWEs [8]

VMWE processing – identification in running text

- PARSEME shared task on automatic identification of verbal MWEs [6]

Annotating MWEs in corpus



PARSEME multilingual corpus of verbal MWEs

International cooperation [8, 6]

- collaborative effort of 14 language teams (20 in edition 1.1)
- unified terminology, typology and annotation guidelines
- corpus of 14 languages, 5,500,000 words, 68,500 annotated VMWEs

Language groups

- **Balto-Slavic**: Polish (PL), Edition 1.1: also BG, HR, LT, SL, CZ
- **Germanic**: German (DE), Swedish (SV) Edition 1.1: also EN
- **Romance**: French (FR), Italian (IT), Romanian (RO), Brazilian Portuguese (PT) Edition 1.1: also ES
- **Others**: Greek (EL), Basque (EU), Gaelic (GA), Hebrew (HE), Hindi (HI), Turkish (TR), Chinese (ZH) Edition 1.1: also AR, FA, HU, MT

VMWE typology

Universal categories (all languages)

- verbal idioms (**VID**)
  - EN *to **call it a day***
- light-verb constructions (**LVCs**)
  - EN *to **give** a **lecture*** (LVC.full)
  - EN *to **grant rights*** (LVC.cause)

Quasi-universal categories (many languages)

- inherently reflexive verbs (**IRVs**)
  - EN *to **help oneself*** 'to take something freely'
- verb-particle constructions (**VPCs**)
  - EN *to **do in*** 'to kill' (VPC.full)
  - EN *to **eat up*** (VPC.semi)
- multi-verb constructions (**MVCs**)
  - HI ***kar le-na*** 'do take.INF'⇒'to do something (for one's own benefit)'

VMWE typology

Language-specific categories

- inherently clitic verbs (**LS.ICV**) [4]

  IT ***prenderle*** 'to take it'⇒'to be beaten'

Unified multilingual annotation guidelines [version 1.2]

the fate of the republic rests on your shoulders

Annotation exercise

- Step 1: identify the candidate and its canonical form: *rests on your shoulders*
- Step 2: determine the lexicalized components
  - ***rests on*** *your/our* ***shoulders***, ***rests on*** *the* ***shoulders*** *of the deputies*, etc.
- Follow the decision tree
  - S.1 [1HEAD] (YES): *rests* is the only verbal head of the whole phrase
  - S.2 [1DEP] (YES): *on shoulders* is the only lexicalized dependent of *rests*
  - S.3 [LEX-SUBJ] (NO): *on* shoulders is not the subject of *rests*
  - S.4 [CATEG] (extended NP): *on shoulders* is a prepositional phrase
  - LVC.0 [N-ABS] (NO): *shoulders* is not abstract
  - VID.1 [CRAN] (NO): all components function also as stand-alone words
  - VID.2 [LEX] (YES): #*remains on your shoulders*, #*rests on your back/arms/head*
- Outcome: **VID**

Inter-annotator agreement

What is IAA?

A measure meant to assess:

- hardness of the annotation task
- quality of the annotation methodology
- quality of the resulting annotations

Popular IAA measure: Cohen's κ

- Setting: two raters classify **N items** into C **mutually exclusive categories**
- Measure:

$$\kappa = \frac{P_O - P_E}{1 - P_E}$$

  - P_O - observed agreement
  - P_E - expected (chance) agreement

Challenges for IAA

Challenges features of VMWE annotation

- *Unitising*, i.e. identifying the boundaries of a VMWE in the text
- *Categorisation*, i.e. assigning each identified VMWE to one of the pre-defined categories
- *Sporadicity*, i.e. the fact that not all text tokens are subject to annotation (unlike in part-of-speech annotation, for instance);
- *Free overlap*, e.g. in (CS) ***ukládal*** *různé* ***sankce*** *a* ***penále*** 'put various sanctions and penalties', where two LVCs share a light verb;

Challenges for IAA

- What are the atomic units (Cohen's items) of annotation?
  - text **tokens** ⇒ categories are not mutually exclusive due to ovelaps
  - text **spans** ⇒ two annotators may end up with different sets of units ⇒ unitising is part of the IAA measure
- In unitising IAA: What is the chance agreement?

Three IAA measures in the PARSEME corpus

F_{span}

- F-measure of annotator A1 prediction wrt. A2 (MWE-based or token-based)

κ_{span}

- Task simplification: For each verb v, decide if v belongs to a VMWE or not.
- Cohen's κ in which the chance agreement is based on the number of verbs

κ_{cat}

- Cohen's κ for VMWEs on which both annotators agree on the span

IAA: data newly annotated in the PARSEME corpus v 1.2

	S	A_1	A_2	F_{span}	κ_{span}	κ_{cat}
Greek	874	293	394	$0.652_{(0.694)}$	$0.608_{(0.665)}$	$\mathbf{0.715}_{(0.673)}$
Irish	800	312	270	0.715	0.663	0.835
Polish	900	252	296	$\mathbf{0.774}_{(0.619)}$	$\mathbf{0.732}_{(0.568)}$	$\mathbf{0.907}_{(0.882)}$
Br. Portug.	1251	253	232	$0.672_{(0.713)}$	$0.640_{(0.684)}$	$\mathbf{0.928}_{(0.837)}$
Swedish	700	364	257	0.734	0.671	0.847
Chinese	3953	883	840	0.584	0.544	0.833

- S = nb. of sentence
- A_1, A_2 = VMWEs per annotator
- subscripts = IAA in edition 1.1 (on different samples)

Format and split

CUPT: extension of the CoNLL-U format

1	-	-	PUNCT	_	_	4	punct	_	_	*
2	si	si	SCONJ	_	_	4	mark	_	_	*
3	vous	il	PRON	_	_	4	nsubj	_	_	*
4	présentez	présenter	VERB	_	_	0	root	_	_	1:LVC.full
5	ou	ou	CCONJ	_	_	8	cc	_	_	*
6	avez	avoir	AUX	_	_	8	aux	_	_	*
7	récemment	récemment	ADV	_	_	8	advmod	_	_	*
8	présenté	présenter	VERB	_	_	4	conj	_	_	2:LVC.full
9	un	un	DET	_	_	10	det	_	_	*
10	saignement	saignement	NOUN	_	_	4	obj	_	_	1;2

Corpus split

- Motivation: **unseen VMWEs** are critically and to identify automatically
- Strategy: split into `train/dev/test` so that `test` has at least 300 unseen VMWEs and the unseen ration is "realistic")

PARSEME corpus applications

PARSEME shared task on automatic identification of VMWEs

- 3 editions, dozens of teams, 22 languages in total
- Training and evaluation based on the PARSEME corpus

Corpus studies

- Characterizing the morpho-syntactic variability of the most frequent VMWEs in French [5]
- Quantifying and characterizing **literal readings** of VMWEs [9]
- Evaluating coverage of a formal grammar with encoded MWEs [10]

Future work

- Extending the annotation guideines to new MWE categories: named entities, nominal, adjectival, adverbial, prepositional MWEs, ...
  - nominal MWEs: non-compositional NPs (*hot dog*), named entities (*Red Sea*), complex terms (*recurrent neural network*)
  - adjectival MWEs: *crystal clear*, *as busy as a bee*
  - adverbial: *all of a sudden*
  - functional: *in front of*, *even if*
- Unifying PARSEME and UD annotation guidelines
- Including new languages and language families
- Continuous corpus enhancements (regular releases)
- Unified multilingual reference datasets with **MWE-annotated corpora** and NLP-oriented **MWE lexicons**.

Keep an ear to the ground 'keep informed'

MWE community

- PARSEME - European network on parsing and MWEs
- MWE section of SIGLEX (special interest group at the ACL) - join both

Keep an ear to the ground 'keep informed'

MWE events

- Yearly MWE workshop co-located with major NLP conferences
  - Joint event with the Linguistic Annotation Workshop community (LAW-MWE-CxG at COLING 2018)
  - Joint event with the WordNet community (MWE-WN at ACL 2019)
  - Joint event with the European Lexicographic Infrastructure (MWE-LEX at COLING 2020)
- PARSEME shared task on automatic identification of VMWEs
  - Editions 1.0, 1.1 and 1.2
  - New edition planned for 2022 (new languages and MWE categories)
- Yearly EUROPHRAS conferences
- MUMTTT workshops (on MWEs in MT)

Keep your nose to the wind 'keep informed'

Book series

Phraseology and Multiword Expressions, at Language Science Press, Berlin

- 2 volumes out, 3 others in the pipeline

MWE resources

- DIMSUM shared task dataset
- SIGLEX-MWE resource list
- PARSEME corpus of verbal MWEs - open-ended project:
  - Serbian is more than welcome! First contat with Cvetana...
  - New MWE categories (adverbials, nominals, ...) will be addressed
- PARSEME annotation guidelines
- PARSEME surveys
  - On MWE annotation in treebanks
  - On lexical resources of MWEs
  - On multilingual MWE resources

*Why do we **eat, sleep and breathe** MWEs?*

'Why are we so enthusiastic and passionate about MWEs?'

- MWEs are fascinating!
  - They convey messages succinctly and efficiently
  - They hide traces of history, stereotypes, and surprising connotations
  - They can be very funny
- MWEs are challenging
  - They are hard to understand for non-native speakers
  - They are signs of a speaker's fluency
  - They behave differently than regular combinations of words
  - They are hard to tokenize, identify, parse, translate automatically
- They are prevalent

Bibliography I

Baldwin, T., and Kim, S. N.
Multiword expressions.
In *Handbook of Natural Language Processing*, N. Indurkhya and F. J. Damerau, Eds., 2 ed. CRC Press, Taylor and Francis Group, Boca Raton, FL, USA, 2010, pp. 267–292.

Gross, M., and Senellart, J.
Nouvelles bases statistiques pour les mots du français.
In *Proceedings of JADT'98, Nice 1998* (1998), pp. 335–349.

Kracht, M.
Compositionality: The very idea.
Research on Language and Computation 5, 3 (2007), 287–308.

Monti, J., Cordeiro, S. R., Ramisch, C., Sangati, F., Savary, A., and Vincze, V.
Advances in Multiword Expression Identification for the Italian language: The PARSEME shared task edition 1.1.
In *Proceedings of Fifth Italian Conference on Computational Linguistics (CLiC-it)* (2018).

Pasquer, C.
Expressions polylexicales verbales : étude de la variabilité en corpus (verbal MWEs : a corpus-based study of variability).
In *Actes des 24ème Conférence sur le Traitement Automatique des Langues Naturelles. 19es REncontres jeunes Chercheurs en Informatique pour le TAL (RECITAL 2017)* (Orléans, France, 6 2017), ATALA, pp. 161–174.

Bibliography II

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., and Xu, H.
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* (online, Dec. 2020), Association for Computational Linguistics, pp. 107–118.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D.
Multiword Expresions: A Pain in the Neck for NLP.
In *Proceedings of CICLING'02* (2002), Springer.

Savary, A., Candito, M., Mititelu, V. B., Bejček, E., Cap, F., Čéplö, S., Cordeiro, S. R., Eryiğit, G., Giouli, V., van Gompel, M., HaCohen-Kerner, Y., Kovalevskaitė, J., Krek, S., Liebeskind, C., Monti, J., Escartín, C. P., van der Plas, L., QasemiZadeh, B., Ramisch, C., Sangati, F., Stoyanova, I., and Vincze, V.
PARSEME multilingual corpus of verbal multiword expressions.
In *Multiword expressions at length and in depth: Extended papers from the MWE 2017 workshop*, S. Markantonatou, C. Ramisch, A. Savary, and V. Vincze, Eds. Language Science Press., Berlin, 2018, pp. 87–147.

Savary, A., Cordeiro, S. R., Lichte, T., Ramisch, C., nurrieta, U. I., and Giouli, V.
Literal Occurrences of Multiword Expressions: Rare Birds That Cause a Stir.
The Prague Bulletin of Mathematical Linguistics 112 (April 2019), 5–54.

Bibliography III

Savary, A., Petitjean, S., Lichte, T., Kallemeyer, L., and Waszczuk, J.
Object-oriented lexical encoding of multiword expressions: Short and sweet.
Lexique 27 (2020), 87–120.