Annotation and automatic identification of light verb constructions in the PARSEME framework

Agata Savary

Université Paris-Saclay, France

Greek SVC workshop, University of Oxford, 5 September 2023

PARSEME

Network

- COST Action on **Parsing and Multiword Expressions** (MWEs) funded by European Commission in **2013-2017**, still **active**
- 31 countries, 30 languages and 6 dialects from 10 language genera
- Outcomes: publications, resources, tutorials, methodologies, PMWE book series

MWE corpora (https://gitlab.com/parseme/corpora/-/wikis/)

- **Collaborative** effort: 26 language teams, 35 language leaders, 200 annotators
- Annotation **guidelines** for **verbal** MWEs **unified** across 26 languages
- Corpora **manually** annotated for MWEs: **26 languages**, open licenses
- **Continuous enhancements** of the guidelines and corpora

Multiword expressions

*The **prime time** speech made 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

*The **prime time** speech made 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**.*

A definition

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

Idiosyncrasy

A mode of behaviour or a property which is **particular** to an (few) individual(s). An **unusual** feature.

Major idiosyncrasy 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'

Major idiosyncrasy 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'

Challenge

Semantic non-compositionality is **hard to test directly**.

Inflexibility: a proxy for semantic non-compositionality

Hypothesis

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

Regular construction	**MWE**	**MWE property**
warm soup $\approx$[1] *hot soup* $\approx$ *warm stew*	***hot dog*** vs. *#warm dog* vs. *#hot terrier*	Lexical inflexibility
to throw meat to the lions $\approx$ *to throw meat to the lion*	*to* ***throw*** *someone* ***to the lions*** vs. *#to throw someone to the lion*	Morphological inflexibility
the die is stolen $\approx$ *someone stole the die*	***the die is cast*** vs. *#someone cast the die*	Syntactic inflexibility

[1] '$\approx$' means that the meaning shift is predictable from the formal change

Focus on **verbal** MWEs – some challenges

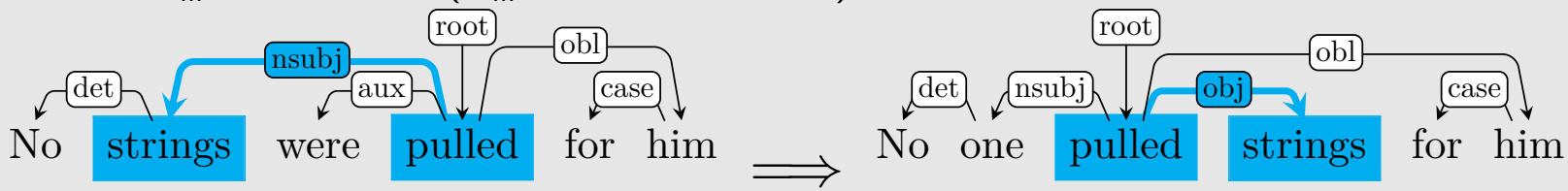
- Discontinuity:
  - EN *Trying hard to **bear** all these more or less important indications **in mind***
- Interleaving:
  - EN ***take** the fact that I **gave up** **into account***
- Multiword tokens
  - DE ***auf**/**machen*** (lit. '*out*/*make*') 'open' vs. *macht auf*
- Flexibility: morphological, syntactic, lexical
  - EN *he **broke** my **fall*** vs. *both of my **falls** were hard to **break***

Neutralizing flexibility

Canonical form

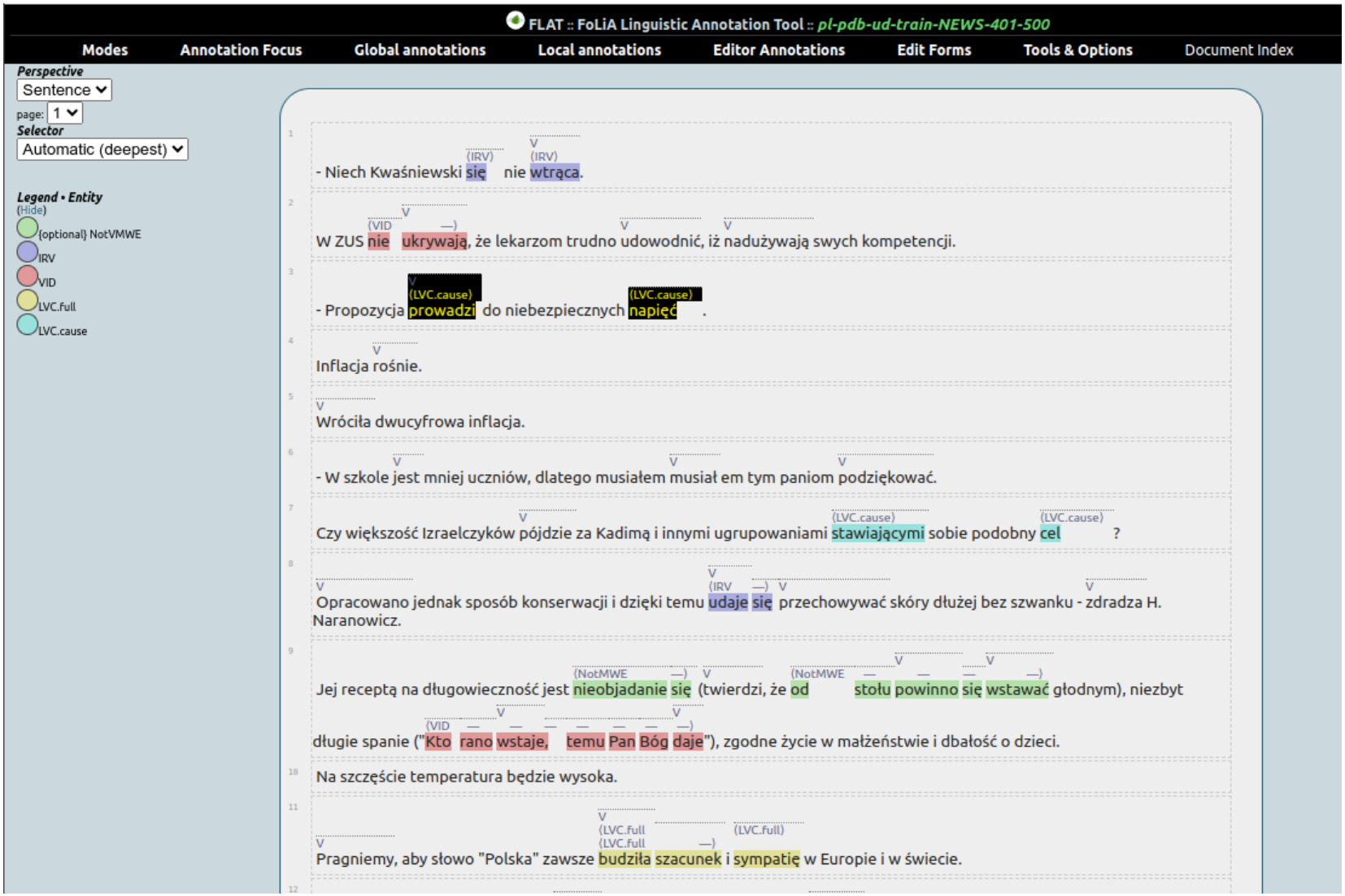
Least syntactically marked syntactic variant which preserves the idiomatic reading.

finite verb $<_m$ infinitive/participle; active voice $<_m$ passive v.; non-negated form $<_m$ negated f.; no extraction $<_m$ extraction, ... ($<_m$ = less marked than)



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

Annotating MWEs in a corpus



PARSEME annotation guidelines

(https://parsemefr.lis-lab.fr/parseme-st-guidelines/1.3)

Objectives

- Formalise idiomaticity in a **cross-linguistically unified** and **computationally tractable** way
- Unify what is truly **similar**, emphasise what is **language-specific**
- Make the annotation **reproducible**

VMWE typology (v. 1.3)

- **Universal** categories (valid for all languages):
  - light verb constructions (**LVCs**)
    - **LVC.full**: EN *to **give** a **lecture***
    - **LVC.cause**: EN *to **grant rights***
  - verbal idioms (**VIDs**)

    EN *to **call it a day***
- **Quasi-universal** categories (valid for many languages):
  - inherently reflexive verbs (**IRVs**)

    FR ***s'évanouir*** 'to faint'
  - verb-particle constructions (**VPCs**)
    - **VPC.full** EN *to **do in*** 'to kill'
    - **VPC.semi** EN *to **eat up*** 'to eat completely'
  - multi-verb constructions (**MVCs**)

    HI ***kar le-na*** (lit. *'do take.INF'*) 'to do something (for one's own benefit)'
- **Experimental** (optional) category
  - inherently adpositional verbs (**IAVs**)

    EN *to **come across** sth/sb*, *to **rely on** sth/sb*

Towards reproducibility – guidelines as decision diagrams

If you are annotating **Italian** or **Hindi**, go to the Italian-specific decision tree or Hindi-specific decision tree. I

- ↳ Apply test S.1 - [**1HEAD**: Unique verb as functional syntactic head of the whole?]
  - ↳ **NO** ⇒ Apply the VID-specific tests ⇒ *VID tests positive?*
    - ↳ **YES** ⇒ Annotate as a VMWE of category **VID**
    - ↳ **NO** ⇒ It is not a VMWE, **exit**
  - ↳ **YES** ⇒ Apply test S.2 - [**1DEP**: *Verb v has exactly one lexicalized dependent d?*]
    - ↳ **NO** ⇒ Apply the VID-specific tests ⇒ *VID tests positive?*
      - ↳ **YES** ⇒ Annotate as a VMWE of category **VID**
      - ↳ **NO** ⇒ It is not a VMWE, **exit**
    - ↳ **YES** ⇒ Apply test S.3 - [**LEX-SUBJ**: *Lexicalized subject?*]
      - ↳ **YES** ⇒ Apply the VID-specific tests ⇒ *VID tests positive?*
        - ↳ **YES** ⇒ Annotate as a VMWE of category **VID**
        - ↳ **NO** ⇒ It is not a VMWE, **exit**
      - ↳ **NO** ⇒ Apply test S.4 - [**CATEG**: *What is the morphosyntactic category of d?*]
        - ↳ **Reflexive clitic** ⇒ Apply IRV-specific tests ⇒ *IRV tests positive?*
          - ↳ **YES** ⇒ Annotate as a VMWE of category **IRV**
          - ↳ **NO** ⇒ It is not a VMWE, **exit**
        - ↳ **Particle** ⇒ Apply VPC-specific tests ⇒ *VPC tests positive?*
          - ↳ **YES** ⇒ Annotate as a VMWE of category **VPC.full** or **VPC.semi**
          - ↳ **NO** ⇒ It is not a VMWE, **exit**
        - ↳ **Verb with no lexicalized dependent** ⇒ Apply MVC-specific tests ⇒ *MVC tests positive?*
          - ↳ **YES** ⇒ Annotate as a VMWE of category **MVC**
          - ↳ **NO** ⇒ Apply the VID-specific tests ⇒ *VID tests positive?*
            - ↳ **YES** ⇒ Annotate as a VMWE of category **ID**
            - ↳ **NO** ⇒ It is not a VMWE, **exit**
        - ↳ **Extended NP** ⇒ Apply LVC-specific decision tree ⇒ *LVC tests positive?*
          - ↳ **YES** ⇒ Annotate as a VMWE of category **LVC**
          - ↳ **NO** ⇒ Apply the VID-specific tests ⇒ *VID tests positive?*
            - ↳ **YES** ⇒ Annotate as a VMWE of category **VID**

VID-specific decision diagram

↳ Apply test VID.1 - [**CRAN**: *Candidate contains cranberry word?*]
- ↳ **YES** ⇒ It is a VID, exit.
- ↳ **NO** ⇒ Apply test VID.2 - [**LEX**: *Regular replacement of a component ⇒ unexpected meaning shift?*]
  - ↳ **YES** ⇒ It is a VID, exit.
  - ↳ **NO** ⇒ Apply test VID.3 - [**MORPH**: *Regular morphological change ⇒ unexpected meaning shift?*]
    - ↳ **YES** ⇒ It is a VID, exit.
    - ↳ **NO** ⇒ Apply test VID.4 - [**MORPHSYNT**: *Regular morphosyntactic change ⇒ unexpected meaning shift?*]
      - ↳ **YES** ⇒ It is a VID, exit.
      - ↳ **NO** ⇒ Apply test VID.5 - [**SYNT**: *Regular syntactic change ⇒ unexpected meaning shift?*]
        - ↳ **YES** ⇒ It is a VID, exit.
        - ↳ **NO** ⇒ It is not a VID, exit

LVC-specific decision diagram

↳ Apply test LVC.0 - [**N-ABS**: *Is the noun abstract?*]
- ↳ **NO ⇒** It is not an LVC, exit
- ↳ **YES or UNSURE ⇒** Apply test LVC.1 - [**N-PRED**: *Is the noun predicative?*]
  - ↳ **NO ⇒** It is not an LVC, exit
  - ↳ **YES or UNSURE ⇒** Apply test LVC.2 - [**V-SUBJ-N-ARG**: *Is the subject of the verb a semantic argument of the noun?*]
    - ↳ **YES or UNSURE ⇒** Apply test LVC.3 - [**V-LIGHT**: *The verb only adds meaning expressed as morphological features?*]
      - ↳ **NO ⇒** It is not an LVC, exit
      - ↳ **YES ⇒** Apply test LVC.4 - [**V-REDUC**: *Can a verbless NP-reduction refer to the same event/state?*]
        - ↳ **NO ⇒** It is not an LVC, exit
        - ↳ **YES ⇒** It is an **LVC.full**
    - ↳ **NO ⇒** Apply test LVC.5 - [**V-SUBJ-N-CAUSE**: *Is the subject of the verb the cause of the noun?*]
      - ↳ **NO ⇒** It is not an LVC, exit
      - ↳ **YES ⇒** It is an **LVC.cause**

Annotation - decision flow ▸ [FLAT] ▸ [guidelines]

the fate of the republic rests on your shoulders (sentence 4)

Annotation - decision flow [FLAT] [guidelines]

the fate of the republic rests on your shoulders (sentence 4)

- 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**

Annotation - decision flow ▸ [FLAT] ▸ [guidelines]

I hate to put a little pressure on you (sentence 4)

Annotation - decision flow [FLAT] [guidelines]

I hate to put a little pressure on you (sentence 4)

- Step 1: identify the candidate and its canonical form: *put a little pressure on you*
- Step 2: determine the lexicalized components
  - ***put*** *a little* ***pressure*** *on you*, *put more/no/a lot of pressure*, etc.
- Follow the [decision tree]
  - S.1 [1HEAD] (YES): *put* is the only verbal head of the whole phrase
  - S.2 [1DEP] (YES): *pressure* is the only lexicalized dependent of *put*
  - S.3 [LEX-SUBJ] (NO): *pressure* is not the subject of *put*
  - S.4 [CATEG] (extended NP): *pressure* is a nominal phrase
  - LVC.0 [N-ABS] (YES): *pressure* is abstract
  - LVC.1 [N-PRED] (YES): 2 semantic arguments: (i) the person putting pressure, (ii) the person subject to the pressure
  - LVC.2 [V-SUBJ-N-ARG] (YES): I is the subject of *put* and the agent of *pressure*
  - LVC.3 [V-LIGHT] (YES): ***put pressure*** $\approx$ force
  - LVC.4 [V-REDUC] (YES): *my pressure on you*
- Outcome: **LVC.full**

Annotation exercise - decision flow ▸ [FLAT] ▸ [guidelines]

This will put new limits on the nature of the environmental changes
(sentence 54)

Annotation exercise - decision flow ▸ [FLAT] ▸ [guidelines]

This will put new limits on the nature of the environmental changes
(sentence 54)

- Step 1: identify the candidate and its canonical form: *this puts a new limit*
- Step 2: determine the lexicalized components
  - **sets/**puts** a new/strong/unexpected **limit***, etc.
- Follow the ▸ decision tree
  - S.1 (YES) → S.2 (YES) → S.3 (NO) → S.4 (extended NP) →
  - LVC.0 [N-ABS] (YES): *limit* is abstract
  - LVC.1 [N-PRED] (YES): 1 semantic arguments: (i) the thing being limited
  - LVC.2 [V-SUBJ-N-ARG] (NO): *This* is the subject of *put* but not a semantic argument of *limit* (a limit can exist without anything setting it).
  - LVC.5 [V-SUBJ-N-CAUSE] (YES): the ***limit*** originates from *this*
- Outcome: **LVC.cause**

Challenges from LVCs

- LVCs are the **gray zone** between idiomatic and productive expressions
  - The noun usually keeps its original sense
  - The verb may be:
    - specific to few nouns: ***pay visit/attention***
    - shared by many nouns but not easily intechangeable: ***do***/**make a* ***job/research***, ***make***/**do* ***effort***
    - shared by a larger number of nouns: ***bring peace/stability/conflict/...***
- We didn't manage to draw the line between idiomatic and productive LVCs
- We include all LVCs into MWEs in a **reproducible** way
- Most PARSEME tests for LVCs are **semantic** rather than morpho-syntactic (see tests LVC.0–3)
- Test LVC.4 hypothesises **more** flexibility in LVCs than in regular constructions.

PARSEME corpus (v 1.3) – main results [Savary et al.(2023)]

Annotations

Sentences	Tokens	VMWEs	VID	IRV	LVC.full	LVC.cause	**VPC.full**	VPC.semi	IAV	MVC
455,629	9,264,811	127,498	26,214	29,062	**40,933**	3,238	9,164	6,443	7,375	5,032

Facts

- **Diversity**: 26 languages from 13 genera
  - AR, BG, CS, DE, EL, EN, ES, EU, FA, FR, GA, HE, HI, HR, HU, IT, LT, MT, PL, PT, RO, SL, SV, SR, TR, ZH
  - Baltic, Basque, Celtic, Chinese, Germanic, Greek, Indic, Iranian, Romance, Semitic, Slavic, Turkic, Ugric
- **"universality" of LVCs and VIDs** is confirmed
- quantitative and qualitative **importance of IRVs** is discovered
- overlapping and nesting is very rare

Greek

Sentences	Tokens	VMWEs	VIDs	IRVs	LVC.full	LVC.cause	VPC.full	MVC
26,175	698,424	8,508	2,841	1	**5,293**	**179**	143	51

Greek in PARSEME

- One of the biggest corpora:
  - 5th (# tokens), and 4th (# VMWEs), 1st (# LVC.full)
- Large and very active language team (**Voula Giouli**, Aggeliki Fotopoulou, Vassiliki Foufi, Sevasti Louizou, **Stella Markantonatou**, Stella Papadelli, Natasa Theoxari)
- Important roles in the MWE community (MWE section representative at SIGLEX, volume editors, working group leaders, task leaders, . . .)

Corpus studies – findings about LVCs

- **LVC.full** is the **largest** category
- LVCs are **shorter** and more **discontinuous** than VIDs; discontinuity outliers are German and Arabic [Savary et al.(2018), Hadj Mohamed et al.(2022)]
- FR LVCs exhibit much higher morphosyntactic variability than in VIDs [Pasquer(2017)]
  - *il **rend** les derniers **hommages*** 'he pays the last tributes'
- MT LT: Some verbs in LVCs (MT *ta* 'to give' LT *sudaryti* 'to make') connect with **many nouns**, others (MT *talab* 'ask', LT *duoti* 'to give') with **few**. Most predicative nouns combine with few light verbs, a few combine with many [Savary et al.(2018)].
- DE EL EU PL PT **Literal readings** of LVCs (and any VMWEs) occur very rarely in corpora [Savary et al.(2019)]:
  - PL *Zdarzenie **miało miejsce** w minioną sobotę* (lit. '*Event had place in last Saturday*') 'The event took place last Saturday'
  - PL *Łódź miała miejsce postoju na przystani* (lit. '*Boat had place of.parking on harbor*') 'The boat had its parking lot in the harbor'
- AR Some LVCs show semantic **dupplication**: the LV and the noun have the same root: خرج خروجا (lit. '*he exited the exit*') 'he went out' [Hadj Mohamed et al.(2022)]

PARSEME corpus infrastructure

- PARSEME [wiki] - extensive documentation of corpora and tools
- Language leaders guide
- User guides
- Gitlab repositories for all languages [language table]
- Corpus validators, converters, filters, release automation ...
- Data quality tools
  - Consistency checks [e.g. for Greek]
- Corpus browser [Grew-match]

MWE identification (MWEI) [Constant et al.(2017)]

- INPUT: text with **morpho-syntactic annotations**
- OUTPUT: text annotated with MWEs

PARSEME shared task on automatic identification of VMWEs [Savary et al.(2017), Ramisch et al.(2018), Ramisch et al.(2020)]

Goal

Automatically identify all VMWE occurrences in running text.

Multilingual framework

- 14–20 languages from 10–13 genera
- Software authors have access to an annotated **corpus** (PARSEME **training** subcorpus)
- Software systems **learn regularities** of VMWEs from the annotated corpus
- They automatically **reproduce** annotation on new, non-annotated texts (PARSEME **test** subcorpus).

Evaluation measures for MWE identification

True entities (annotated by a linguist)

*The **prime time speech made** 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**.*

Positives (identified by a system)

The prime time speech made 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.

Precision, recall, F-measure

	MWE-based measures (only full matches count)	Token-based measures (partial matches count)
$\|T\|$	8	20
$\|P\|$	7	19
$\|TP\|$	4	16
Précision: $P = \frac{\|TP\|}{\|P\|}$	$\frac{4}{6} = 0.67$	$\frac{16}{19} = 0.84$
Recall: $R = \frac{\|TP\|}{\|T\|}$	$\frac{4}{8} = 0.5$	$\frac{16}{20} = 0.8$
F-measure: $F = \frac{2*P*R}{\|P\|+\|R\|}$	$\frac{2*0.67*0.5}{0.67+0.5} = 0.57$	$\frac{2*0.84*0.8}{0.84+0.8} = 0.82$

Evaluation dimensions

- Precision, recall and F1-measure
- **Precise-span** (MWE-based) measure vs. **partial-match** (token-based) measure
- **Per-language** scores vs. **cross-lingual** macro-averages
- **General** measures (all VMWEs) vs. **phenomenon-specific** measures (e.g. VMWEs unseen in the)

Results

Cross-lingual macro-averages

Best systems	#Lang	**Unseen** MWE-based				Global MWE-based				Global Token-based			
		P	R	F1	#	P	R	F1	#	P	R	F1	#
ERMI	14/14	25.3	27.2	26.2	1	64.8	52.9	58.2	2	73.7	54.5	62.6	2
Seen2Seen	14/14	36.5	00.6	**01.1**	2	76.2	58.6	**66.2**	1	78.6	57.0	**66.1**	1
MTLB-STRUCT	14/14	36.2	41.1	**38.5**	1	71.3	69.1	**70.1**	1	77.7	70.9	**74.1**	1
TRAVIS-multi	13/14	28.1	33.3	30.5	2	60.7	57.6	59.1	3	70.4	60.1	64.8	2
TRAVIS-mono	10/14	24.3	28.0	26.0	3	49.5	43.5	46.3	4	55.9	45.0	49.9	4

Per-language scores

System	**Global MWE-based F-score**													
	DE	**EL**	**EU**	**FR**	**GA**	**HE**	**HI**	**IT**	**PL**	**PT**	**RO**	**SV**	**TR**	**ZH**
ERMI	0.52	0.61	0.73	0.61	0.20	0.31	0.60	0.44	0.69	0.64	0.84	0.63	0.64	0.61
MTLB-STRUCT	0.76	0.73	0.80	0.79	0.30	0.48	0.74	0.64	0.81	0.73	0.90	0.72	0.69	0.70
Seen2Seen	0.69	0.67	0.77	0.79	0.27	0.43	0.54	0.65	0.82	0.73	0.82	0.71	0.63	0.49
TRAVIS-mono	0.71	0.13		0.83			0.05	0.61	0.82		0.91	0.67	0.71	0.72
TRAVIS-multi-	0.67	0.72	0.75	0.77	0.07	0.42	0.51	0.59	0.79		0.87	0.69	0.69	0.70

Results for LVCs

System	LVC.full MWE-based F-score													
	DE	**EL**	**EU**	**FR**	**GA**	**HE**	**HI**	**IT**	**PL**	**PT**	**RO**	**SV**	**TR**	**ZH**
ERMI	0.18	0.66	0.75	0.52	0.10	0.35	0.60	0.29	0.57	0.68	0.78	0.48	0.68	0.36
HMSid				0.83										
MTLB-STRUCT	0.56	0.74	0.80	0.76	0.24	0.51	0.71	0.53	0.73	0.74	0.86	0.58	0.72	0.61
Seen2Seen	0.50	0.71	0.77	0.71	0.12	0.48	0.48	0.67	0.71	0.71	0.88	0.62	0.63	0.34
TRAVIS-mono	0.52	0.09		0.75			0.07	0.51	0.75		0.90	0.52	0.72	0.58
TRAVIS-multi	0.40	0.74	0.76	0.70	0.00	0.44	0.55	0.50	0.70		0.80	0.50	0.71	0.56

Summary

- The PARSEME annotations guidelines for **verbal** are **unified** accross 26 languages (including modern Greek), with relatively few **language-specific** sections
- Annotation follows a **decision diagram** (unique starting point), for the sake of **reproducibility**
- Tests are driven by the **syntactic structure**
- Non-compositionality is a matter of **scale** but decisions must be **binary**
- **Semantic non-compositionality** is the major property to capture but is **hard to test directly**
- Lexical and morpho-syntactic **inflexibility** is considered a **proxy** for semantic non-compositionality
- **LVCs** are exceptional: LVC-specific tests are semantic or assuming a **larger syntactic flexibility** than regular constructions
- LVCs are more **flexible** and **discontinuous** than other VMWEs
- VMWE identification is a still **unsolved NLP task**; previously unseen VMWEs are particularly challenging
- LVC identification – globally as hard as for all VMWEs:
  - **simplicity**: frequent light verbs, predictable structure, predicative nouns
  - **hardness**: morpho-syntactic variability, discontinuity

Bibliography I


Constant, M., Eryiğit, G., Monti, J., van der Plas, L., Ramisch, C., Rosner, M., and Todirascu, A. (2017).
Multiword Expression Processing: A Survey.
Computational Linguistics, 43(4), 837–892.


Hadj Mohamed, N., Khelil, C. B., Savary, A., Keskes, I., Antoine, J.-Y., and Hadrich, L. B. (2022).
Annotating Verbal Multiword Expressions in Arabic: Assessing the Validity of a Multilingual Annotation Procedure.
In 13th Conference on Language Resources and Evaluation (LREC 2022), pp. 1839–1848, Marseille, France.


Pasquer, C. (2017).
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), pp. 161–174, Orléans, France. ATALA.


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., and 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), pp. 222–240. Association for Computational Linguistics.

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. (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, pp. 107–118, online. Association for Computational Linguistics.

Savary, A., Ramisch, C., Cordeiro, S., Sangati, F., Vincze, V., QasemiZadeh, B., Candito, M., Cap, F., Giouli, V., Stoyanova, I., and Doucet, A. (2017).
The PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions.
In Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017), pp. 31–47, Valencia, Spain. Association for Computational Linguistics.

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. (2018).
PARSEME multilingual corpus of verbal multiword expressions.
In S. Markantonatou, C. Ramisch, A. Savary, and V. Vincze, eds., Multiword expressions at length and in depth: Extended papers from the MWE 2017 workshop, pp. 87–147. Language Science Press., Berlin.

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

Bibliography III

Savary, A., Ben Khelil, C., Ramisch, C., Giouli, V., Barbu Mititelu, V., Hadj Mohamed, N., Krstev, C., Liebeskind, C., Xu, H., Stymne, S., Güngör, T., Pickard, T., Guillaume, B., Bejček, E., Bhatia, A., Candito, M., Gantar, P., Iñurrieta, U., Gatt, A., Kovalevskaite, J., Lichte, T., Ljubešić, N., Monti, J., Parra Escartín, C., Shamsfard, M., Stoyanova, I., Vincze, V., and Walsh, A. (2023).
PARSEME corpus release 1.3.
In Proceedings of the 19th Workshop on Multiword Expressions (MWE 2023), pp. 24–35, Dubrovnik, Croatia. Association for Computational Linguistics.