Light Verb Constructions in Universal Dependencies for South Asian Languages

Abishek Stephen, Daniel Zeman

May 25, 2024

Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

unless otherwise stated

Light Verb Constructions

- The LVCs belong to the class of complex predicates with a wide range of combinatorial potential where a verb (VERB) can combine with adjectives (ADJ), adverbs (ADV) or nouns (NOUN).

Light Verb Constructions

- The LVCs belong to the class of complex predicates with a wide range of combinatorial potential where a verb (`VERB`) can combine with adjectives (`ADJ`), adverbs (`ADV`) or nouns (`NOUN`).

- The current annotations in the treebanks of many languages in Universal Dependencies (UD) treat the LVCs as combinations of lexemes that morphosyntactically behave as single words and mark them using the dependency relation `compound` or its subtype `compound:lvc`.

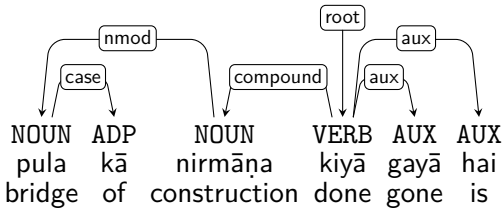
Light Verb Constructions

- The LVCs belong to the class of complex predicates with a wide range of combinatorial potential where a verb (`VERB`) can combine with adjectives (`ADJ`), adverbs (`ADV`) or nouns (`NOUN`).

- The current annotations in the treebanks of many languages in Universal Dependencies (UD) treat the LVCs as combinations of lexemes that morphosyntactically behave as single words and mark them using the dependency relation `compound` or its subtype `compound:lvc`.

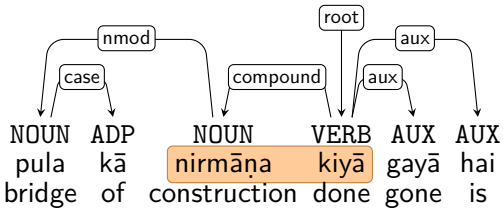
- In the case of South Asian languages this is problematic given the surface-identical noun incorporations and object-verb sequences.

Examples from Hindi



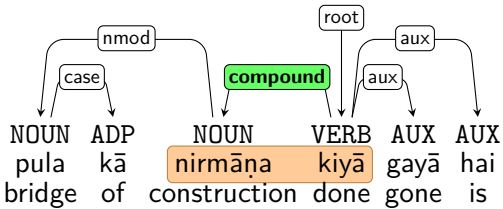
'The bridge has been constructed.'

Examples from Hindi



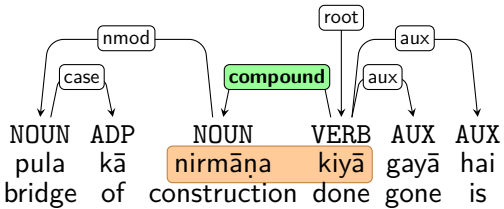
'The bridge has been constructed.'

Examples from Hindi

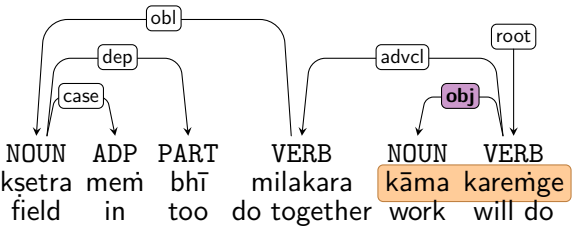


'The bridge has been constructed.'

Examples from Hindi

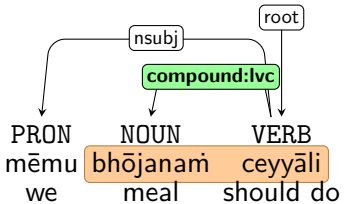


'The bridge has been constructed.'



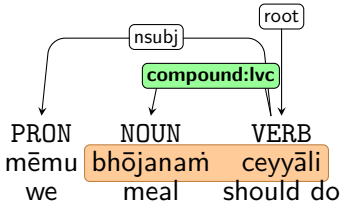
'(We) will work together in the field too.'

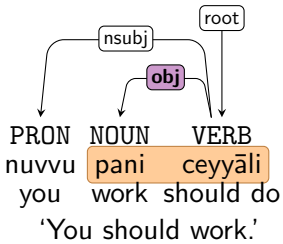
Examples from Telugu



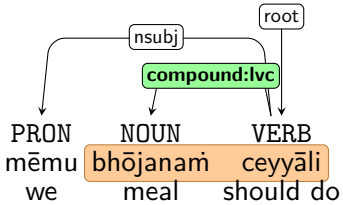
‘We should eat.’

Examples from Telugu

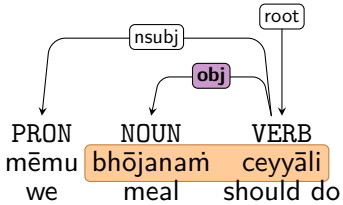




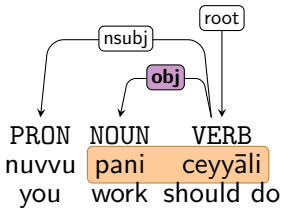
Examples from Telugu



'We should eat.'

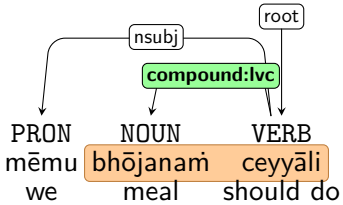


'We should eat.'

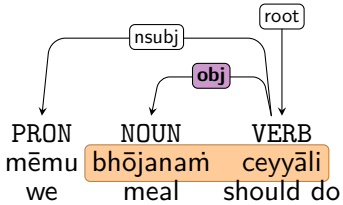


'You should work.'

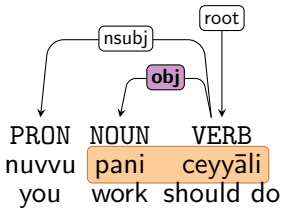
Examples from Telugu



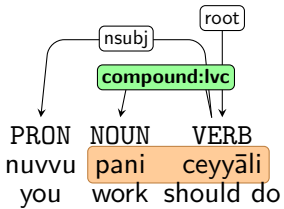
‘We should eat.’



‘We should eat.’



‘You should work.’



‘You should work.’

Data

We use the treebanks of Indo-Aryan and Dravidian languages from UD (de Marneffe et al., 2021) version 2.13.

Language	Treebank	Sentences	Words
Sanskrit	Vedic	3,997	27,117
Sanskrit	UFAL	230	1,843
Hindi	HDTB	16,649	351,704
Hindi	PUD	1,000	23,829
Urdu	UDTB	5,130	138,077
Kangri	KDTB	288	2,514
Bhojpuri	BHTB	357	6,665
Bengali	BRU	56	320
Marathi	UFAL	466	3,847
Sinhala	STB	100	880
Telugu	MTG	1,328	6,465
Tamil	TTB	600	9,581
Tamil	MWTT	534	2,584
Malayalam	UFAL	218	2,403

Data

We use the treebanks of Indo-Aryan and Dravidian languages from UD (de Marneffe et al., 2021) version 2.13.

Language	Treebank	Sentences	Words
Sanskrit	Vedic	3,997	27,117
Sanskrit	UFAL	230	1,843
Hindi	HDTB	16,649	351,704
Hindi	PUD	1,000	23,829
Urdu	UDTB	5,130	138,077
Kangri	KDTB	288	2,514
Bhojpuri	BHTB	357	6,665
Bengali	BRU	56	320
Marathi	UFAL	466	3,847
Sinhala	STB	100	880
Telugu	MTG	1,328	6,465
Tamil	TTB	600	9,581
Tamil	MWTT	534	2,584
Malayalam	UFAL	218	2,403

Data

We use the treebanks of Indo-Aryan and Dravidian languages from UD (de Marneffe et al., 2021) version 2.13.

Language	Treebank	Sentences	Words
Sanskrit	Vedic	3,997	27,117
Sanskrit	UFAL	230	1,843
Hindi	HDTB	16,649	351,704
Hindi	PUD	1,000	23,829
Urdu	UDTB	5,130	138,077
Kangri	KDTB	288	2,514
Bhojpuri	BHTB	357	6,665
Bengali	BRU	56	320
Marathi	UFAL	466	3,847
Sinhala	STB	100	880
Telugu	MTG	1,328	6,465
Tamil	TTB	600	9,581
Tamil	MWTT	534	2,584
Malayalam	UFAL	218	2,403

Indo-Aryan Languages I

Most of these treebanks use the dependency label `compound` to mark the verbo-nominal compounds or LVCs but the Bengali, Marathi, and Sinhala treebanks use the language-specific dependency sub-type label `compound:lvc`.

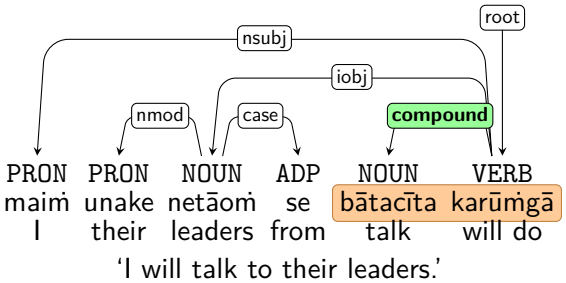


Figure 1: Compound analysis in Hindi (HDTB).

Indo-Aryan Languages II

Bengali, Bhojpuri and Kangri also present a similar picture where the verbs ‘to do’ and ‘to be’ persistently head such constructions.

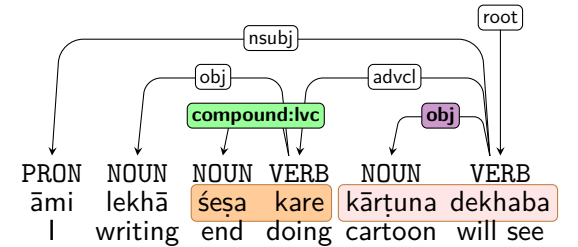


Figure 2: Compound analysis in Bengali (BRU).

Indo-Aryan Languages III

Sinhala happens to be the only Indo-Aryan language in UD to select the noun as a head for LVCs (Liyanage et al., 2023).

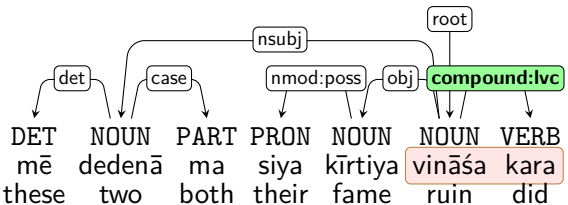


Figure 3: A verbo-nominal compound in Sinhala (STB), headed by the nominal node.

Dravidian Languages

For LVCs, only the compounds with the do-verb *ceyyuka* are labeled as `compound:lvc` in the Malayalam UFAL treebank (Stephen and Zeman, 2023).

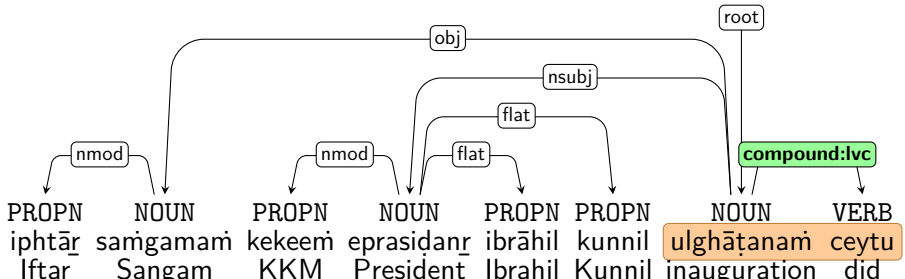


Figure 4: A verbo-nominal compound in Malayalam (UFAL), headed by the nominal node.

Compounding I

The UD taxonomy has a more relaxed definition of compounds: it states that the `compound` relation should be used for combinations of lexemes that morphosyntactically behave as single words, and lexicalization or semantic idiomaticity should not be a criterion for identifying compounds.

Compounding II

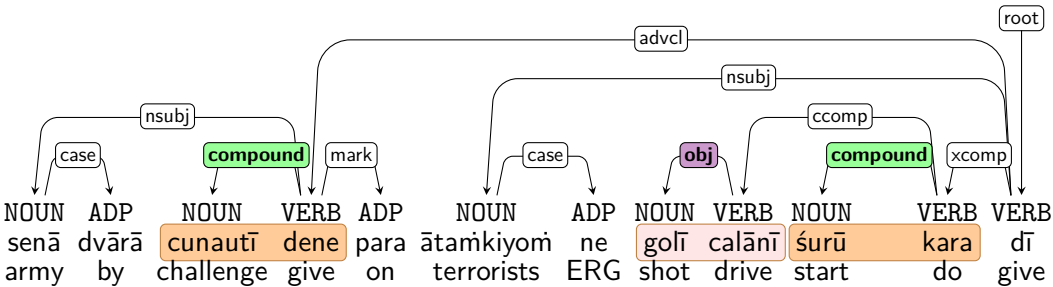


Figure 5: Multiple noun-verb pairs in Hindi (HDTB)

Compounding III

Expressions that would qualify should have a single argument structure or in other words, the syntactic head of an LVC should select all the required arguments and the dependent noun should neither be modified nor have an argument structure of its own. But in the case of the Indo-Aryan languages, this does not seem to be the case!!

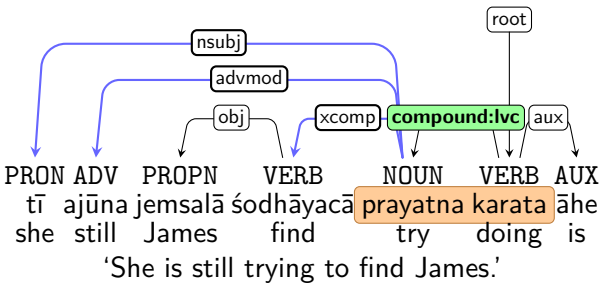


Figure 6: A verbo-nominal compound in Marathi (UFAL), arguments attached to the nominal node.

Compounding IV

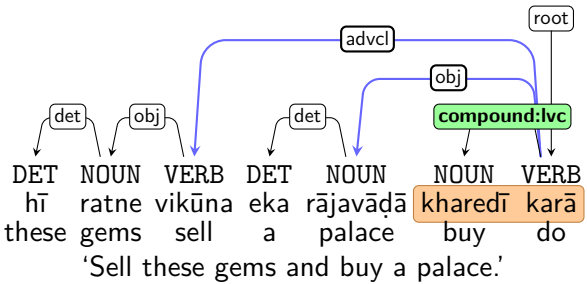


Figure 7: A verbo-nominal compound in Marathi (UFAL), arguments attached to the verbal node.

Noun Incorporation

- According to Haspelmath (2023a), incorporation is an event-denoting noun-verb compound construction in which the noun occupies an argument slot of the verb and occurs in a position where nominal patient arguments cannot occur.

Noun Incorporation

- According to Haspelmath (2023a), incorporation is an event-denoting noun-verb compound construction in which the noun occupies an argument slot of the verb and occurs in a position where nominal patient arguments cannot occur.

- In most Indo-Aryan languages, verbo-nominal predicates must be analyzed as a lexical category but paradoxically enough, the noun is on par with a syntactically independent argument (Mohanan, 1995).

Noun Incorporation

- According to Haspelmath (2023a), incorporation is an event-denoting noun-verb compound construction in which the noun occupies an argument slot of the verb and occurs in a position where nominal patient arguments cannot occur.

- In most Indo-Aryan languages, verbo-nominal predicates must be analyzed as a lexical category but paradoxically enough, the noun is on par with a syntactically independent argument (Mohanan, 1995).

- Therefore, even though noun incorporation is a type of compounding of a syntactic object with the verb, both the object and the verb can have their own argument structures.

Noun Incorporation

- According to Haspelmath (2023a), incorporation is an event-denoting noun-verb compound construction in which the noun occupies an argument slot of the verb and occurs in a position where nominal patient arguments cannot occur.

- In most Indo-Aryan languages, verbo-nominal predicates must be analyzed as a lexical category but paradoxically enough, the noun is on par with a syntactically independent argument (Mohanan, 1995).

- Therefore, even though noun incorporation is a type of compounding of a syntactic object with the verb, both the object and the verb can have their own argument structures.

- Incorporated nouns do not take case or plural markers and external modifiers, they are morphosyntactically different from the regular object nouns.

Case Marking

- Hindi, Urdu, and some other Indo-Aryan languages follow a split-ergative pattern. Animate direct objects use the postposition *ko*. Inanimate direct objects may omit the postposition *ko* in the accusative case.

Case Marking

- Hindi, Urdu, and some other Indo-Aryan languages follow a split-ergative pattern. Animate direct objects use the postposition *ko*. Inanimate direct objects may omit the postposition *ko* in the accusative case.
- Bhojpuri uses the same postposition *(ke)* for accusative, dative, and genitive, making it less obvious when it is selected by the nominal and not the verb.

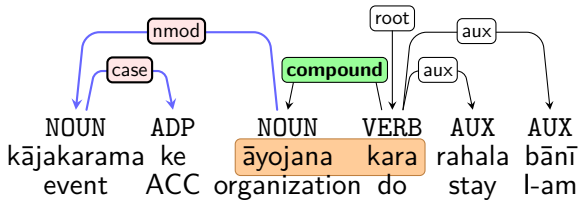


Figure 8: A verbo-nominal compound in Bhojpuri (BHTB) where the nominal conjunct *āyojana* 'organizing' selects the argument *kājakarama* 'event' case marked using the postposition *ke* '.'

Agreement

- Agreement with the verb in transitive-perfective clauses is another signal that the nominal of an LVC candidate is an object rather than part of a compound.

Agreement

- Agreement with the verb in transitive-perfective clauses is another signal that the nominal of an LVC candidate is an object rather than part of a compound.

- And it can also attest to the opposite: In the Hindi example, *mere pitā ne pūjā śurū kar dī hai* 'my father has started the prayer', the verb has a feminine form, agreeing with *pūjā*, while both *pitā* 'father' and *śurū* 'start' are masculine.

Agreement

- Agreement with the verb in transitive-perfective clauses is another signal that the nominal of an LVC candidate is an object rather than part of a compound.

- And it can also attest to the opposite: In the Hindi example, *mere pitā ne pūjā śurū kar dī hai* 'my father has started the prayer', the verb has a feminine form, agreeing with *pūjā*, while both *pitā* 'father' and *śurū* 'start' are masculine.

- In Telugu, the verb agrees with the subject when it is in the nominative case, whereas when there is a dative "subject", the verb agrees with the incorporated noun (Nadimpalli and Lakshmi, 2022).

Agreement

- Agreement with the verb in transitive-perfective clauses is another signal that the nominal of an LVC candidate is an object rather than part of a compound.
- And it can also attest to the opposite: In the Hindi example, *mere pitā ne pūjā śurū kar dī hai* 'my father has started the prayer', the verb has a feminine form, agreeing with *pūjā*, while both *pitā* 'father' and *śurū* 'start' are masculine.
- In Telugu, the verb agrees with the subject when it is in the nominative case, whereas when there is a dative "subject", the verb agrees with the incorporated noun (Nadimpalli and Lakshmi, 2022).
- In many instances of noun-verb sequences agreement between the noun and the verb is observed and represents a deviation from typical compound behavior.

Modification

- One of the signs of compounds is that their parts (and especially the dependent part) cannot be modified individually.

Modification

- One of the signs of compounds is that their parts (and especially the dependent part) cannot be modified individually.
- In Kangri, the nominal *galla* ‘matter’ is modified by the determiner *isadī* ‘this’, suggesting that *galla mannī* is not a compound.

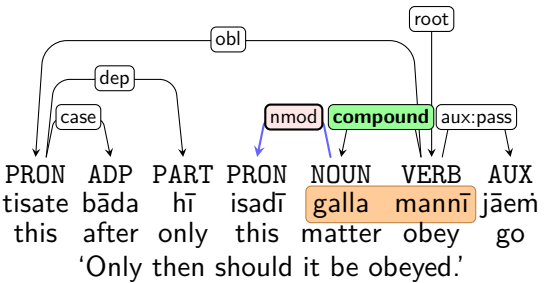


Figure 9: Compound analysis in Kangri (KDTB).

The Current State

- Noun-verb compounds are very frequent in the current UD treebanks of South Asian languages.

The Current State

- Noun-verb compounds are very frequent in the current UD treebanks of South Asian languages.

- In Hindi HDTB, there are 6187 such compounds with the 5 most common verbs alone (out of which 4159 occurrences belong just to *karanā* 'to do'). A similar pattern is found in the smaller Urdu treebank: 3542 occurrences with the top 5 verbs, including 2346 with *krnā* 'to do'.

The Current State

- Noun-verb compounds are very frequent in the current UD treebanks of South Asian languages.
- In Hindi HDTB, there are 6187 such compounds with the 5 most common verbs alone (out of which 4159 occurrences belong just to *karanā* 'to do'). A similar pattern is found in the smaller Urdu treebank: 3542 occurrences with the top 5 verbs, including 2346 with *krnā* 'to do'.
- For example, Hindi *bāta karanā* 'to talk' is a relatively frequent expression and it is usually annotated as `compound` (118 instances), though occasionally it is annotated as `obj` (25 instances).

The Current State

- Noun-verb compounds are very frequent in the current UD treebanks of South Asian languages.
- In Hindi HDTB, there are 6187 such compounds with the 5 most common verbs alone (out of which 4159 occurrences belong just to *karanā* 'to do'). A similar pattern is found in the smaller Urdu treebank: 3542 occurrences with the top 5 verbs, including 2346 with *krnā* 'to do'.
- For example, Hindi *bāta karanā* 'to talk' is a relatively frequent expression and it is usually annotated as `compound` (118 instances), though occasionally it is annotated as `obj` (25 instances).
- We can conclude that in the present versions of the treebanks of South Asian languages, the treatment of noun-verb sequences or LVCs as compounds is not consistent because the interplay of surface level similarities between real noun-verb compounds and noun incorporations somehow weigh down the morphosyntatic cues.

What could be done?

- There should not be a problem if noun-verb compounds satisfying the UD guidelines are marked as `compound:lvc` just to differentiate it from other type of compounds.

What could be done?

- There should not be a problem if noun-verb compounds satisfying the UD guidelines are marked as `compound:lvc` just to differentiate it from other type of compounds.
- This would also handle most of the noun incorporations, but once the nominal participant is case marked, modified or triggering verbal agreement, the sequence should be analyzed differently.

What could be done?

- There should not be a problem if noun-verb compounds satisfying the UD guidelines are marked as `compound:lvc` just to differentiate it from other type of compounds.
- This would also handle most of the noun incorporations, but once the nominal participant is case marked, modified or triggering verbal agreement, the sequence should be analyzed differently.
- One of the solutions could be to label the relation **`obj:lvc`**, modifying Vincze et al. (2017)'s proposal to fit the current UD version.

What could be done?

- There should not be a problem if noun-verb compounds satisfying the UD guidelines are marked as `compound:lvc` just to differentiate it from other type of compounds.
- This would also handle most of the noun incorporations, but once the nominal participant is case marked, modified or triggering verbal agreement, the sequence should be analyzed differently.
- One of the solutions could be to label the relation **`obj:lvc`**, modifying Vincze et al. (2017)'s proposal to fit the current UD version.
- By doing so, there will be a three-way distinction.

What could be done?

- There should not be a problem if noun-verb compounds satisfying the UD guidelines are marked as `compound:lvc` just to differentiate it from other type of compounds.
- This would also handle most of the noun incorporations, but once the nominal participant is case marked, modified or triggering verbal agreement, the sequence should be analyzed differently.
- One of the solutions could be to label the relation **`obj:lvc`**, modifying Vincze et al. (2017)'s proposal to fit the current UD version.
- By doing so, there will be a three-way distinction.
  - Noun incorporations (with a single argument structure) marked as `compound:lvc`.

What could be done?

- There should not be a problem if noun-verb compounds satisfying the UD guidelines are marked as `compound:lvc` just to differentiate it from other type of compounds.
- This would also handle most of the noun incorporations, but once the nominal participant is case marked, modified or triggering verbal agreement, the sequence should be analyzed differently.
- One of the solutions could be to label the relation **`obj:lvc`**, modifying Vincze et al. (2017)'s proposal to fit the current UD version.
- By doing so, there will be a three-way distinction.
  - Noun incorporations (with a single argument structure) marked as `compound:lvc`.
  - Object-verb sequences marked as `obj`.

What could be done?

- There should not be a problem if noun-verb compounds satisfying the UD guidelines are marked as `compound:lvc` just to differentiate it from other type of compounds.
- This would also handle most of the noun incorporations, but once the nominal participant is case marked, modified or triggering verbal agreement, the sequence should be analyzed differently.
- One of the solutions could be to label the relation **`obj:lvc`**, modifying Vincze et al. (2017)'s proposal to fit the current UD version.
- By doing so, there will be a three-way distinction.
  - Noun incorporations (with a single argument structure) marked as `compound:lvc`.
  - Object-verb sequences marked as `obj`.
  - Noun-incorporations with individual noun and verb argument structures as `obj:lvc`.

Conclusion

- We have presented morphosyntactic clues for identifying light verb constructions in South Asian languages, which could prove instrumental in achieving consistent annotations of `compound` and `compound:lvc` dependency relations.

Conclusion

- We have presented morphosyntactic clues for identifying light verb constructions in South Asian languages, which could prove instrumental in achieving consistent annotations of `compound` and `compound:lvc` dependency relations.

- While LVCs as semantically idiosyncratic constructions are widespread in these languages, we have shown that in many cases their syntactic behavior is transparent or very close to standard object-verb constructions.

Conclusion

- We have presented morphosyntactic clues for identifying light verb constructions in South Asian languages, which could prove instrumental in achieving consistent annotations of `compound` and `compound:lvc` dependency relations.
- While LVCs as semantically idiosyncratic constructions are widespread in these languages, we have shown that in many cases their syntactic behavior is transparent or very close to standard object-verb constructions.
- Their compound analysis should be reconsidered and the annotation could be changed to `obj` or `obj:lvc` based on the type of argument sharing.

Thank you!

stephen,zeman@ufal.mff.cuni.cz

References I

Avery D. Andrews. The major functions of the noun phrase. In Timothy Shopen, editor, *Language Typology and Syntactic Description*, page 132–223. Cambridge University Press, Cambridge, UK, 2007.

Rafiya Begum, Karan Jindal, Ashish Jain, Samar Husain, and Dipti Misra Sharma. Identification of conjunct verbs in Hindi and its effect on parsing accuracy. In Alexander F. Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing - 12th International Conference, CICLing 2011, Tokyo, Japan, February 20-26, 2011. Proceedings, Part I*, volume 6608 of *Lecture Notes in Computer Science*, pages 29–40. Springer, 2011. doi: 10.1007/978-3-642-19400-9_3. URL https://doi.org/10.1007/978-3-642-19400-9_3.

Miriam Butt. The light verb jungle. *Harvard Working Papers in Linguistics*, 9, 01 2003.

Marie-Catherine de Marneffe, Christopher Manning, Joakim Nivre, and Daniel Zeman. Universal Dependencies. *Computational Linguistics*, 47(2):255–308, 2021. ISSN 1530-9312. URL https://doi.org/10.1162/coli_a_00402.

Puneet Dwivedi and Daniel Zeman. The forest lion and the bull: Morphosyntactic annotation of the Panchatantra. *Computación y Sistemas*, 22(4):1377–1384, 2018. ISSN 2007-9737. doi: 10.13053/CyS-22-4-3076. URL https://www.cys.cic.ipn.mx/ojs/index.php/CyS/article/view/3076.

Rita Finkbeiner and Barbara Schlücker. *Compounds and multi-word expressions in the languages of Europe*, pages 1–44. De Gruyter, 01 2019. ISBN 9783110632446. doi: 10.1515/9783110632446-001.

Martin Haspelmath. Compound and incorporation constructions as combinations of unexpandable roots, July 2023a. URL https://doi.org/10.5281/zenodo.8137251.

Martin Haspelmath. Defining the word. *WORD*, 69(3):283–297, 2023b. doi: 10.1080/00437956.2023.2237272. URL https://doi.org/10.1080/00437956.2023.2237272.

Oliver Hellwig, Salvatore Scarlata, Elia Ackermann, and Paul Widmer. The treebank of vedic Sanskrit. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 5137–5146, Marseille, France, May 2020. European Language Resources Association. ISBN 979-10-95546-34-4. URL https://aclanthology.org/2020.lrec-1.632.

Aravind K. Joshi. 483 Tree-Adjoining Grammars. In *The Oxford Handbook of Computational Linguistics*. Oxford University Press, 01 2005. ISBN 9780199276349. doi: 10.1093/oxfordhb/9780199276349.013.0026. URL https://doi.org/10.1093/oxfordhb/9780199276349.013.0026.

Sylvain Kahane, Kim Gerdes, and Marine Courtin. Multi-word annotation in syntactic treebanks: Propositions for universal dependencies. In *16th international conference on Treebanks and Linguistic Theories (TLT)*, 2018.

References II

Parameswari Krishnamurthy and Kengatharaiyer Sarveswaran. Towards building a modern written Tamil treebank. In *Proceedings of the 20th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2021)*, pages 61–68, Sofia, Bulgaria, December 2021. Association for Computational Linguistics. URL https://aclanthology.org/2021.tlt-1.6.

Bhadriraju Krishnamurti. Syntax. In *The Dravidian Languages*, Cambridge Language Surveys, pages 420–469. Cambridge University Press, Cambridge, UK, 2003.

Chamila Liyanage, Kengatharaiyer Sarveswaran, Thilini Nadungodage, and Randil Pushpananda. Sinhala dependency treebank (STB). In Loïc Grobol and Francis Tyers, editors, *Proceedings of the Sixth Workshop on Universal Dependencies (UDW, GURT/SyntaxFest 2023)*, pages 17–26, Washington, D.C., March 2023. Association for Computational Linguistics. URL https://aclanthology.org/2023.udw-1.3.

Francesca Masini. Multi-word expressions and morphology, 09 2019. URL https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-611.

Tara Mohanan. Wordhood and lexicality: Noun incorporation in Hindi. *Natural Language & Linguistic Theory*, 13(1):75–134, 1995. ISSN 0167806X, 15730859. URL http://www.jstor.org/stable/4047888.

Tara Mohanan. *Grammatical and Light Verbs*, pages 1–27. John Wiley & Sons, Ltd, 2017. ISBN 9781118358733. doi: https://doi.org/10.1002/9781118358733.wbsyncom016. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118358733.wbsyncom016.

Stefan Müller. Complex predicates: Structure, potential structure and underspecification. In *Linguistic Issues in Language Technology (LiLT) 16*, 2019. URL https://aclanthology.org/2019.lilt-17.3.

Satish Kumar Nadimpalli and Bh VN Lakshmi. Is there noun incorporation in Telugu? *Journal of Language and Linguistic Studies*, 18(2):895–903, 2022.

Joakim Nivre and Veronika Vincze. Light verb constructions in universal dependencies. In *Poster at the 5th PARSEME meeting, Iasi, Romania*, 2015.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106, 2005.

Martha Palmer, Rajesh Bhatt, Bhuvana Narasimhan, Owen Rambow, Dipti Misra Sharma, and Fei Xia. Hindi syntax: Annotating dependency, lexical predicate-argument structure, and phrase structure. In *The 7th International Conference on Natural Language Processing*, pages 14–17, 2009.

Taraka Rama and Sowmya Vajjala. A dependency treebank for Telugu. In *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories (TLT16)*, pages 119–128, Prague, Czechia, 2018. URL https://aclanthology.org/W17-7616.pdf.

References III

Vinit Ravishankar. A Universal Dependencies treebank for Marathi. In Jan Hajič, editor, *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories*, pages 190–200, Prague, Czechia, 2017. URL https://aclanthology.org/W17-7623.

Agata Savary, Cherifa Ben Khelil, Carlos Ramisch, Voula Giouli, Verginica Barbu Mititelu, Najet Hadj Mohamed, Cvetana Krstev, Chaya Liebeskind, Hongzhi Xu, Sara Stymne, Tunga Güngör, Thomas Pickard, Bruno Guillaume, Eduard Bejček, Archna Bhatia, Marie Candito, Polona Gantar, Uxoa Iñurrieta, Albert Gatt, Jolanta Kovalevskaite, Timm Lichte, Nikola Ljubešić, Johanna Monti, Carla Parra Escartín, Mehrnoush Shamsfard, Ivelina Stoyanova, Veronika Vincze, and Abigail Walsh. PARSEME corpus release 1.3. In Archna Bhatia, Kilian Evang, Marcos Garcia, Voula Giouli, Lifeng Han, and Shiva Taslimipoor, editors, *Proceedings of the 19th Workshop on Multiword Expressions (MWE 2023)*, pages 24–35, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.mwe-1.6. URL https://aclanthology.org/2023.mwe-1.6.

Mojgan Seraji, Filip Ginter, and Joakim Nivre. Universal dependencies for Persian. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2361–2365, 2016.

Abishek Stephen and Daniel Zeman. Universal Dependencies for Malayalam. *The Prague Bulletin of Mathematical Linguistics*, 120:31–46, 2023. URL https://ufal.mff.cuni.cz/pbml/120/art-stephen-zeman.pdf.

Karumuri V. Subbarao. *South Asian Languages: A Syntactic Typology*. Cambridge University Press, Cambridge, UK, 2012. ISBN 978-0-521-86148-9.

Anuradha Sudharsan. *A minimalist account of null subjects in Kannada*. PhD thesis, University of Hyderabad, 1998.

Juhi Tandon, Himani Chaudhry, Riyaz Ahmad Bhat, and Dipti Sharma. Conversion from Paninian karakas to Universal Dependencies for Hindi dependency treebank. In Annemarie Friedrich and Katrin Tomanek, editors, *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*, pages 141–150, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/W16-1716. URL https://aclanthology.org/W16-1716.

Gopal Thakur. *A Grammar of Bhojpuri*. LINCOM studies in Indo-European linguistics. LINCOM GmbH, 2021. ISBN 9783969392331. URL https://books.google.cz/books?id=lBaCzgEACAAJ.

Ashwini Vaidya, Owen Rambow, and Martha Palmer. Light verb constructions with 'do' and 'be' in Hindi: A tag analysis. In *Proceedings of Workshop on Lexical and Grammatical Resources for Language Processing*, pages 127–136, 2014.

Ashwini Vaidya, Sumeet Agarwal, and Martha Palmer. Linguistic features for Hindi light verb construction identification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1320–1329, 2016.

References IV

Veronika Vincze, Katalin Ilona Simkó, Zsolt Szántó, and Richárd Farkas. Universal Dependencies and morphology for Hungarian – and on the price of universality. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 356–365, Valencia, Spain, 2017. Association for Computational Linguistics. URL `https://aclanthology.org/E17-1034.pdf`.

Daniel Zeman. Core arguments in Universal Dependencies. In *Proceedings of the fourth international conference on dependency linguistics (DepLing 2017)*, pages 287–296, Pisa, Italy, 2017. URL `https://aclanthology.org/W17-6532.pdf`.