Linguistic disciplines analyzed and linguistic phenomena

	Linguistic level					Discovery method			
	Morphology	Syntax	Semantics	Discourse	Specific linguistic phenomena studied	Language(s) analyzed (ISO 639 codes)	Feature attribut.	Example-based	Probing	Analysis of arch.
Abdou et al. [1]		✓	✓		low-frequency vocabulary, lexical ambiguity, and syntactic complexity	EN			✓	
Acs et al. [2]	✓	✓				42 languages			✓	
Aghazadeh et al. [3]			✓		metaphors	EN, FA, RU, ES			✓	
Alajrami and Aletras [4]	✓	✓	✓		various	EN			✓	
Alleman et al. [5]		✓				EN	✓			
Amini et al. [6]	✓	✓				ES	✓	✓	✓	✓
Aoyama and Schneider [7]	✓	✓	✓			EN		✓	✓	
Arps et al. [8]		✓	✓			EN		✓	✓	
Auyespek et al. [9]		✓				EN			✓	
Beloucif and Biemann [10]			✓		correlation between semantic attributes and their values.	EN	✓			✓
Bölücü and Can [12]		✓	✓		dependency, constituency and semantic parsing	EN, DE, FR, TR				✓
Buijtelaar and Pezzelle [13]			✓		compounds ('sunlight', 'handgun')	EN	✓			
Cai et al. [14]	✓					EN				✓
Cassani et al. [15]			✓		semantic representation	EN				✓
Celikkanat et al. [16]		✓	✓		passivization and negation	EN, DE | EN (monolingual), EN-DE, EN-DE+EL (multilingual)			✓	✓
Chersoni et al. [17]			✓		semantic features (based on Binder et al. [11]) encoded in contextual embeddings	EN			✓	
Chi et al. [18]		✓				AR, ZH, CS, EN, FA, FI, FR, DE, ID, LV, ES			✓	

	Linguistic level					Discovery method			
	Morphology	Syntax	Semantics	Discourse	Specific linguistic phenomena studied	Language(s) analyzed (ISO 639 codes)	Feature attribut.	Example-based	Probing	Analysis of arch.
Chistyakova and Kazakova [19]	✓				adjectives' gender; nouns' number and case; verbs' aspect, person and tense	RU			✓	
Chizhikova et al. [20]			✓			EN			✓	✓
Choenni and Shutova [21]	✓	✓	✓		25 specific phenomena (such as definite and indefinite articles, SVNegO, position of negative morphemes in SVO languages...), based on the field of Linguistic Tipology, and centered around the following general categories: Word order, Nominal and Verb categories, and Simple clauses.	RU, UK, DA, SV, CS, PL, PT, ES, HI, MR, MK, BG, IT, FR			✓	✓
Chronis and Erk [22]			✓			EN			✓	✓
Chrupała and Alishahi [23]		✓			syntax trees	EN				✓
Clark et al. [24]		✓			correference	EN			✓	✓
Conia and Navigli [25]		✓	✓			EN, ZH			✓	
Dai et al. [26]		✓				EN, RU, DE, ZH			✓	
Dankers et al. [27]			✓		non-compositionality of idioms	EN, NL, DE, SV, DA, FR, IT, ES			✓	✓
Davis and van Schijndel [28]		✓		✓	coreference resolution	EN		✓		
Derby et al. [29]			✓		activate lexico-semantic knowledge similarly to humans	EN	✓			✓
Dufter and Schütze [30]		✓	✓							✓
Durrani et al. [31]	✓	✓	✓			EN			✓	✓
Elazar et al. [32]		✓	✓			EN			✓	✓
Ethayarajh [33]			✓		polysemy	EN				✓
Fayyaz et al. [34]					–	EN			✓	
Garcia et al. [35]			✓			EN, PT		✓		✓
Garí Soler and Apidianaki [36]			✓		word senses	EN, FR, ES, EL				✓

	Linguistic level					Discovery method			
	Morphology	Syntax	Semantics	Discourse	Specific linguistic phenomena studied	Language(s) analyzed (ISO 639 codes)	Feature attribut.	Example-based	Probing	Analysis of arch.
Glavaš and Vulić [37]		✓	✓		Tests on whether injection of UDeps improves LU	EN (and zero-shot lang transfer)			✓	
Guarasci et al. [38]		✓			omissibility of the subject syntactic phenomenon	EN, IT, FR			✓	
Guarasci et al. [39]		✓			null-subject and subject-verb agreement	IT			✓	
Gupta et al. [40]		✓				DE, EN, ES, FR, ID, IT, JA, KO, PT (Brazilian), SV			✓	
Hao et al. [41]					(general methods for Transformer-based PLMs interpretability; attention heads-specific)	EN				✓
Hernandez and Andreas [42]	✓	✓				EN			✓	
Hessel and Schofield [43]	✓	✓	✓			EN	✓			✓
Hewitt et al. [44]		✓	✓			EN			✓	
Hewitt and Manning [45]		✓				EN			✓	
Hou and Sachan [46]		✓	✓		linguistic graphs	EN			✓	
Huber et al. [47]				✓	coherence between clauses, discourse relations	EN		✓	✓	
Jo and Myaeng [48]		✓	✓			EN			✓	
Kahardipraja et al. [49]		✓			correference resolution	EN			✓	
Kasthuriarachchy et al. [50]		✓	✓			EN			✓	
Kauf et al. [51]			✓		generalized event knowledge	EN			✓	
Klafka and Ettinger [52]		✓	✓			EN			✓	

	Linguistic level					Discovery method			
	Morphology	Syntax	Semantics	Discourse	Specific linguistic phenomena studied	Language(s) analyzed (ISO 639 codes)	Feature attribut.	Example-based	Probing	Analysis of arch.
Koto et al. [53]				✓		EN, ZH, DE, ES			✓	
Kovaleva et al. [54]		✓	✓			EN				✓
Krasnowska-Kieraś and Wróblewska [55]	✓	✓	✓		many phenomena (some surface-form-related, such as sentence length, others in syntax and morphology such as grammatical number, dependency depth or tense)	EN, PL			✓	
Kulmizev et al. [56]		✓			Universal Dependencies, Surface-Syntactic Universal Dependencies	AR, ZH, EN, EU, FI, HE, HI, IT, JA, KO, RU, SV, TR			✓	
Kunz and Kuhlmann [57]		✓	✓		word level representation	EN			✓	
Kunz and Kuhlmann [58]		✓				EN			✓	
Kunz and Kuhlmann [59]		✓			POS; syntactic ancestors	EN, CS, FI, DE, HE, SV, TR	✓			
Kuznetsov and Gurevych [60]			✓		role semantics	EN, DE			✓	
Lasri et al. [61]	✓	✓			grammatical number	EN			✓	
Lee and Shin [62]		✓			garden-path, transitivity, plausibility	EN				✓
Li et al. [63]		✓				FR			✓	
Li et al. [64]	✓	✓				EN			✓	
Li et al. [65]	✓	✓	✓			EN			✓	✓
Limisiewicz and Mareček [66]		✓	✓		dependency syntax, lexical hypernymy, position in a sentence, random structures	EN			✓	
Limisiewicz et al. [67]		✓			dependency trees	EN, DE, FR, CS, FI, ID, TR, KO, JA				✓
Lin et al. [68]		✓			sbj-verb agreement; anaphor-antecedent reps.	EN			✓	✓
Liu et al. [69]		✓	✓		Many	EN			✓	

	Linguistic level					Discovery method			
	Morphology	Syntax	Semantics	Discourse	Specific linguistic phenomena studied	Language(s) analyzed (ISO 639 codes)	Feature attribut.	Example-based	Probing	Analysis of arch.
Liu et al. [70]	✓	✓	✓		various (noun-verb agreement, POS, conjunction acceptability, among others)	EN			✓	
Loureiro et al. [71]			✓			EN			✓	
Loureiro et al. [72]			✓			EN				✓
Loureiro et al. [73]			✓		lexical ambiguity	EN				✓
Lovering et al. [74]		✓			concordance, polarity items,	EN		✓	✓	
Luo [75]		✓			constituency gammar	EN				✓
Ma et al. [76]		✓				EN			✓	
Mareček and Rosa [77]		✓			syntactic phrases	EN, FR, DE			✓	✓
Maudslay and Cotterell [78]		✓	✓			EN	✓		✓	
Maudslay et al. [79]		✓				AR, EU, CS, EN, FI, JA, KO, TA, TR			✓	
Miaschi et al. [80]	✓	✓			many (77): tense, mood, person, number, distribution of verbal roots and verbal heads, depth of the whole syntactic tree, etc.	IT			✓	
Miaschi et al. [81]	✓	✓			order of elements, morpho-syntactic information (POS), use of subordination, syntactic relations, global and local parsed tree structures, inflectional morphology, verbal predicate structure	IT			✓	
Miaschi et al. [82]	✓	✓	✓		68 NLP-like (UD) grammatical phenomena	EN			✓	
Miaschi et al. [83]	✓	✓	✓		various, they check many morpho-syntactic features in a context with learners errors	IT			✓	
Miaschi and Dell'Orletta [84]	✓	✓				EN			✓	

	Linguistic level					Discovery method			
	Morphology	Syntax	Semantics	Discourse	Specific linguistic phenomena studied	Language(s) analyzed (ISO 639 codes)	Feature attribut.	Example-based	Probing	Analysis of arch.
Miaschi et al. [85]	✓	✓				IT			✓	
Miaschi et al. [86]	✓	✓				IT			✓	
Michael et al. [87]		✓	✓			EN			✓	
Mickus et al. [88]		✓	✓		various	EN				✓
Mikhailov et al. [89]	✓	✓			syntactic and inflectional sentence perturbations	EN, FR, DE, RU		✓	✓	
Mikhailov et al. [90]		✓	✓			RU, EN			✓	✓
Mohebbi et al. [91]		✓	✓		grammatical number and tense information; word-level-inversion; phrasal-level inversion	EN	✓		✓	
Mueller et al. [92]		✓			subject-verb concordance	EN, FR, DE, NL, FI		✓		✓
Müller-Eberstein et al. [93]		✓	✓			EN			✓	
Mysiak and Cyranka [94]		✓				BE, HR, CS, LV, LT, PL, RU, SK, SL, UK			✓	
Newman et al. [95]		✓			subject-verb (S/V) number agreement	EN				
Nikolaev and Padó [96]			✓		frame semantics	EN, KO			✓	✓
Nikoulina et al. [97]		✓	✓			EN			✓	✓
Niu et al. [98]		✓	✓			EN			✓	
Niu et al. [99]	✓	✓	✓			EN			✓	✓
Oba et al. [100]					domain-specific specialized neurons (non-linguistic analysis, with an *a posteriori* linguistic knowledge attribution)	EN				✓

	Linguistic level					Discovery method			
	Morphology	Syntax	Semantics	Discourse	Specific linguistic phenomena studied	Language(s) analyzed (ISO 639 codes)	Feature attribut.	Example-based	Probing	Analysis of arch.
Oota et al. [101]		✓	✓		surface, sensitivity to word order, depth of syntactic tree, sequence of top-level constituents in the syntax tree, tense, subject number, sensitivity to random replacement of a noun/verb, random swapping of coordinated clausal conjuncts	EN			✓	✓
Ormerod et al. [102]			✓		compositional semantics	EN			✓	✓
Otmakhova et al. [103]	✓	✓			flexibility of word order, head directionality, morphological type, presence of grammatical gender, and morphological richness	EN, KO, RU	✓		✓	
Paganelli et al. [104]			✓			EN				✓
Pande et al. [105]	✓	✓				EN				✓
Papadimitriou et al. [106]	✓	✓			morphosyntactic alignment	HI, UR, EU, FI, HE, LA, ET, JA, ZH, LV, SR, FR, SK, NO, PL, RU, HR, FA, CS, DE, EN, ID, ES, SL			✓	
Papadimitriou et al. [107]	✓	✓			ergativity	EN			✓	
Papadimitriou et al. [108]		✓	✓		word order	EN			✓	
Phang et al. [109]		✓				EN				✓
Pimentel et al. [110]	✓	✓				EU, EN, FI, MT, TR			✓	
Pimentel et al. [111]		✓			PoS labelling	EU, CS, EN, FI, ID, KO, MT, TA, TE, TR, UR			✓	
Pimentel et al. [112]		✓				EU, EN, TA, TR			✓	
Pimentel et al. [113]		✓				EN, EU, TA, TR			✓	
Proietti et al. [114]			✓		proto-role information	EN			✓	

	Linguistic level					Discovery method			
	Morphology	Syntax	Semantics	Discourse	Specific linguistic phenomena studied	Language(s) analyzed (ISO 639 codes)	Feature attribut.	Example-based	Probing	Analysis of arch.
Puccetti et al. [115]	✓	✓				EN			✓	
Raganato and Tiedemann [116]		✓	✓			EN, CS, DE, ET, FI, RU, TR, ZH			✓	✓
Rama et al. [117]			✓			100 languages			✓	
Ravishankar et al. [118]	✓	✓	✓			FR, DE, ES, RU, TR, FI			✓	
Reif et al. [119]		✓	✓			EN			✓	
Richardson et al. [120]			✓		semantic fragments—systematically generated datasets that each target a different semantic phenomenon	EN	✓			
Sajjad et al. [121]	✓	✓	✓		various (what they call Morphology, Semantics and Syntactic concepts)	EN				✓
Schneidermann et al. [122]			✓		hyperboles	EN			✓	
Schuster and Hegelich [123]		✓	✓			EN			✓	
Sevastjanova et al. [124]		✓	✓			EN	✓			✓
Sevastjanova et al. [125]			✓		focus on contextualization, degree of contextualization of function vs. content words	EN				✓
Seyffarth et al. [126]			✓		causativity of events denoted by verbs	EN			✓	✓
Shapiro et al. [127]	✓	✓			many (166 of them), depending on each language: part of speech, number, gender, case, tense...	AF, HR, FI, HE, KO, ES, TR, AR, ZH, MR, SL, TL, YO			✓	
Sinha et al. [128]		✓	✓		word ordering	EN		✓	✓	
Sinha et al. [129]		✓				EN, ZH (Mandarin Chinese)	✓			
Song et al. [130]			✓		unsupervised extraction of keyphrases from documents	EN				✓
Sorodoc et al. [131]	✓		✓		pronominal anaphora	EN			✓	

	Linguistic level						Discovery method			
	Morphology	Syntax	Semantics	Discourse	Specific linguistic phenomena studied	Language(s) analyzed (ISO 639 codes)	Feature attribut.	Example-based	Probing	Analysis of arch.
Stańczak et al. [132]	✓	✓				AR, EN, FI, PL, PT, RU			✓	✓
Taktasheva et al. [133]		✓				EN, SV, RU		✓	✓	
Talmor et al. [134]			✓			EN			✓	
Tan and Jiang [135]			✓		idiomatic expressions (PIE)	EN			✓	
Tenney et al. [136]	✓	✓	✓		various	EN			✓	
Tian et al. [137]		✓		✓	disfluency	EN				✓
Timmapathini et al. [138]		✓	✓		coreference resolution	EN				✓
Tucker et al. [139]		✓				EN		✓	✓	
Tucker et al. [140]		✓				EN		✓	✓	
Varda and Marelli [141]	✓	✓			agreement violations	EN, DE, FR, HE, RU			✓	✓
Vulić et al. [142]			✓			EN, DE, RU, FI, ZH, TR			✓	
Wallat et al. [143]			✓		coreference resolution and name entity recognition	EN			✓	
Wang et al. [144]	✓				word structure (word segmentation)	ZH			✓	✓
Wei et al. [145]		✓			subject-verb agreement	EN	✓			
Weissweiler et al. [146]		✓	✓		comparative correlative (CC)	EN		✓	✓	
Weissweiler et al. [147]		✓	✓		comparative correlative (CC)	EN (potentially any language)		✓	✓	
White et al. [148]		✓				EU, EN, FI, KO, TA, TR			✓	
Wu et al. [149]		✓		✓		EN	✓	✓		✓
Xia et al. [150]			✓			EN			✓	✓

	Linguistic level					Discovery method			
	Morphology	Syntax	Semantics	Discourse	Specific linguistic phenomena studied	Language(s) analyzed (ISO 639 codes)	Feature attribut.	Example-based	Probing	Analysis of arch.
Xu et al. [151]	✓	✓			cross-lingual transfer of syntactic knowledge	AR, BG, DE, EL, EN, ES, ET, FA, FI, FR, HE, HI, IT, JA, KO, LV, NL, PL, PT, RO, RU, TR, VI, ZH			✓	
Yi et al. [152]		✓			alternations in which a verb may participate is taken to be a lexical property of the verb // syntactic frames an individual verb can participate in	EN			✓	
Zanzotto et al. [153]		✓			universal syntactic interpretations	EN	✓			
Zhang et al. [154]		✓				EN, DE, FR, RO				✓
Zhang et al. [155]		✓	✓			EN			✓	
Zhao et al. [156]			✓			EN			✓	
Zhao and Bethard [157]		✓	✓		negation (more specifically, negation scope detection)	EN				✓
Zheng and Liu [158]	✓	✓	✓		language identity and language typology	36 languages			✓	
Zheng and Liu [159]		✓				ZH			✓	
Zheng and Sun [160]	✓				word structure	ZH (Ancient Chinese)				
Zhu et al. [161]				✓	Rhetorical Structure Theory (RST)	EN			✓	

References

[1] Mostafa Abdou, Artur Kulmizev, Felix Hill, Daniel M. Low, and Anders Søgaard. 2019. Higher-order Comparisons of Sentence Encoder Representations. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, Hong Kong, China, 5838–5845. https://doi.org/10.18653/v1/D19-1593

[2] Judit Acs, Endre Hamerlik, Roy Schwartz, Noah A Smith, and Andras Kornai. 2023. Morphosyntactic probing of multilingual BERT models. Natural Language Engineering (2023), 1–40.

[3] Ehsan Aghazadeh, Mohsen Fayyaz, and Yadollah Yaghoobzadeh. 2022. Metaphors in Pre-Trained Language Models: Probing and Generalization Across Datasets and Languages. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 2037–2050. https://doi.org/10.18653/v1/2022.acl-long.144

[4] Ahmed Alajrami and Nikolaos Aletras. 2022. How does the pre-training objective affect what large language models learn about linguistic properties?. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 131–147. https://doi.org/10.18653/v1/2022.acl-

short.16

[5] Matteo Alleman, Jonathan Mamou, Miguel A Del Rio, Hanlin Tang, Yoon Kim, and SueYeon Chung. 2021. Syntactic Perturbations Reveal Representational Correlates of Hierarchical Phrase Structure in Pretrained Language Models. In Proceedings of the 6th Workshop on Representation Learning for NLP (RepL4NLP-2021), Anna Rogers, Iacer Calixto, Ivan Vulić, Naomi Saphra, Nora Kassner, Oana-Maria Camburu, Trapit Bansal, and Vered Shwartz (Eds.). Association for Computational Linguistics, Online, 263–276. https://doi.org/10.18653/v1/2021.repl4nlp-1.27

[6] Afra Amini, Tiago Pimentel, Clara Meister, and Ryan Cotterell. 2023. Naturalistic Causal Probing for Morpho-Syntax. Transactions of the Association for Computational Linguistics 11 (2023), 384–403. https://doi.org/10.1162/tacl_a_00554

[7] Tatsuya Aoyama and Nathan Schneider. 2022. Probe-Less Probing of BERT's Layer-Wise Linguistic Knowledge with Masked Word Prediction. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop, Daphne Ippolito, Liunian Harold Li, Maria Leonor Pacheco, Danqi Chen, and Nianwen Xue (Eds.). Association for Computational Linguistics, Hybrid: Seattle, Washington + Online, 195–201. https://doi.org/10.18653/v1/2022.naacl-srw.25

[8] David Arps, Younes Samih, Laura Kallmeyer, and Hassan Sajjad. 2022. Probing for Constituency Structure in Neural Language Models. In Findings of the Association for Computational Linguistics: EMNLP 2022, Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 6738–6757. https://doi.org/10.18653/v1/2022.findings-emnlp.502

[9] Temirlan Auyespek, Thomas Mach, and Zhenisbek Assylbekov. 2021. Hyperbolic Embedding for Finding Syntax in BERT.. In DP@ AI* IA. 58–64.

[10] Meriem Beloucif and Chris Biemann. 2021. Probing Pre-trained Language Models for Semantic Attributes and their Values. In Findings of the Association for Computational Linguistics: EMNLP 2021, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Punta Cana, Dominican Republic, 2554–2559. https://doi.org/10.18653/v1/2021.findings-emnlp.218

[11] Jeffrey R Binder, Lisa L Conant, Colin J Humphries, Leonardo Fernandino, Stephen B Simons, Mario Aguilar, and Rutvik H Desai. 2016. Toward a brain-based componential semantic representation. Cognitive neuropsychology 33, 3-4 (2016), 130–174.

[12] Necva Bölücü and Burcu Can. 2022. Analysing Syntactic and Semantic Features in Pre-trained Language Models in a Fully Unsupervised Setting. In Proceedings of the 19th International Conference on Natural Language Processing (ICON), Md. Shad Akhtar and Tanmoy Chakraborty (Eds.). Association for Computational Linguistics, New Delhi, India, 19–31. https://aclanthology.org/2022.icon-main.3

[13] Lars Buijtelaar and Sandro Pezzelle. 2023. A Psycholinguistic Analysis of BERT's Representations of Compounds. In Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, Andreas Vlachos and Isabelle Augenstein (Eds.). Association for Computational Linguistics, Dubrovnik, Croatia, 2230–2241. https://doi.org/10.18653/v1/2023.eacl-main.163

[14] Xingyu Cai, Jiaji Huang, Yuchen Bian, and Kenneth Church. 2021. Isotropy in the contextual embedding space: Clusters and manifolds. In International conference on learning representations.

[15] Giovanni Cassani, Fritz Günther, Giuseppe Attanasio, Federico Bianchi, and Marco Marelli. 2023. Meaning Modulations and Stability in Large Language Models: An Analysis of BERT Embeddings for Psycholinguistic Research. (2023).

[16] Hande Celikkanat, Sami Virpioja, Jörg Tiedemann, and Marianna Apidianaki. 2020. Controlling the Imprint of Passivization and Negation in Contextualized Representations. In Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, Afra Alishahi, Yonatan Belinkov, Grzegorz Chrupała, Dieuwke Hupkes, Yuval Pinter, and Hassan Sajjad (Eds.). Association for Computational Linguistics, Online, 136–148. https://doi.org/10.18653/v1/2020.blackboxnlp-1.13

[17] Emmanuele Chersoni, Enrico Santus, Chu-Ren Huang, Alessandro Lenci, et al. 2021. Decoding word embeddings with brain-based semantic features. Computational Linguistics 47, 3 (2021), 663–698.

[18] Ethan A. Chi, John Hewitt, and Christopher D. Manning. 2020. Finding Universal Grammatical Relations in Multilingual BERT. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 5564–5577. https://doi.org/10.18653/v1/2020.acl-main.493

[19] Ksenia E Chistyakova and Tatiana B Kazakova. 2023. Grammar In Language Models: Bert Study. Technical Report. National Research University Higher School of Economics.

[20] Anastasia Chizhikova, Sanzhar Murzakhmetov, Oleg Serikov, Tatiana Shavrina, and Mikhail Burtsev. 2022. Attention Understands Semantic Relations. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis (Eds.). European Language Resources Association, Marseille, France, 4040–4050. https://aclanthology.org/2022.lrec-1.430

[21] Rochelle Choenni and Ekaterina Shutova. 2022. Investigating language relationships in multilingual sentence encoders through the lens of linguistic typology. Computational Linguistics 48, 3 (2022), 635–672.

[22] Gabriella Chronis and Katrin Erk. 2020. When is a bishop not like a rook? When it's like a rabbi! Multi-prototype BERT embeddings for estimating semantic relationships. In Proceedings of the 24th Conference on Computational Natural Language Learning, Raquel Fernández and Tal Linzen (Eds.). Association for Computational Linguistics, Online, 227–244. https://doi.org/10.18653/v1/2020.conll-1.17

[23] Grzegorz Chrupała and Afra Alishahi. 2019. Correlating Neural and Symbolic Representations of Language. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Anna Korhonen, David Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, Florence, Italy, 2952–2962. https://doi.org/10.18653/v1/P19-1283

[24] Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D. Manning. 2019. What Does BERT Look at? An Analysis of BERT's Attention. In Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, Tal Linzen, Grzegorz Chrupała, Yonatan Belinkov, and Dieuwke Hupkes (Eds.). Association for Computational Linguistics, Florence, Italy, 276–286. https://doi.org/10.18653/v1/W19-

4828

[25] Simone Conia and Roberto Navigli. 2022. Probing for Predicate Argument Structures in Pretrained Language Models. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 4622–4632. https://doi.org/10.18653/v1/2022.acl-long.316

[26] Yuqian Dai, Marc de Kamps, and Serge Sharoff. 2022. BERTology for Machine Translation: What BERT Knows about Linguistic Difficulties for Translation. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis (Eds.). European Language Resources Association, Marseille, France, 6674–6690. https://aclanthology.org/2022.lrec-1.719

[27] Verna Dankers, Christopher Lucas, and Ivan Titov. 2022. Can Transformer be Too Compositional? Analysing Idiom Processing in Neural Machine Translation. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 3608–3626. https://doi.org/10.18653/v1/2022.acl-long.252

[28] Forrest Davis and Marten van Schijndel. 2020. Discourse structure interacts with reference but not syntax in neural language models. In Proceedings of the 24th Conference on Computational Natural Language Learning, Raquel Fernández and Tal Linzen (Eds.). Association for Computational Linguistics, Online, 396–407. https://doi.org/10.18653/v1/2020.conll-1.32

[29] Steven Derby, Paul Miller, and Barry Devereux. 2021. Representation and Pre-Activation of Lexical-Semantic Knowledge in Neural Language Models. In Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, Emmanuele Chersoni, Nora Hollenstein, Cassandra Jacobs, Yohei Oseki, Laurent Prévot, and Enrico Santus (Eds.). Association for Computational Linguistics, Online, 211–221. https://doi.org/10.18653/v1/2021.cmcl-1.25

[30] Philipp Dufter and Hinrich Schütze. 2020. Identifying Elements Essential for BERT's Multilinguality. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 4423–4437. https://doi.org/10.18653/v1/2020.emnlp-main.358

[31] Nadir Durrani, Hassan Sajjad, Fahim Dalvi, and Yonatan Belinkov. 2020. Analyzing Individual Neurons in Pre-trained Language Models. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 4865–4880. https://doi.org/10.18653/v1/2020.emnlp-main.395

[32] Yanai Elazar, Shauli Ravfogel, Alon Jacovi, and Yoav Goldberg. 2021. Amnesic probing: Behavioral explanation with amnesic counterfactuals. Transactions of the Association for Computational Linguistics 9 (2021), 160–175.

[33] Kawin Ethayarajh. 2019. How Contextual are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, Hong Kong, China, 55–65. https://doi.org/10.18653/v1/D19-1006

[34] Mohsen Fayyaz, Ehsan Aghazadeh, Ali Modarressi, Hosein Mohebbi, and Mohammad Taher Pilehvar. 2021. Not All Models Localize Linguistic Knowledge in the Same Place: A Layer-wise Probing on BERToids' Representations. In Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, Jasmijn Bastings, Yonatan Belinkov, Emmanuel Dupoux, Mario Giulianelli, Dieuwke Hupkes, Yuval Pinter, and Hassan Sajjad (Eds.). Association for Computational Linguistics, Punta Cana, Dominican Republic, 375–388. https://doi.org/10.18653/v1/2021.blackboxnlp-1.29

[35] Marcos Garcia, Tiago Kramer Vieira, Carolina Scarton, Marco Idiart, and Aline Villavicencio. 2021. Probing for idiomaticity in vector space models. In Proceedings of the 16th conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics (ACL), 3551–3564.

[36] Aina Garí Soler and Marianna Apidianaki. 2021. Let's play mono-poly: BERT can reveal words' polysemy level and partitionability into senses. Transactions of the Association for Computational Linguistics 9 (2021), 825–844.

[37] Goran Glavaš and Ivan Vulić. 2021. Is Supervised Syntactic Parsing Beneficial for Language Understanding Tasks? An Empirical Investigation. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Paola Merlo, Jorg Tiedemann, and Reut Tsarfaty (Eds.). Association for Computational Linguistics, Online, 3090–3104. https://doi.org/10.18653/v1/2021.eacl-main.270

[38] Raffaele Guarasci, Stefano Silvestri, Giuseppe De Pietro, Hamido Fujita, and Massimo Esposito. 2022. BERT syntactic transfer: A computational experiment on Italian, French and English languages. Computer Speech & Language 71 (2022), 101261.

[39] Raffaele Guarasci, Stefano Silvestri, Giuseppe De Pietro, Hamido Fujita, and Massimo Esposito. 2023. Assessing BERT's ability to learn Italian syntax: A study on null-subject and agreement phenomena. Journal of Ambient Intelligence and Humanized Computing 14, 1 (2023), 289–303.

[40] Vikram Gupta, Haoyue Shi, Kevin Gimpel, and Mrinmaya Sachan. 2022. Deep clustering of text representations for supervision-free probing of syntax. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36. 10720–10728.

[41] Yaru Hao, Li Dong, Furu Wei, and Ke Xu. 2021. Self-attention attribution: Interpreting information interactions inside transformer. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35. 12963–12971.

[42] Evan Hernandez and Jacob Andreas. 2021. The Low-Dimensional Linear Geometry of Contextualized Word Representations. In Proceedings of the 25th Conference on Computational Natural Language Learning, Arianna Bisazza and Omri Abend (Eds.). Association for Computational Linguistics, Online, 82–93. https://doi.org/10.18653/v1/2021.conll-1.7

[43] Jack Hessel and Alexandra Schofield. 2021. How effective is BERT without word ordering? Implications for language understanding and data privacy. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 204–211. https://doi.org/10.18653/v1/2021.acl-short.27

[44] John Hewitt, Kawin Ethayarajh, Percy Liang, and Christopher Manning. 2021. Conditional probing: measuring usable information beyond a baseline. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 1626–1639. https://doi.org/10.18653/v1/2021.emnlp-main.122

[45] John Hewitt and Christopher D. Manning. 2019. A Structural Probe for Finding Syntax in Word Representations. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, Minneapolis, Minnesota, 4129–4138. https://doi.org/10.18653/v1/N19-1419

[46] Yifan Hou and Mrinmaya Sachan. 2021. Bird's Eye: Probing for Linguistic Graph Structures with a Simple Information-Theoretic Approach. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 1844–1859. https://doi.org/10.18653/v1/2021.acl-long.145

[47] Laurine Huber, Chaker Memmadi, Mathilde Dargnat, and Yannick Toussaint. 2020. Do sentence embeddings capture discourse properties of sentences from Scientific Abstracts?. In CODI 2020-EMNLP 1st Workshop on Computational Approaches to Discourse.

[48] Jae-young Jo and Sung-Hyon Myaeng. 2020. Roles and Utilization of Attention Heads in Transformer-based Neural Language Models. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 3404–3417. https://doi.org/10.18653/v1/2020.acl-main.311

[49] Patrick Kahardipraja, Olena Vyshnevska, and Sharid Loáiciga. 2020. Exploring Span Representations in Neural Coreference Resolution. In Proceedings of the First Workshop on Computational Approaches to Discourse, Chloé Braud, Christian Hardmeier, Junyi Jessy Li, Annie Louis, and Michael Strube (Eds.). Association for Computational Linguistics, Online, 32–41. https://doi.org/10.18653/v1/2020.codi-1.4

[50] Buddhika Kasthuriarachchy, Madhu Chetty, Adrian Shatte, and Darren Walls. 2021. From general language understanding to noisy text comprehension. Applied Sciences 11, 17 (2021), 7814.

[51] Carina Kauf, Anna A Ivanova, Giulia Rambelli, Emmanuele Chersoni, Jingyuan Selena She, Zawad Chowdhury, Evelina Fedorenko, and Alessandro Lenci. 2023. Event knowledge in large language models: the gap between the impossible and the unlikely. Cognitive Science 47, 11 (2023), e13386.

[52] Josef Klafka and Allyson Ettinger. 2020. Spying on Your Neighbors: Fine-grained Probing of Contextual Embeddings for Information about Surrounding Words. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 4801–4811. https://doi.org/10.18653/v1/2020.acl-main.434

[53] Fajri Koto, Jey Han Lau, and Timothy Baldwin. 2021. Discourse Probing of Pretrained Language Models. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (Eds.). Association for Computational Linguistics, Online, 3849–3864. https://doi.org/10.18653/v1/2021.naacl-main.301

[54] Olga Kovaleva, Alexey Romanov, Anna Rogers, and Anna Rumshisky. 2019. Revealing the Dark Secrets of BERT. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, Hong Kong, China, 4365–4374. https://doi.org/10.18653/v1/D19-1445

[55] Katarzyna Krasnowska-Kieraś and Alina Wróblewska. 2019. Empirical Linguistic Study of Sentence Embeddings. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Anna Korhonen, David Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, Florence, Italy, 5729–5739. https://doi.org/10.18653/v1/P19-1573

[56] Artur Kulmizev, Vinit Ravishankar, Mostafa Abdou, and Joakim Nivre. 2020. Do Neural Language Models Show Preferences for Syntactic Formalisms?. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 4077–4091. https://doi.org/10.18653/v1/2020.acl-main.375

[57] Jenny Kunz and Marco Kuhlmann. 2020. Classifier Probes May Just Learn from Linear Context Features. In Proceedings of the 28th International Conference on Computational Linguistics, Donia Scott, Nuria Bel, and Chengqing Zong (Eds.). International Committee on Computational Linguistics, Barcelona, Spain (Online), 5136–5146. https://doi.org/10.18653/v1/2020.coling-main.450

[58] Jenny Kunz and Marco Kuhlmann. 2021. Test Harder than You Train: Probing with Extrapolation Splits. In Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, Jasmijn Bastings, Yonatan Belinkov, Emmanuel Dupoux, Mario Giulianelli, Dieuwke Hupkes, Yuval Pinter, and Hassan Sajjad (Eds.). Association for Computational Linguistics, Punta Cana, Dominican Republic, 15–25. https://doi.org/10.18653/v1/2021.blackboxnlp-1.2

[59] Jenny Kunz and Marco Kuhlmann. 2022. Where Does Linguistic Information Emerge in Neural Language Models? Measuring Gains and Contributions across Layers. In Proceedings of the 29th International Conference on Computational Linguistics, Nicoletta Calzolari, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na

(Eds.). International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 4664–4676. https://aclanthology.org/2022.coling-1.413

[60] Ilia Kuznetsov and Iryna Gurevych. 2020. A matter of framing: The impact of linguistic formalism on probing results. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 171–182. https://doi.org/10.18653/v1/2020.emnlp-main.13

[61] Karim Lasri, Tiago Pimentel, Alessandro Lenci, Thierry Poibeau, and Ryan Cotterell. 2022. Probing for the Usage of Grammatical Number. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 8818–8831. https://doi.org/10.18653/v1/2022.acl-long.603

[62] Jonghyun Lee and Jeong-Ah Shin. 2023. Decoding bert's internal processing of garden-path structures through attention maps. Korean Journal of English Language and Linguistics 23 (2023), 461–481.

[63] Bingzhi Li, Guillaume Wisniewski, and Benoit Crabbé. 2022. How distributed are distributed representations? An observation on the locality of syntactic information in verb agreement tasks. In 60th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 501–507.

[64] Bai Li, Zining Zhu, Guillaume Thomas, Yang Xu, and Frank Rudzicz. 2021. How is BERT surprised? Layerwise detection of linguistic anomalies. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 4215–4228. https://doi.org/10.18653/v1/2021.acl-long.325

[65] Jiaoda Li, Ryan Cotterell, and Mrinmaya Sachan. 2022. Probing via Prompting. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (Eds.). Association for Computational Linguistics, Seattle, United States, 1144–1157. https://doi.org/10.18653/v1/2022.naacl-main.84

[66] Tomasz Limisiewicz and David Mareček. 2021. Introducing Orthogonal Constraint in Structural Probes. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 428–442. https://doi.org/10.18653/v1/2021.acl-long.36

[67] Tomasz Limisiewicz, David Mareček, and Rudolf Rosa. 2020. Universal Dependencies According to BERT: Both More Specific and More General. In Findings of the Association for Computational Linguistics: EMNLP 2020, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 2710–2722. https://doi.org/10.18653/v1/2020.findings-emnlp.245

[68] Yongjie Lin, Yi Chern Tan, and Robert Frank. 2019. Open Sesame: Getting inside BERT's Linguistic Knowledge. In Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, Tal Linzen, Grzegorz Chrupała, Yonatan Belinkov, and Dieuwke Hupkes (Eds.). Association for Computational Linguistics, Florence, Italy, 241–253. https://doi.org/10.18653/v1/W19-4825

[69] Nelson F. Liu, Matt Gardner, Yonatan Belinkov, Matthew E. Peters, and Noah A. Smith. 2019. Linguistic Knowledge and Transferability of Contextual Representations. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, Minneapolis, Minnesota, 1073–1094. https://doi.org/10.18653/v1/N19-1112

[70] Zeyu Liu, Yizhong Wang, Jungo Kasai, Hannaneh Hajishirzi, and Noah A. Smith. 2021. Probing Across Time: What Does RoBERTa Know and When?. In Findings of the Association for Computational Linguistics: EMNLP 2021, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Punta Cana, Dominican Republic, 820–842. https://doi.org/10.18653/v1/2021.findings-emnlp.71

[71] Daniel Loureiro, Alípio Mário Jorge, and Jose Camacho-Collados. 2022. LMMS reloaded: Transformer-based sense embeddings for disambiguation and beyond. Artificial Intelligence 305 (2022), 103661.

[72] Daniel Loureiro, Kiamehr Rezaee, Mohammad Taher Pilehvar, and Jose Camacho-Collados. 2020. Language models and word sense disambiguation: An overview and analysis. arXiv preprint arXiv:2008.11608 (2020).

[73] Daniel Loureiro, Kiamehr Rezaee, Mohammad Taher Pilehvar, and Jose Camacho-Collados. 2021. Analysis and evaluation of language models for word sense disambiguation. Computational Linguistics 47, 2 (2021), 387–443.

[74] Charles Lovering, Rohan Jha, Tal Linzen, and Ellie Pavlick. 2021. Predicting inductive biases of pre-trained models. In International Conference on learning representations.

[75] Ziyang Luo. 2021. Have Attention Heads in BERT Learned Constituency Grammar?. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop, Ionut-Teodor Sorodoc, Madhumita Sushil, Ece Takmaz, and Eneko Agirre (Eds.). Association for Computational Linguistics, Online, 8–15. https://doi.org/10.18653/v1/2021.eacl-srw.2

[76] Weicheng Ma, Brian Wang, Hefan Zhang, Lili Wang, Rolando Coto-Solano, Saeed Hassanpour, and Soroush Vosoughi. 2023. Improving Syntactic Probing Correctness and Robustness with Control Tasks. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 402–415. https://doi.org/10.18653/v1/2023.acl-short.35

[77] David Mareček and Rudolf Rosa. 2019. From Balustrades to Pierre Vinken: Looking for Syntax in Transformer Self-Attentions. In Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, Tal Linzen, Grzegorz Chrupała, Yonatan Belinkov, and Dieuwke Hupkes (Eds.). Association for Computational Linguistics, Florence, Italy, 263–275. https://doi.org/10.18653/v1/W19-4827

[78] Rowan Hall Maudslay and Ryan Cotterell. 2021. Do Syntactic Probes Probe Syntax? Experiments with Jabberwocky Probing. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (Eds.). Association for Computational Linguistics, Online, 124–131. https://doi.org/10.18653/v1/2021.naacl-main.11

[79] Rowan Hall Maudslay, Josef Valvoda, Tiago Pimentel, Adina Williams, and Ryan Cotterell. 2020. A Tale of a Probe and a Parser. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 7389–7395. https://doi.org/10.18653/v1/2020.acl-main.659

[80] Alessio Miaschi, Chiara Alzetta, Dominique Brunato, Felice Dell'Orletta, and Giulia Venturi. 2021. Probing tasks under pressure. (2021).

[81] Alessio Miaschi, Chiara Alzetta, Dominique Brunato, Felice Dell'Orletta, and Giulia Venturi. 2023. Testing the Effectiveness of the Diagnostic Probing Paradigm on Italian Treebanks. Information 14, 3 (2023), 144.

[82] Alessio Miaschi, Dominique Brunato, Felice Dell'Orletta, and Giulia Venturi. 2020. Linguistic Profiling of a Neural Language Model. In Proceedings of the 28th International Conference on Computational Linguistics, Donia Scott, Nuria Bel, and Chengqing Zong (Eds.). International Committee on Computational Linguistics, Barcelona, Spain (Online), 745–756. https://doi.org/10.18653/v1/2020.coling-main.65

[83] Alessio Miaschi, Dominique Brunato, Felice Dell'Orletta, and Giulia Venturi. 2022. On Robustness and Sensitivity of a Neural Language Model: A Case Study on Italian L1 Learner Errors. IEEE/ACM Transactions on Audio, Speech, and Language Processing 31 (2022), 426–438.

[84] Alessio Miaschi and Felice Dell'Orletta. 2020. Contextual and Non-Contextual Word Embeddings: an in-depth Linguistic Investigation. In Proceedings of the 5th Workshop on Representation Learning for NLP, Spandana Gella, Johannes Welbl, Marek Rei, Fabio Petroni, Patrick Lewis, Emma Strubell, Minjoon Seo, and Hannaneh Hajishirzi (Eds.). Association for Computational Linguistics, Online, 110–119. https://doi.org/10.18653/v1/2020.repl4nlp-1.15

[85] Alessio Miaschi, Gabriele Sarti, Dominique Brunato, Felice Dell'Orletta, and Giulia Venturi. 2020. Italian transformers under the linguistic lens. Computational Linguistics CLiC-it 2020 (2020), 310.

[86] Alessio Miaschi, Gabriele Sarti, Dominique Brunato, Felice Dell'Orletta, and Giulia Venturi. 2022. Probing linguistic knowledge in italian neural language models across language varieties. IJCoL. Italian Journal of Computational Linguistics 8, 8-1 (2022).

[87] Julian Michael, Jan A Botha, and Ian Tenney. 2020. Asking without telling: Exploring latent ontologies in contextual representations. arXiv preprint arXiv:2004.14513 (2020).

[88] Timothee Mickus, Denis Paperno, and Mathieu Constant. 2022. How to dissect a Muppet: The structure of transformer embedding spaces. Transactions of the Association for Computational Linguistics 10 (2022), 981–996.

[89] Vladislav Mikhailov, Oleg Serikov, and Ekaterina Artemova. 2021. Morph Call: Probing Morphosyntactic Content of Multilingual Transformers. In Proceedings of the Third Workshop on Computational Typology and Multilingual NLP, Ekaterina Vylomova, Elizabeth Salesky, Sabrina Mielke, Gabriella Lapesa, Ritesh Kumar, Harald Hammarström, Ivan Vulić, Anna Korhonen, Roi Reichart, Edoardo Maria Ponti, and Ryan Cotterell (Eds.). Association for Computational Linguistics, Online, 97–121. https://doi.org/10.18653/v1/2021.sigtyp-1.10

[90] Vladislav Mikhailov, Ekaterina Taktasheva, Elina Sigdel, and Ekaterina Artemova. 2021. RuSentEval: Linguistic Source, Encoder Force!. In Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing, Bogdan Babych, Olga Kanishcheva, Preslav Nakov, Jakub Piskorski, Lidia Pivovarova, Vasyl Starko, Josef Steinberger, Roman Yangarber, Michał Marcińczuk, Senja Pollak, Pavel Přibáň, and Marko Robnik-Šikonja (Eds.). Association for Computational Linguistics, Kiyv, Ukraine, 43–65. https://aclanthology.org/2021.bsnlp-1.6

[91] Hosein Mohebbi, Ali Modarressi, and Mohammad Taher Pilehvar. 2021. Exploring the Role of BERT Token Representations to Explain Sentence Probing Results. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 792–806. https://doi.org/10.18653/v1/2021.emnlp-main.61

[92] Aaron Mueller, Yu Xia, and Tal Linzen. 2022. Causal Analysis of Syntactic Agreement Neurons in Multilingual Language Models. In Proceedings of the 26th Conference on Computational Natural Language Learning (CoNLL), Antske Fokkens and Vivek Srikumar (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid), 95–109. https://doi.org/10.18653/v1/2022.conll-1.8

[93] Max Müller-Eberstein, Rob van der Goot, Barbara Plank, and Ivan Titov. 2023. Subspace Chronicles: How Linguistic Information Emerges, Shifts and Interacts during Language Model Training. In Findings of the Association for Computational Linguistics: EMNLP 2023, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 13190–13208. https://doi.org/10.18653/v1/2023.findings-emnlp.879

[94] Aleksandra Mysiak and Jacek Cyranka. 2023. Is German secretly a Slavic language? What BERT probing can tell us about language groups. In Proceedings of the 9th Workshop on Slavic Natural Language Processing 2023 (SlavicNLP 2023), Jakub Piskorski, Michał Marcińczuk, Preslav Nakov, Maciej Ogrodniczuk, Senja Pollak, Pavel Přibáň, Piotr Rybak, Josef Steinberger, and Roman Yangarber (Eds.). Association for Computational Linguistics, Dubrovnik, Croatia, 86–93. https://doi.org/10.18653/v1/2023.bsnlp-1.11

[95] Benjamin Newman, Kai-Siang Ang, Julia Gong, and John Hewitt. 2021. Refining Targeted Syntactic Evaluation of Language Models. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (Eds.). Association for Computational Linguistics, Online, 3710–3723. https://doi.org/10.18653/v1/2021.naacl-main.290

[96] Dmitry Nikolaev and Sebastian Padó. 2023. The argument–adjunct distinction in BERT: A FrameNet-based investigation. In Proceedings of the 15th International Conference on Computational Semantics, Maxime Amblard and Ellen Breitholtz (Eds.). Association for Computational Linguistics, Nancy, France, 233–239. https://aclanthology.org/2023.iwcs-1.23

[97] Vassilina Nikoulina, Maxat Tezekbayev, Nuradil Kozhakhmet, Madina Babazhanova, Matthias Gallé, and Zhenisbek Assylbekov. 2021. The rediscovery hypothesis: Language models need to meet linguistics. Journal of Artificial Intelligence Research 72 (2021), 1343–1384.

[98] Jingcheng Niu, Wenjie Lu, Eric Corlett, and Gerald Penn. 2022. Using Roark-Hollingshead Distance to Probe BERT's Syntactic Competence. In Proceedings of the Fifth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, Jasmijn Bastings, Yonatan Belinkov, Yanai Elazar, Dieuwke Hupkes, Naomi Saphra, and Sarah Wiegreffe (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid), 325–334. https://doi.org/10.18653/v1/2022.blackboxnlp-1.27

[99] Jingcheng Niu, Wenjie Lu, and Gerald Penn. 2022. Does BERT Rediscover a Classical NLP Pipeline?. In Proceedings of the 29th International Conference on Computational Linguistics, Nicoletta Calzolari, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na (Eds.). International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 3143–3153. https://aclanthology.org/2022.coling-1.278

[100] Daisuke Oba, Naoki Yoshinaga, and Masashi Toyoda. 2021. Exploratory Model Analysis Using Data-Driven Neuron Representations. In Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, Jasmijn Bastings, Yonatan Belinkov, Emmanuel Dupoux, Mario Giulianelli, Dieuwke Hupkes, Yuval Pinter, and Hassan Sajjad (Eds.). Association for Computational Linguistics, Punta Cana, Dominican Republic, 518–528. https://doi.org/10.18653/v1/2021.blackboxnlp-1.41

[101] Subba Reddy Oota, Manish Gupta, and Mariya Toneva. 2023. Joint processing of linguistic properties in brains and language models. In NeurIPS 2023. https://www.microsoft.com/en-us/research/publication/joint-processing-of-linguistic-properties-in-brains-and-language-models/

[102] Mark Ormerod, Jesús Martínez del Rincón, and Barry Devereux. 2024. How is a "kitchen chair" like a "farm horse"? Exploring the representation of noun-noun compound semantics in transformer-based language models. Computational Linguistics 50, 1 (2024), 49–81.

[103] Yulia Otmakhova, Karin Verspoor, and Jey Han Lau. 2022. Cross-linguistic Comparison of Linguistic Feature Encoding in BERT Models for Typologically Different Languages. In Proceedings of the 4th Workshop on Research in Computational Linguistic Typology and Multilingual NLP, Ekaterina Vylomova, Edoardo Ponti, and Ryan Cotterell (Eds.). Association for Computational Linguistics, Seattle, Washington, 27–35. https://doi.org/10.18653/v1/2022.sigtyp-1.4

[104] Matteo Paganelli, Donato Tiano, and Francesco Guerra. 2023. A multi-facet analysis of BERT-based entity matching models. The VLDB Journal 33, 4 (Nov. 2023), 1039–1064. https://doi.org/10.1007/s00778-023-00824-x

[105] Madhura Pande, Aakriti Budhraja, Preksha Nema, Pratyush Kumar, and Mitesh M Khapra. 2021. The heads hypothesis: A unifying statistical approach towards understanding multi-headed attention in bert. In Proceedings of the AAAI conference on artificial intelligence, Vol. 35. 13613–13621.

[106] Isabel Papadimitriou, Ethan A. Chi, Richard Futrell, and Kyle Mahowald. 2021. Deep Subjecthood: Higher-Order Grammatical Features in Multilingual BERT. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Paola Merlo, Jorg Tiedemann, and Reut Tsarfaty (Eds.). Association for Computational Linguistics, Online, 2522–2532. https://doi.org/10.18653/v1/2021.eacl-main.215

[107] Isabel Papadimitriou, Ethan A Chi, Richard Futrell, and Kyle Mahowald. 2021. Multilingual BERT, ergativity, and grammatical subjecthood. Society for Computation in Linguistics 4, 1 (2021).

[108] Isabel Papadimitriou, Richard Futrell, and Kyle Mahowald. 2022. When classifying grammatical role, BERT doesn't care about word order... except when it matters. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 636–643. https://doi.org/10.18653/v1/2022.acl-short.71

[109] Jason Phang, Shikha Bordia, Samuel R Bowman, et al. 2019. Do Attention Heads in BERT Track Syntactic Dependencies?. In NY Academy of Sciences NLP, Dialog, and Speech Workshop.

[110] Tiago Pimentel, Naomi Saphra, Adina Williams, and Ryan Cotterell. 2020. Pareto Probing: Trading Off Accuracy for Complexity. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 3138–3153. https://doi.org/10.18653/v1/2020.emnlp-main.254

[111] Tiago Pimentel, Josef Valvoda, Rowan Hall Maudslay, Ran Zmigrod, Adina Williams, and Ryan Cotterell. 2020. Information-Theoretic Probing for Linguistic Structure. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 4609–4622. https://doi.org/10.18653/v1/2020.acl-main.420

[112] Tiago Pimentel, Josef Valvoda, Niklas Stoehr, and Ryan Cotterell. 2022. The Architectural Bottleneck Principle. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 11459–11472. https://doi.org/10.18653/v1/2022.emnlp-main.788

[113] Tiago Pimentel, Josef Valvoda, Niklas Stoehr, and Ryan Cotterell. 2022. Attentional Probe: Estimating a Module's Functional Potential. 11459–11472. https://doi.org/10.18653/v1/2022.emnlp-main.788

[114] Mattia Proietti, Gianluca Lebani, and Alessandro Lenci. 2022. Does BERT Recognize an Agent? Modeling Dowty's Proto-Roles with Contextual Embeddings. In Proceedings of the 29th International Conference on Computational Linguistics, Nicoletta Calzolari, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na (Eds.). International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 4101–4112. https://aclanthology.org/2022.coling-1.360

[115] Giovanni Puccetti, Alessio Miaschi, and Felice Dell'Orletta. 2021. How Do BERT Embeddings Organize Linguistic Knowledge?. In Proceedings of Deep Learning Inside Out (DeeLIO): The 2nd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures, Eneko Agirre, Marianna Apidianaki, and Ivan Vulić (Eds.). Association for Computational Linguistics, Online, 48–57. https://doi.org/10.18653/v1/2021.deelio-1.6

[116] Alessandro Raganato and Jörg Tiedemann. 2018. An analysis of encoder representations in transformer-based machine translation. In Proceedings of the 2018 EMNLP workshop BlackboxNLP: analyzing and interpreting neural networks for NLP. 287–297.

[117] Taraka Rama, Lisa Beinborn, and Steffen Eger. 2020. Probing Multilingual BERT for Genetic and Typological Signals. In Proceedings of the 28th International Conference on Computational Linguistics, Donia Scott, Nuria Bel, and Chengqing Zong (Eds.). International Committee on Computational Linguistics, Barcelona, Spain (Online), 1214–1228. https://doi.org/10.18653/v1/2020.coling-main.105

[118] Vinit Ravishankar, Memduh Gökırmak, Lilja Øvrelid, and Erik Velldal. 2019. Multilingual Probing of Deep Pre-Trained Contextual Encoders. In Proceedings of the First NLPL Workshop on Deep Learning for Natural Language Processing, Joakim Nivre, Leon Derczynski, Filip Ginter, Bjørn Lindi, Stephan Oepen, Anders Søgaard, and Jörg Tidemann (Eds.). Linköping University Electronic Press, Turku, Finland, 37–47. https://aclanthology.org/W19-6205

[119] Emily Reif, Ann Yuan, Martin Wattenberg, Fernanda B Viegas, Andy Coenen, Adam Pearce, and Been Kim. 2019. Visualizing and measuring the geometry of BERT. Advances in neural information processing systems 32 (2019).

[120] Kyle Richardson, Hai Hu, Lawrence Moss, and Ashish Sabharwal. 2020. Probing natural language inference models through semantic fragments. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34. 8713–8721.

[121] Hassan Sajjad, Nadir Durrani, Fahim Dalvi, Firoj Alam, Abdul Khan, and Jia Xu. 2022. Analyzing Encoded Concepts in Transformer Language Models. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (Eds.). Association for Computational Linguistics, Seattle, United States, 3082–3101. https://doi.org/10.18653/v1/2022.naacl-main.225

[122] Nina Schneidermann, Daniel Hershcovich, and Bolette Pedersen. 2023. Probing for Hyperbole in Pre-Trained Language Models. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop), Vishakh Padmakumar, Gisela Vallejo, and Yao Fu (Eds.). Association for Computational Linguistics, Toronto, Canada, 200–211. https://doi.org/10.18653/v1/2023.acl-srw.30

[123] Carolin M. Schuster and Simon Hegelich. 2022. From BERT's Point of View: Revealing the Prevailing Contextual Differences. In Findings of the Association for Computational Linguistics: ACL 2022, Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 1120–1138. https://doi.org/10.18653/v1/2022.findings-acl.89

[124] Rita Sevastjanova, A Kalouli, Christin Beck, Hanna Hauptmann, and Mennatallah El-Assady. 2022. LMFingerprints: Visual explanations of language model embedding spaces through layerwise contextualization scores. In Computer Graphics Forum, Vol. 41. Wiley Online Library, 295–307.

[125] Rita Sevastjanova, Aikaterini-Lida Kalouli, Christin Beck, Hanna Schäfer, and Mennatallah El-Assady. 2021. Explaining contextualization in language models using visual analytics. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 464–476.

[126] Esther Seyffarth, Younes Samih, Laura Kallmeyer, and Hassan Sajjad. 2021. Implicit representations of event properties within contextual language models: Searching for "causativity neurons". In Proceedings of the 14th International Conference on Computational Semantics (IWCS), Sina Zarrieß, Johan Bos, Rik van Noord, and Lasha Abzianidze (Eds.). Association for Computational Linguistics, Groningen, The Netherlands (online), 110–120. https://aclanthology.org/2021.iwcs-1.11

[127] Naomi Shapiro, Amandalynne Paullada, and Shane Steinert-Threlkeld. 2021. A multilabel approach to morphosyntactic probing. In Findings of the Association for Computational Linguistics: EMNLP 2021, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Punta Cana, Dominican Republic, 4486–4524. https://doi.org/10.18653/v1/2021.findings-emnlp.382

[128] Koustuv Sinha, Robin Jia, Dieuwke Hupkes, Joelle Pineau, Adina Williams, and Douwe Kiela. 2021. Masked Language Modeling and the Distributional Hypothesis: Order Word Matters Pre-training for Little. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 2888–2913. https://doi.org/10.18653/v1/2021.emnlp-main.230

[129] Koustuv Sinha, Prasanna Parthasarathi, Joelle Pineau, and Adina Williams. 2021. UnNatural Language Inference. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 7329–7346. https://doi.org/10.18653/v1/2021.acl-long.569

[130] Mingyang Song, Yi Feng, and Liping Jing. 2022. Utilizing BERT Intermediate Layers for Unsupervised Keyphrase Extraction. In Proceedings of the 5th International Conference on Natural Language and Speech Processing (ICNLSP 2022), Mourad Abbas and Abed Alhakim Freihat (Eds.). Association for Computational Linguistics, Trento, Italy, 277–281. https://aclanthology.org/2022.icnlsp-1.32

[131] Ionut-Teodor Sorodoc, Kristina Gulordava, and Gemma Boleda. 2020. Probing for referential information in language models. In Jurafsky D, Chai J, Schluter N, Tetreault J, editors. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; 2020 Jul 5-10; Stroudsburg, USA. Stroudsburg (PA): ACL; 2020. p. 4177-89. ACL (Association for Computational Linguistics).

[132] Karolina Stańczak, Lucas Torroba Hennigen, Adina Williams, Ryan Cotterell, and Isabelle Augenstein. 2023. A latent-variable model for intrinsic probing. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 37. 13591–13599.

[133] Ekaterina Taktasheva, Vladislav Mikhailov, and Ekaterina Artemova. 2021. Shaking Syntactic Trees on the Sesame Street: Multilingual Probing with Controllable Perturbations. In Proceedings of the 1st Workshop on Multilingual Representation Learning, Duygu Ataman, Alexandra Birch,

Alexis Conneau, Orhan Firat, Sebastian Ruder, and Gozde Gul Sahin (Eds.). Association for Computational Linguistics, Punta Cana, Dominican Republic, 191–210. https://doi.org/10.18653/v1/2021.mrl-1.17

[134] Alon Talmor, Yanai Elazar, Yoav Goldberg, and Jonathan Berant. 2020. oLMpics-on what language model pre-training captures. Transactions of the Association for Computational Linguistics 8 (2020), 743–758.

[135] Minghuan Tan and Jing Jiang. 2021. Does BERT understand idioms? A probing-based empirical study of BERT encodings of idioms. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021), Virtual Conference, September. 1–3.

[136] Ian Tenney, Dipanjan Das, and Ellie Pavlick. 2019. BERT Rediscovers the Classical NLP Pipeline. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Anna Korhonen, David Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, Florence, Italy, 4593–4601. https://doi.org/10.18653/v1/P19-1452

[137] Ye Tian, Tim Nieradzik, Sepehr Jalali, and Da-shan Shiu. 2021. How does BERT process disfluency?. In Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue, Haizhou Li, Gina-Anne Levow, Zhou Yu, Chitralekha Gupta, Berrak Sisman, Siqi Cai, David Vandyke, Nina Dethlefs, Yan Wu, and Junyi Jessy Li (Eds.). Association for Computational Linguistics, Singapore and Online, 208–217. https://doi.org/10.18653/v1/2021.sigdial-1.22

[138] Hariprasad Timmapathini, Anmol Nayak, Sarathchandra Mandadi, Siva Sangada, Vaibhav Kesri, Karthikeyan Ponnalagu, and Vijendran Gopalan Venkoparao. 2021. Probing the SpanBERT Architecture to interpret Scientific Domain Adaptation Challenges for Coreference Resolution.. In SDU@ AAAI.

[139] Mycal Tucker, Tiwalayo Eisape, Peng Qian, Roger Levy, and Julie Shah. 2022. When Does Syntax Mediate Neural Language Model Performance? Evidence from Dropout Probes. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (Eds.). Association for Computational Linguistics, Seattle, United States, 5393–5408. https://doi.org/10.18653/v1/2022.naacl-main.394

[140] Mycal Tucker, Peng Qian, and Roger Levy. 2021. What if this modified that? syntactic interventions via counterfactual embeddings. arXiv preprint arXiv:2105.14002 (2021).

[141] Andrea Gregor de Varda and Marco Marelli. 2023. Data-driven cross-lingual syntax: An agreement study with massively multilingual models. Computational Linguistics 49, 2 (2023), 261–299.

[142] Ivan Vulić, Edoardo Maria Ponti, Robert Litschko, Goran Glavaš, and Anna Korhonen. 2020. Probing Pretrained Language Models for Lexical Semantics. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 7222–7240. https://doi.org/10.18653/v1/2020.emnlp-main.586

[143] Jonas Wallat, Fabian Beringer, Abhijit Anand, and Avishek Anand. 2023. Probing BERT for ranking abilities. In European Conference on Information Retrieval. Springer, 255–273.

[144] Yile Wang, Leyang Cui, and Yue Zhang. 2020. Does Chinese BERT Encode Word Structure?. In Proceedings of the 28th International Conference on Computational Linguistics, Donia Scott, Nuria Bel, and Chengqing Zong (Eds.). International Committee on Computational Linguistics, Barcelona, Spain (Online), 2826–2836. https://doi.org/10.18653/v1/2020.coling-main.254

[145] Jason Wei, Dan Garrette, Tal Linzen, and Ellie Pavlick. 2021. Frequency Effects on Syntactic Rule Learning in Transformers. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 932–948. https://doi.org/10.18653/v1/2021.emnlp-main.72

[146] Leonie Weissweiler, Valentin Hofmann, Abdullatif Köksal, and Hinrich Schütze. 2022. The better your Syntax, the better your Semantics? Probing Pretrained Language Models for the English Comparative Correlative. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 10859–10882. https://doi.org/10.18653/v1/2022.emnlp-main.746

[147] Leonie Weissweiler, Valentin Hofmann, Abdullatif Köksal, and Hinrich Schütze. 2023. Explaining pretrained language models' understanding of linguistic structures using construction grammar. Frontiers in Artificial Intelligence 6 (2023), 1225791.

[148] Jennifer C. White, Tiago Pimentel, Naomi Saphra, and Ryan Cotterell. 2021. A Non-Linear Structural Probe. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (Eds.). Association for Computational Linguistics, Online, 132–138. https://doi.org/10.18653/v1/2021.naacl-main.12

[149] Zhiyong Wu, Yun Chen, Ben Kao, and Qun Liu. 2020. Perturbed Masking: Parameter-free Probing for Analyzing and Interpreting BERT. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 4166–4176. https://doi.org/10.18653/v1/2020.acl-main.383

[150] Tingyu Xia, Yue Wang, Yuan Tian, and Yi Chang. 2021. Using prior knowledge to guide bert's attention in semantic textual matching tasks. In Proceedings of the web conference 2021. 2466–2475.

[151] Ningyu Xu, Tao Gui, Ruotian Ma, Qi Zhang, Jingting Ye, Menghan Zhang, and Xuanjing Huang. 2022. Cross-Linguistic Syntactic Difference in Multilingual BERT: How Good is It and How Does It Affect Transfer?. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 8073–8092. https://doi.org/10.18653/v1/2022.emnlp-main.552

[152] David Yi, James Bruno, Jiayu Han, Peter Zukerman, and Shane Steinert-Threlkeld. 2022. Probing for Understanding of English Verb Classes and Alternations in Large Pre-trained Language Models. In Proceedings of the Fifth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, Jasmijn Bastings, Yonatan Belinkov, Yanai Elazar, Dieuwke Hupkes, Naomi Saphra, and Sarah Wiegreffe (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid), 142–152. https://doi.org/10.18653/v1/2022.blackboxnlp-1.12

[153] Fabio Massimo Zanzotto, Andrea Santilli, Leonardo Ranaldi, Dario Onorati, Pierfrancesco Tommasino, and Francesca Fallucchi. 2020. KERMIT: Complementing Transformer Architectures with Encoders of Explicit Syntactic Interpretations. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 256–267. https://doi.org/10.18653/v1/2020.emnlp-main.18

[154] Jingyi Zhang, Gerard de Melo, Hongfei Xu, and Kehai Chen. 2023. A Closer Look at Transformer Attention for Multilingual Translation. In Proceedings of the Eighth Conference on Machine Translation, Philipp Koehn, Barry Haddow, Tom Kocmi, and Christof Monz (Eds.). Association for Computational Linguistics, Singapore, 496–506. https://doi.org/10.18653/v1/2023.wmt-1.45

[155] Xiongyi Zhang, Jan-Willem van de Meent, and Byron Wallace. 2021. Disentangling Representations of Text by Masking Transformers. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 778–791. https://doi.org/10.18653/v1/2021.emnlp-main.60

[156] Mengjie Zhao, Philipp Dufter, Yadollah Yaghoobzadeh, and Hinrich Schütze. 2020. Quantifying the Contextualization of Word Representations with Semantic Class Probing. In Findings of the Association for Computational Linguistics: EMNLP 2020, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 1219–1234. https://doi.org/10.18653/v1/2020.findings-emnlp.109

[157] Yiyun Zhao and Steven Bethard. 2020. How does BERT's attention change when you fine-tune? An analysis methodology and a case study in negation scope. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 4729–4747. https://doi.org/10.18653/v1/2020.acl-main.429

[158] Jianyu Zheng and Ying Liu. 2022. Probing language identity encoded in pre-trained multilingual models: a typological view. PeerJ Computer Science 8 (2022), e899.

[159] Jianyu Zheng and Ying Liu. 2023. What does Chinese BERT learn about syntactic knowledge? PeerJ Computer Science 9 (2023).

[160] Jianyu Zheng and Jin Sun. 2023. Exploring the Word Structure of Ancient Chinese Encoded in BERT Models. In 2023 16th International Conference on Advanced Computer Theory and Engineering (ICACTE). IEEE, 41–45.

[161] Zining Zhu, Chuer Pan, Mohamed Abdalla, and Frank Rudzicz. 2020. Examining the rhetorical capacities of neural language models. In Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, Afra Alishahi, Yonatan Belinkov, Grzegorz Chrupała, Dieuwke Hupkes, Yuval Pinter, and Hassan Sajjad (Eds.). Association for Computational Linguistics, Online, 16–32. https://doi.org/10.18653/v1/2020.blackboxnlp-1.3