

/instituut voor de Nederlandse taal/

Stage PARSEME-NL corpus

Contactpersoon: [Carole Tiberius](#)

[PARSEME](#) is een internationaal initiatief waarbinnen meertalige corpora worden ontwikkeld en onderhouden die geannoteerd zijn voor [meerwoorduitdrukkingen](#) (*MWEs*). PARSEME beperkt zich voornamelijk tot verbale MWEs en tot nu toe zijn er corpora geannoteerd in 26 verschillende talen. Voor het annoteren zijn universele [annotatierichtlijnen](#) opgesteld, die door alle talen die onderdeel uitmaken van het PARSEME-initiatief gebruikt worden.

Op dit moment wordt er gewerkt aan het realiseren van een PARSEME-NL corpus. Dit is een samenwerkingsverband tussen de Universiteit Groningen, de Universiteit Utrecht, en de Universiteit Leiden en het Instituut voor de Nederlandse Taal. Voor dit project zijn wij op zoek naar een stagiair(e) die gaat meewerken aan de annotatie van verbale MWEs in het Nederlandse corpus.

Voor meer informatie over PARSEME, zie o.a.:

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, Archana Bhatia, Marie Candito, Polona Gantar, Uxo Inurrieta, Albert Gatt, et al.. 2023. [PARSEME corpus release 1.3](#). In *Proceedings of the 19th Workshop on Multiword Expressions (MWE 2023)*, pages 24–35, Dubrovnik, Croatia. Association for Computational Linguistics.

Werkzaamheden bestaan uit:

- het kritisch bestuderen van de PARSEME-richtlijnen en het toevoegen van Nederlandse voorbeelden aan deze richtlijnen
- het annoteren van verbale MWEs volgens de PARSEME-richtlijnen in de [FLAT](#)-annotatietool
- het controleren van automatische annotaties
- het bijdragen aan het opstellen van taalspecifieke richtlijnen voor het Nederlands voor zover dit van toepassing is.

Vereiste kennis en vaardigheden:

- voltooide bachelor studie Nederlandse taal of een andere taalkundige studie
- goede beheersing van de Nederlandse taal (moedertaalspeker)
- goede beheersing van de Nederlandse syntaxis
- ervaring met het werken met corpora is een pre

Praktische informatie:

Het Instituut voor de Nederlandse Taal biedt een werkplek, computer, en begeleiding. Het werk kan indien gewenst in overleg deels thuis worden uitgevoerd.

Het Instituut voor de Nederlandse Taal (INT) in Leiden is hét kennisinstituut voor het Nederlands. Het is een breed toegankelijk wetenschappelijk instituut dat alle aspecten van de Nederlandse taal bestudeert, waaronder de woordenschat, grammatica en taalvariatie.