



Universiteit Leiden

The role of questions in disinformation

Applicants

| Supervisor Name | Discipline |
|------------------|---|
| Matthijs Westera | Assistant professor of Humanities & AI |
| Jenny Doetjes | Professor of Semantics & Language Variation |

Project description

There are many ways to ask a question. In our past and ongoing research, we investigate the form, meaning and use of different types of questions, in relation to phenomena such as uncertainty, bias and rhetoric. The goal of the proposed traineeships is to approach these topics specifically from the angle of *disinformation*, aimed at the following research question:

Main question: How are different types of questions used for misleading an audience, in different linguistic domains (e.g., political debates vs. fake news)?

The trainees will address this question in three stages:

1. Typology of misleading questions: In close collaboration with the supervisors, the trainees will develop a theoretically sound typology of misleading questions. This will require integrating a currently fragmented, multi-disciplinary literature, borrowing from existing question typologies in education sciences, linguistic work on bias, presupposition and rhetoric, and sociological and psychological findings on disinformation.

2. Data collection and annotation: Using the developed typology of misleading questions, the trainees will extract and annotate misleading questions in a language of choice (Dutch, French or English) from two domains, such as political debates, twitter, or fake news, to enable a between-domain comparison. Each trainee will focus on one domain. The language and domains investigated will be determined in collaboration, taking into account the interests and backgrounds of the trainees.

3. Qualitative and quantitative analysis: Qualitative analysis will consist in devising theoretical explanations (syntax, semantics, pragmatics) for phenomena encountered. Quantitative analysis will consist in fitting statistical (including machine learning) models to assess inter-annotator agreement and to compare the use of different misleading question types between linguistic domains.

The traineeships will culminate in one joint academic paper submitted to a fitting conference.

Research Trainee Profile

We are looking for Ma or ResMa students with a background in Linguistics, Natural Language Processing, or Journalism. The following additional features are a plus, but not strict prerequisites:

- Demonstrable courses or term papers on relevant topics, especially pragmatics, semantics, questions, rhetoric, or disinformation.
- Demonstrable familiarity with relevant research methods, such as linguistic annotation, corpus analysis and natural language processing (e.g., Python).
- Fluency in at least one of the languages of choice (Dutch, French or English).

Each of the three stages can hold academic challenges, depending on the trainees' backgrounds. The main challenge in stage 1 will be the interdisciplinarity of the research topic, and the fragmented and multi-disciplinary nature of the literature on misleading questions. The main challenge in stage 2 can be either technical (data processing and filtering) or linguistic (annotation). Stage 3 will especially test the trainees' ability to interpret data and formulate and test hypotheses.

Collaboration

The proposed traineeships will augment the initial phase of a recently started, interdisciplinary collaboration between the applicants, on questions and disinformation. Matthijs Westera's ongoing research focuses on the role of explicit and implicit questions in discourse structure, relying on quantitative methods, big data and machine learning. Jenny Doetjes was one of the applicants of the recently finished NWO-project *Understanding questions*, which examined the relation between prosody, processing, syntax and semantics in French and Mandarin by means of theoretically driven experiments. Our new project on questions and disinformation will be a data-driven project with a strong theoretical emphasis, in which the strengths of both applicants come together. The proposed traineeships will exhibit the same interdisciplinarity.

Our collaboration and the proposed traineeships will also increase synergy between groups within LUCL that are not normally working together. Given the topic of our project and its combination of methodologies, the results will be relevant for both theoretical and computational linguists, as well as researchers studying disinformation from the perspective of language use and journalism, such as the group behind Nieuwscheckers.nl and the recently awarded EDMO-project, with whom we are in contact. These researchers have not yet looked specifically at the role of questions in disinformation, and we are looking forward to sharing our findings with them, as well as to benefitting from their expertise on disinformation more generally.

Deliverables

The project is intended to result in the following:

- A reusable, comprehensive typology of ways of misleading with questions, none of which currently exists despite the surge of research on disinformation.
- An annotated dataset of questions (in either Dutch, French or English) from two domains (such as political debates, twitter or fake news), annotated according to the aforementioned typology of misleading question types.
- One academic paper reporting on the foregoing items as well as the resulting analysis, submitted to a fitting conference such as the International Pragmatics Conference, Semantics and Pragmatics of Dialogue (SemDial), Linguistic Resources and Evaluation (LREC), various computational venues (ACL, EMNLP, CoLing, IWCS), more theoretical venues (SALT, Sinn und Bedeutung) or one of the many student research conferences.

Planning

| Month | Activities | Hours* |
|-------|--|--------|
| 1 | Warming-up. Initial exploration of literature and existing datasets. Decide which language and domains to investigate. | 25 |
| 2 | Stage 1: Typology of misleading questions. The main activities are literature review and theoretical synthesis. | 25 |
| 3 & 4 | Stage 2: Data collection and annotation. The annotation will proceed in stages, where after each stage we compare inter-annotator agreement and where necessary revise the typology. | 50 |
| 5 & 6 | Stage 3: Qualitative and quantitative analysis. The trainees will develop and pursue their own hypotheses and evaluation methods, depending on their interests and skills. | 50 |
| 7 | Finalize paper. During each stage the trainees will already work on a first draft of a corresponding section for the paper, which will be completed and submitted in the final month. | 25 |

*Approximate working hours *per trainee*, based on 7 months × 0.15fte = 173 hours per trainee.

Student Application

Please submit your application, consisting of a cover letter and resume in a single PDF file, to m.westera@hum.leidenuniv.nl.