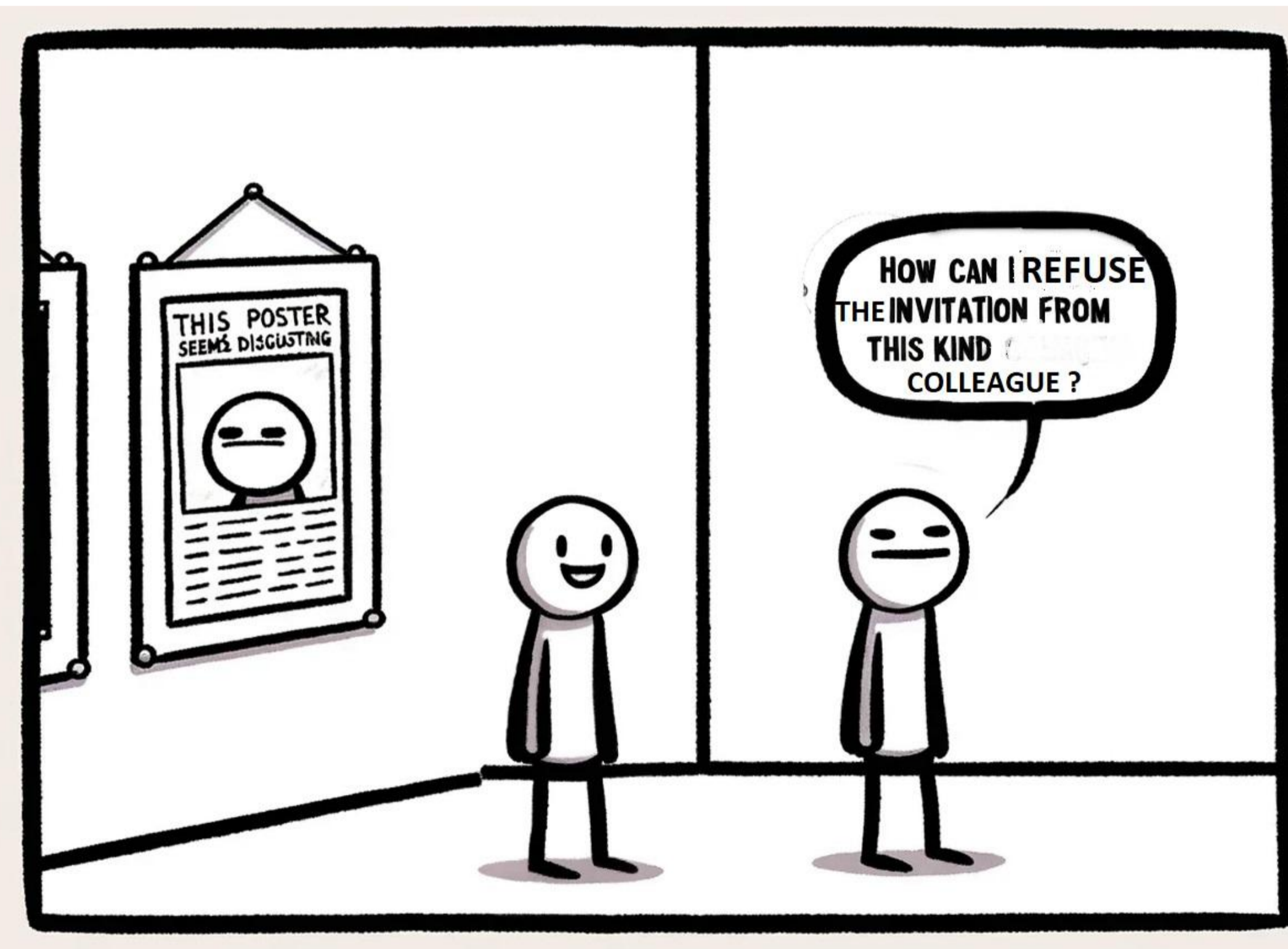


CLARIAH-AT-Roadshow
PLUS Salzburg 13.09.2024
Funding Call 2022: Interoperability and
Reusability of DH Data and Tools

Nicola Brocca
Universität Innsbruck
Institut für Fachdidaktik

Intro

- Corpus pragmatics: potential benefits for contrastive pragmatics, foreign language didactics, and theoretical and applied linguistics
- Challenges in pragmatic annotation: addressing the demands for accurate and efficient analysis



Steps

1. Designing an annotation scheme:

The scheme is derived from taxonomies based on speech-act theory and previously used in intercultural pragmatic studies.

LadderWeb: a machine learning-based web app for pragmatic annotation achieving 0.7 accuracy*



Based on a test conducted on July 26, 2024,
involving 12 cancellations from the DisDir Corpus:
Accuracy on Sub-acts: 0.7
Accuracy on Sub-act specifications: 0.66
Accuracy on Modifiers: 1



Scan to
try the app.
Or go to:
<https://ifd-ladderweb.uibk.ac.at/>

References:

Brocca, N., Nuzzo, E., Cortés Velásquez, D., & Rudigier, M. (2023). Linguistic politeness across Austria and Italy: Backing out of an invitation with an instant message. *Journal of Pragmatics*, 209, 56–70. <https://doi.org/10.1016/j.pragma.2023.02.018>

Brocca, N., & Nuzzo, E. (2024). Exploring Request Strategies in Austrian Italian Learners: Pragmatic Transfer Insights. *Journal of Pragmatics*, 220, 32–46. <https://doi.org/10.1016/j.pragma.2023.11.012>

Nuzzo, E., Brocca N. (in review) Raising the (meta)pragmatic awareness of non-native pre-service teachers of L2 Italian with a telecollaborative data-driven learning project on formal email writing. *System*

Project presentation: <https://clariah.at/ladderweb>

Students working with Ladder Data: <https://ladder.hypotheses.org>

The complete Ladder and DisDir corpora are also archived in ARCHE:
<https://hdl.handle.net/21.11115/0000-0011-83CC-3> => LADDER
<https://hdl.handle.net/21.11115/0000-0011-83CD-2> => DisDir

3. Executing manual annotation

Data for manual annotation come from annotated **corpora** DisDir and Ladder, focusing on **cancellations** and **requests** in **Italian L1**, **Italian L2**, and **German**.



3. Creating an AI-based routine:

Utilizing decision trees and statistical methods, the routine categorizes words based on annotation and features. The application integrates the Corpus Query Processor (CQP) of the IMS Open Corpus Workbench, offering a user-friendly interface for accessing annotated texts. Meta-information enables the creation of subcorpora based on speech act type, social context, medium, languages, participant data, etc.

3. Training the routine

The routine is trained with the aforementioned manually annotated data, including (to date) approximately 800 cancellations and 200 requests.

3. Manually correcting the results

While the model can autonomously tag new cancellations and requests with reasonable accuracy*, the results require manual fine-tuning for improved precision.