

# Using Weak Signal in NLP

*Malvina Nissim*

University of Groningen

m.nissim@rug.nl

## ABSTRACT

Treebanking requires substantial expert labour in the annotation of a variety of language phenomena. Possibly to a lesser extent for phenomena where laypeople can also contribute, the need for assigning manual labels nevertheless characterises almost all language processing tasks, since they are usually best solved by supervised models. Such models are indeed accurate, but we also know that they lack portability, as they are bound to languages, genres, and even specific datasets. Having spent years dealing with annotation issues and label acquisition for various semantic and pragmatic tasks, in this talk I take a radically different perspective, which hopefully can yield interesting reflections over treebanking, too. I will show various ways to cheaply obtain and exploit weaker signal in supervised learning, even venturing on the suggestion to reduce existing strong, accurate signal in order to enhance portability. I will do so via discussing three case studies in three different classification tasks, all focused on social media.

---

**KEYWORDS:** supervised learning, weak signal, social media.

---