Sign Language Annotation: How to explore the terra incognita?

The lack of experienced human annotators is a limiting factor for the size of annotated sign language corpora. In fact, many sign language corpora around the world have only a small portion of the collected data fully annotated and thus fully accessible.

In order to have some idea where to search for phenomena outside the annotated parts of a corpus, it is common practice to have (rough) translations into a spoken language for most parts of the corpus. This at least gives a semantic approach to the signed content which is useful e.g. to identify signs for less-frequent concepts. While computer vision approaches hold promise for searching diverse form aspects (without the detour through translations into spoken language), the current state of the art indicates that automatic measures may not match the annotation quality delivered by skilled human annotators for the foreseeable future. For applications like sign spotting, exploring not only the top-ranked but also lower-ranked outputs of machine learning models is crucial. Where computational models are time-consuming to compute, this requires storing the results of such analyses to become available to the corpus users.

However, traditional multimodal annotation schemes assign a specific value to a combination of a time span on the video and a tier. Even if we know that human annotation is error-prone, and quality assurance is, for cost reasons, less effective than e.g. independent double annotation, any annotation made is considered truth. From automatic processing, a number of alternatives compete for the same timespan, with their confidence levels fluctuating over time.

In this paper, we present a hybrid model integrating these fundamentally different concepts.

We also discuss meta measures such as the degree of interactivity of a conversation and demonstrate the advantages of using definitions that can be computed automatically on annotated data and can be approximated on unannotated data by means of machine learning models such as action detection from pose data. For corpus linguistic measures such as lexical frequency rank, we follow approaches from the literature to interpret confidence levels as probabilities in order to compensate the distortions introduced by spot annotation islands in terra incognita.