

Disentangling linguistic intelligence: automatic generalisation of structure and meaning across languages

All speakers can understand a sentence never heard before, or derive the meaning of a word or a sentence from its parts. Children can learn any language they are exposed to. And yet, these basic linguistic skills have proven very hard to reach by computational models. The current reported success of machine learning architectures is based on computationally (and environmentally) expensive algorithms and prohibitively large amounts of data that are available for only a few, non-representative languages. This limits the access to natural language processing technology to a few dominant languages and modalities and leads to the development of systems that are not human-like, with great potential for unfairness and bias.

To reach better, possibly human-like, abilities in neural networks' abstraction and generalisation, we need to move beyond the simple language modelling tasks currently used and develop tasks and data that train the networks to more complex and compositional linguistic skills. In this project, we set the challenging goals of achieving higher-level linguistic abilities in machines, while training in more realistic settings. We identify these abilities as the intelligent ability to infer patterns of regularities in unstructured data, generalise from few examples, using abstractions that are valid across possibly very different languages. We study if current neural network architectures have the same properties of learning, generalisation and abstraction when processing language. Specifically, we ask (i) Can they learn the underlying generative structure of complex data? (ii) Under what conditions can they learn from zero or few examples? (iii) Do their learning patterns exhibit cross-linguistically valid abstractions?

To achieve these goals, we concentrate on one of the core building blocks of any language: verbs and their argument structure, the *who* did *what* to *whom* expressing core events and actions. Argument structure is defined by specific combinations of elements (the arguments and the predicate) in different templates (the subcategorisation frames), that form higher-order patterns of similarity across sentences (the alternations) and are defined at a high level of abstraction across languages (the semantic roles).

We aim to learn disentangled representations of these components of argument structure. A disentangled representation encodes information about the salient factors of variation in the data independently. We have developed, and demonstrated in pilot work, a new set of **progressive matrix tasks** inspired by IQ intelligence tests. These tasks are developed specifically for language and learn disentangled linguistic representations of underlying linguistic rules of grammar.

We apply this novel method to the three main questions of our investigation

(i) To demonstrate learning of the generative components of argument structure from data and facets of verb meaning, we develop new data sets and new versions of the progressive matrix task for argument structure concentrating specifically on learning argument alternations (*John loaded the truck with hay/John loaded Hay onto the truck*), on the gradation of compositionality (obligatory arguments vs optional adjuncts) and the compositionality of complex clauses.

(ii) To investigate the conditions that make learning from few examples possible, we hypothesize that structured categorisation reduces the data sample size needed for learning. We also take inspiration from human learning biases, and develop models of increasing complexity and size to study how structure and data size interact.

(iii) To study if truly abstract representations of verb argument structure and semantic roles emerge, we create novel cross-lingual learning tasks and novel cross-lingual artificial complex data and study learning and generalisation of complex morphological paradigms across distant languages.

This kind of investigation, based on tasks that require a mixture of linguistic knowledge and higher-level linguistic reasoning has never been tried before for natural language processing. Current pilot studies for simple linguistic problems by the PI and her team show that the method is promising. We plan to extend the method here to two very challenging areas of natural language processing. On the one hand, we tackle complex linguistic data that tap into the core semantics of clauses, a well-studied, but so far unsolved, core problem for natural language understanding. On the other hand, we study complex morphological paradigms and systems, a very little studied cross-linguistically important problem.

If successful, this research could lead to a significant methodological shift. These investigations can lead to three beneficial improvements of methods and practices: (i) deep, compositional representations would be learnt, thus reducing needs in data size; (ii) current machine learning methods would be extended to low-resources languages or low-resource modalities and scenarios; (iii) higher-level abstractions would be learnt, avoiding the use of superficial, associative cues that are the cause of so much bias and potential harm in the representations learned by current artificial intelligence and natural language processing systems.