

Semantic subcategorisation for creative generation of light verb constructions

Lin de Huybrecht, Computational Creativity Lab @ AI Lab VUB, Geraint Wiggins

We need transparent language models, based on linguistics & cognitive science

Knowing which expressions are correct is hard for second language learners (2LLs).

"*grab a snack*"

"**grab a beef stew*"

*not semantically idiomatic

application: Personalised technology can help 2LLs with mastering these expressions

I am extending the DisCoCat framework for creative natural language generation

Distributional semantics

Compositionality

Category theory

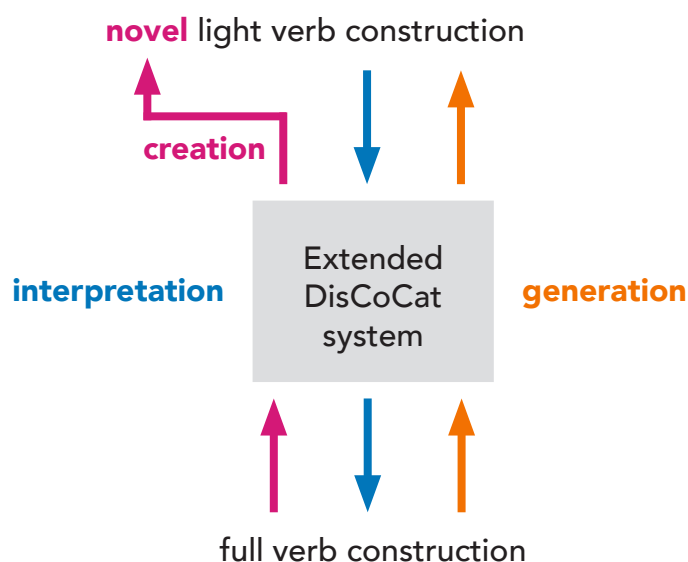
language = syntax × semantics

sentence embedding

word type

word embedding

I am developing a system that generates novel light verb constructions



Classical part-of-speech categories are insufficient for text generation

We want:

Alice read a paper.

Alice read a table.

Alice tabled a paper.

Alice talks.

We don't want:

*The dog read a table.

*The paper read a cat.

*Alice talked a paper.

*The table talks.

This research will give us a deeper insight on how semantic spaces interact.

This will contribute towards more transparent and explainable technology.

We need to subcategorise word types to account for selectional restrictions

