Testing the Developmental Plausibility of BERT by Capturing the Role of Verb-Event Structure in Early Word Sense Disambiguation

Francesco Cabiddu

Cardiff University
CabidduF@cardiff.ac.uk

Talk outline

- Research questions
- Experimental data
- Computational modelling

- Questions

BERT as a usage-based learner

Domain-general processes that drive language acquisition

- Associative learning
    - Context-dependent knowledge
    - language exemplars

- Analogy
    - Context-independent knowledge
    - linguistic structures

(Abbot-Smith & Tomasello, 2006; Ambridge, 2019; 2020; Bybee, 2010)

BERT as a usage-based learner

BERT can
- be used cross-modally

  (Lu et al., 2019; Qi et al., 2020; Sun et al., 2019)

- form context-dependent representations (i.e., each token as a vector representation)

- form context-independent abstractions:
  - Parse trees, Subject-verb relations, Attention heads attend to different syntactic relations
  - Semantic features and roles

  (Clark et al., 2019; Goldberg, 2019; Hewitt and Manning, 2019; Jawahar et al., 2019; Manning et al., 2020; Tenney et al., 2019)

BERT as a usage-based learner

- It can be applied to a multitude of tasks

- It can learn from naturalistic speech

Previous work:

High-level descriptions / limited set of artificial tasks:
- E.g., Verb-event profiles in a Bayesian model from syntactic patterns and semantic properties
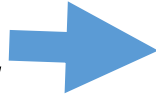 (Alishahi and Stevenson, 2010).

Research questions

Can BERT (Devlin et al., 2019)
1. capture aspects of child language processing?
   - Word learning: Modest results (Jr. & Licato, 2021)
   - Use of verb-event structures in word sense disambiguation

2. capture developmental differences between children and adults?

Word sense disambiguation in children
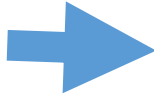
Resolution depends on sentence parsing →

Elmo watched a movie with a castle and a princess. That was a funny **night**
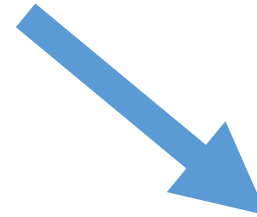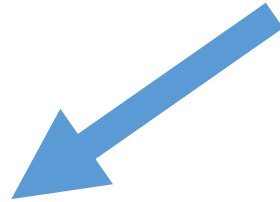
(Rabagliati et al., 2013)

Word sense disambiguation in children

Resolution depends on sentence parsing

Elmo watched a movie with a castle and a princess. That was a funny night

(Rabagliati et al., 2013)

Word sense disambiguation in children
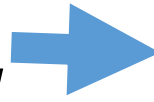
Resolution depends on sentence parsing → *Elmo watched a movie with a castle and a princess. That was a funny **night***
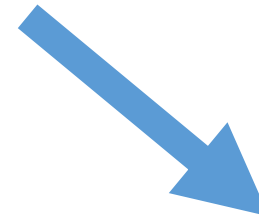
(Rabagliati et al., 2013)

low-level cues
(Snedeker & Yuan, 2008)

high-level cues
(Trueswell & Gleitman, 2007)

*e.g., "castle" and "princess" frequently co-occur with "**knight**".*

*e.g., it is plausible to assume that the speaker watched the movie at **night**.*

The role of verb-event structure

Early word processing
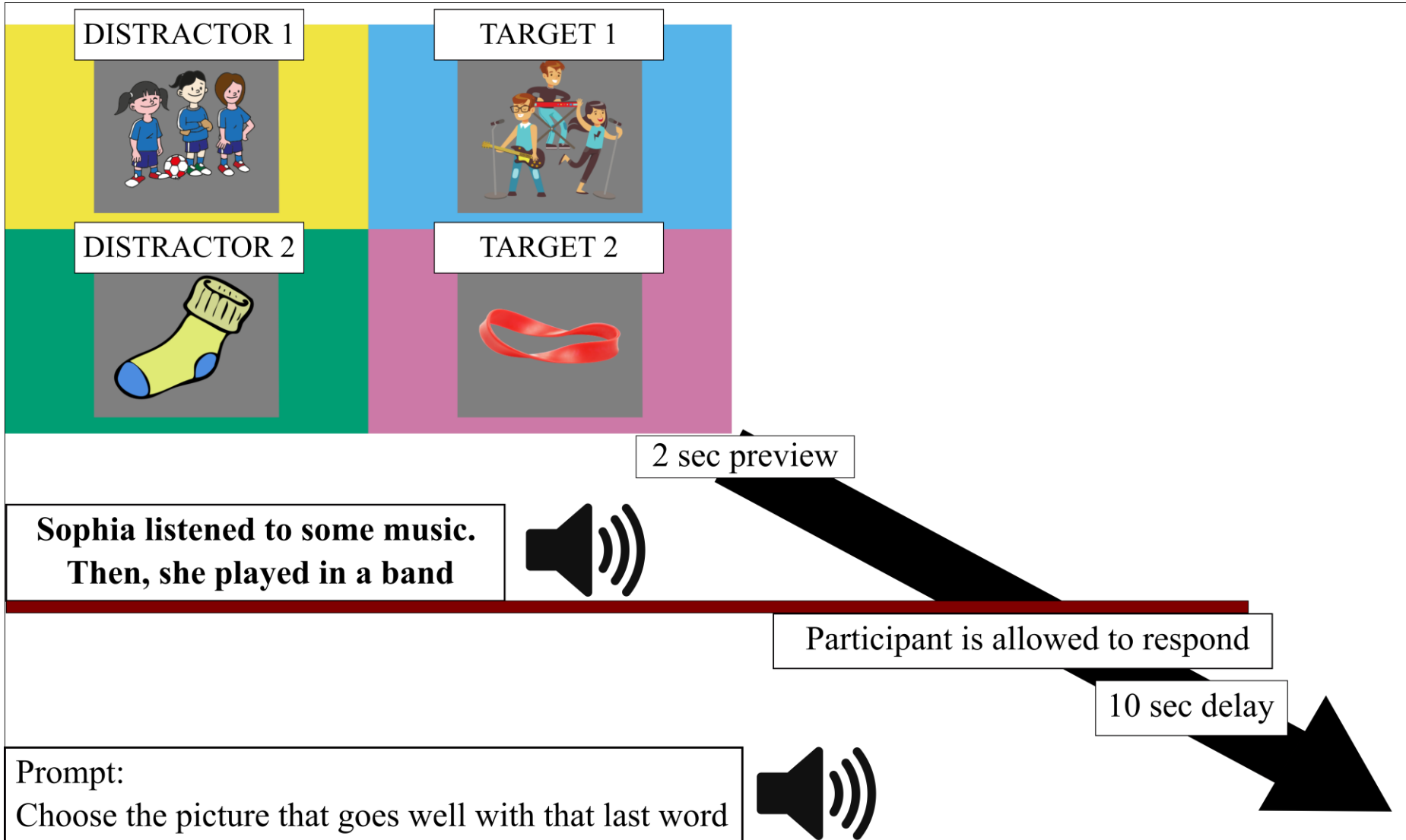### (Andreu et al., 2013)

Pushing *the…*

The role of verb-event structure
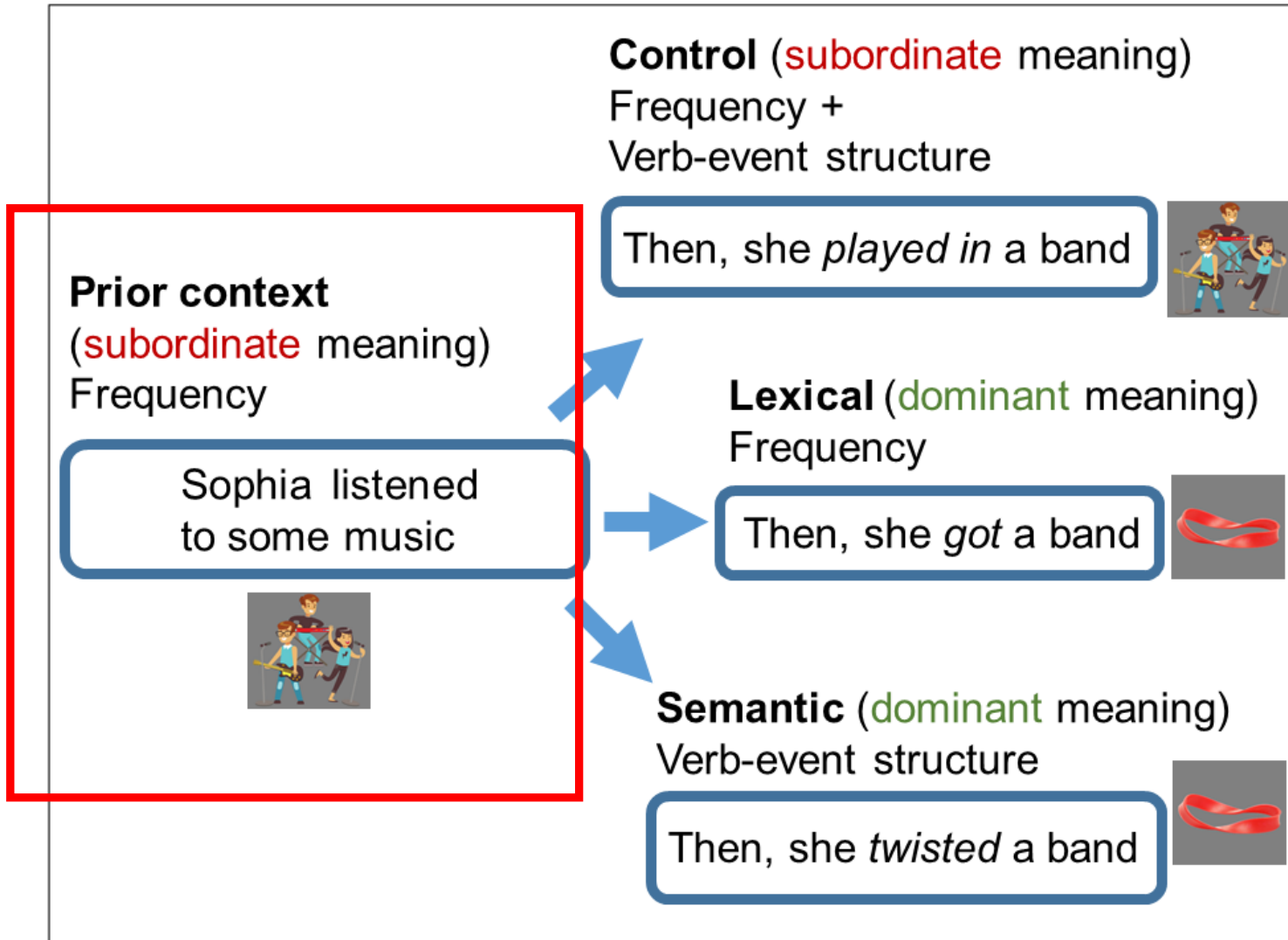
Can young children use verb-event structures
to resolve lexical ambiguities?

Adult and children study
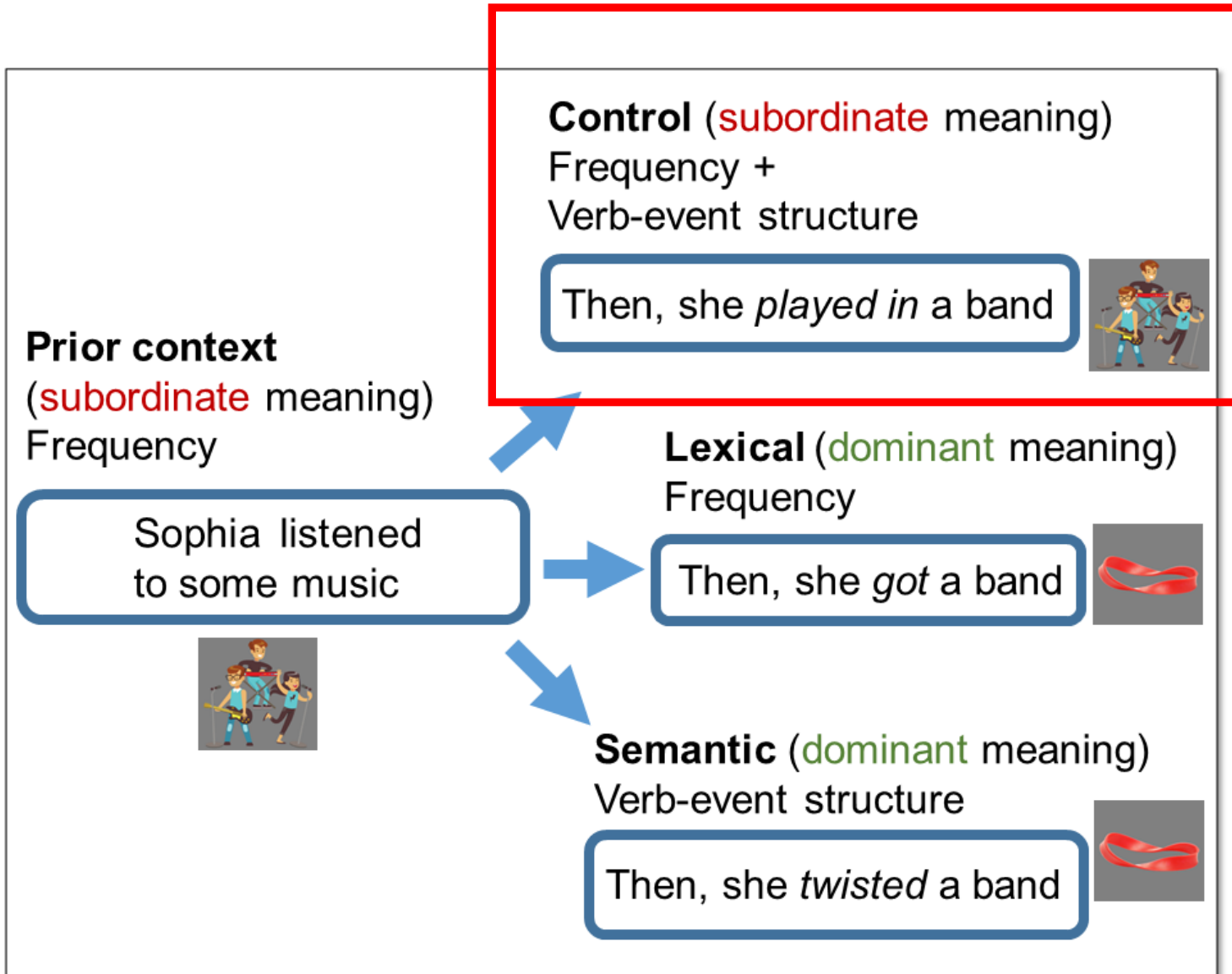

DISTRACTOR 1

TARGET 1

DISTRACTOR 2

TARGET 2

2 sec preview

Sophia listened to some music. Then, she played in a band

Participant is allowed to respond

10 sec delay

Prompt:
Choose the picture that goes well with that last word

Adult and children study

Prior context
(subordinate meaning)
Frequency

> Sophia listened
> to some music

Control (subordinate meaning)
Frequency +
Verb-event structure

> Then, she *played in* a band

Lexical (dominant meaning)
Frequency

> Then, she *got* a band

Semantic (dominant meaning)
Verb-event structure

> Then, she *twisted* a band

Adult and children study

Prior context
(subordinate meaning)
Frequency

Sophia listened to some music

Control (subordinate meaning)
Frequency +
Verb-event structure

Then, she *played in* a band

Lexical (dominant meaning)
Frequency

Then, she *got* a band

Semantic (dominant meaning)
Verb-event structure

Then, she *twisted* a band

Adult and children study

Control (subordinate meaning)
Frequency +
Verb-event structure

Then, she *played in* a band

Prior context
(subordinate meaning)
Frequency

Sophia listened
to some music

Lexical (dominant meaning)
Frequency

Then, she *got* a band

Semantic (dominant meaning)
Verb-event structure

Then, she *twisted* a band

Adult and children study
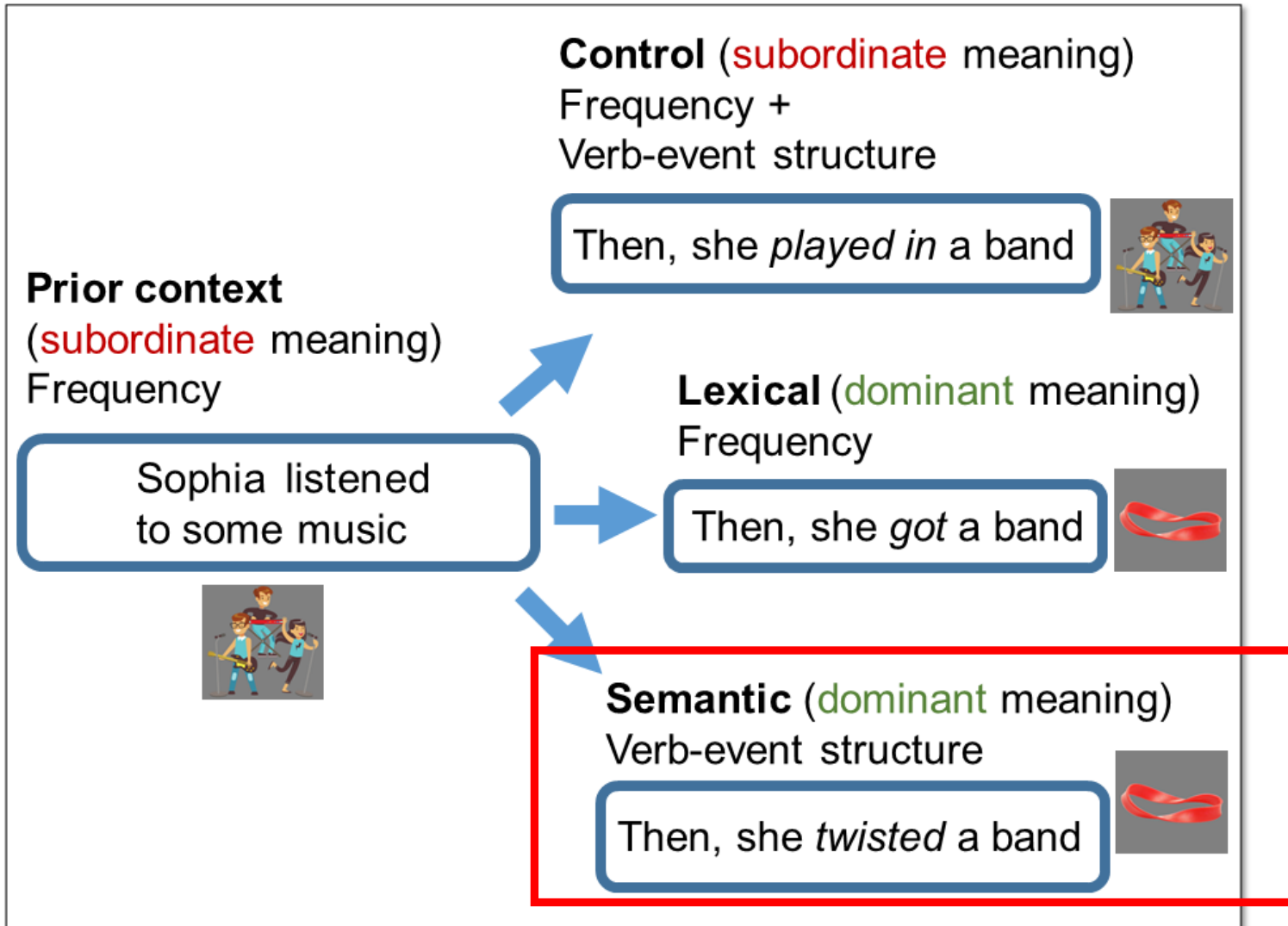
Adult and children study

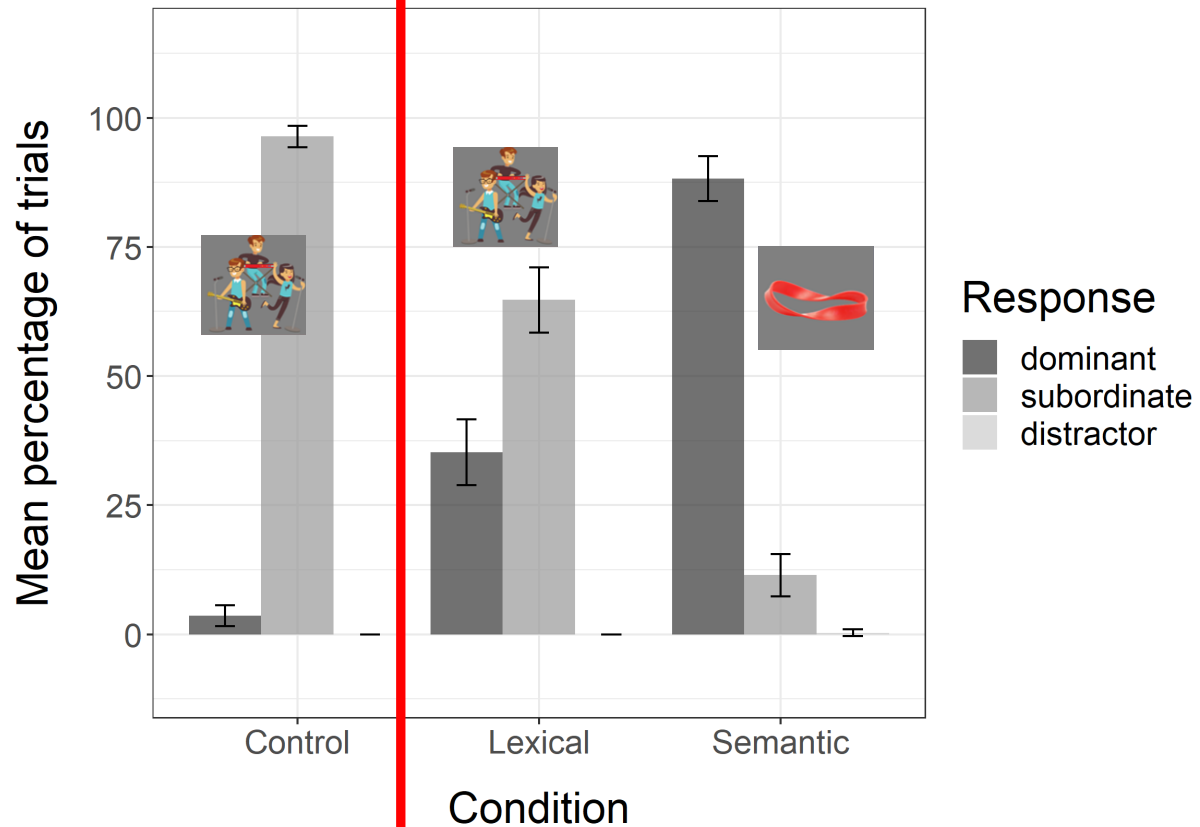**83** adults age: *M* = 23 years, *SD* = 5 years; 62 female

45 4-year-olds: *M* = 52 months, *SD* = 3 months;
21 third gender, 9 male, 8 female, 7 prefer not to say
English-speaking children (10 bilingual)
SES representative of England-Wales population

Adult performance

Played in a band
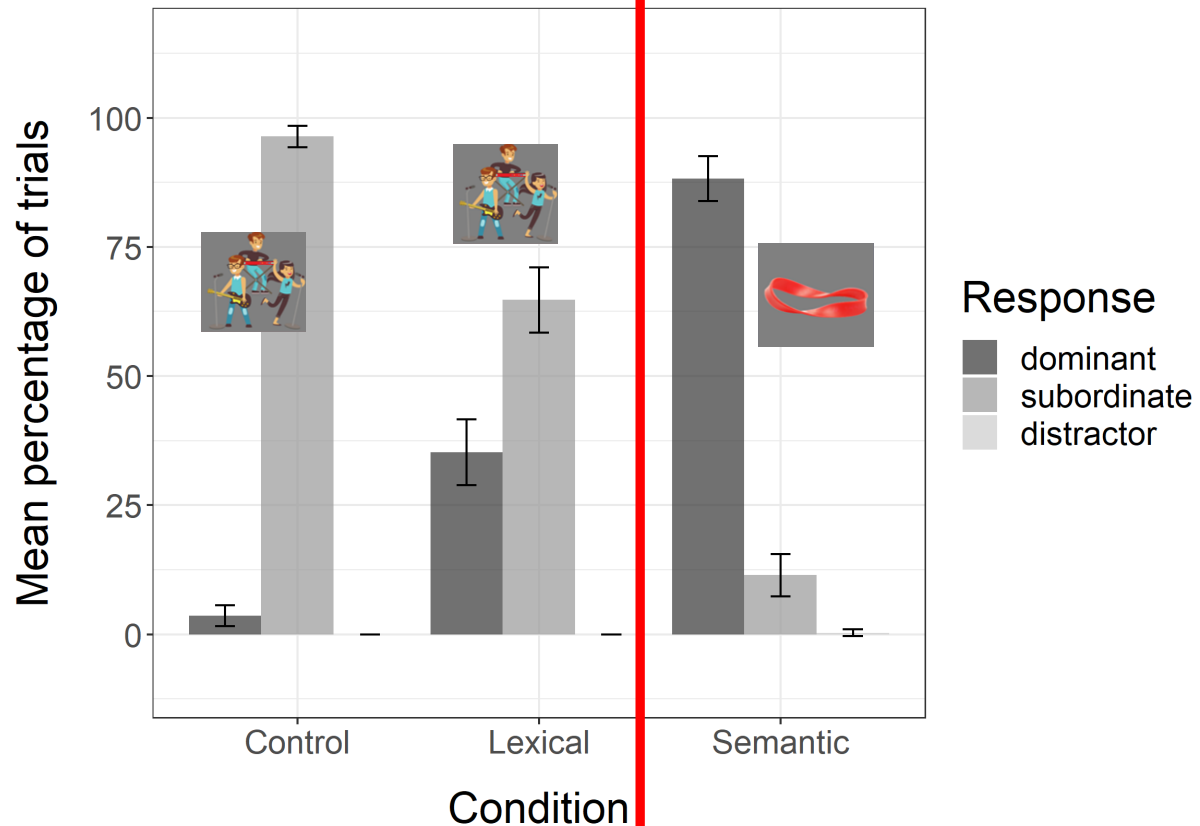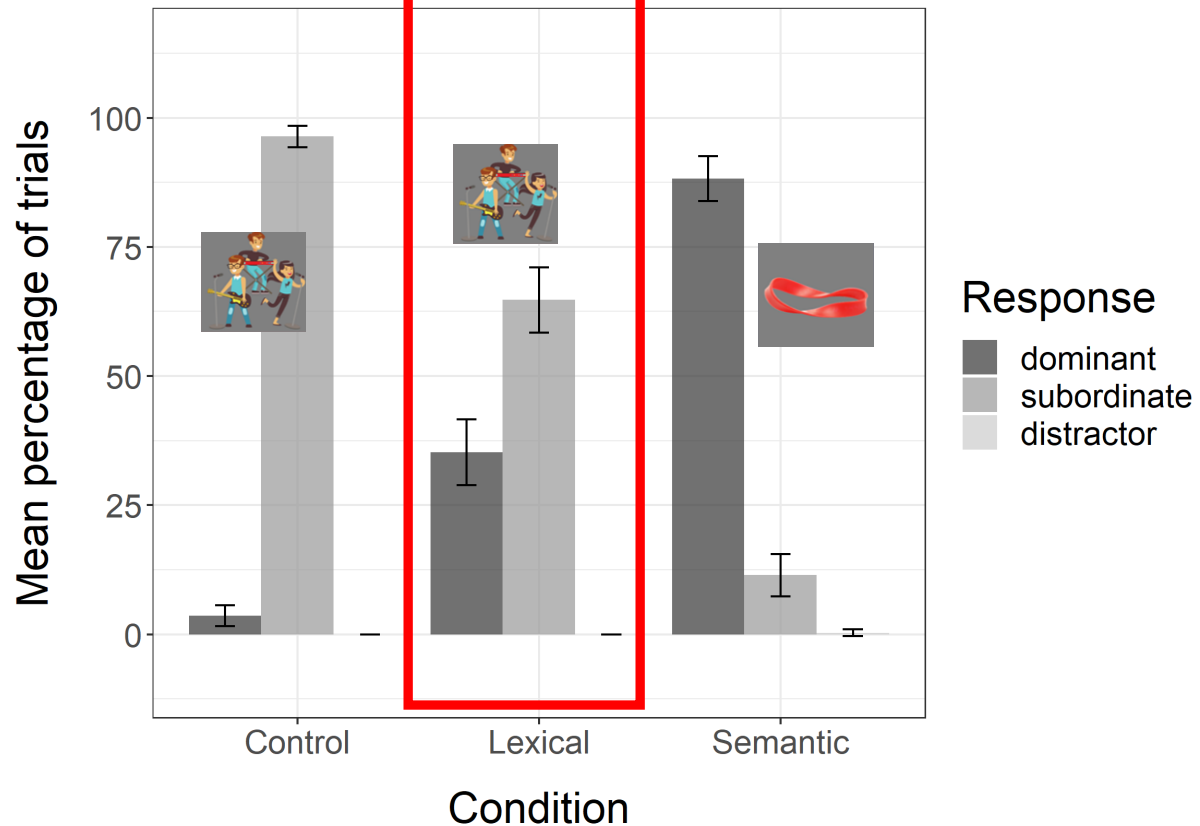
got a band

twisted a band

Sophia listened to some music. Then, she…

3a

Adult performance

Played in a band

got a band

twisted a band

Sophia listened to some music. Then, she…

Mean percentage of trials

100

75

50

25

0

Control   Lexical   Semantic

Condition

Response

dominant
subordinate
distractor

Adult performance

Played in a band

got a band

twisted a band

Sophia listened to some music. Then, she…

Child performance

Played in a band

got a band

twisted a band

Sophia listened to some music. Then, she…
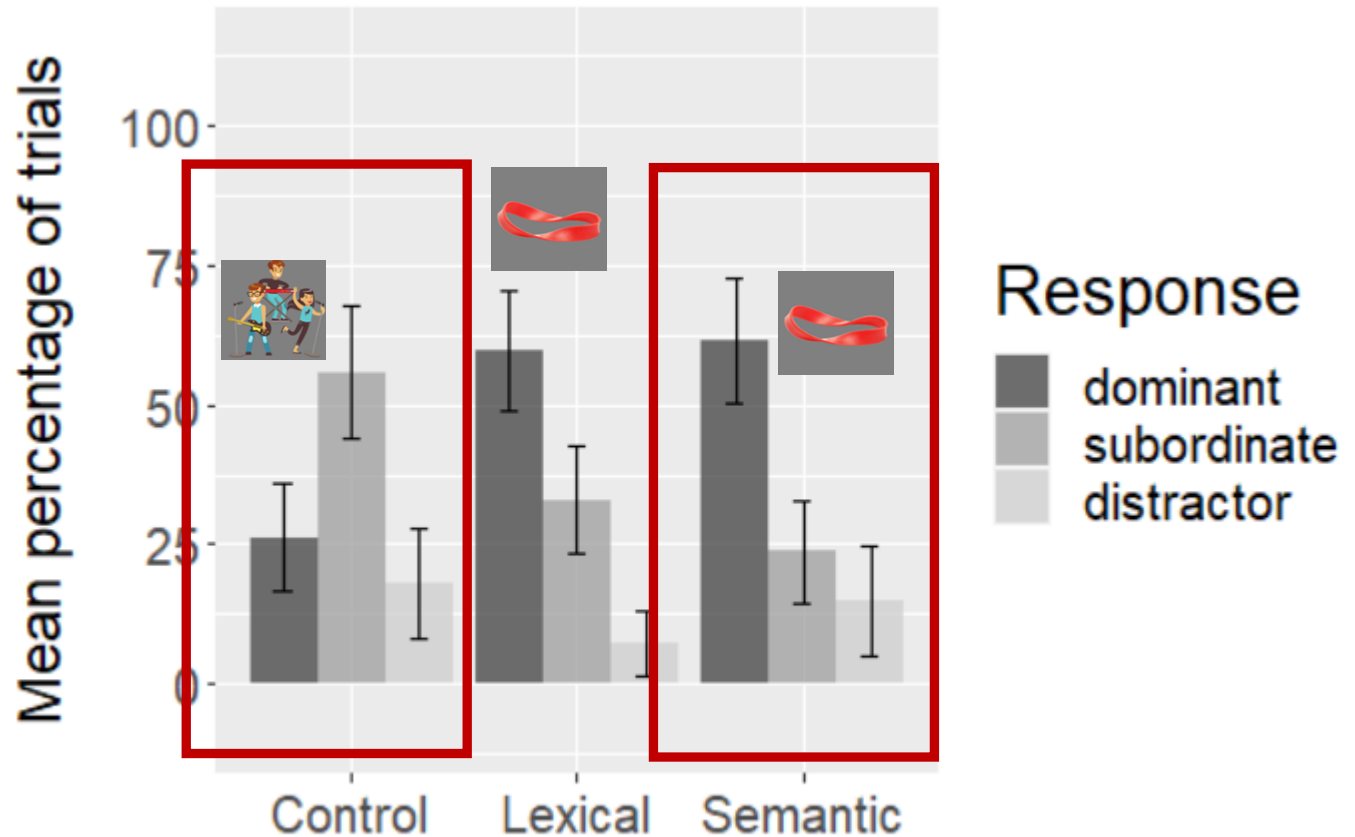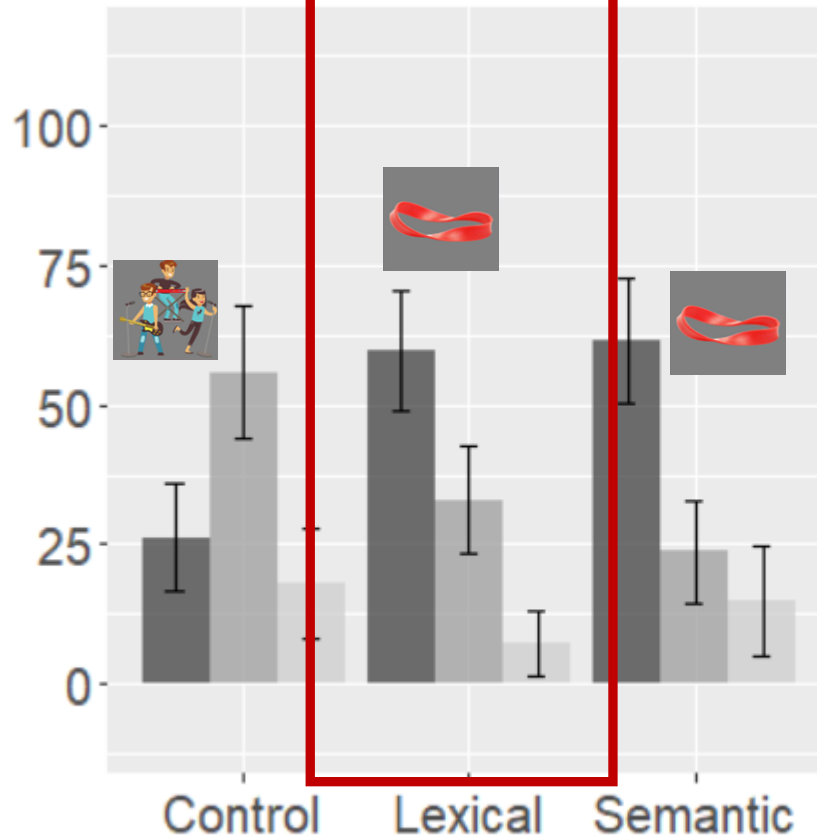
Child performance

Played in a band

got a band

twisted a band

Sophia listened to some music. Then, she…

Discussion

First evidence that verb-event structure might be a significant cue that helps children resolve lexical ambiguities, in line with a cue-validity account (Trueswell & Gleitman, 2007)

Discussion

First evidence that verb-event structure might be a significant cue that helps children resolve lexical ambiguities, in line with a cue-validity account (Trueswell & Gleitman, 2007)

Evidence of qualitative differences between adults and children.

BERT on word sense disambiguation
(Loureiro et al., 2021; 2022)

- Transformer-based models tend to dominate most WSD benchmarks

- BERT can capture aspects of adult word sense knowledge (Nair et al., 2020; Loureiro et al., 2021)

- Loureiro et al., 2021; 2022: framework for analysing and evaluating language models in word sense disambiguation for English

BERT on word sense disambiguation
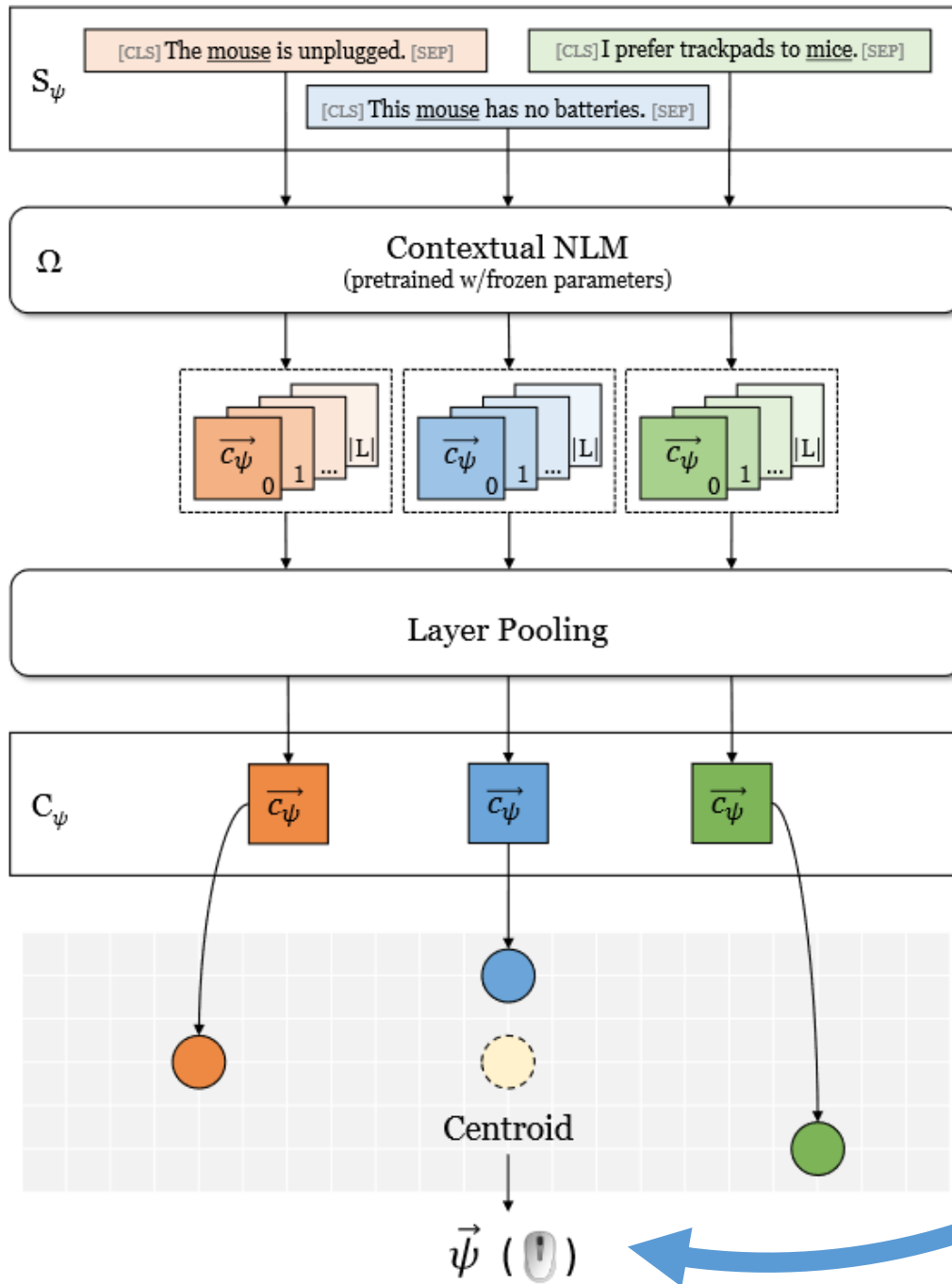**(Loureiro et al., 2021; 2022)**

- Aim: understanding to what extent pre-trained language models encode sense knowledge and their practical disambiguation capabilities.
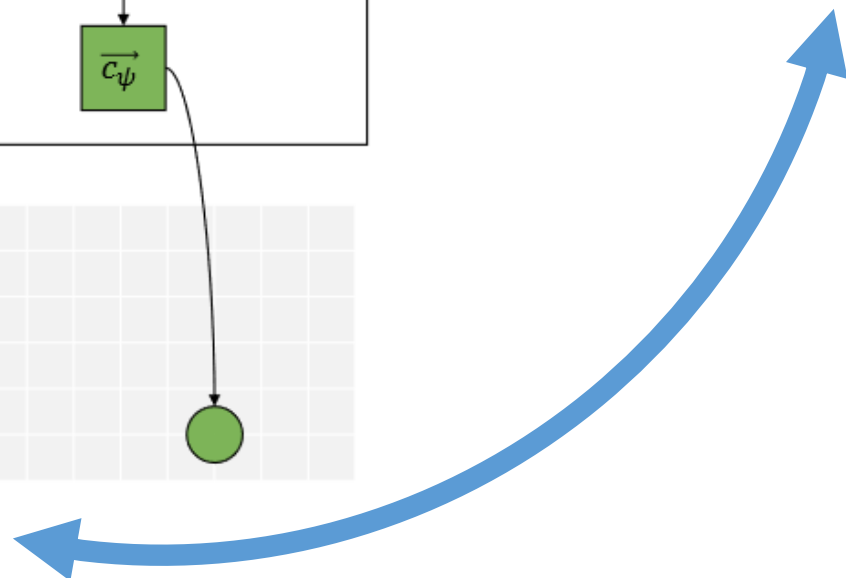
BERT on word sense disambiguation
(Loureiro et al., 2021; 2022)

- Paradigms: knowledge-based or **supervised**. The latter outperforms the former and it is based on training using sense-annotated corpora.

- Approaches:
  - fine-tuning for WSD classification.
  - nearest neighbour classifiers based on **feature extraction**:
    - It protects from Most Frequent Sense bias.
    - It works well with scarce training data (from 3 instances per sense)

(Loureiro et al., 2022)

Ernie saw a little animal on his desk. It was chewing fast on his **mouse**

BERT on word sense disambiguation
(Loureiro et al., 2021; 2022)

Sense-annotated corpora:

- datasets mostly based on WordNet inventory (Fellbaum, 1998) and using **SemCor** (Miller et al., 1993) as training dataset. [fine-grained];

- **CoarseWSD-20** (Loureiro et al., 2021): dataset of Wikipedia sentences with 20 nouns [coarse-grained];

- **ChiSense-12** (Cabiddu et al., 2022): 12 ambiguous nouns familiar to children + verb-noun tags

BERT on word sense disambiguation
(Loureiro et al., 2021; 2022)

BabyBERTa (Huebner et al., 2021):
- 8M parameters
- It can gain strong grammatical knowledge even when pre-trained on a small corpus of child-directed speech (30M words: 6000x less than RoBERTa-base)
  - American-English speech directed to 1-6-year-olds

BERT pre-trained on spoken transcripts?

Conclusion

This project is an opportunity to

- Test core aspects of a usage-based theory by exploiting recent developments in NLP

- Test (and potentially generate) hypotheses on the learning mechanisms that might underly performance in early word sense disambiguation

Thank you!

CabidduF@cardiff.ac.uk