WAX: A New Dataset for Word Association eXplanations

Chunhua Liu Trevor Cohn Simon De Deyne Lea Frermann

THE UNIVERSITY OF MELBOURNE

Word associations
When I hear the word dog, the first word that comes to my mind is ___

Motivation
1. **Why** do people associate certain words with others?
2. Existing work captures the associations, but not the **underlying reasons.**
3. We present a dataset WAX to **better understand human associations** and **improve models of common sense**.

WAX Data Collection Framework

Phase 1: Word Associations

Given the cue, which words spontaneously come to mind?

cue → *association*

bagpipe → **instrument**, **wood**, **kilt**

Phase 2: Explanations

In a short sentence, explain why you linked the cue with your response.

explanation

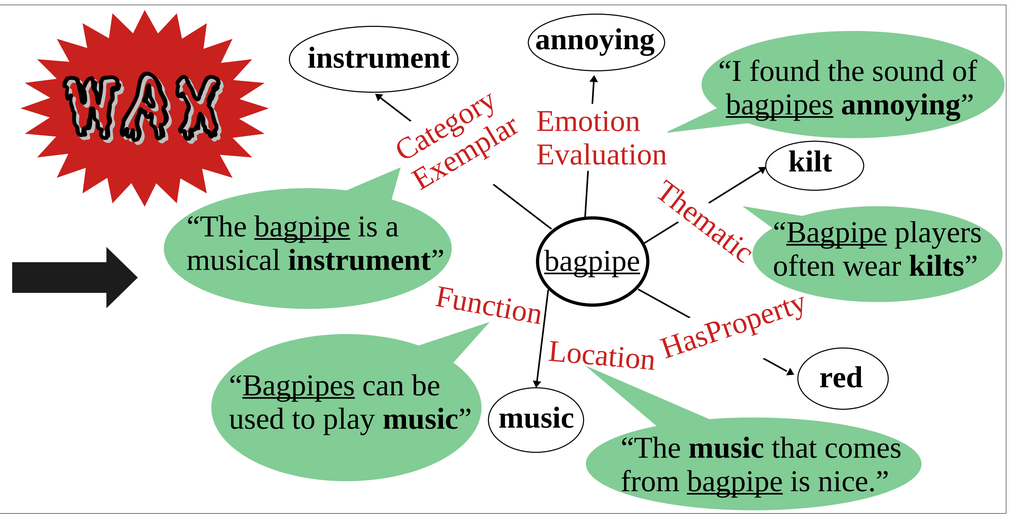
The bagpipe is a nice **instrument**
The bagpipe is made of **wood**
Men playing bagpipes wear **kilts**

Phase 3: Relations

Assign the most appropriate label to the cue-response relation expressed in the explanation.

relation

CategoryExemplar
MadeOf
Thematic

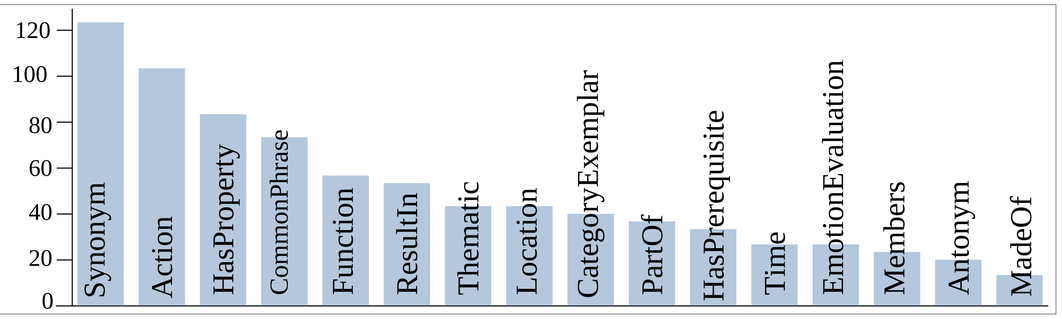


WAX consists of **19K explanations** for **1.1K cues. 1.5K** explanations are **relation-labelled**.

16 Word Association Relations

Relation distribution over the labelled set

Dominated by taxonomic relations (Synonym,..) and situational relations (Action, Function)



Association Ambiguity

WAX allows us to study whether two words are consistently associated with The same relation type (relation unambiguous; Ex1) or with different relation types (relation ambiguous; Ex 2)

Ex 1: pocket → **wallet**
- Many men put their **wallets** In their back pockets. → Location
- I put my **wallet** in my pocket. → Location

Ex 2: disaster → **hurricane**
- A **hurricane** can cause a large disaster. → ResultIn
- A **hurricane** is a good example of a natural disaster. → Category Exemplar

Evaluation 1: Explanation Generation

We fine-tuned BART on our data set to generate an explanation given a (cue, association, relation) triple.

Example generated outputs:

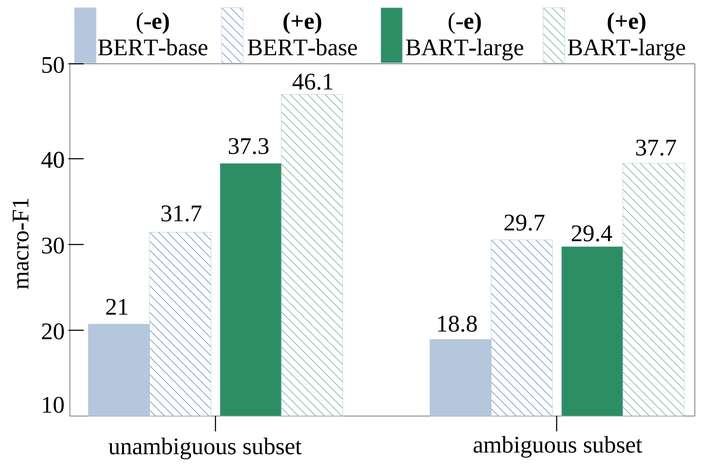
pocket, **wallet**, Location → Most men keep their **wallet** in their back pocket.

disaster, **hurricane**, ResultIn → A **hurricane** creates a major disaster when it hits the coast.

disaster, **hurricane**, CategoryExemplar → A **hurricane** is a type of disaster.

Evaluation 2: Relation Classification

Predict the most likely relation given
- a (cue, association) pair (**-e**)
- a (cue, association, explanation) tuple (**+e**)



Results:
(1) access to explanations improves performance (**+e** > **-e**)
(2) the ambiguous subset is harder to predict

Find our WAX dataset here
https://github.com/ChunhuaLiu596/WAX