



Large Language Models for Lexical Resource Enhancement: Multiple Hypernymy Resolution in WordNet



Dimitar Hristov, Institute for Bulgarian Language, Bulgarian Academy of Sciences. dimitar@dcl.bas.bg

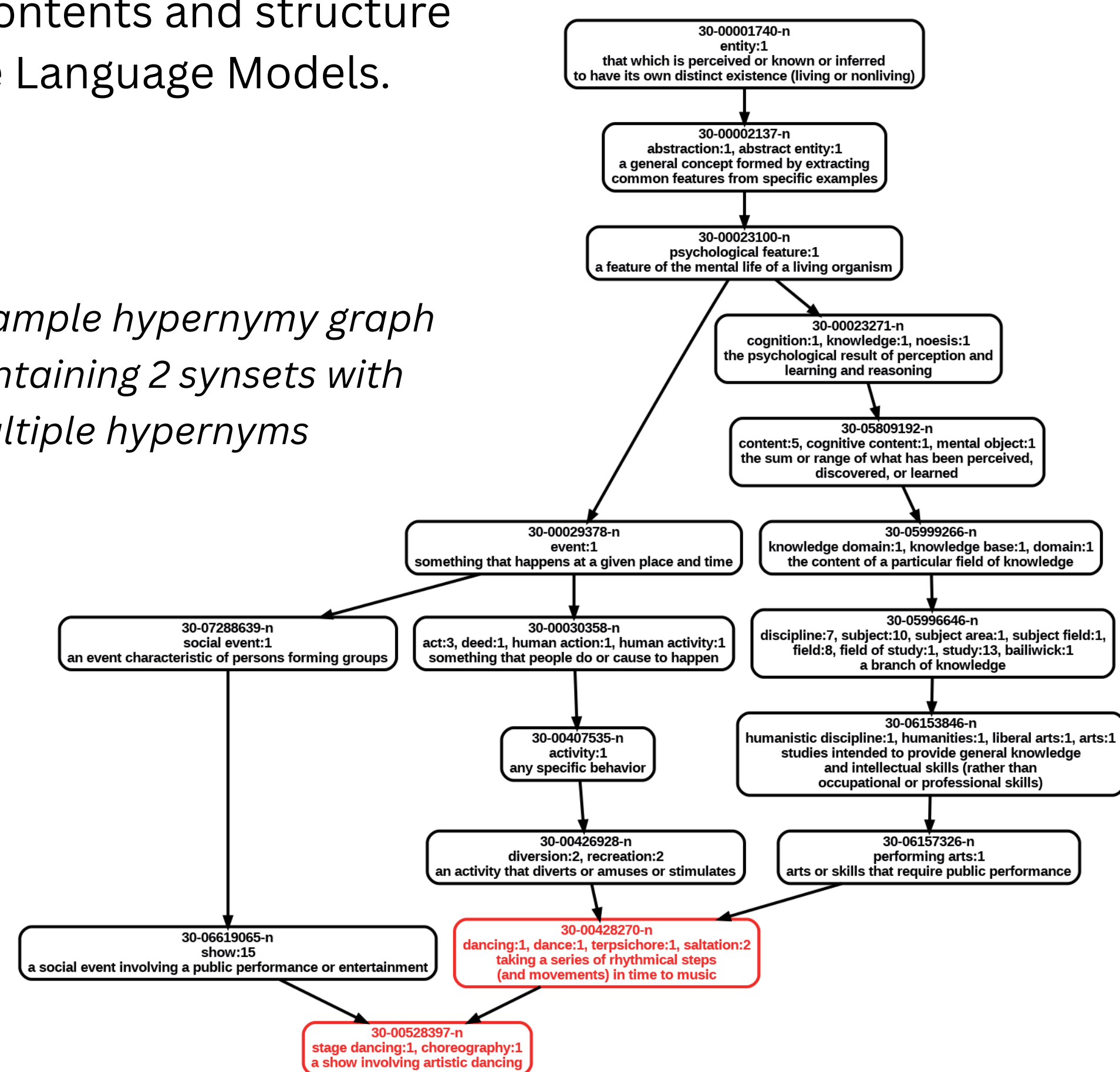
Motivation

Context: WordNet is an ontology-based structured language resource representing the interconnectedness of language concepts.

Problem: Differences of language perception, ambiguity and other factors may cause errors in both the lexical data and the structure. Multiple hypernymy may be such a phenomenon.

Task: Explore methods for evaluating and aiding improvement of the WordNet contents and structure using Large Language Models.

Example hypernymy graph containing 2 synsets with multiple hypernyms



Data

Data selection and evaluation based on manual hypernymy resolution proposal by Koeva and Hristov (2023).

- Total count of noun synsets in WordNet 3.0: **82,115**
- Synsets with multiple hypernymy in WordNet 3.0: **1,421**
- Selected one of current hypernyms in manual proposal: **1,344**
- Selected synsets as examples for 1-shot and 5-shot: **5**
 - 3 examples with 2 hypernyms (example 1 used also for 1-shot)
 - 2 examples with 3 hypernyms (used only for 5-shot)
- Remaining test set size: **1,339**

Svetla Koeva and Dimitar Hristov. *Resolving multiple hyperonymy*. GWC 2023

Evaluation

Agreement among proposals

Main measure defined as:

- LLM with manual - the ratio of synsets, for which the LLM has proposed the same hypernym as in the manual proposal.
- LLM with LLM - the ratio of synsets for which the two LLMs have assigned the same hypernym.

0-shot	Manual	Gemma 3 4B	Llama 3.1 8B	Mistral 7B	Phi-4 14B
Manual		45.4%	50.4%	50.9%	54.1%
Gemma 3 4B	45.4%		55.4%	52.7%	57.6%
Llama 3.1 8B	50.4%	55.4%		71.5%	70.9%
Mistral 7B	50.9%	52.7%	71.5%		64.3%
Phi-4 14B	54.1%	57.6%	70.9%	64.3%	

1-shot	Manual	Gemma 3 4B	Llama 3.1 8B	Mistral 7B	Phi-4 14B
Manual		53.3%	49.2%	48.7%	48.9%
Gemma 3 4B	53.3%		67.1%	65.8%	63.9%
Llama 3.1 8B	49.2%	67.1%		77.6%	76.9%
Mistral 7B	48.7%	65.8%	77.6%		76.4%
Phi-4 14B	48.9%	63.9%	76.9%	76.4%	

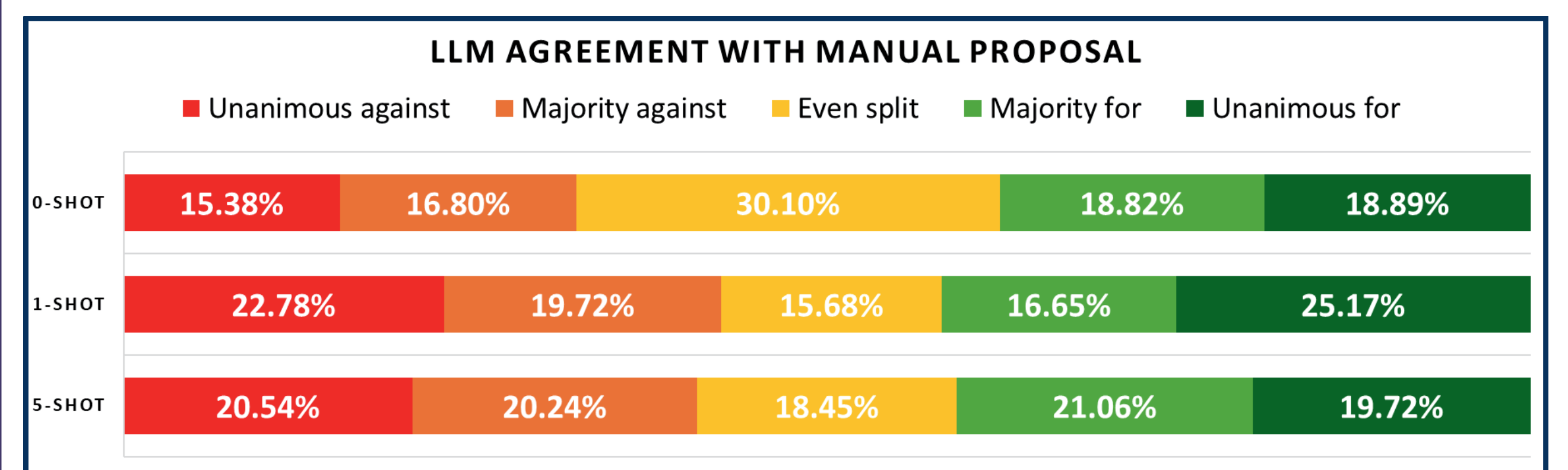
5-shot	Manual	Gemma 3 4B	Llama 3.1 8B	Mistral 7B	Phi-4 14B
Manual		51.0%	47.4%	50.1%	49.2%
Gemma 3 4B	51.0%		62.0%	63.6%	58.6%
Llama 3.1 8B	47.4%	62.0%		75.8%	71.9%
Mistral 7B	50.1%	63.6%	75.8%		69.7%
Phi-4 14B	49.2%	58.6%	71.9%	69.7%	

Agreement	M or U	Gemma 3 4B	Llama 3.1 8B	Mistral 7B	Phi-4 14B	
0-shot		69.9%	56.9%	64.4%	60.6%	62.1%
1-shot		84.3%	66.8%	79.1%	78.1%	76.9%
5-shot		81.6%	63.0%	75.7%	75.4%	70.9%

Unanimity and majority

Significance of LLM results per synset is based on agreement:

- Unanimity - all 4 LLMs agree (have proposed the same hypernym).
- Majority - 3 out of 4 LLMs agree while the last while has proposed a different hypernym.



Methodology

Base prompt

You are a WordNet expert. Your task is to evaluate hypernymy relations between semantic concepts. Each semantic concept is represented by a group of words with common meaning. This group is called a synset. If concept A is a hypernym of concept B, then concept B is a type of concept A, and concept A is a more generic version of concept B.

Each synset is presented by its ID, group of words and meaning. You will be given a synset and its hypernyms and will be instructed to choose a single hypernym.

Reply only with the chosen hypernym synset ID with format 30-<8 digits>-n and no other words. Do not give any reasoning and do not generate other text.

Task definition

You are given the following synsets:

- ID (ID_1) with words ($words_1$) and meaning ($definition_1$)
- ...
- ID (ID_n) with words ($words_n$) and meaning ($definition_n$)

Which of the synsets (ID_1)... and (ID_n) is most likely to be the hypernym of synset (ID_{hypo}) defined as:

- ID (ID_{hypo}) with words ($words_{hypo}$) and meaning ($definition_{hypo}$)

Full prompt structure

(Base prompt)

EXAMPLE [1..5]

(Example task definition)
(Example result)
...

TASK

(Main task definition)

Example definitions with EXAMPLE... TASK wrapper for 1-shot and few-shot experiments

Experiments were performed with 0, 1 and 5-shot approach on 4 LLMs.

Discussion

Possible uses:

- Starting point for hypernymy resolution based on high inter-LLM agreeability
- Evaluation of manual work with focus on unanimous and majority disagreeing with manual proposal

Extensions:

- More and more diverse LLMs, manual evaluation
- Synset grouping based on common hypernyms