

Ukrainian Visual Word Sense Disambiguation Benchmark



Yurii Laba
(Speaker)



Oles
Dobosevych



Rostyslav
Hryniv

Ukrainian Visual Word Sense Disambiguation Benchmark



Hanna
Dydyk-Meush



Halyna
Kyryleyza



Yaryna
Mohytych



Ivanna
Rogulya

Visual-WSD task definition

The main goal of the Visual-WSD task is to identify, with minimal contextual information, the most appropriate representation of a given ambiguous word (homonym) from a set of ten images.

Visual-WSD task definition

The main goal of the Visual-WSD task is to identify, with minimal contextual information, the most appropriate representation of a given ambiguous word (homonym) from a set of ten images.

Ambiguous word: Koca

1. Braid
2. An agricultural hand tool for mowing grass or harvesting crops
3. Landform
4. Spleen
5. ...



(a)



(b)



(c)



(d)

Minimal context: Волосся (en: hair)

Visual-WSD task definition

The main goal of the Visual-WSD task is to identify, with minimal contextual information, the most appropriate representation of a given ambiguous word (homonym) from a set of ten images.

Ambiguous word: Koca

1. An agricultural hand tool for mowing grass or harvesting crops
2. Landform
3. Braid
4. Spleen
5. ...



(a)



(b)



(c)



(d)

Minimal context: Волосся (en: hair)

Visual-WSD task definition

The main goal of the Visual-WSD task is to identify, with minimal contextual information, the most appropriate representation of a given ambiguous word (homonym) from a set of ten images.

Ambiguous word: Koca

1. An agricultural hand tool for mowing grass or harvesting crops
2. Landform
3. Braid
4. Spleen
5. ...



(a)



(b)



(c)



(d)

Minimal context: Волосся (en: hair)

U-VWSD benchmark overview

Each entry in the benchmark includes:

1. **A target word** - homonym in a particular sense
 - a. **Коса - Braid**
2. **One or multiple trigger words (the context)** - words that are enough to identify correct the definition of target word when considering the association between target and image
 - a. **Волосся (en. Hair)**

U-VWSD benchmark overview

Each entry in the benchmark includes:

1. **A target word** - homonym in a particular sense
 - a. Коса - Braid
 - b. Коса - An agricultural hand tool for mowing grass or harvesting crops
2. **One or multiple trigger words (the context)** - word(s) that are enough to identify correct the definition of target word when considering the association between target and image
 - a. Волосся (en. Hair)
 - b. Знаряддя праці (en. Tool)

U-VWSD benchmark overview

Each entry in the benchmark includes:

1. **A target word** - homonym in a particular sense
 - a. Koca - Braid
 - b. Koca - An agricultural hand tool for mowing grass or harvesting crops
 - c. **Koca - Landform**
2. **One or multiple trigger words (the context)** - word(s) that are enough to identify correct the definition of target word when considering the association between target and image
 - a. Волосся (en. Hair)
 - b. Знаряддя праці (en. Tool)
 - c. **Географія (en. Geography)**

U-VWSD benchmark overview

Each entry in the benchmark includes:

3. Positive Sample



U-VWSD benchmark overview

Each entry in the benchmark includes:

4. Alternative interpretations of the ambiguous target word (3 images per entry)



U-VWSD benchmark overview

Each entry in the benchmark includes:

5. Similar concepts within the domain (3 images per entry)



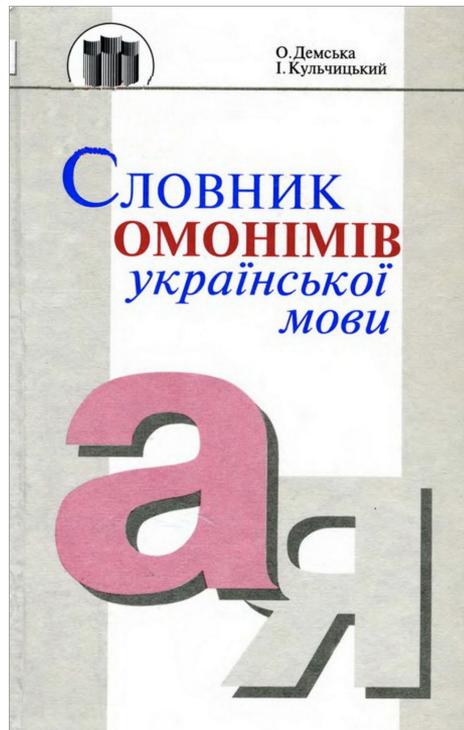
U-VWSD benchmark overview

Each entry in the benchmark includes:

6. Randomly selected concepts (3 images per entry)



The methodology for constructing the benchmark



The dictionary of homonyms of the Ukrainian language

- Used the dictionary as data source to select targets by domain experts
- The selected homonyms are nouns (to optimize the search for visual complementary material), of high usage frequency in the modern Ukrainian language

The methodology for constructing the benchmark



ВІКІПЕДІЯ
Вільна енциклопедія

Ukrainian wikipedia

- Used it as a source of positive and negative images
- Positive image is the title picture of the target from its Wikipedia article
- Negative images were collected using Wikipedia articles from the same domain as the target word

The methodology for constructing the benchmark

Сторінки в категорії «Зачіски»

Показано 49 сторінок цієї категорії (із 49).

<ul style="list-style-type: none">• Список зачісок	<ul style="list-style-type: none">• Дредлоки	
<p>*</p> <ul style="list-style-type: none">• Зачіска	З <ul style="list-style-type: none">• Шаблон:Зачіски	
А <ul style="list-style-type: none">• Айдар (чуб)• Алонж (історія костюма)• Амасунзу• Андеркат• Асиметрія (зачіска)• Афро (зачіска)	І <ul style="list-style-type: none">• Ірокез (зачіска)	
Б <ul style="list-style-type: none">• Бабетта (зачіска)• Балаяж• Біксі• Бокс (зачіска)	К <ul style="list-style-type: none">• Каре (зачіска)• Каскад (зачіска)• Кінський хвіст (зачіска)• Коса (зачіска)• Коса вінок (зачіска)• Кочівницька коса• Кучма (чуприна)	П <ul style="list-style-type: none">• Паж (зачіска)• Пасмо юності• Пейси• Перукар• Перукарня• Піксі (зачіска)• Проділ• Пучок (зачіска)
Г <ul style="list-style-type: none">• Гаврош (зачіска)• Гарсон (зачіска)	Л <ul style="list-style-type: none">• Лоб (зачіска)	С <ul style="list-style-type: none">• Свевський вузол• Сессон (зачіска)
Д <ul style="list-style-type: none">• Джерегеля• Драбинка (зачіска)	М <ul style="list-style-type: none">• Майданчик (стрижка)• Маллет (зачіска)• Міксі	Ф <ul style="list-style-type: none">• Філірування волосся
	Н <ul style="list-style-type: none">• Напівбокс• Нарощування волосся	Ц <ul style="list-style-type: none">• Цезар (зачіска)
		Ч <ul style="list-style-type: none">• Чуб (зачіска)• Чубчик
		Ш <ul style="list-style-type: none">• Шапочка (зачіска)• Шеггі (зачіска)• Шімада (зачіска)

The methodology for constructing the benchmark



ВІКІПЕДІЯ
Вільна енциклопедія

Ukrainian wikipedia

- Used it as a source of positive and negative images
- Positive image is the title picture of the target from its Wikipedia article
- Negative images were collected using Wikipedia articles from the same domain as the target word
- **We collected 40 negative samples for each definition**
- **Domain experts reviewed them and kept only the 9 most suitable images**

The methodology for constructing the benchmark

- Domain experts were asked to provide trigger words (the context)
- Trigger words were chosen to be challenging so they don't reveal the image's meaning on their own. The target word is usually needed to understand the full context.

Evaluation process

Target word: Коса + **Trigger word:** Волосся (en: hair)

List of 10 images



Evaluation process

Target word: Коса + **Trigger word:** Волосся (en: hair)

1. [0.2, -0.3, ..., 0.7]

Embedding calculation

List of 10 images



1. [0.2, -0.3, ..., 0.7]
2. [-0.15, 0.24, ..., 0.03]
...
10. [-0.32, -0.05, ..., 0.68]

Evaluation process

Target word: Коса + **Trigger word:** Волосся (en: hair)

List of 10 images



Embedding calculation

1. [0.2, -0.3, ..., 0.7]

1. [0.2, -0.3, ..., 0.7]
2. [-0.15, 0.24, ..., 0.03]
...
10. [-0.32, -0.05, ..., 0.68]

Calculate cosine similarity

1. 0.39
2. 0.60
...
10. 0.44

Evaluation process

Target word: Коса + **Trigger word:** Волосся (en: hair)

List of 10 images



Embedding calculation

1. [0.2, -0.3, ..., 0.7]

1. [0.2, -0.3, ..., 0.7]
2. [-0.15, 0.24, ..., 0.03]
...
10. [-0.32, -0.05, ..., 0.68]

Calculate cosine similarity

2. 0.60
10. 0.44
...
1. 0.39

Retrieved Image:



Evaluation metrics

1. Mean Reciprocal Rank (MRR)

$$\text{MRR} = \frac{1}{n} \sum_{i=1}^n \frac{1}{r_i} \times 100\%,$$

2. HIT@1

$$\text{HIT@1} = \frac{1}{n} \sum_{i=1}^n \text{correct}(r_i) \times 100\%,$$

Benchmark size

- At the time of evaluation, we had **87** homonyms (at least two pairs per homonym)
- Now we have **131** homonyms
- Our work continues

Evaluation results

Model	HIT@1	MRR
XLM-Roberta-Large-Vit-B-16Plus	42.78	60.30
XLM-Roberta-Large-Vit-L-14	40.21	58.65
XLM-Roberta-Large-Vit-B-32	39.69	57.69
GPT4-Vision	38.50	45.29
LABSE-Vit-L-14	35.57	54.37
clip-ViT-B-32-multilingual-v1	32.99	52.46
GCP Multimodal Embeddings	22.68	41.74
LLaVA-1.5	14.43	33.03
clip-ViT-B-32-multilingual-v1 (baseline on English language)	60.48	73.88

Yes, we tested GPT-4o

Model	HIT@1	MRR
XLM-Roberta-Large-Vit-B-16Plus	42.78	60.30
GPT-4o	42.52	43.75
XLM-Roberta-Large-Vit-L-14	40.21	58.65
XLM-Roberta-Large-Vit-B-32	39.69	57.69
GPT4-Vision	38.50	45.29
LABSE-Vit-L-14	35.57	54.37
clip-ViT-B-32-multilingual-v1	32.99	52.46
GCP Multimodal Embeddings	22.68	41.74
LLaVA-1.5	14.43	33.03
clip-ViT-B-32-multilingual-v1 (baseline on English language)	60.48	73.88

Contributions

- Introduced the benchmark for the Visual-WSD task in the Ukrainian language that integrates textual and visual modalities
 - 87 homonyms at the time of evaluation
 - 131 homonyms at this point
- Assessed various suitable Multilingual and Multimodal LLMs using the compiled benchmark



Next steps

- Enrich the homonym list by adding units that haven't been documented in Ukrainian neo-lexicography yet but have become integral to modern Ukrainian speech, whether academic or informal:
 - бот (en: bot, transl: bot) in meaning програмний агент (en: program agent, transl: prohramnyy ahent),
 - град (en: hail, transl: hrad) in meaning бойова машина (en: combat vehicle, transl: boyova mashyna)
- Integrate the compiled benchmark for the Ukrainian language into existing benchmarks for other languages
- We aim to release the benchmark along with a compiled list of homonyms and their definitions as online resources. This will include an API for accessing the materials.

Many Thanks for Your Time

QA