Image-to-Sense Alignment Using AI Tools

Andrej Perdih[1], Dejan Gabrovšek[1], Janoš Ježovnik[1]

[1]ZRC SAZU, Fran Ramovš Institute of the Slovenian Language,
Novi trg 2, 1000 Ljubljana, Slovenia
E-mail: andrej.perdih@zrc-sazu.si, dejan.gabrovsek@zrc-sazu.si, janos.jezovnik@zrc-sazu.si

Abstract

This paper evaluates the results of using GPT-4o mini language model batch processing with image recognition capability to align 1,572 images of 398 polysemous nouns in the Dictionary of the Slovenian Standard Language (second edition) to their specific dictionary senses, and it compares them to the results of the manual image-to-sense alignment process. The images were manually assigned to entries in a previous task, but no sense information was provided at the time. The language model showed relatively high overall agreement with the human annotator (i.e., 85.1%). In cases in which multiple senses were selected per image in both manual and automated annotation, the agreement was even slightly higher (i.e. in 89.4% of all sense evaluations). The agreement rate was higher when the language model evaluated only the matching senses and lower when it also evaluated the non-matching senses within the entry.

Keywords: images; lexicography; image-to-sense alignment; image recognition

1. Introduction

This pilot study evaluated the results of using the GPT-4o mini language model with image recognition capability to align images to senses in the second edition of the Dictionary of the Slovenian Standard Language (SSKJ2), focusing on noun entries. The images were linked to dictionary entries prior to this study during the compilation of the Franček Slovenian language educational portal (Perdih et al., 2021; 2024). Manual alignment to specific senses is currently an ongoing activity. To evaluate the reliability of the GPT-4o mini language model in the image-to-sense alignment process, the manual alignment results were compared against the GPT-4o mini alignment results.

The flexibility of the digital environment facilitated developments in providing images in various dictionary types. In terms of prioritizing activities in the dictionary-making process, the inclusion of images in dictionaries can be considered, as Lišková et al. (2024: 164) put it, a "nice-to-have", not a "must-have". With the increasing number of dictionaries available online, the question of including images has gained greater attention recently (Biesaga, 2016; 2017a; 2017b; Dziemianko, 2022; 2024; Kallas et al., 2024; Lew, 2010; Lew et al., 2018; Liu, 2015). The most open in this regard have been dictionaries aimed at children or foreign-language learners, but often with a limitation to a certain number of items presented (Biesaga, 2016: 100).

In building the Franček Slovenian language educational portal, a different approach was adopted: to provide images for as many dictionary entries as possible, using images already available. In the process, various types of images including drawings and photos, were selected. A total of 64,063 images were obtained from the web via the Bing search engine (license-free images) and from the Slovenian Ethnographic Museum's photo gallery. Images were manually selected for inclusion in 19,760 dictionary entries (out of approximately 100,000 entries) for the Franček portal prior to this study, but alignment to specific senses was not conducted at the time. The process was carried out by 2019, prior to the rise of AI image generation tools.

With the motivation of providing a resource for a psycholinguistic word–picture matching experiment and with the long-term motivation of improving lexicographic data, images were aligned to specific dictionary senses in SSKJ2. The dictionary was used as a source for the headword-list compilation and the meaning module of the final dictionary entries in the Franček portal database and, therefore, the compatibility between the dictionary and the portal as a whole is very high. Because one of the portal's current deficiencies is that images of several senses may be shown in random order, the process of aligning images to senses will also make it possible to sort images according to sense order[1].

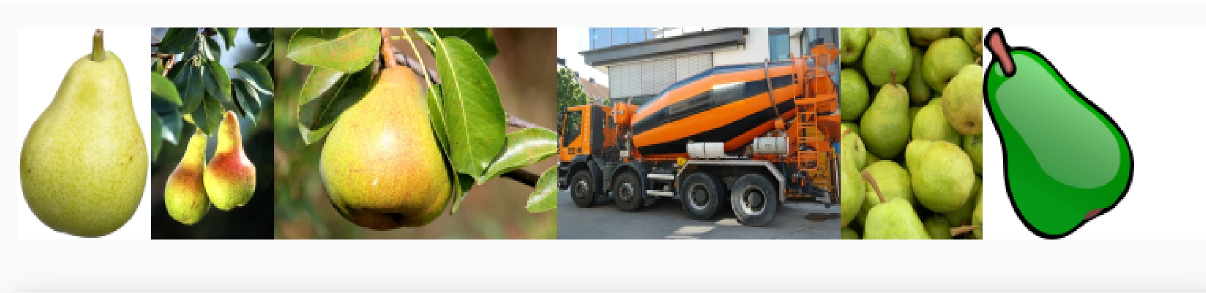

Figure 1: Images of the polysemous entry *hruška* 'pear (tree); cement truck' on the Franček portal

The use of LLMs and ChatGPT has been tested in lexicography for various purposes (cf. de Schryver, 2023; Lew, 2024; and abstracts in Krek, 2024). To conduct the image-to-sense alignment of a large number of images in polysemous entries, the use of the GPT-4o mini language model with image recognition capability was tested and evaluated against manual alignment.

2. Types of Images

On the Franček educational language portal, images serve as additional meaning representation and support understanding, but they do not replace the definitions. Various number of images are presented to the user for a headword, thus possibly providing several images per sense. This reduces the need for the "perfect" image because users can understand the semantic sense from the common features of a set of images.

[1] Another ongoing activity is selecting the most adequate image from the set of images for a specific sense.

Two types of images were deemed acceptable (cf. also Biesaga, 2017b):

1) **Strictly lexical images**: images that closely reflect word senses. This type is the most suitable for representing objects, plants, and animals, especially due to its relative unambiguity.

2) **Associative images**: images that are associatively connected to word senses. Images do not directly correspond to a word sense, but they do support its understanding (e.g., *advent* 'Advent' is represented by an Advent wreath). This approach allows graphic representation of additional groups of words and semantic fields, such as abstract nouns, verbs, adjectives, gerunds, activities, or relationships. It also serves as a workaround in cases in which strictly lexical images were not available due to copyright constrains.

Using this approach, nearly 20% of 100,000 entries were assigned at least one image. Because the focus was on obtaining a large number of images, no further information was added in the selection process.

3. Methodology

The data used comprised the SSKJ2 dictionary, the image files, and the Franček database, which included the SSKJ2 data and the relevant image file names related to each individual SSKJ2 entry. Therefore, the starting point was images linked to specific dictionary entries, but without any sense information provided for the images.

The first step was to establish links for all the nouns, whereas entries belonging to other parts of speech were to be dealt with later. The nouns were first split into two groups: polysemous and monosemous entries (based on the sense description in SSKJ2). A total of 4,499 polysemous noun entries and 9,656 monosemous noun entries were assigned at least one image.

This pilot study focused on 398 selected polysemous noun entries. The entries consisted of four groups, each group containing 100 entries (two entries were subsequently removed due to errors, cf. Footnote 2) starting with different letters: A, E–G, K, and again K. Within each group, nouns were sorted according to alphabetical order, as used in the dictionary. This simple approach was applied to gather semantically diverse nouns, thus sufficiently reducing the word-family effect.

To match each image to one of the senses in the corresponding entry, two image-to-sense activities were carried out: a manual image-to-sense alignment and an AI image-to-sense alignment.

3.1 Manual Image-to-Sense Alignment

The dictionary data are stored in XML format, but the human annotator was provided with a custom-made HTML + Javascript application. Forty-five HTML files were generated, each containing 100 polysemous noun entries (except the last file, which contained the remaining entries) with dictionary sense data and images. Each entry was presented in the form of a table with two columns, each row representing data about a single image. Images were provided in the first column. The second column contained all the sense information. Each sense was presented in a new line that started with a checkbox followed by sense number, labels, definitions, and/or synonyms.

Images in the first column were shown based on the hyperlinks established by concatenating a fixed URL part to the Franček portal and the image file name as recorded in the database. In the second column, the annotator was instructed to select the checkbox next to the corresponding sense. In rare cases, in which an image was assigned to an entry by mistake, the annotator may have selected no checkbox. Some images may have been suitable for representing several senses within the entry. In such cases, several checkboxes may have been selected.

Sense disambiguation of the images in this pilot study was conducted by a professional lexicographer with detailed understanding of the SSKJ2 dictionary and its approach to sense distinction and definitions.

3.2 AI Image-to-Sense Alignment

The same dataset comprising 1,572 images aligned with 398 unique headwords (and altogether 1,110 unique senses)[2] was processed using GPT-4o mini language model batch processing with an output limit of 300 tokens. Because the same headwords were represented with multiple images, some senses were included in multiple evaluations. The input, structured in the JSONL format, comprised prompts in Slovenian that included the image URL, the instruction, and the candidate senses to be evaluated. A prompt example is provided in Table 1[3].

Prompt segments	Description
Word: "ábak"	The headword is provided in its original

[2] The set included 252 unique headwords with 2 candidate senses, 85 headwords with 3 senses, 45 headwords with 4 senses, 24 headwords with 5 senses, 11 headwords with 6 senses, 6 headwords with 7 senses, 4 headwords with 8 senses, 5 headwords with 9 senses, 2 headwords with 10 senses, and 1 headword with 11 senses. Two headwords were excluded from the dataset prior to processing (the headwords were monosemous, but their subheadwords' senses were mistakenly counted in), bringing the number in the input to the 398 headwords mentioned above, down from originally 400.

[3] The originally one-line Slovenian prompt has been translated by the authors, JSONL formatting has been removed and the prompt is broken down with intermittent comments by the authors.

	dictionary form with stress marks included where present.
Analyse the image at "image_url"	The *image_url* variable was provided as a part of the JSONL prompt and pointed to the relevant image, hosted on the Franček portal.
and determine which of the word's given senses the image represents. Keep in mind that the image can represent a single sense, several senses, or none of them.	This part of the prompt was added because the model was found to undergenerate during initial testing.
Also provide an estimate of certainty ranging from 1 (least certain) to 5 (most certain) and a short argumentation for the answer.	Description of the expected results.
In images showing people, do not attempt to identify persons but instead determine which of the given senses the image represents.	This part of the prompt was added because the model refused to analyse images that represented people, explicitly stating that it should not be used to identify people.
Return the answer in the form of a table with the columns custom_id, senseID, senseDef, certainty_estimate, argumentation.	Formatting instruction.
Choose between the following senses: 000013: 1 [with Ancient Greeks and Romans] a plate for mechanical calculation 000014: 2 [art] a covering slab at the top of the capital of a column	Content of the original dictionary definitions including sense IDs, sense numbering, labels, and so on, with minimal formatting changes.

Table 1: Prompt segments and description

4. Results and Discussion

Before analysis, results for 26 pairs of images (i.e., 52 images with 26 their corresponding headwords, altogether comprising 118 senses) were removed from the output because they turned out to be duplicates. The final output thus consisted of the results for 1,520 images, each linked to one headword (398 unique headwords) and with the assessment of one or more dictionary senses (4,260 non-unique senses). The results discussed below do not include the data eliminated.

Based on the results of manual annotation, the expected output was one sense (rarely two[4]) per image with an added estimate of certainty and a short argumentation. However, the model only returned single-sense responses for 357 prompts (all with the high certainty estimates of 5 for 347 responses or 4 for 10 responses); the other 1,163 responses included rankings of all senses provided, where it appears that the model had provided an ordinal adequacy rating instead of a certainty estimate due to misinterpreting the instructions. This is most evident in the argumentation provided (e.g., with certainty_estimate of "1": "No element fitting this definition is visible in the image"; "The image does not include X, so this sense is less likely"; "There are no signs to indicate X"). Examples such as these suggest that the number preceding the argumentation does not represent a certainty estimate, but rather a numerical evaluation of adequacy. Most prompts thus resulted in a table of such rankings for all input senses per image. Because of this, the processing was repeated with clearer instructions. This did result in an output closer to the one expected but, unfortunately, the output format was extremely inconsistent, which greatly inhibited further analysis. That is why the decision was nevertheless made to analyse the results of the first processing, fully realizing that continuing with this approach might result in methodological hindrances due to the two subsets following different manners of rating: nominal (i.e., providing a discrete assessment of individual senses as either adequate or inadequate), vs. ordinal (i.e., providing an assessment of an individual sense's adequacy on a scale from 1 to 5).

The dataset thus comprised two subsets: 357 "best candidate" senses (and, implicitly by omission, 745 inadequate senses) and 1,163 sets of multiple evaluated senses or, in other words, 1,102 (either adequate or inadequate) senses as nominal values and 3,158 senses rated with ordinal values ranging from 1 to 5. To normalize the subsets and to allow for a subsequent comparison with the dataset produced with manual annotation, it was decided to conservatively treat the adequacy values of 4 and 5 in the ordinally ranked subset as adequate senses, and the values of 3 and lower as inadequate senses.

The overall agreement across the combined dataset is presented in Table 2. The automated alignment resulted in 1,559 true positives (36.6%), 2,066 true negatives (48.5%; for a combined agreement in 3,625 alignments or 85.1%), 372 false positives (8.7%), and 263 false negatives (6.2%; for a combined disagreement in 635 alignments or 14.9%). The value of Cohen's kappa (κ) calculated was 0.70, which indicated substantial agreement between manual and automated alignment. Overall precision was 0.807, overall recall was 0.856, and the F1 score was 0.831.

[4] This is mostly the case with metonymic relation between senses.

Automated alignment Manual alignment	Aligned senses (n)	Aligned senses (%)	Non-aligned senses (n)	Non-aligned senses (%)
Aligned senses	1,559	36.6%	263	6.2%
Non-aligned senses	372	8.7%	2,066	48.5%

Table 2: Overall agreement between the manual and automated alignment of senses within the entire dataset (n = 4,260, κ = 0.70, precision = 0.807, recall = 0.856, F1 = 0.831)

A slightly more detailed analysis by each subset is described below.

Looking first at the nominally rated subset, it can be concluded that the model only opted for such responses in cases in which it estimated its certainty as high (mostly 5 or, rarely, 4, as previously noted). Comparing this dataset to the manually annotated set (see Table 3), it can be established that the model correctly marked 331 senses as adequate (30%), while flagging 26 (2.4%) and 39 (3.5%) as false positives and false negatives, respectively. The high ratio of true negatives (64.1%), along with the resulting high level of cumulative agreement (94.1%, κ = 0.87), should not come as surprising given that it was inferred from what the model chose to omit but should nevertheless be taken with a grain of salt because the reasons for the omission are not completely clear. Keeping in mind these limitations, the precision was 0.927, recall was 0.895, and the F1 score was 0.911.

Automated alignment Manual alignment	Aligned senses (n)	Aligned senses (%)	Non-aligned senses (n)	Non-aligned senses (%)
Aligned senses	331	30%	39	3.5%
Non-aligned senses	26	2.4%	706	64.1

Table 3: Agreement between the manual and automated alignment of senses for the nominally rated subset (n = 1,102, κ = 0.87, precision = 0.927, recall = 0.895, F1 = 0.911)

Even though the sets rated 5 and 4 are too disproportionate to be adequately compared, it might still be worth noting that, out of the 10 images for which the certainty estimate for the chosen sense was rated as 4, agreement between manual and automated alignment was 88% (29 senses, including those inferred), whereas the agreement between manual and automated alignment for the 347 images with a certainty estimate of 5 was 94.3% (in 1,008 senses, including those inferred). This seems to correlate with the certainty estimates.

Taking a look at the ordinally ranked subset (see Table 4), it can be established that the model shows a strong preference for extremes because most senses were marked as either fully adequate (5; 34.6%) or fully inadequate (1; 28.3%), suggesting a high degree of confidence in its judgments. This tendency is further supported by the overall distribution of adequacy ratings, which has a mean of approximately 3.16 and a relatively high standard deviation of 1.66. Even though these ratings are ordinal and not ideally suited for parametric analysis, the numerical summaries offer informative insights into the model's behaviour. The mean indicates a general leaning toward higher ratings. However, the high standard deviation and the observed bimodal distribution point to a polarized pattern: ratings tend to cluster at the extremes, with relatively few intermediate values. This suggests that the model frequently makes categorical rather than nuanced judgments of semantic alignment, a pattern that may stem from strong internal confidence or task constraints that discourage equivocal responses.

Adequacy	**No. of senses**	**% of senses**
5	1,088	34.45%
4	486	15.39%
3	311	9.85%
2	378	11.97%
1	895	28.34%
TOTAL	3,165	100%

Table 4: Distribution of adequacy values within the ordinally rated subset ($\bar{x} = 3.16$, SD = 1.66)

As previously noted, to rate the agreement between manual and automated alignment within the ordinally rated subset, it was decided to normalize it by treating the senses rated 4 or 5 as adequate and the senses rated 3, 2, and 1 as inadequate. The resulting agreement is presented in Table 5.

Automated alignment Manual alignment	Aligned senses (n)	Aligned senses (%)	Non-aligned senses (n)	Non-aligned senses (%)
Aligned senses	1,228	38.9%	224	7.1%
Non-aligned senses	346	11.0%	1,360	43.0%

Table 5: Agreement between the manual automated alignment of senses for the ordinally rated subset (n = 3,158, κ = 0.87, precision = 0.78, recall = 0.846, F1 = 0.812)

The model's success rate with the ordinally rated subset was noticeably lower compared to the nominally rated subset, resulting in 1,228 true positives (38.9%) and 1,360 true negatives (43%) for a total agreement of 82%, and 346 false positives (11%) and 224 false negatives (7.1%) for a total disagreement of 18%. Precision was significantly lower (0.78) and recall was slightly lower (0.846) for a combined F1 score of 0.812.

It is also worth noting that although the model was instructed that none of the potential senses might be relevant to a given image, it nevertheless always chose to evaluate at least one of the candidate senses as adequate. In contrast, the manual annotator excluded all candidate senses in 77 images. On the other hand, the human annotator selected more than one candidate sense for 345 images (choosing between 908 candidate senses in total, with 2.6 senses per image on average) and the model did so for 420 images (choosing between 1,198 candidate senses in total, with 2.9 senses per image on average). There was an overlap of 206 images in these two groups (for a total of 556 candidate senses, or on average 2.7 senses per image). The agreement between manual and automated alignment for these overlapping cases is presented in Table 6. Both the manual and automated alignment agreed in 89.4% of all sense evaluations (74.3% true positives and 15.1% true negatives) and disagreed in 10.7% of assessments (5.8% false positives and 4.9% false negatives). The corresponding Cohen's kappa of 0.67 indicates substantial agreement between the human and model evaluations, as do the values of precision (0.928) and recall (0.939), and the F1 score (0.933).

Automated alignment Manual alignment	Aligned senses (n)	Aligned senses (%)	Non-aligned senses (n)	Non-aligned senses (%)
Aligned senses	413	74.3%	27	4.9%
Non-aligned senses	32	5.8%	84	15.1%

Table 6: Agreement between the manual and automated alignment of senses for images in which both manual and automated alignment marked more than one sense ($n = 556$, $\kappa = 0.67$)

4.1 Limitations

A possible limitation of the study lies in the fact that the senses were rated by only one human annotator (Annotator 1), as also pointed out by one of the anonymous reviewers. After the initial study had been conducted, the whole dataset was subsequently annotated by two additional human annotators (Annotator 2, Annotator 3). The results are presented in Table 7.

Annotator 1	Annotator 2	Annotator 3	No.
0	0	0	2,121 (49.8%)
0	0	1	104 (2.4%)
0	1	0	86 (2%)
0	1	1	127 (3%)
1	0	0	133 (3.1%)
1	0	1	139 (3.3%)
1	1	0	116 (2.7%)
1	1	1	1,434 (33.7%)

Table 7: Contingency table representing the agreement between all three human annotators (0 = image does not correspond to the given sense, 1 = image corresponds to the given sense, $n = 4{,}260$, $\kappa_F = 0.77$)

Complete agreement among all three human annotators was achieved for 83.5% of all items (i.e., all three annotators agreed that a certain sense pertained to a given image in 33.7% of cases, or that a certain sense did not pertain to a given image in 49.8% cases). We further assessed inter-annotator agreement using Fleiss' kappa (κ_F), which yielded a value of 0.77, indicating good agreement between annotators.

Comparing the two secondary annotators' alignments to those produced by the model, we find that, for Annotator 2, automated alignment produced 1,542 true positives (36.2%), 2,108 true negatives (49.5%) for a combined agreement of 3,650 alignments or 85.7%, along with 389 false positives (9.1%), and 221 false negatives (5.2%), resulting in a combined disagreement in 610 alignments or 14.3%; the corresponding Cohen's kappa value was 0.71. For Annotator 3, the automated alignment resulted in 1,588 true positives (37.3%), 2,113 true negatives (49.6%) for a combined agreement of 3,701 alignments or 86.9%, and 343 false positives (8.1%), and 216 false negatives (5.1%) for a total disagreement in 559 alignments or 13.1%; the corresponding Cohen's kappa value was 0.73.

A comparison of all four datasets – the manually aligned senses rated by the three human annotators and the automated alignments produced by the model – shows that the overall agreement (positive agreement in 30.8% and negative agreement in 45.9% of all cases, for a total of 76.7%) is somewhat lower than that achieved in comparison with individual manual alignments. This can, of course, be attributed to the differences in the choices made by the manual annotators. Nevertheless, the Fleiss' kappa value of 0.74 still indicates strong overall agreement.

An empirical analysis of the senses where the human annotators' assessments diverged, revealed that the most significant discrepancies occurred in the labelling of metonymic senses. For example, in the case of the entry *akvarel* ('aquarelle', 'watercolour'), the question arises whether the images represent solely 'a painting technique using transparent water-based colours on paper' or also 'a painting created in this technique' – or vice versa. This set of metonymies can most readily be resolved by convention, i.e., by deciding beforehand whether such metonymic meanings are also to be annotated, specifically as "associative images" (cf. Section 2). Another source of discrepancy stems from differing interpretations of what a particular image actually depicts, particularly in cases where the depicted content is not unambiguously identifiable or involves rather abstract senses. The third source of variation, naturally, involves individual annotation errors.

The success rate of the automated image-to-sense alignment may also have been influenced by several factors, including the use of Slovenian for both prompts and sense definitions, the focus on nouns, and the fact that the images had been pre-selected for the Franček portal in a prior project, meaning that only suitable images were included in the evaluation.

5. Conclusion

The comparison between the manual image-to-sense alignment results and the GPT-4o mini language model results revealed a high success rate of the language model's image recognition and sense interpretation. The automated process achieved an overall agreement of 85.1% with the human annotator, with a Cohen's kappa (κ) value of 0.70, which indicates substantial agreement between manual and automated alignment. Subsequent comparison with two additional manually aligned datasets resulted in a comparable total agreement in 85.7% and 86.9% of all items with Cohen's kappa values of 0.71 and 0.73, respectively. The differences between human annotators stemmed mostly from differing decisions as pertaining to metonymic senses.

6. Acknowledgements

This article was written as part of the research project Slovenian Word Prevalence: An Online Mega-Study of Word Knowledge (code J6-50199) and the program Slovene Language in Synchronous and Diachronic Development (code P6-0038) funded by the Slovenian Research and Innovation Agency (ARIS).

7. References

Biesaga, M. (2016). Pictorial Illustration in Dictionaries: The State of Theoretical Art. In T. Margalitadze & G. Meladze (eds.) *Proceedings of the 17th EURALEX International Congress, EURALEX 2016*. Tbilisi: Lexicographic Centre, Ivane Javakhishvili Tbilisi State University, pp. 99–108. http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202016/euralex_2016_007_p99.pdf.

Biesaga, M. (2017a). Pictorial Illustrations in Encyclopaedias and in Dictionaries – a Comparison. In I. Kosem, C. Tiberius, M. Jakubíček, J. Kallas, S. Krek, & V. Baisa (eds.) *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*. Lexical Computing CZ, pp. 221–236. Available at: https://elex.link/elex2017/wp-content/uploads/2017/09/paper13.pdf.

Biesaga, M. (2017b). Dictionary Tradition vs. Pictorial Corpora: Which Vocabulary Thematic Fields Should Be Illustrated? *Lexikos*, 27, pp. 132–151. Available at: https://doi.org/10.5788/27-1-1397.

De Schryver, G.-M. (2023). Generative AI and Lexicography: The Current State of the Art Using ChatGPT. *International Journal of Lexicography*, 36(4), pp. 355–387. Available at: https://doi.org/10.1093/ijl/ecad021.

Dziemianko, A. (2022). The usefulness of graphic illustrations in online dictionaries. *ReCALL*, 34(2), pp. 218–234. Available at: https://doi.org/10.1017/S0958344021000264.

Dziemianko, A. (2024). Pictures in Online Dictionaries: Shall We See Them? *GEMA Online® Journal of Language Studies*, 24(2), pp. 31–53. Available at: https://doi.org/10.17576/gema-2024-2402-03.

Franček. Accessed at: https://www.franček.si/ (7 July 2025)

Kallas, J., Koppel, K., & Tsepelina, K. (2024). The EKI Picture Dictionary. A Multilingual Tool for A1–B1 Learners. In K. Š. Despot, A. Ostroški Anić, & I. Brač (eds.) *Lexicography and Semantics. Book of Abstracts of the XXXI EURALEX International Congress. Institut za hrvatski jezik*, pp. 163–165. Available at: https://euralex.jezik.hr/wp-content/uploads/2021/09/Euralex_boa_20.pdf.

Krek, S. (ed.). (2024). *Book of Abstracts of the Workshop Large Language Models and Lexicography.* [Pre-publication]. Available at: https://www.cjvt.si/wp-content/uploads/2024/10/LLM-Lex_2024_Book-of-Abstracts.pdf.

Lew, R. (2010). Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries. *Lexikos*, 20, pp. 290–306. https://doi.org/10.5788/20-0-144.

Lew, R., Kaźmierczak, R., Tomczak, E., & Leszkowicz, M. (2018). Competition of Definition and Pictorial Illustration for Dictionary Users' Attention: An Eye-Tracking Study. *International Journal of Lexicography*, 31(1), pp. 53–77. Available at: https://doi.org/10.1093/ijl/ecx002.

Lew, R. (2024). Dictionaries and lexicography in the AI era. *Humanities and Social Sciences Communications*, 11(1). https://doi.org/10.1057/s41599-024-02889-7.

Lišková, M., & Šemelík, M. (2024). Show me the meaning of being lonely... Graphic Illustrations in The Academic Dictionary of Contemporary Czech. In K. Š. Despot, A. Ostroški Anić, & I. Brač (eds.) *Lexicography and Semantics. Book of Abstracts of the XXXI EURALEX International Congress. Institut za hrvatski jezik*, pp. 163–165. Available at: https://euralex.jezik.hr/wp-content/uploads/2021/09/Euralex_boa_20.pdf.

Liu, X. (2015). Multimodal Definition: The Multiplication of Meaning in Electronic Dictionaries. *Lexikos*, 25, pp. 210–232. Available at: https://doi.org/10.5788/25-1-1296.

Perdih, A., Ahačič, K., Jakop, N., Ledinek, N., & Petric Žižić, Š. (2024). Semantic Information on the Franček Educational Language Portal for Slovenian. In K. Š. Despot, A. Ostroški Anić, & I. Brač (eds.) *Lexicography and Semantics. Proceedings of the XXI EURALEX International Congress. Institut za hrvatski jezik*, pp. 155–168. Available at: https://euralex.jezik.hr/wp-content/uploads/2021/09/Euralax-XXI-final-web.pdf.

Perdih, A., Ahačič, K., Ježovnik, J., & Race, D. (2021). Building an Educational Language Portal Using Existing Dictionary Data. *Journal of Linguistics/Jazykovedný časopis*, 72(2), pp. 568–578. Available at: https://doi.org/10.2478/jazcas-2021-0052.

SSKJ2: *Slovar slovenskega knjižnega jezika.* (2014). 2nd edition. Ljubljana: Inštitut za slovenski jezik Frana Ramovša ZRC SAZU.