

Cross-level semantic similarity annotation guidelines¹

Two texts can have a similar meaning even if the actual words and phrases they contain are different. In contrast, two texts can have very different meanings despite initially appearing very similar in the choice of words and phrases, as well as overall structure.

Your task is to compare two texts of different lengths and establish the kind of relationship that exists between their meanings, i.e. the messages they convey (or, in other words, between what they say about, or what they refer to in the real world). You will work on text pairs of the following lengths:

- Paragraph – sentence
- Sentence – phrase

In order to complete the task successfully, please imagine the situations described by the texts and compare the content of the first text to the content of the second text.

Do the texts refer to the same person, action, event, idea or thing? Are the texts similar to each other, but at the same time different in terms of minor or major details?

Some tips:

- Try to be precise in assigning similarity scores, and try not to over-rely on a single score (e.g., do not mark most pairs as “somewhat similar” or “somewhat related but not similar”).
- Pay attention to subtle differences between texts that might have an important influence on what is being described or said.
- Ignore any grammar errors or inadequate formulations, as long as they do not obfuscate the meaning conveyed by the text.

Scores and examples

When annotating semantic similarity, you should use a Likert scale ranging from 0 to 4, with the following meaning of the individual scores:²

- 4 – very similar – the two items have very similar meanings and the most important ideas, concepts, or actions in the larger text are represented in the smaller text. Some less important information may be missing, but the smaller text is a very good summary of the larger text
- 3 – somewhat similar – the two items share many of the same important ideas, concepts, or actions, but include slightly different details. The smaller text may use similar but not identical concepts, or may omit a few of the more important ideas present in the larger text
- 2 – somewhat related but not similar – the two items have dissimilar meanings, but share concepts, ideas, and actions that are related. The smaller text may use related but not necessarily similar concepts but should still share some overlapping concepts, ideas, or actions with the larger text
- 1 – slightly related – the two items describe dissimilar concepts, ideas and actions, but may share some small details or domain in common and might be likely to be found together in a longer document on the same topic
- 0 – unrelated – the two items do not mean the same thing and are not on the same topic

Three examples to be used as reference are provided for each mark and for each set of text lengths.

Sentence - paragraph

- (4)
 - The organizers set a 50 percent audience capacity limit at all Games events; access is forbidden to supporters from abroad.
 - The number of supporters at sporting events is limited, but at least some will be there. A limited number of local supporters will be allowed to attend the Tokyo Olympic Games, the organizers announced today. The organizers set a limit of 50 percent capacity, i.e. a maximum of 10,000 spectators at all Games events.
 - The Mayor of Belgrade Zoran Radojičić stated that all 193 public fountains in the capital are working, and that due to the 40 degrees forecast and the "red weather alert" issued today, mobile drinking water tanks have been positioned in the busiest places.
 - Zoran Radojičić, the mayor of Belgrade, addressed the Belgraders. At the same time, he stated that the citizens should not worry about drinking water, because there is a mobile drinking water tank positioned at every corner. - In Belgrade, we have 193 public fountains where Belgraders can refresh themselves, and we have positioned mobile drinking water tanks in the busiest places to make these hot days easier for our fellow citizens, said the mayor of Belgrade, Prof. Zoran Radojičić for TV Prva.
 - The company Nvidia has announced that as of October it will discontinue the support for Windows 7, Windows 8 and Windows 8.1 drivers, as it plans to focus on Windows 10 and more recent versions of the operating system.
 - Starting October 2021, Nvidia's GPU drivers will only be supported on Windows 10, while the older Windows versions will only receive critical security updates. Currently, Nvidia offers its GPU drivers for Windows 7, Windows 8, Windows 8.1 and Windows 10. According to the company, most users have already migrated to Windows 10, and since the development of driver updates takes too much time, the company decided to discontinue its support for older versions of Windows.
- (3)
 - The Italian coach Andrea Trinchieri will continue to coach Bayern Munich.
 - Andrea Trinchieri left the bench of Belgrade's Partizan last year and moved to Germany, where he took over Bayern Munich. He had a great season with them and almost led the team to the Final Four, more precisely to the final EuroLeague tournament. According to Emiliano Carchia, a well-informed Italian journalist, Andrea Trinchieri will continue his career at Bayern, having signed a new contract. The club is expected to officially announce this soon.
 - The singer Dragica Radosavljević Cakana found herself in an awkward situation when, as she announced, her phone was hacked.
 - Cakana informed her friends and colleagues via social media to be careful if they receive a suspicious message from her number. She also stated having reported the case to the police. - I want to inform all of my contacts and friends that my phone number has been hacked, and to contact me if they receive any inappropriate content via Viber, WhatsApp or SMS, from my number or another number under my name.
 - The Czech Republic seeks compensation from Russia

- Prague asked Moscow for full compensation for the damage after the explosions in the military warehouse in the village of Vrbetice in 2014. Deputy Foreign Affairs Minister Martin Smolek handed the corresponding note to the Russian Ambassador to Prague, Alexander Zmejevski, reports "Mladá fronta Dnes". The Czech authorities have estimated the damage at 650 million crowns, or 25.5 million euros.
- (2)
 - The Finance Minister Siniša Mali participated in the closing ceremony of the 15th Annual and 8th Electoral Assembly of NALED entitled "Agenda 2025: A Vision for the Future".
 - At the 15th Annual and 8th Electoral Assembly of the National Alliance for Local Economic Development (NALED), Mali said that Serbia's public debt is still under control and will remain below 60 percent by the end of the year. "The growth rate will definitely be above six percent, and this rate is currently the biggest challenge for all of the world's economies, where and how to find resources for growth despite the crisis and stagnation," the minister pointed out.
 - The Mayor of Pirot Vladan Vasić, MA, together with his associates, deputy Miloš Colić and assistant Milica Golubović, received today in his office the Ambassador of the Kingdom of Belgium to Serbia Koen Adam, with whom he discussed, among other things, questions of ecology and environmental protection.
 - The Mayor of Novi Sad Miloš Vučević met the Ambassador of the People's Republic of China to Serbia HE Chen Bo at the City Hall. On that occasion, they expressed great satisfaction at the traditionally good cooperation and connection between the two nations. They discussed the importance of implementing large infrastructure projects that contribute to a change in the economic situation and to overall progress.
 - China's rover landed on Mars
 - The Chinese National Space Administration (CNSA) published a video showing the movements of the Zhurong rover on the surface of Mars. The video was recorded by a wireless camera placed on the surface by a robot. However, one detail that was recorded aroused the viewers' curiosity.
- (1)
 - The Bishop of Bačka Irinej Bulović assessed, on the occasion of the adoption of the Law on Gender Equality, that "a new radical reform and arbitrary transformation of the Serbian language is being imposed and implemented in practice".
 - The coalition NADA, composed of POKS, DSS and 19 citizens' associations, announced today that they will propose to the National Parliament a Draft law on the protection of the Cyrillic alphabet, given that more than 85 percent of public signs in Serbia are written in the Latin alphabet, and the number increases by another percent every year. "If this trend continues, in 15 years Cyrillic signs will completely disappear from Serbia, and this is why we will propose a law that will put an end to this, and will put the Cyrillic alphabet back to its rightful place."
 - The Raiffeisen bank has enabled all its Visa card users to use the Garmin Pay service.
 - This is a joint project prepared by NIS, in cooperation with the company Visa and with the Poštanska štedionica bank. With the introduction of the new service, the three partner companies have once again proven that they remain committed to implementing innovations, and that meeting the customers' requirements and expectations remains a priority in their business.

- There is great interest in the allocation of empty houses
- "Let us try not to empty any more houses, to make people from Kosovo and Metohija, Southern Serbia, Medveđa, Crna Trava not run away to Belgrade or even further, but to make that circle go in the opposite direction, try to keep them there in those houses, and even bring some back from here", said Krkobabić as a guest on Pink Television. He added that this call is not just about a salary and a secure job, but also about a secured roof over people's heads, and that is the key issue today.
- (0)
 - A woman from Florida revealed that one billion dollars had been paid into her bank account.
 - The Israeli Ambassador Jahel Vilan said, at a forum entitled "Water Treatment and Efficient Wastewater Management", that a CCIS office will be opened in Jerusalem, and he called on Serbian and Israeli businessmen to form a network. Vesović said that the office will help Serbian and Israeli businessmen to establish cooperation in a faster and easier way, reports Tanjug. He expressed his expectation that the cooperation between Serbia and Israel will be raised to a higher level.
 - The capital of Turkmenistan, Ashgabat, is the most expensive city in the world for foreign workers.
 - One of the recommendations to employers is to consider the possibility of interrupting outdoor work between 12 and 16 o'clock, and to continue working after 16 o'clock. - If the work process is such that this cannot be applied, the Ministry recommends that employers should move the start of working hours for those working outdoors to six o'clock in the morning - the Federal Ministry of Labour and Social Policy announced.
 - Momčilo Otašević says that he engages in acting solely out of love for the profession, and that he does not pay too much attention to popularity.
 - Four passengers have been taken to the Emergency Centre after the collision of two city buses near the National Library of Serbia in Belgrade. The City's Emergency Room service told the Beta agency that they had minor injuries. The RTS reported that a line 401 City Transport Company bus and a line 78 bus by a private carrier collided at the stop around 3 pm.

Phrase – sentence

- (4)
 - Change of conditions for travelling to Egypt
 - There has been a change in regulations when it comes to the entry of Serbian citizens into Egypt, the Egyptian ambassador to Serbia said today.
 - Parliament on pensions
 - The Parliament is initiating a new extraordinary session to consider the Draft Law on modification and completion of the Law on Pension and Disability Insurance.
 - Large fire at a London railway station
 - A large fire broke out today at an underground station in central London.
- (3)
 - Serbia's green bonds

- The Government of Serbia made a decision on establishing an intersectoral workgroup for the development and implementation of a framework document for issuing Serbia's green bonds in the international financial market, reports the latest issue of the Official Gazette.
- A new national holiday: Juneteenth
- The American Congress passed a Draft Law declaring 19 June a holiday to commemorate the end of slavery and the liberation of the last slaves in 1865 in the state of Texas.
- A major archaeological discovery near Varvarin
- During earthworks on the route of the Morava Corridor, in the village of Maskare, near Varvarin, a tomb was found containing well-preserved human skeletons, most likely from the 8th century according to the experts' first estimates.
- (2)
 - Solutions to the combined final examination
 - Primary school graduates took the combined final examination today, which consisted of a total of 20 questions from five subjects - biology, history, geography, physics and chemistry.
 - A long list of consumers without electricity today
 - The public company Waterworks and Sewerage will interrupt water supplying for the following consumers today.
 - A major problem for Portugal
 - Bad news arrives for Portugal just before the start of the European Championship.
- (1)
 - A green light for Medijana's summer of culture
 - For the fourth year in a row, the municipality of Medijana and the Stanimirović Jewellery Store awarded the best students of the secondary schools located within the territory of this municipality, introducing in this way a tradition of valuing hard work and its results.
 - A pre-wedding argument
 - Mirko Šijan and Bojana Rodić are expecting their first child soon, and their wedding is being prepared.
 - Wine Royal Serbia in 12 stories
 - During her visit to Kragujevac, the Minister of Trade, Tourism and Telecommunications in the Government of the Republic of Serbia, Tatjana Matić, visited the premises of the Military Technical Institute and the House of Đuro Jakšić, where the Balkan Wine Network project was presented.
- (0)
 - A reception for the most successful students of the Valjevo Gymnasium
 - The search for mushrooms starts at dawn, and pickers spend up to several hours in the woods, armed with knowledge and patience.
 - A silo opening in Zrenjanin
 - Maja Žeželj, TV presenter, told the story of how some time ago she nearly died.

- Bitter strawberries from Spain
- Milonjić was a member of the Legal Council of the Metropolitanate of Montenegro and the Littoral in the negotiations related to the disputed Law on freedom of religion.

¹ This document is the English translation of the guidelines originally written in Serbian (apart from score definitions, see note 2). In the translation of examples an attempt was made to remain as faithful to the original text as possible. For the Serbian originals, see:

[https://vukbatanovic.github.io/CLSS.news.sr/blob/main/Annotation guidelines - Serbian.pdf](https://vukbatanovic.github.io/CLSS.news.sr/blob/main/Annotation%20guidelines%20-%20Serbian.pdf).

² Score definitions are taken from: Jurgens, D., Pilehvar, M. T., and Navigli, R. (2014). SemEval-2014 Task 3: Cross-Level Semantic Similarity. In *Proceedings of the Eighth International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, Ireland, Association for Computational Linguistics, pp. 17-26.