

## Kratek opis usposabljanja mladega raziskovalca (*Short description of the Young Researcher's training*)

1. Raziskovalna organizacija (*Research organisation*):

Univerza v Ljubljani, Filozofska fakulteta

2. Ime, priimek in elektronski naslov mentorja (*Mentor's name, surname and email*):

doc. dr. Apolonija Gantar, apolonija.gantar@ff.uni-lj.si

3. Šifra in naziv raziskovalnega področja (*Research field*):

6.05 Jezikoslovje

4. Kratek opis usposabljanja mladega raziskovalca (*Short description of the Young Researcher's training*):

Navedite tudi morebitne druge zahteve, vezane na usposabljanje mladega raziskovalca (npr. znanje angleškega jezika, izkušnje z laboratorijskim delom, potrebne licence za usposabljanje...).

### Vsebina raziskovalne tematike

Raziskovalno usposabljanje mlade raziskovalke ali mladega raziskovalca bo potekalo na področju pomenoslovnih raziskav slovenščine s poudarkom na formalizaciji in medjezikovnem povezovanju znanja. V ospredju bo kontekstualna analiza realnega jezikovnega stanja in medsebojna odvisnost pomenskih in slovničnih lastnosti leksikalnih enot. Tako zasnovane pomenoslovne raziskave se bodo neposredno vključevale v raziskovalni program Slovenski jezik – bazične, kontrastivne in aplikativne raziskave. In sicer se bodo v okviru sklopa jezikovnih analiz in jezikovnih opisov slovenščine raziskovalne aktivnosti osredotočale tako na poznavanje in aktualizacijo obstoječih metodoloških in teoretičnih pristopov v pomenoslovnih raziskavah slovenščine kot na analizo pristopov v razumevanju pomena, ki se v novejšem času vzpostavljajo v digitalnem svetu. Raziskovalne aktivnosti bodo vključene tudi v jezikoslovne raziskave projekta Nova slovnica sodobne standardne slovenščine: viri in metode ter projekta Pomenoslovne raziskave slovenščine kot osnova za formalizacijo in medjezikovno povezovanje znanja.

Raziskovalno usposabljanje kandidatke ali kandidata bo temeljilo na korpusni metodologiji ter kontekstualni in diskurzni manifestaciji pomena v jeziku. Metode dela bodo vključevale analitično-sintetične pristope na ravni kvantitativnih in kvalitativnih jezikoslovnih analiz, kar vključuje samostojno opravljanje zahtevnih analitičnih del, sintezo področnih znanj in njihovo uporabo pri reševanju konkretnih problemov. Posamezen raziskovalni problem bo razdelan na podlagi realnih jezikovnih podatkov ter umeščen v teoretični in metodološki okvir, ki temelji na kontekstualni in komunikacijski realizaciji pomena, ter ovrednoten z vidika različnih analitičnih dognanj. Kandidatkino ali kandidatovo usposabljanje bo usmerjeno tudi v samostojno uporabo metodoloških orodij, zlasti v izvajanje in organiziranje raziskav ter uporabo in nadaljnji razvoj raziskovalnih, zlasti interdisciplinarnih metod in tehnik.

### Terminski plan aktivnosti mladega raziskovalca

Okvirna tema kandidatkine ali kandidatove doktorske naloge bo zasnovana že na začetku usposabljanja, temu pa bo nato v čim večji meri prilagojena tako izbira temeljnih predmetov in izbirnih doktorskih seminarjev kot tudi vključitev v tekoče projektne in programske raziskave.

### **1. letnik**

V prvem letniku bo kandidatka ali kandidat v okviru študijskega usposabljanja opravil\_a zahtevane študijske obveznosti iz naslova organiziranih študijskih oblik (40 ECTS), in sicer 2 temeljna predmeta s področja jezikoslovja oz. slovenistike ter 2 izbirna doktorska seminarja v okviru izbranega študijskega programa 3. stopnje. Raziskovalni del usposabljanja bo zajemal metodološko teoretični pregled semantičnih raziskav skozi različne vidike: zgodovinski pregled, razvoj teorij in njihova aplikacija v slovenskem prostoru. Kandidatka ali kandidat bo pripravil\_a sintetični pregled vseh relevantnih in obstoječih slovenskih virov, ki vključujejo pomenske in pomenskoskladenske podatke, od referenčnih korpusov za slovenščino, specializiranih korpusov, slovarjev do leksikalnih podatkovnih baz za slovenščino in nekatere tuje jezike. V okviru raziskovalnega usposabljanja bo kandidatka ali kandidat aktivno vključen\_a v posamezne faze tekočih projektov, kjer se bo seznanil\_a z osnovnimi metodami raziskovalnega dela, uporabo jezikovnotehnoloških orodij ter samostojno pripravo empiričnih podatkov za nadaljnje pomenoslovne in leksikološke analize.

### **2. letnik**

V okviru študijskega usposabljanja bo kandidatka ali kandidat opravil\_a predvidene študijske obveznosti, in sicer zunanji izbirni predmet (10 ECTS), ter izdelal\_a in zagovarjal\_a dispozicijo doktorske disertacije (15 + 15 ECTS). Preostale ETCS (20) bo kandidatka ali kandidat pridobil\_a z raziskovalnim delom, ki bo vključeval: sistematični študij relevantne tuje literature ter pregled in ovrednotenje sistemov za reprezentacijo semantičnih podatkov, npr. FrameNet, WordNet, Corpus Pattern Analysis, Vallex ipd., ter jezikovnopriročni oz. slovarsko orientiranih leksikalnih podatkovnih baz. Kandidatka ali kandidat bo identificiral\_a konceptualne vidike semantičnih modelov, ki jih bo mogoče uporabiti pri izdelavi semantične baze za slovenščino, ter definiral\_a osnovne leksikografske elemente, ki vključujejo semantične informacije, kot so npr. vsebinski elementi pomena, vsebinski in strurni elementi definicij, eksplisitne in implicitne pragmatične informacije, slovnične informacije, vezane na pomen, besednozvezne in stavčne vzorce, ki razdvoumljajo pomene, semantične vloge udeležencev in medpomenska razmerja. V tem segmentu bodo izdelane tudi prve semantične analize na podlagi semantično označenih učnih korpusov za slovenščino ter identificirana nekatera ključna raziskovalna vprašanja. Pri tem se bo kandidatka ali kandidat usposobil\_a za napredno uporabo orodij za jezikovno analizo, kot je npr. Sketch Engine, skladenjski razčlenjevalnik in morfosintaktični označevalnik za slovenščino, ter se seznanil\_a z različnimi programi za pomensko in pomenskoskladenjsko analizo korpusov in besedil v digitalni obliki.

### **3. letnik**

Raziskovalni del usposabljanja kandidatke ali kandidata bo usmerjen v (a) pripravo izvirnega znanstvenega prispevka – na podlagi identifikacije in kategorizacije problematičnih pomenskih vsebin kot rezultata predhodno izdelanih analiz korpusnih in semantičnih virov za slovenščino – ter (b) izdelavi doktorske disertacije. V tem času je predvideno tudi usposabljanje kandidatke ali kandidata v tujini v obliki študijskih projektov na tujih univerzah oziroma raziskovalnih institutih ter aktivna udeležba na domačih in mednarodnih delavnicah ter konferencah, povezanih z raziskovalno tematiko kandidatke ali kandidata.

Poleg zainteresiranosti za raziskovanje pomenoslovne problematike jezika in zavedanja nujnosti interdisciplinarnih pristopov na tem področju se od kandidatke ali kandidata pričakuje dobro znanje angleškega jezika ter spremnost pri uporabi računalniških orodij za analizo besedil.

### **Description of research topic**

The research training of the young researcher will be carried out in the field of Slovene semantics with an emphasis on the formalization and interlingual linking of knowledge. The training will focus on the contextual analysis of actual language use and the interdependence of semantic and grammatical properties of lexical units. The semantic research will be directly

involved in the research program *Slovenian Language - basic, contrastive and applied research*. In the context of Slovene language analyses and language description, the research activities of the candidate will focus both on the knowledge and actualization of existing methodological and theoretical approaches in Slovene semantics as well as on the analysis of approaches to understanding semantics within the context of the digital environment. The research activities will also be included in the linguistic research of the project New Grammar of Contemporary Standard Slovene: Resources and Methods and the project Semantic Research of Slovene as a Basis for the Formalization and Interlingual Linking of Knowledge.

The candidate's training will follow a corpus methodology focusing on contextual and discourse manifestations of meaning in language. The methodology will include analytical and synthetic approaches to quantitative and qualitative language analysis, e.g. independent research in complex language analyses, a synthetic overview of specialized knowledge and its use when solving concrete problems. Each individual research problem will be elaborated based on actual language data and placed within a theoretical and methodological purview based on contextual and communicational realizations of meaning. The candidate's training will also focus on an independent use of methodological tools with emphasis on the organization of research and development of interdisciplinary research methods.

### **Plan of activities for the young researcher**

The outline of the topic of the candidate's doctoral dissertation will be made at the beginning of their training. To the greatest extent, the choice of subjects and elective doctoral seminars will follow the topic of the dissertation, as will the candidate's activities in current projects and research programs.

#### **Year 1**

In their first year, the candidate's training will include the required academic obligations (40 ECTS), namely 2 basic subjects from the field of (Slovene) linguistics and 2 elective doctoral seminars within the chosen doctoral study program. The research part of the training will include a methodological-theoretical overview of semantics in its various aspects: a historic overview, the development of various theories and their application in Slovene linguistics. The candidate will prepare a synthetic overview of all relevant existing Slovene resources that include semantic and semantic-syntactic data: from Slovene reference and specialized corpora to dictionaries and lexical databases (both Slovene and those based on foreign languages). As part of their research training, the candidate will be actively included in individual phases of ongoing research projects, which will allow them to become familiar with basic research methods, the use of language technology tools and independent preparation of empirical data for further semantic and lexicological analyses.

#### **Year 2**

As part of their training, the candidate will cover the required study obligations, which include an external elective subject (10 ECTS) and the preparation and defence of the outline of their doctoral dissertation (15 + 15 ECTS). The other 20 ECTS will be achieved through research, which will include a systematic study of relevant related work and an overview as well as an evaluation of various systems for the presentation of semantic data, e.g. FrameNet, WordNet, Corpus Pattern Analysis, Vallex. In addition, the candidate will analyse the structure of language reference books and dictionary-based lexical databases. The candidate will identify conceptual aspects of semantic models, which can later be used to construct a semantic database for Slovene, and define basic lexicographic elements that include semantic information, e.g. elements expressing meaning, structural and content elements of definitions, explicit and implicit pragmatic information, grammatical semantic-related information, word-phrase- and sentence-level patterns that disambiguate meanings, semantic roles and intersemantic relations. In this segment, the first semantic analyses will be carried out on the basis of semantically manually annotated training corpora for Slovene in order to identify a series of key research questions.

The candidate will be trained for advanced use of language analysis tools, e.g. Sketch Engine, syntactic parser and morphosyntactic tagger for Slovene, as well as various programs for semantic and semantic-syntactic analysis of corpora and texts in digital format.

**Year 3**

The research part of the candidate's training will focus on (a) the preparation of an original scientific paper – based on the identification and categorization of problematic semantic content as a result of previous analyses of corpus and semantic resources for Slovene, and (b) the writing of the doctoral dissertation. During this time, the candidate will also be trained abroad in the form of study projects on foreign universities and research institutions, and will also actively participate in national and international workshops and conferences relevant to the candidate's research topic.

The candidate is expected to show interest in the research of semantics in language and to be aware that interdisciplinary approaches in this field are necessary. In addition, the candidate is expected to have a good grasp of English as well as good computer skills, in particular with text analysis tools.