Co-funded by the
Erasmus+ Programme
of the European Union

Research-Based Teaching: Guidelines and Best Practices

UPSKILLS Intellectual output 2.1

Compiled by:
Marko Simonović[*], Boban Arsenijević[*], Iulianna van der Lek[†], Stavros Assimakopoulos[‡], Louis ten Bosch[†], Darja Fišer[†], Tihana Kraš[§], , Paul Marty[‡], Maja Miličević Petrović[**], Stefan Milosavljević[*], Marc Tanti[‡], Lonneke van der Plas[‡], Margherita Pallottino[††], Genoveva Puskas[††], Tanja Samadžić[‡‡]

[*] University of Graz
[†]CLARIN ERIC
[‡]University of Malta
[§]University of Rijeka
[**]University of Bologna
[††]University of Geneva
[‡‡]University of Zurich

June 30 2023

UPSKILLS: UPgrading the SKIlls of Linguistics and Language Students

Erasmus+ Programme
Key Action 2: Cooperation for Innovation and the Exchange of Good Practices Action
KA203: Strategic Partnerships for Higher Education

Grant Agreement Number: 2020-1-MT01-KA203-074246

Document status: Final
Dissemination level: Public

UPSKILLS Consortium:

&

with financial support from

https://doi.org/10.5281/zenodo.8176220

Table of contents

Executive Summary

This document constitutes the central deliverable of the IO2 'Best practices and guidelines for research-based teaching'. It addresses the topic of the research-teaching nexus and provides guidelines for individual lecturers meant to assist them in developing both well-elaborated views on the research-teaching nexus and putting these views into practice through the inclusion of their own ongoing research into ongoing teaching in the form of research-based teaching courses.

The document consists of two main parts and five appendices.

Part 1: Research-based teaching and UPSKILLS introduces the topic area and the rationale of these guidelines and best practices, as well as the general structure of research-based courses described in the remaining two parts of the document.

In Section 1: Introduction, we present the general concepts of research-teaching nexus and research-based teaching, discuss the challenges in implementing research-based teaching and outline how the present guidelines and best practices can be used to overcome these challenges.

Section 2: RBT in UPSKILLS discusses research-based teaching in the broader context of the UPSKILLS project, especially focusing on how research-based teaching can be beneficial to the employability of graduates of language-related programmes while also enabling them to make informed career choices.

Section 3: Rationale and layout of an UPSKILLS RBT course presents topics covered in a typical UPSKILLS RBT course grouped into three topic blocks (Research design, Research infrastructures & techniques and Subject-specific aspects), the rationale behind this grouping, as well as the ideal time distribution across the topic blocks.

In Section 4: Choosing/developing a course subject, we provide specific advice regarding the identification of the domain within the lecturer's area of expertise that can be optimally covered in an RBT course, with special attention for the transferable skills that need to be acquired by the students.

In Section 5: Learning outcomes of RBT courses, we address what students learn in an RBT course, devoting special attention to general research-related learning outcomes. The learning outcomes are organised in a structure that matches the structure of RBT topics presented in Section 3. This section concludes Part 1 of these guidelines and best practices.

With the necessary groundwork in place, in Part 2: Guidelines we lay down specific recommendations for the creation and implementation of an RBT course.

In Section 6: Making a course outline we present a step-by-step roadmap from the selection of an RBT course subject to designing a preliminary course plan.

Section 7: Organising the work in an RBT course tackles the workflows in an RBT course, the optimal use of the research report in framing and organising the students' work and the recommendations for writing (research) instructions for students.

In Section 8: Supervision, assessment and evaluation we address the optimal division between the portions of students' work on which they only receive feedback and those that contribute to their final grade, as well as the importance of communicating this distinction to the students in order to foster creativity and encourage innovative thinking.

Section 9: After the course focuses on the issues of data reusability, giving the final feedback to students and debriefing them, as well as on the ways of receiving feedback from the students about the RBT course. This section concludes the main parts of these guidelines and best practices.

Annex 1: Detailed overview of general topics in RBT courses is a template that gives a taxonomy of likely types of content of an RBT course. It is meant to be used by teachers in the development of a concrete course, whereby they can simply concretise each of the general types given in the template. The example RBT courses in Annex 5 have their respective topics given and their intended use illustrated in this template.

Annex 2: The research report template provides a general structure for a student's research report within an RBT course. Its purpose is to drive the students' work, and once filled in, to be the main component of the final evaluation.

Annex 3: Detailed overview of learning outcomes per topic block in RBT courses is a template with a taxonomy of likely types of learning outcomes of an RBT course. It is meant to be used by teachers in the development of a concrete course, whereby they simply concretise each of the general types given in the template. As with the topics from Annex 1, the example RBT courses in Annex 5 have their respective learning outcomes given in this template and illustrate its intended use.

Annex 4: Survey for course evaluation by students contains the editable form that can be used for obtaining students' feedback after RBT courses.

Annex 5: Examples of RBT courses contains detailed descriptions of topics and learning outcomes of the RBT courses piloted within the Intellectual Output 2.1. They are presented using the templates from Annex 1 and Annex 3, and also instrumentalize the research report template from Annex 2.

List of abbreviations

Abbreviation	Definition
BA	Bachelor of Arts
EC	European Commission
MA	Master of Arts
NLP	Natural language processing
PhD	Doctor of Philosophy, Doctorate
RBT	research-based teaching

List of Figures

List of Tables

PART ONE: Research-based teaching and UPSKILLS

1. Introduction

Teaching and **research** constitute the two core activities in academia. However, these two activities are typically planned, performed and evaluated separately. It, therefore, comes as no surprise that the connection between them, usually discussed in the literature under the rubric of the **research-teaching nexus**, receives little attention in the reality of most academic institutions (for a notable exception, and a rich collection of experiences and ideas, see Bastiaens, van Tilburg & van Merriënboer 2017).

The near absence of attention for the research-teaching nexus at the institutional level has ramifications for the ways academics see these two activities. Existing research on the research-teaching nexus indicates that academics conceptualise research and teaching separately. In a study of the nexus in humanities, Visser-Wijnveen (2009: 141) found that "academics' conceptions of the research-teaching nexus are related to their conceptions of teaching and not to their conceptions of research and knowledge". Moreover, she found that "the conceptions of research and knowledge are more closely related to each other than to the conception of teaching". The explanation offered by the author is that the conceptions of research and knowledge come from a different source than those related to teaching. For most academics, conceptions of research and knowledge are formed in the course of their research schooling (presumably during their postgraduate studies), whereas their conceptions related to teaching originate from their earlier experiences as students, as confirmed in studies on beginning teachers (e.g. Feiman-Nemser & Remillard, 1996).

Addressing the research-teaching nexus explicitly is desirable at all levels, from individual to team, institutional and national. While higher levels are also crucial, our main goal in these guidelines is to assist **individual lecturers**[1] in developing both well-elaborated views and putting these views into practice in the research-teaching nexus. We trust that the most logical and rewarding first step is the inclusion of one's own **ongoing research** into **ongoing teaching** in the form of **research-based teaching** (henceforth, RBT). While the other direction, the advancement of research based on teaching, should not be neglected (e.g. Giraud & Saulpic 2019), the focus of the UPSKILLS project is on the enhancement of students' skills in ways that will enhance their employability. For this reason, the present document will concentrate on research-based teaching alone.

In the model proposed in Healey (2005) and refined in Jenkins et al. (2007), represented in Figure 1 below, research-based teaching is one of the four possible

[1] Throughout this text, the term lecturer is used for anyone who teaches courses at the university.

implementations of the research-teaching nexus. Research-based teaching **focuses on the process of research** rather than on the results of research and involves **students as active participants** rather than as audience. As such, research-based teaching introduces students to the role of researchers to a higher extent than the other three implementations shown in Figure 1.

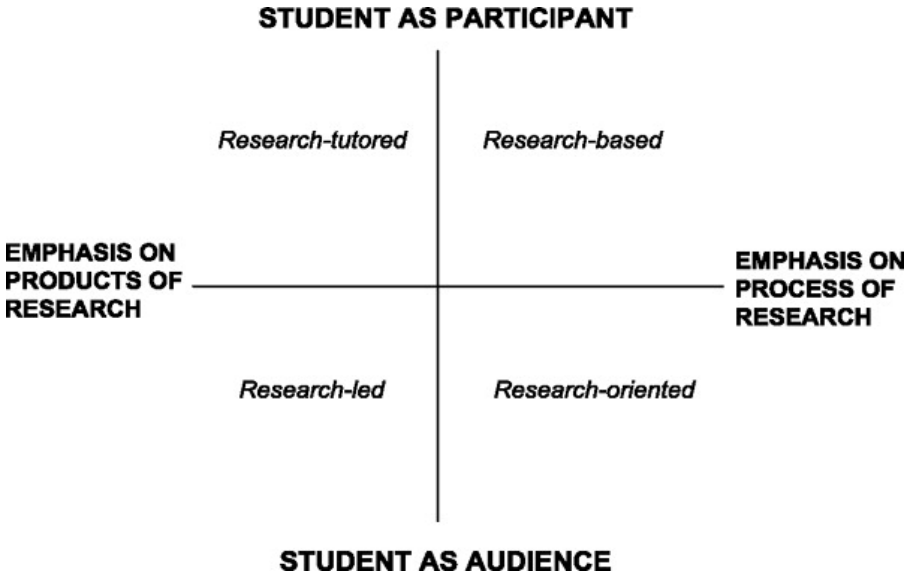

Figure 1. Healey's (2005, Jenkins et. al 2007) adapted model of the research-teaching nexus (from Visser-Wijnveen et al. 2010).

The rationale for focusing on research-based teaching in addressing the research-teaching nexus is threefold. First, exposing students to the actual research practices helps them acquire **aspects of disciplinary knowledge** (including the disciplinary conceptualisations of knowledge and research discussed above) **that otherwise remain implicit and unaddressed**. As previous research shows, "research awareness is stimulated by a carefully organised close look at their teacher's research" (Visser-Wijnveen 2009: 141-142). Second, we trust that the **everyday research practice** of academics involves **transferable skills** which do not get taught to students during regular courses. Finally, for the lecturers, research-based teaching has the potential of **obtaining valid research output** from their students but also of **discovering teachable aspects** of their ongoing research activities that students will find useful or even necessary for their future (academic and non-academic) careers.

According to Dekker & Wolf (2016), the four manifestations of the research-teaching nexus in Figure 1 are ordered on a scale of the importance of the role of research as in Figure 2.

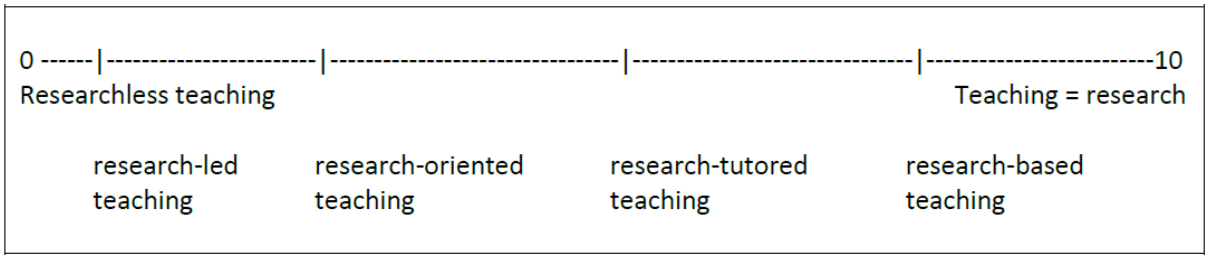

Figure 2. Dekker & Wolf (2016)'s scale of the research-teaching nexus.

A further classification of the research-teaching nexus, which does not perfectly overlap with the one shown above, is that into five profiles describing what the lecturer does in a course. These are I) teach research results, II) make research known, III) show what it means to be a researcher, IV) help to conduct research, and V) provide research experience (Visser-Wijnveen 2009: 131). Our implementation of research-based teaching involves elements of all profiles but its overlap with profile V) is the greatest, while that with profile I) is the most limited, with all the other profiles falling somewhere in between these two extremes.

In courses with a strong research-based component of the kind that will be discussed in these best-practice guidelines, students are **involved in activities of ongoing research projects**, using the same **research tools and infrastructures** that the lecturer uses in their own research. Students can then **acquire and actively apply problem-solving and analytical skills** to **actual data**, using **approaches implemented in actual research**. In some of the scenarios, when the lecturer is involved in this type of research activities, students can also be exposed to **industry-based research**, which they are likely to encounter in their post-university professional careers. In this setting, industry-based research refers to research that companies conduct as part of their business, rather than with the aim of advancing knowledge, as academics typically do, and comprises a number of tasks that graduates are bound to encounter in their day-to-day workflow when they gain employment in the industry (for more details, see van der Plas & Assimakopoulos 2023).

The concept of research-based teaching is often met with ambivalence by lecturers. Most first experiences with research-based teaching seem to be marked by **considerable challenges and frustrations**. An important part of these frustrations stems from underestimating how much additional work (but also what kind of additional work) is required in order to set up and run a research-based course. In many cases, research-based teaching is initially perceived as the intersection of two well-known domains (one's own teaching and one's own research) and therefore requires minimal additional effort. However, the actual implementation can involve setbacks that are not usually encountered in non-RBT courses.

Marko's[2] testimony

In the first RBT MA course I taught, the whole concept seemed to be failing already at Class 3. The plan for this class was to discuss how data obtained from a corpus would be further annotated and processed. The students were already working on their specific projects, so I was going to use data from "their" projects as an illustration. From my perspective, the only previous knowledge the students needed was a minimum of Excel skills and some commonsensical notion of how to organise data in a table. What I was discovering at every step is that, as students pointed out, I was using "technical terms without defining them", something I normally take pride in never doing when I teach. When I asked what these technical terms were, they gave examples like "annotating" and "organising data", notions which I believed were perfectly transparent for an MA student. We reached an impasse and I wasn't sure how to proceed. I felt like it was wrong to spend multiple classes introducing such general notions which have nothing to do with linguistics, let alone with the topic of the course. I also had the feeling that I would actually not be teaching them new concepts, but the way we are used to talking about them in the field. I discussed the options with the students. I was open to giving the students a less active role and presenting more research results to them instead, basically making the course research-led. However, the students expressed the wish to follow the original plan, just allowing much more time for additional explanations. I should admit that an important part of my motivation for honouring the students' wish lay in the fact that we were planning the UPSKILLS project application and I wanted to gather RBT experience. I ended up teaching a whole different course and I was quite satisfied with the results in the end. I realised that I needed to reconsider what I take for granted and plan my RBT courses more flexibly.

Lonneke's testimony

I decided to include an industry-based research component for a unit called Multilingual Computing, after a visit of a terminologist from the European Commission (EC) to our Institute at the University of Malta (UM). The aim of her visit was among others to strengthen the links between the translation unit at the EC and the UM by working on projects that are beneficial to both. The project the students worked on was concerned with extracting medical terminology available on

[2] We disclose the names of the instructors, but not of the students, as some of the students preferred to remain anonymous.

the IATE Public Database in English and finding a suitable Maltese equivalent by identifying the various Maltese variants which have been used so far in translations of EU legislation. They used computational methods and provided term variants that were then evaluated by the terminologist. I reserved one of the lectures for a talk by a person from the translation unit of the EC with the aim of showing the students a real use case of the things covered in my lectures, after which the students started working on the project. Considering that this was my first experience, it worked pretty well in my view. The terminologist was happy with the results the students provided and could improve the resources for Maltese based on them. I gave the students clear handles (step-by-step procedure) to get to the results. I think they got an idea of how the methods we were discussing during the lectures could be used in practice. It was quite satisfying to see that the results from their work were actually useful for the terminologist. What I think was wrong in my approach is that I left very little to the initiative of the students. It was not a research experience as such. It was a practical task that was linked to a use case, but it did not give them much opportunity to acquire and actively apply problem-solving and analytical skills. In order for that to happen, I would have needed to spend more time with the students, allowing them to come up with a solution for the problem together. I reserved too little time for that and just gave them a list of steps to follow. I think that this can be a pitfall of industry-based research. The industrial partner involved has expectations and the academic would like to come with useful results. This pressure might lead to the lecturer preferring to give a lot of guidance at the expense of the learning experience for the students.

The challenging aspects of these experiences are arguably traceable to the **systematic lack of a well-elaborated view of the research-teaching nexus**, despite clear conceptualisations of research and of teaching. As already previewed above, one of the main goals of these best-practice guidelines is to facilitate the establishment of such a well-elaborated view.

The ambivalent attitude of lecturers towards research-based teaching does not mean that there is no recognition that this form of teaching is useful and necessary. As transpired from the survey we held within the preliminary needs analysis of our project, a bit over 50% of lecturers indicated that they integrate research into teaching, while around 75% indicated that they would be happy to follow dedicated training focusing on this. It is therefore safe to conclude that there is much room for expanding the domain of research-based teaching in the individual repertoires of the lecturers. Also, Visser-Wijnveen (2009: 141-142) found that students enrolled in courses with a strong research component report more learning outcomes on academic disposition and research awareness than their teachers had aimed for in their course designs.

Congruent signals are coming from the industry. The **lack of training in research skills, data acquisition and data handling skills** were cited by our informants from the industry as one of the major obstacles to integrating linguists into collaborative workflows (Miličević Petrović et al. 2021).

Against the background sketched above, these best-practice guidelines have the goal of **giving lecturers a headstart in developing their elaborated view of the research-teaching nexus** and **developing or improving their own research-based teaching**. We complement the existing literature in at least three ways. First, our best-practice guidelines have a **disciplinary scope**, focusing only on courses in **language-related disciplines.** Second, they also take into account **the employability of graduates** of language-related programmes (see next section). Finally, they have a **broad testing ground**: they are based on experiences from sixteen courses taught by six consortium members.

These best-practice guidelines are especially targeting individual lecturers who are willing to integrate their ongoing research into their teaching. Clearly, they can also be used by research teams, departments etc. to create more variegated RBT courses. While not explicitly addressed, we believe that our best-practice guidelines will also be useful to administrators and policy makers at the university and beyond in articulating their agenda when it comes to research-teaching nexus.

We created these best-practice guidelines in an attempt to establish a general course-design framework that can be easily adapted to most university settings in Europe. However, we are aware that some of the contexts are less welcoming to research-based teaching than others. **It is our hope that administrations and policy makers who get in touch with these best-practice guidelines will help take the necessary steps in adopting research-based teaching on a larger scale.**

In the following sections, we discuss how RBT fits the goals of the UPSKILLS project.

2. RBT in UPSKILLS

The central goal of the UPSKILLS project is to identify **skills gaps and mismatches in the curricula of language-related programmes** and tackle these skills gaps and mismatches through the **development of a new curriculum component and supporting materials** to be embedded in existing linguistics and language-related programmes. These measures are intended to **enhance students' employability** by providing them with the skills needed to compete for a wider range of positions in the labour market.

We trust that explicit attention to the research-teaching nexus in general and research-based teaching in particular are **beneficial not only to the employability of**

graduates of language-related programmes but also to their ability to make informed career choices.

One aspect that is often forgotten in the discussions of the employability of graduates is that **universities are employers as well** and that there is no guarantee that a university will be inclined to hire its own graduates on a research project. At academic institutions where the gap between teaching and research is large and graduates have limited research skills, project leaders may have a tendency to hire graduates from other institutions or even other disciplines. Multiple consortium members can provide anecdotal evidence of this type, where especially graduates of programmes like psychology or computer science seem to have a headstart over graduates of language-related programmes, not because the position requires disciplinary knowledge related to psychology or computer science but because they are more versed in general research skills like **data collection, data processing or statistics**.

A further negative consequence of a lack of involvement of students in actual ongoing research is that graduates are likely to **develop a misguided notion of what research is**. In study programs where lecturing and broad yet shallow handbooks predominate in teaching methods, students mistake introductions and overviews for the immediate results of research. This can for instance be observed when students are asked to come up with topics for research for a course paper or a BA or MA thesis, and instead of a research question, they come up with a domain that corresponds to a chapter in a textbook. “I want to investigate secondary predicates'”, or “My plan is to conduct research on the genitive case”. This indicates that students tend to think of research as a purely positivist enterprise in which encyclopedic knowledge is accumulated. This knowledge is then viewed as presentable in a new version of the textbook. Misconceptions of this type may lead the students to avoid a career path in research or research-related pathways or, paradoxically, to strive for one on the false grounds. Experience with real-life academic research is an important antidote to this, enabling students to make informed career choices.

What is true of university-based career paths can be for a good part extended to other career paths. This is because most of the skills identified by the industry as lacking in graduates of language-related programmes in our needs analysis (Miličević Petrović et al. 2021) are **essential to research activities performed at their universities** and performed by their lecturers. As argued above, these skills are present, but not entirely transferable because there is a lack of attention for the research-teaching nexus.

While we trust that involvement in actual ongoing research is beneficial as such, we did develop a specific approach largely influenced by our disciplinary focus and the realities of our academic institutions. As already previewed, we are focussing on research-based teaching as the approach which enables the most direct involvement in research practices. A key aspect of our implementation of research-based teaching is the use of a **research report** as a dominant format of the final deliverable by the student. As will be elaborated below, the

research report should have as many characteristics as possible of an 'actual' research report (in the sense of being presentable/publishable at an actual academic venue) and it should also be evaluated as such.

As mentioned above, these guidelines are mainly based on our experience with RBT courses. The main instrument of quality assurance was piloting and obtaining feedback from both instructors and students. In the course of the project, sixteen courses were run. We have thus used and will continue to use the feedback from both students and lecturers to reshape the guidelines in order to ensure the broadest possible applicability.

3. Rationale and layout of an UPSKILLS RBT course

In this section, we present the most general layout of UPSKILLS RBT courses. Many teachers are aware that the involvement of students in specific research projects may carry the risk of being too specific so that the students get a narrow perspective on research, failing to get acquainted with its full width and depth regarding topics, methods and techniques. This is a problem that riddles most involvement of students in internships and student assistantships: students perform tasks that are useful to the project, but without really getting the perspective of the entire research design in which they are participating. In order to avoid this pitfall, our research-based teaching guidelines envision a **general introduction to research design as part of all RBT courses**. Once the general background is in place, specific research topics can become the testing ground on which analytical and problem-solving skills can grow.

In Table 1, we present the general overview of the topics applicable to any UPSKILLS RBT course. The overview contains three main topic blocks. The order of the topic blocks reflects the order of appearance of the general topic in a majority of UPSKILLS courses (first we discuss research design, then infrastructures and techniques and only then do we spend substantial time on the disciplinary issues specific to the course). Also, the relative order of the topics within the block is the one shown in Table 1 for a majority of the UPSKILLS courses. However, in most RBT courses, topic blocks will overlap in time, i.e. some specific topics from the first block will be introduced after some of the topics from the second and the third block. (A more detailed overview of topics can be found in Annex 1). In a detailed course plan, the colour-coding of the different topics can help the lecturer to keep track of the topic blocks and the time dedicated to each of them.

Topic block	Topics
Research design	General research design: research problems, research questions, hypotheses, predictions, tests, variables, conditions, methods, statistical analysis…
	Adapting general research design to the specific topic of interest
	Adapting the research design to the available research infrastructures
	Research reporting
Research infrastructures & techniques	For obtaining literature
	For obtaining, sharing and managing data
	For analysing data
Subject-specific aspects	General question
	Particular questions, tasks and skills

Table 1. Overview of topics in an UPSKILLS course

The time dedicated to each of the topic blocks will naturally vary from course to course, but in Figure 3 below we show the ideal distribution of time over the three topic blocks. In the description of the topics below, we also point out the topics that in our experience often turn out to require more time than initially planned, so we recommend planning some extra time for them.

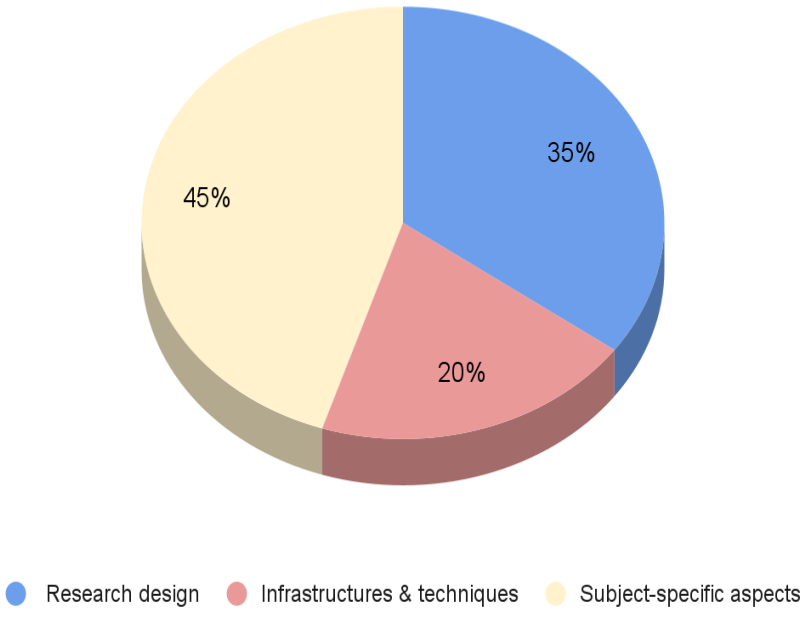

Figure 3. Ideal distribution of time in an UPSKILLS course

We now turn to each of the topic blocks. The topic block ***Research design*** contains topics that are common and essential to any research. In a RBT course outline, it should account for a substantial amount of course duration (ideally 35%, no less than 20% of time), since this block contains the most transferable skills inherent to conducting research. As discussed above, in our experience some of these skills may be so self-evident that lecturers may become aware of them only during the course. This is why, especially when the course is taught for the first time, we recommend keeping some room for developing these 'covert' skills in the course outline.

As stressed above, it is essential for each RBT course to contain a segment dedicated to *General research design.* An issue commonly encountered in teaching on this topic is that students have very different levels of previous knowledge and different previous research experience. An additional problem is that it often feels face-threatening to many students to inquire about the very basic aspects, as this is experienced as admitting lacking a clear idea of how research works. One resource which we found useful in shaping the classes dedicated to this topic is the course "Introduction to Research in Linguistics: Theory, Logic, Method", which was created by some of the consortium members in a previous project[3]. This online course (or parts of it) can be used as preparatory reading followed by a discussion session in

[3]Available at https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/course/, developed within a project whose description can be found at here.

which all the basic concepts are covered in order to arrive at a common ground which can be further elaborated in class.

Once an understanding of the general research design is established, we can move on to *Adapting general research design to the specific topic of interest*, where some initial specific research designs are discussed as applying to the subject matter of the course. Our experience speaks in favour of planning ample time for this topic, as it is one of the topics where lecturers often discover that ways of thinking they take for granted need to be explicated at some length.

> Tihana's testimony
>
> *One of the most challenging aspects of the RBT MA course that I have taught for my students was identifying and formulating the research problem within the research topic that they were dealing with. Students tended to confuse the research problem (an important gap in the literature that the study aims to fill) with the research question(s) and with the aim(s) of the study. It took a lot of practice and learning by trial and error for the students to learn the distinction between the three concepts and how to formulate the research problem in an adequate way. They didn't seem to realise at the beginning that it takes more than a sentence to describe the research problem and that using declarative sentences rather than interrogative ones was more suitable for this purpose. What helped the students, apart from my feedback, was providing them with ample examples of formulations of research problems in published research studies and asking them to find such examples themselves. For more details, see the course description in II.III.*

While it is useful to consider research designs as if there were no limitations on what can actually be done, once this step is completed, it is important to move on to *Adapting the research design to the available research infrastructures.* This is a further topic which in our experience usually requires more time than initially expected and should be planned accordingly.

Finally, once an idea of a research design is in place, we can turn to *Research reporting*. For this aim, we used the general research-report template designed by two UPSKILLS consortium members and presented in Annex 2. As foreshadowed above, completing this template will lead to the final deliverable of the course, on which also a substantial part of the evaluation will be based. It is therefore instrumental for students to create their own copy of this template and start filling it in as soon as it is introduced to them. As discussed at some length in Section 7, in adapting the research report template for a particular course, instructors inevitably face a degree of tension between creating a report that closely mirrors the conventions of the discipline (thus resembling a real-world research paper), and one that

is as clear and transparent as possible by documenting each step in separate sections. In this case as well, it is helpful to encourage students to ask questions and seek clarification regarding both the disciplinary conventions and the documentation requirements.

> Student testimony (course *Tense in Bosnian / Croatian / Serbian*)
>
> *I found the course very stimulating, especially because it had some open questions and we had the possibility to work on them for real, with the teacher not only as a moderator but also as a participant. In my learning process, I like to solve quizzes and exercises. However, I believe it is more useful to realise that research is not something that will lead us to a perfect formula, but rather a process that will evolve gradually and still leave us with uncertainty (while hopefully giving us some more knowledge). So, personally, it was very useful to see that the instructor was not just giving us information but also helping us build an attitude towards a discipline.*

The second topic block, ***Infrastructures[4] & techniques***, focuses on the specific tools the students will use in order to implement their research design. This also includes identifying suitable datasets in repositories and using data responsibly. The way the topics in this block are structured is highly dependent on the specific infrastructures and techniques used by the lecturer in their research. However, even in cases where some of the tools are already provided, (e.g. all literature is provided beforehand, so *Infrastructures & techniques for obtaining literature* could be skipped) we recommend spending some time on each of these topics. For instance, if no literature search is necessary for the specific domain, a minor literature-search assignment can be added on a closely related domain. This theme block also has an important potential when it comes to transferable skills. Working knowledge of tools used in actual research is not only an important asset, but also an important springboard for learning how to use other tools.

> Student testimony (course *Deverbal derivations*)
>
> *It was a leap in understanding the research in linguistics for me to learn to design, prepare and administer online a lexical decision task experiment, and in particular to statistically analyse the data. I believe that the effect would have been even better if I had a better general picture of the theory of lexicon and morphology, because I feel that I do not entirely understand the significance of the results, and of the questions they address.*
>
> Marko's testimony

[4] Research infrastructures are defined as services, facilities and resources used during research.

In May 2022 I taught an RBT course at the University of Novi Sad (see II.V for the course description) in which a lot of students' work took place in a single table on Google Drive, from which various counts and simple statistics were drawn automatically. Two colleagues were sitting in the classes. In December 2022, one of the colleagues told me that they were teaching a course in which, as they said, they implemented my principles and that they were enthusiastic about the results. Once they shared their course outline and materials with me, I realised that the main aspect that they took over was the specific way of using Google sheets with a single big document that many students have simultaneous access to. They did not learn about any of the specific technical possibilities from my course, but just seeing the specific familiar infrastructure in use made them rethink and reorganise their course.

The final topic block, ***Subject-specific aspects***, is used as an umbrella block for all topics which are strictly disciplinary. In all RBT courses a *General question* can be identified, whose answering typically requires addressing further *Particular questions, tasks and skills*. This is the domain in which the lecturer is an expert. This block is essential, but it should not cover more than 60% of the entire course outline (ideally 45%). With this in mind, we can now turn to the ways in which an optimal course subject can be selected.

4. Choosing/developing a course subject

As previewed above, a common feature of all RBT courses is that they are centered around a **subject that the lecturer is currently working on** or **has extensive research experience with**. To be sure, this does not exclude course scenarios where the students select the narrower topic for their projects. However, these topics should fall within the area of expertise of the lecturer.

The reason for insisting on the research expertise of the lecturer lies in the fact that, as emphasised in previous chapters, the actual research practices of the lecturer are an important and underexploited resource. While designing and teaching an RBT course, the lecturer **identifies research and transferable skills** that can be taught to students.

Whereas on the side of the lecturer we recommend sticking to subjects within the area of expertise, on the side of the students no extensive background should be a prerequisite. Here again, the expertise of the lecturer is what guarantees that the necessary background will be provided (e.g., in the form of the best available overview of the state-of-the-art, short presentation, etc.).

In the selection of the course subject, we recommend considering **which steps in answering the research question can be performed by the students** without extensive assistance by the lecturer and making sure that such steps constitute a majority of the research trajectory. In many cases, there are steps that need to be performed by the lecturer or with extensive help by the lecturer (e.g., the final formulation of the research question, coming up with a data set or performing a complex statistical analysis). In such cases, these steps can be turned into classroom activities, where the students can appreciate what these steps entail and learn how the knowledge lacking at the moment can be acquired in order to perform this step independently.

Maja's testimony

For several years I taught a BA-level (fourth year) Second Language Acquisition course that was not really RBT, but that involved projects that students did in small groups, on topics of their own choosing (mostly related to the acquisition of specific lexical or grammatical phenomena). The students had regular in-class presentations during which they received feedback from me as well as their fellow students, especially in terms of project feasibility. Overall, things tended to work out quite well, and the students seemed to really enjoy the project activities and seeing themselves as researchers. The two most difficult aspects were rather consistently the formulation of specific research questions within a broader topic and the identification of research variables. E.g., many students wanted to work on the acquisition of English articles or the verbal aspect in Slavic languages, but when asked to narrow down these immense topics, they really struggled and did not actually seem to understand what they were being asked to do. And this was after the same students had already worked, in their third year, on similar projects in a Psycholinguistics course, where the questions were given to them. So my overall impression is that even if RBT is applied, the students should be given an opportunity to work on formulating the questions, even if these will eventually be discarded and new ones imposed by the lecturer.

A further beneficial aspect to be considered when choosing a suitable subject for an RBT course is the possibility of **triangulation**. If multiple methodological approaches can be applied in addressing the same research question, this is an important asset to be used at different steps of the research trajectory. Groups of students can work using different methods and exchange their experiences and results.

Boban's testimony

On the one hand, our study programs (Slavic studies) mainly attract students interested in humanities and inclining towards interpretive and other less formal methods. On the other, the general orientation of the UPSKILLS project is to address the demands of industry and general transferable research skills (analytic skills, statistics, use of the infrastructure). In my experience, these two were clashing: our students were often put off by the formal, computational, experimental components in the course descriptions. Overall, the same slots in the curriculum had two to three times more students when I taught them in a traditional fashion than when I used them to pilot research-based teaching.

Finally, many research questions can be addressed by applying the same methodology to **different samples**. In this case, groups of students can work with "their" samples throughout the course and their results can be combined in a single conclusion. While we recommend using the type of data that the lecturer has experience with, the data should not be previously published or analysed (at least using the given methodology). The real-life-like aspect of an RBT course is exactly about the outcome being unknown before the course.

5. Learning outcomes of RBT courses

In this section, we address what students learn in an RBT course with special attention to general research-related outcomes. The topics listed in Table 1 map onto learning outcomes which we provide in Table 2. A more detailed list of specific learning outcomes is available in Annex 3. As will be discussed in the following section, we suggest using this more detailed list when creating an RBT course. Unsurprisingly, subject-specific learning outcomes (the yellow zone) are defined only in a very general way in this overview because they depend on the specific subject of the course.

Topic block	Learning outcomes per topic
Research design	Students will be able to sketch the general research design.
	Students will be able to create a suitable research design for the specific topic of interest.
	Students will be able to adapt a research design to the available research infrastructures.
	Students will be able to report on their performed research in accordance with standards and conventions in the field.
Research infrastructures & techniques	Students will be able to identify and use suitable infrastructures & techniques for obtaining literature
	Students will be able to identify and use suitable infrastructures & techniques for obtaining, sharing and managing data
	Students will be able to identify and use suitable infrastructures & techniques for analysing data
Subject-specific aspects	Students will be able to address a general subject-specific question
	Students will be able to address particular subject-specific questions, carry out particular subject-specific tasks and develop particular subject-specific skills

Table 2. Learning outcomes per topic block in an UPSKILLS course

A more detailed list of specific learning outcomes is available in Annex 3. As will be discussed in the following section, we suggest using this more detailed list when creating an RBT course. Unsurprisingly, subject-specific learning outcomes (the yellow zone) are defined only in a very general way in this overview because they depend on the specific subject of the course.

The learning outcomes in the first two topic blocks reflect **competences necessary in research on which the course was based**, ranging **from very general research skills to**

familiarity with specific techniques and infrastructures. Each of these skills has application beyond the immediate academic context in which it was initially presented.

Well-defined learning outcomes naturally connect to **evaluation tools**. As already mentioned above, we suggest using the **research report as the main deliverable by the student**, on which an important part of the final grade is based. In our experience, the complete report allows for **evaluating the students' work on all of the course outcomes**. However, it does occur that students omit some of the information from the research report necessary for the evaluation. It is therefore commendable to plan at least one round of feedback before the final submission of the research report, so that specific aspects can be further elaborated. We discuss the assessment and evaluation in an UPSKILLS course in further detail in Section 8.

Concluding our discussion of the preliminaries, we now turn to the specific guidelines for creating an RBT course.

PART TWO: GUIDELINES

6. Making a course outline

We open these guidelines with a quick roadmap to making an RBT course outline. In our experience, the best way to use this roadmap is to open a separate document and start taking notes at each step, moving on to the next step only after completing the previous one. Of course, previous steps can be revised based on insights at later steps.

Based on our experiences with RBT courses we also formulated some DOs (👍) and/or DON'Ts (👎) for each step of the trajectory.

Step 1: Having considered Section 3 and suggestions in Section 4, define a course subject.

👍 Pick a subject that allows spending less than 60% of the time on strictly subject-specific issues.

👍 Pick a subject which is of interest to what you are currently working on or have recently worked on.

👍 Pick a subject you would like to do more research on.

Step 2: Having considered the research-report outline (Annex 2), consider what kind of data you would like the students to work on and what techniques they should apply.

👍 Choose data that can realistically be obtained, processed, analysed and interpreted by students (after instruction that you can reasonably provide). Especially consider the technical skills that the students can be assumed to have. We advise being conservative in these estimations. It is better to be pleasantly surprised than the opposite, which will result in you having to heavily support the students or them failing/feeling demotivated due to initial technical obstacles.

👍 Decide early on if any steps in the process need to be performed by you. For instance, you might decide that you have to provide students with data. This is okay, just make sure such steps are not many and can be presented to the students in a meaningful way.

👎 Working on previously unanalysed data is great, but it is better not to leave students with randomly assigned different and unpredictable datasets for the whole course without

building in a possibility of reorganisation, redistribution of work etc. Different datasets may lead to large discrepancies in the workload for different (groups of) students and influence the students' motivation and course achievement.

👎 It may seem tempting to build in some problems into the data initially presented to the students (e.g., data sets from different populations, wrong annotation, problematic variables) and to make it the students' task to fix these problems. This type of twist may be a great exercise in courses where the role of students is less active, but it is not very effective in RBT courses, where the students take the role of a collaborator and should have all the information available to the instructor.

Step 3: Using the detailed overview of topics in Annex 1, make an overview of topics you would like to cover in your course.

👍 The blue zone should be at least 20% of the overview.

👍 The yellow zone should be at most 60% of the overview.

👍 When planning the course, leave approximately 20% of the time for "implicit" aspects that you might take for granted at the beginning but actually need to be addressed explicitly.

Step 4: Using the detailed overview of learning outcomes in Annex 3, make an overview of the learning outcomes you would like students to achieve in your course.

👍 Some of the "implicit" aspects that you discovered are likely to turn into outcomes, so don't make your list too packed.

Step 5: Having the learning outcomes in place, go over and, if necessary, adapt the research-report outline (Annex 2).

👍 Make sure that students get a chance to work individually and in groups, but make the final report individual. Explicitly address how other student's work should be acknowledged in the report.

👍 Ideally, research reports compiled by the students should look like research reports in your discipline. The version that we prepared addresses all the major steps in the process in

separate chapters. We found this useful when working with students with little experience with research reporting. When teaching a more advanced group, consider lumping together some of the steps in the report, following the mores in your field.

In the following sections, we address various further aspects of an RBT course, starting from the organisation of the work.

7. Organising the work in an RBT course

RBT courses lend themselves excellently for an **outcome-based** and **goal-oriented** organisation. The final deliverable by the students, the **research report**, is experienced by the students as a **tangible result of their efforts**, but also as an important **milestone in their academic development**. As a consequence, students find activities much more meaningful if they view them as **a step toward completing their own project report**. We therefore encourage introducing the research-report outline at the outset of the course and referencing the specific parts of the outline whenever possible.

> Student testimony (course *Language acquisition in Slavic*)
>
> *It was much easier and less stressful to write my research report using the provided template in several ways. First of all, I would not remember to include all the aspects of the research to be reported without the kind of check-list that the template presents. Second, in the very process of research, there would be more points at which I would not be sure what to do next. And then also, the structure of the report with its structure is informative about the principles and goals of scientific research.*

Since the research report provides the general **framework within which all specific activities are situated**, we recommend devoting special attention to the **adaptation of our research report template** (**Annex 2**) to the specific course. As mentioned in Section 6: Making a course outline, the version prepared by us has **dedicated chapters** for all the major steps in the research process. We found this useful when working with students who have little experience with research reporting, as it helps them focus on the content of the report without having to address different steps in the same chapter. When teaching a more advanced course, there is more room for making the research report similar to an **actual research paper** by lumping together some of the steps in the report, following the mores in the relevant field.

While the research report is a useful final goal, more 'local' goals and outcomes also have a role to play in the organisation of the work and motivation of the students. In most RBT courses, both teaching and the actual research activities are **divided between homework and in-class activities**. In both cases, the purpose of the specific activities and their function in the 'big picture' of the course goals may be much less obvious to the students than the lecturer might expect. We therefore recommend **starting and closing each activity with an explicit reference to the learning outcomes**. For example, if students are supposed to learn to extract a certain type of data from a corpus, and the lecturer provides a brief presentation on how to do that, this outcome should be explicitly mentioned before and after the presentation. At the next step, when the students are extracting data for their own projects, e.g., as part of their homework, this outcome should be referred to before and after this activity. While this may sound overly formal and potentially tedious, it takes less than a minute and our experience is that this helps students to **appreciate the type and goal of the activity they are engaging in,** as well as to make the most of it. In the absence of explicit tags, some students tend to miss the information on how general the skill is that they are acquiring or disregard the role of this skill in the more general research design.

In most RBT courses, a good part of the steps in the research process is taken by the students as part of their homework. In this case, we recommend formulating **the instructions** on **what needs to be done** and **how it needs to be documented** in the research report. If possible, these instructions should be presented to the students in class and some time (10 to 15 minutes) should be allowed for a Q&A in which any ambiguity is discussed and cleared out of the way. It is actually useful to revise the instructions during the Q&A as necessary, so the same instructions can be reused for another group.

Ideally, the **work submitted by students** consists of **parts of the research report** (together with possible accompanying output, such as annotations, lists, descriptive statistics etc.). These are short enough to make it possible to give a sizable amount of feedback in class. In our experience, many comments apply to work submitted by multiple students. Here as well, we recommend **keeping track of common misinterpretations of the instructions** (e.g., terms which have a specific meaning in the context of the research report). When revising the instructions, the lecturer can consider **clearing the confusing formulations out of the way** by making the instructions or the rubrics in the research-report outline more explicit.

Boban's testimony

In an RBT BA course I taught, a seminar on thematic roles, it turned out that the tasks were appropriately demanding, but that there was a problem with the instructions. After a couple of attempts to improve the instructions, I realized that this won't solve the problem. The simple instructions weren't sufficiently informative,

the detailed ones were too complicated, i.e. included many terms that the students weren't familiar with and opened room for wrong interpretations. The students asked if it was possible that they begin with the task and send me the materials for feedback, and only then finish the task. This worked very well, so for the rest of the course, the assignments were always split into a pilot part to be submitted by the following class, and the rest for one or two weeks later. The pilot was then always discussed in the next class, the unclarities and misunderstandings were taken care of, and then the students could complete the task properly. This is an option that should always be kept in mind - especially at lower levels of study, and implemented at the first sign of troubles with the instructions.

An important advantage of the students' work being centered around a single final deliverable, the research report, is that it also allows both the lecturer and the students to **keep track of the progress** made during the course. For the student, the research report can also serve as a useful indication of the number of completed steps and of the steps that still lie ahead.

Clearly, all research-based courses require the lecturer to **provide some necessary background**, both on general research-related issues and on subject-specific issues. In our experience, students tend to find it useful if such less active parts of learning are clearly delineated and if the need for them is explicated. One option we recommend is providing the necessary background through a short unit that includes the following parts:

- preparatory reading,
- short presentation by the teacher,
- Q&A
- short (but obligatory) quiz.

While many of our courses have been taught to small groups and, in extreme cases, to single students, we have had some positive experiences with **group work** in RBT courses. In many courses, a sizable number of steps in the research process can be performed in a group. In this case, the groups are asked to submit a **common report in which they also specify the specific contribution of the group members**. These reports can then easily be included in the individual research reports.

Quite independently from the amount of work done in groups, we recommend securing a **common work space** for the entire class both physically and virtually. Furthermore, a cloud or physical **repository**, where students can share their results, give each other feedback, etc. is a useful addition to any RBT course.

8. Supervision, assessment and evaluation

The main bulk of individual supervision in an RBT course takes the form of **feedback on intermediate versions** of the research report. At least in some academic cultures, it is unusual for students to get a lot of feedback on the piece of work that eventually gets evaluated as the final deliverable of the course, so it is useful to be explicit on **how the final grade is calculated** and what contributes to it. This is also a way to ensure that the students feel free to try out approaches they are not perfectly confident about in intermediate versions of their research reports.

As previewed above, the **main deliverable** which also plays a central role in the assessment and evaluation is the **final research report**. Throughout the course, this research report grows with regular feedback provided both by the lecturer and by peers.

In most RBT courses, there are **other essential activities that can be evaluated**, e.g., quizzes, group reports, intermediate reports etc. As mentioned above, evaluating intermediate versions of (parts of) the research report carries the inherent risk that the students may not be as creative as in the scenario where they just get feedback on intermediate versions. We therefore recommend including intermediate versions as **necessary milestones towards the final research report** without evaluating them separately (or evaluating them with a pass/fail mark in systems where this is common).

In Figure 4, we give an overview of the general **evaluation strategy**.

Rubric	Weighing
Participation incl. homework (initiative, forward-thinking, problem solving, critical thinking, organisational skills, time management)	30%
Outputs based on the final research report ○ oral presentation ○ written report	70%

Figure 4. General evaluation strategy for RBT courses

The first part of the mark reflects the **lecturer's general impression of the student's participation in the course**. Especially if there is a natural midterm evaluation point, we recommend giving **intermediate feedback** on this aspect to the students.

For output based on the research report, we suggest using the following evaluation form, which also includes questions to be answered in giving the specific mark.

Take into consideration each of the criteria below that apply to work produced in your course. 👍 Make sure that at least two of the criteria predominantly involve transferable skills (e.g., the students choose and apply appropriate statistical tests, build a database, interview informants) and mark those criteria. Give those criteria a higher weight in the evaluation. 👍 Make sure to keep track of what the students learned in the course (so you don't expect them to know things you never taught them). Give the final mark.	
Criterion	Grade
General	
Comprehensibility Is the output comprehensible for audiences who did not participate in its creation (e.g., can a group of students present their research in a way comprehensible to other students)?	
Coherence Is the output internally coherent (logical flow of ideas, no contradictions, consistent terminology, consistent referencing style, etc.)?	
Exploiting conventions Does the output make use of the conventions typically used in the field (structure, terminology, artwork, a referencing style, etc.)?	
Research-related	
General understanding of the research design Does the output reflect a clearly defined and appropriate research design?	
Formulation of research questions and hypotheses Are the research questions and hypotheses clearly formulated?	
Formulation of predictions of the null hypothesis and alternative hypothesis	

Are the predictions of the null hypothesis and alternative hypothesis clearly defined?	
Validity of the employed research design Does the employed research design adequately address the research question(s)?	
Familiarity with the relevant resources, tools and infrastructures presented in class Does the output reflect sufficient familiarity with the relevant resources, tools and infrastructures presented in class?	
Selection of data sources and research techniques Is the selection of data sources and research techniques sufficiently justified?	
Selection of the data analysis method Is the selection of the data analysis method sufficiently justified?	
Inferring theoretical consequences from the specific data analysis results Does the output reflect the ability to infer broader theoretical consequences from the analysis results?	

Table 3. Evaluation for outputs based on the final report in an RBT course

9. After the course

One of the important results of many (ideally all) RBT courses are **reusable research results**. A way to make sure that these results get properly reused (and the student's work gets credited) is having the students **store their data in a repository as the final course activity**. The lecturers who would like to know more about teaching the students how to deposit their dataset in a repository will profit from the guidelines, *Integrating Research Infrastructures into Teaching: Recommendations and Best Practices* (van der Lek, I. and Fišer, D., 2023) and the accompanying learning content on Moodle, *Introduction to Language Data: Standards and Repositories*.

Offering **feedback on the final research report** (in addition to the final mark) means additional work, but students tend to appreciate it. One tool that can be used for this is the **evaluation form in Table 3 with answers supported by examples from the report**. One advantage of using this tool over writing comments in the report is that positive aspects are elaborated on as well.

We also recommend **debriefing the students** after the course ends and (if applicable) informing them about **the wider academic context of the course**, e.g., the UPSKILLS project, further RBT courses etc.

Once the students have received final feedback, it is useful to see how **they perceived the course**. We created a **survey for course evaluation by students**, using Google Forms, available at https://forms.gle/gebgGK3vt9RWFUXU8 and included in **Annex 4** which can be used for these purposes. Note that you need to create The survey can be further adapted and combined with assessment forms used at the institution where the course was taught.

Finally, we welcome **feedback from the users of these guidelines**. We created a dedicated survey to be filled in upon the completion of a RBT course for which the guidelines were used. The survey is available at https://forms.gle/aPUCn6wY4dP52Ldg9.

You can also send any type of feedback to Marko Simonovic from the University of Graz at marko.simonovic@uni-graz.at.

References

Bastiaens E., van Tilburg J., van Merriënboer J. (eds.) (2016). *Research-Based Learning: Case Studies from Maastricht University*. Professional Learning and Development in Schools and Higher Education, vol 15. Cham: Springer. https://doi.org/10.1007/978-3-319-50993-8

Dekker, H., & Wolf, W.S. (2016). Re-inventing research-based teaching and learning. Paper presented at the *European Forum for Enhanced Collaboration in Teaching of the European University association,* Brussels.

Feiman-Nemser, S., & Remillard, J. (1996). Perspectives on learning to teach. In F.B. Murray (ed.), *The Teacher Educator's Handbook* (pp. 63-91). San Francisco, CA: Jossey-Bass.

Giraud, F. and Saulpic, O. (2019). Research-based teaching or teaching-based research: Analysis of a teaching content elaboration process. *Qualitative Research in Accounting & Management*, 16(4), 563-588. https://doi.org/10.1108/QRAM-10-2017-0097

Healey, M. (2005). Linking research and teaching exploring disciplinary spaces and the role of inquiry-based learning. In Barnett, R.(ed.) *Reshaping the University: New Relationships between Research, Scholarship and Teaching* (pp. 67-78). Maidenhead: Open University Press.

Jenkins, A., Healey, M., & Zetter, R. (2007). Linking teaching and research in departments and disciplines. York, UK: The Higher Education Academy.

Miličević Petrović, M., Bernardini, S., Ferraresi, A., Aragrande, G., & Barrón-Cedeño, A. (2021). Language data and project specialist: A new modular profile for graduates in language-related disciplines. *UPSKILLS task report*. https://doi.org/10.5281/zenodo.5030929

van der Lek, I. & Fišer, D. (2023) Integrating Research Infrastructures into Teaching: Recommendations and Best Practices. https://doi.org/10.5281/zenodo.8114407

van der Plas, L. & Assimakopoulos, S. (2023) Integrating industry-based research projects into teaching: Opportunities and pitfalls. *UPSKILLS task report*. https://doi.org/10.5281/zenodo.5030929

Visser-Wijnveen, G. J. (2009). *The research-teaching nexus in the humanities: Variations among academics*. Phd thesis. Leiden University.

Visser-Wijnveen, G. J., Van Driel, J. H., Van der Rijst, R. M., Verloop, N., & Visser, A. (2010). The ideal research-teaching nexus in the eyes of academics: building profiles. *Higher Education Research & Development*, 29(2), 195-210. https://doi.org/10.1080/07294360903532016

Annex 1 - Detailed overview of general topics in RBT courses

A: Research design
A1: General research design
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Adapting the general research design to the specific topic of interest
A2.1: Formulation of questions and hypotheses in terms of variables
A2.2: Formulation of predictions of H0 and H1
A2.3: Selection of appropriate research techniques, selection and creation of corresponding data sources • Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading) • Developing and exploiting databases and corpora (e.g. manual data annotation, computing inter-annotator agreement)
A2.4: Identifying the optimal data analysis method
A2.5: Inferring theoretical consequences from the specific data analysis results
A3: Adapting the research design to the available research infrastructures
A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3) • data compilation, data analysis, data archiving (e.g., XML, XLS), data reuse; • understanding, selecting and performing optimal statistical tests and models
A4: Research reporting

A4.1 Presentation modes for research reporting (short oral presentation, poster, squib, report, article etc.)

A4.2 Established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/squib/report,
- organization of the presentation,
- amount of text and graphical items on a poster (including text size),
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions

B: Infrastructures & techniques

B1: For obtaining literature

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, ROA

B2: For obtaining, sharing and managing data

B2.1: Definition of research infrastructures are, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**

B2.2: Platforms and repositories

- **General-purpose repositories** and **disciplinary repositories**
  - [GENERAL-PURPOSE REPOSITORY] Zenodo, FigShare
  - [DISCIPLINARY REPOSITORY] CLARIN, The Language Archive

B2.3:Identifying, collecting, creating and/or using relevant data for research projects

- Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them
  - [DISCIPLINARY REPOSITORY] CLARIN, ELRC-SHARE, the Language Archive
- Citing linguistic data sets as appropriate
- Depositing research data in a **certified repository** and selecting an appropriate licence for sharing their data
- The **versioning** policy of repositories
- Familiarity with online survey tools

B2.4: **Data management plan**

- Understanding the **data lifecycle**
- Understanding how to generate data, analyse and handle it

■ Understanding the **legal and ethical issues** around data generation and use (e.g. licensing, GDPR compliance, anonymisation, the importance of FAIR principles and Open Access) ■ Secure storage and backup of research data ■ Documenting workflows and what metadata to use to describe the nature of the data based on existing standards ■ What data needs to be destroyed, preserved in a data repository and made available for reuse
B3: For analysing data
B3.1: Software for statistical tests
B3.2: Software for conducting a phonological analysis (e.g. Praat)
B3.3: Concordancers for the analysis of corpora

Annex 2 - The research report template

This research report template is adapted from the one created by Tanja Samardžić, Maja Miličević Petrović and Genoveva Puskas within the course "Introduction to research in linguistics: theory, logic, method"[5].

---

1. Introduction
- a brief definition/description of the phenomenon
- formulation of a) the research problem (a gap in the literature): what the other studies got wrong or missed or can't agree on about the phenomenon and your study will fix/find out, or b) the engineering problem - something that should work, but it doesn't
- significance of your study: why does the world (or at least your research community) need to know about your study, why is it important

2. The goal/aim
- a clear and precise statement of what you wish to achieve in your study
- what is the challenge and what is the significance of your contribution in the narrow light of the research problem

3. State of the art
- theoretical background (relevant theories and concepts)
- a critical review of previous relevant research (what is already known about the phenomenon and in what way the previous studies are limited)
- further elaboration of the research problem (a gap in the literature): a detailed description of what the other studies got wrong or missed or can't agree on and your study will fix/find out

4. Present approach, or the proposed solution for the engineering problem
- your research question(s): what do you want to find out about the phenomenon
- expected or possible answers (hypotheses), i.e. rationale behind the proposal for the engineering problems

5. Data and methods
- what information (=data) you collect
- where from or from whom you collect it
- how you collect it
- how you organise the collected data
- what analyses you apply

[5] The course is available at https://elearn.mnf.uzh.ch/courses/course-v1:PHIL+Movetia101+2022/about

- what tools you use to perform the analyses

6. Findings
- counts, percentages, outcomes of statistical tests, often given in tables or graphs
- descriptions of these tables and graphs
- and/or clear statements containing new facts established in your study

7. Interpretation
- relating the findings to the starting expectations/hypotheses and/or research question(s)
- how your findings improve the knowledge about the phenomenon you studied

8. Discussion
- relating your findings to the relevant theories and previous empirical findings (pointing to similarities and differences between the two sets of findings)
- obstacles encountered
- alternative explanations - speculations
- limitations of your study
- directions for future research
- broader relevance - theoretical and practical implications of the findings

9. Conclusion
- synthesis, the main message of your study

10. References

Annex 3 - Detailed overview of learning outcomes per topic block in RBT courses

A: Research design
A1: Students will be able to make an overview of the general research design.
UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Students will be able to create a suitable research design for the specific topic of interest.
A2.1: Students will be able to formulate questions and hypothesis in terms of variables
A2.2: Students will be able to formulate H0 and H1
A2.3: Students will be able to select optimal research techniques, and create corresponding data sources ● Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading) ● Developing and exploiting databases and corpora (e.g., manual data annotation, computing inter-annotator agreement)
A2.4: Students will be able to select and implement the optimal data analysis method
A2.5: Students will be able to infer theoretical consequences from the specific data analysis results.
A3: Students will be able to adapt a research design to the available research infrastructures.
A3.1 Students will be able to select of optimal research techniques, select and create corresponding data sources (see also A2.3) ● data compilation, data analysis, data archiving (e.g., XML, XLS), data reuse; ● understanding, selecting and performing optimal statistical tests and models.

A4: Students will be able to report on their performed research in accordance with standards and conventions in the field.
A4.1 Students will be able to select and implement different presentation modes for research reporting (short oral presentation, poster, squib, report, article etc.)
A4.2 Students will be able to implement established procedures and conventions in research reporting, such as: • the ordering of thematic units in an article/squib/report, • organization of the presentation, • amount of text and graphical items on a poster (including text size), • amount of text and graphical items on a slide/handout, • terminology, • citing conventions.
B: Infrastructures & techniques
B1: Students will be able to identify and apply suitable infrastructures & techniques for obtaining literature
[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, Rutgers Optimality Archive.
B2: Students will be able to identify and apply suitable infrastructures & techniques for obtaining, sharing and managing data
B2.1: Students will understand what research infrastructures are, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**.

B2.2: Students will be able to identify suitable platforms and repositories.
• Understand the difference between **general-purpose repositories** and **disciplinary repositories**
  ○ [GENERAL-PURPOSE REPOSITORY] Zenodo, FigShare
  ○ [DISCIPLINARY REPOSITORY] CLARIN, The Language Archive

B2.3: Students will be able to identify, collect, create and/or use relevant data for their research projects • Search, identify and select relevant corpora from language resources platforms and repositories hosting them   ○ [DISCIPLINARY REPOSITORY] CLARIN, ELRC-SHARE, the Language Archive. • Cite linguistic data sets as appropriate. • Deposit their research data in a **certified repository** of their choice and select an appropriate licence for sharing their data

- Understand the **versioning** policy of the repository
- Use online survey tools

B2.4: Students will be able to create a **data management plan**

- Understand the **data lifecycle**
- Understand how to generate data, analyse and handle it
- Understand the **legal and ethical issues** around data generation and use (e.g. licensing, GDPR compliance, anonymisation, the importance of FAIR principles and Open Access).
- Know how to securely store and backup their research data
- Know how to document their workflows and what metadata to use to describe the nature of the data based on existing standards.
- Know what data needs to be destroyed, preserved in a data repository and made available for reuse.

B3: Students will be able to identify and apply suitable infrastructures & techniques for analysing data.

B3.1: Students will be able to select and use softwares for statistical tests

B3.2: Students will be able to select and use softwares for conducting a phonological analysis (e.g. Praat)

B3.3: Students will be able to select and use concordancers for the analysis of corpora

Annex 4 - Survey for course evaluation by students

The survey is available at https://forms.gle/gebgGK3vt9RWFUXU8
For the users' convenience we copy the contents of the survey below.

Research and analytic skills: Survey for students

Section 1 of 6

Questionnaire about your research and problem-solving skills upon completion of an UPSKILLS course

The course you completed was part of an Erasmus+ strategic partnership project UPSKILLS (UPgrading the SKIlls of Linguistics and Language Students). One of the goals was improving the general research and problem-solving skills in students of language-related areas.

This questionnaire takes between 5 and 10 minutes to complete and consists of the following parts.

A. Questions about your background
B. Information on the UPSKILLS course
C. Assessment of your progress in the area of research and problem-solving skills
D. Assessment of the generalisability of the acquired research and problem-solving skills
E. End of survey

The aim of this questionnaire is to collect some insights about the courses which integrate research into teaching. The insights will be used in the UPSKILLS project (an ERASMUS+ project) to develop guidelines about "Best practices for integrating research into teaching."

Further information about the project can be found at https://upskillsproject.eu/. If you have questions about the survey, please contact Marko Simonovic from the University of Graz at rkicma@gmail.com.

The information provided by you in this questionnaire will be used for research purposes. It will not be used in a manner which would allow the identification of your individual responses. Anonymised research data will be archived in order to make it available to other researchers in UPSKILLS in line with the current data sharing practices.

Section 2 of 6

A. General questions

This section includes general questions about your background.

A1. Which area is the main focus of your studies? (Check all boxes that apply)*

- Cultural Studies
- Literary Studies
- Linguistics
- Language and/or speech technology
- Linguistic data management
- Language teaching
- Translation & Interpreting Studies
- Translation
- My main focus area is not language-related.
- Other…

A2. If there are specific academic areas in which you are specialising, please list them below.

A3. In which country/countries do you study?*

A4. Are you interested in pursuing a career in academia?*

Yes
No
Maybe
Other…

Section 3 of 6

B. The UPSKILLS course

This section includes a set of questions about the UPSKILLS course that you have taken.

B1. At what level did you take the UPSKILLS course? (Check all boxes that apply)*

BA

MA
PhD
Postgraduate course
Other…

B2. What were some of the topics that the course focused on? (Check everything that applies.)*

Anthropological Linguistics
Applied Linguistics
Clinical Linguistics
Cognitive Science
Computational Linguistics
Computer-Aided Translation
Digital Humanities
Discourse Analysis
Forensic Linguistics
General Linguistics
Historical Linguistics
Language Acquisition
Language Technology
Language Documentation
Lexicography
Literature
Morphosyntax
Neurolinguistics
Philosophy of Language
Phonetics/Phonology
Psycholinguistics
Semantics/Pragmatics
Sociolinguistics
Speech Technology
Text/Corpus Linguistics
Translation
Translation Studies
Typology
Writing Systems
Other…

B3. What general grade would you give the UPSKILLS course?*

I did not like the course at all
1
2

3
4
5
6
7
8
9
10
I liked the course very much

Please add any further comment(s) you may have regarding the previous question.

Section 4 of 6

C. Assessment of your progress in the area of research and problem-solving skills

This section includes questions about your assessment of your own progress in the area of research and problem-solving skills.

Please evaluate your progress on each of these skills on a scale from 1 to 10.

C1. Identifying relevant literature*

I did not learn anything new.
1
2
3
4
5
6
7
8
9
10
I learned much more than I expected.

C2. Identifying relevant infrastructures and digital tools for data collection and processing (corpora, databases, experimental stimuli etc).*

I did not learn anything new.
1
2
3
4
5
6

7
8
9
10
I learned much more than I expected.

C3. General research design*
I did not learn anything new.
1
2
3
4
5
6
7
8
9
10
I learned much more than I expected.

C4. Formulation of research questions and hypothesis in terms of variables*
I did not learn anything new.
1
2
3
4
5
6
7
8
9
10
I learned much more than I expected.

C5. Formulation of predictions of the null hypothesis and alternative hypothesis*
I did not learn anything new.
1
2
3
4
5
6

7
8
9
10
I learned much more than I expected.

C8. Selection of optimal data sources and research techniques*
I did not learn anything new.
1
2
3
4
5
6
7
8
9
10
I learned much more than I expected.

C7. Identifying the optimal data analysis method*
I did not learn anything new.
1
2
3
4
5
6
7
8
9
10
I learned much more than I expected.

C8. Inferring theoretical consequences from the specific data analysis results*
I did not learn anything new.
1
2
3
4
5
6

7
8
9
10
I learned much more than I expected.

C9. Oral presentation of research output*
I did not learn anything new.
1
2
3
4
5
6
7
8
9
10
I learned much more than I expected.

C10. Written presentation of research output*
I did not learn anything new.
1
2
3
4
5
6
7
8
9
10
I learned much more than I expected.

Please add any further comment(s) about the questions in this section.

Section 5 of 6

D. Generalisability of the research and problem-solving skills

This section includes questions about the generalisability of the research and problem-solving skills acquired in the UPSKILLS course beyond the topic of the course.

Please mark to what extent you agree with the following statements.

D1. I can use the research and problem-solving skills acquired in this course for other language-related courses and activities in my studies.

*

I fully disagree.

1

2

3

4

5

6

7

8

9

10

I fully agree.

D2. I can use the research and problem-solving skills acquired in this course for non-language-related subjects and activities in my studies. (Skip if all subjects and actifivities in your studies are language-related.)

I fully disagree.

1

2

3

4

5

6

7

8

9

10

I fully agree.

D3. The research and problem-solving skills acquired in this course may improve my employability in language-related jobs.*

I fully disagree.

1

2

3

4

5

6

7
8
9
10
I fully agree.

D4. The research and problem-solving skills acquired in this course may improve my employability beyond language-related jobs.*

I fully disagree.
1
2
3
4
5
6
7
8
9
10
I fully agree.

D5. Please add any further comment(s) about the questions in this section.

Section 6 of 6

E. End of the questionnaire and follow-up

This concludes the questionnaire!

Please click on the 'Submit' button at the end of this page to register your answers.

Thank you very much for your input!

Annex 5 - Examples of RBT courses

This section presents sixteen fully developed courses (at all levels, from BA to PhD) that were taught during the UPSKILLS project. They can be used as examples of best practices for the implementation of the above course outline with specific course topics. Furthermore, the lecturers who would like to target thematic domains related to those targeted by our courses can use these courses as their starting point in creating their own courses. In the topic overview, the link between the general topic and the specific topic is described in green. All the course descriptions are based on courses taught by the UPSKILLS project partners in the course of the project. The table below contains the UPSKILLS course description template.

i.	**Name of the course**
ii.	**Level of the course**
iii.	**Workload**
iv.	**Institution**
v.	**Course instructor(s)**
vi.	**Brief course description (Suggested structure: Why? What? How?)**
vii.	**Research related subject**
viii.	**Tools and data the students work with**
ix.	**Topics**
x.	**Learning outcomes**
xi.	**Evaluation**
xii.	**Further information on the course that the instructor considers relevant (assessment, background, reading materials, detailed weekly plan, career paths etc.)**

Table 4. Course description template for UPSKILLS RBT courses

The first part describes courses taught at the University of Graz which were tightly connected with the Austrian-Slovenian research project "Hyperspacing the Verb: The interplay between prosody, morphology and semantics in the Western South Slavic verbal domain". This project has been carried out between 2019 and 2023 at the Slavic department of the University of Graz (in collaboration with the University of Nova Gorica, Slovenia). The specific course topics illustrated in the description are therefore related to Slavic morphophonology and morphosyntax. In the second part, we focus on courses which are not directly related to the specific project and focus on subjects related to research of the individual instructors.

I. UPSKILLS meets Hypespacing: RBT courses and a specific research project

Hyperspacing the verb is a research project aimed at informing the theory of verbal semantics, syntax, morphology and phonology in Western South Slavic languages (Slovenian and Bosnian / Croatian / Montenegrin / Serbian (BCMS), henceforth WSS) from quantitative empirical insights at the lexical level (i.e. targeting those properties that may be taken to hold for a particular verb). One of the deliverables of the project is the database WeSoSlaV (the Western South Slavic Verb Database) consisting of over 8000 WSS verbs annotated for a range of relevant properties, including aspect, argument structure, conjugation classes, affixes, prosody. As indicated in the UPSKILLS project description, in order to develop and exemplify some ways in which research projects can be instrumentalized in research-based teaching, we developed a series of research-based courses which centrally employ WeSoSlaV in its research orientation.

There are essentially three ways in which WeSoSlaV was employed in these courses. One is through research activities conducted to supplement it with the data it does not yet contain. Such is the case with the course Correlates of thematic roles in BCMS, in which the students are trained in the skills of corpus retrieval, annotation of linguistic material, inter-annotator agreement and resolution, while annotating the verbs for a set of thematic roles - information previously not included in the database. While the course also goes into the theoretical consequences of the annotated properties and their interaction with other properties in WeSoSlaV, the main interaction with the database is in enriching it with new information.

The second role that WeSoSlaV played was that of the main source of information to be computed. Here, the students were instructed to derive quantitative insights from the relations between various properties annotated in the database, applying various statistical tools. No course was solely based on this way of using WeSoSlaV, but it did play the main role in several of them, such as the course Secondary imperfectivisation in Bosnian/Croatian/Serbian, in

which the students observed other relevant properties of those verbs that were annotated as secondarily imperfectivized in the database (but also for some properties which are not included in WeSoSlaV).

Finally, there were also courses which took WeSoSlaV as the starting point, drew from it some relevant quantitative insights, and then developed in a new direction which does not speak back to the database - whether by focusing on theoretical issues, or by going into new quantitative domains. Such was, for instance, the course Theoretical and methodological basics: Multipurpose Suffixes, in which the database provided material only for some uses of the multipurpose suffixes discussed, and then further investigation provided data about other uses of the same suffixes, leading together to a theoretical model.

All these three implementations share some content - that related to the use of the database (e.g., retrieval, various different representations of the data and reorganizations of their representation, techniques for excerpting and deriving quantitative information as well as for its statistical computation), yet they all also involve different embedding of skills, and different additional research practices.

WeSoSlaV plays an important role in providing readily available material that would otherwise take too much time and effort to prepare for the type of courses that are taught at the BA and MA levels, and any building on top of that would be impossible. It also gave the students the opportunity to work in a real research environment, participating in the entire process as pursued in the project, as well as a chance to interact with the developers in the process of the development of the database. Finally, especially in those courses which also contributed to the enrichment of the database, or to deriving non-trivial insights, the students had the experience of seeing the results of their research materialized in an ongoing research project.

Part of our intention in designing the courses was exactly to diversify all these various aspects, and illustrate the broad range of ways in which a research result as simple as an annotated database of verbs from two languages can be instrumentalized to serve a variety of different purposes, i.e. in a variety of different research skills and techniques, of different levels of grammar and individual phenomena, and also of various types of empirical and theoretical reasoning and argumentation.

I.I. Course Description 1: Secondary imperfectivisation in Bosnian/ Croatian/Serbian

i.	**Name of the course** Secondary imperfectivisation in Bosnian/Croatian/Serbian (BCS)
ii.	**Level of the course** MA, PhD
iii.	**Workload**

5 ECTS

iv. Institution

University of Graz

v. Course instructor(s)

Boban Arsenijevic, Marko Simonovic

vi. Brief course description

The course provides a state-of-the-art overview of the literature and ongoing research of one of the most hotly debated issues in Slavic linguistics. The course enhances the problem-solving and data-analysis skills, thus preparing the students for a wide range of possible careers. It also provides the students with first-hand scientific research experience.

A collection of state-of-the-art articles on the topic are read and discussed in each class. This discussion amounts to approximately 50% of the class.

For the second half of the class, the students are asked to focus on one of the claims made in the article read for the respective class (if the article is not about BCS, the students verify whether the claim about the other Slavic language applies to BCS) and present the results of their small-size empirical research testing this claim. To this end, the students are given access to the database of the project *Hyperspacing the verb*, where 5000+ verbs are annotated for various phonological, morphological, syntactic and semantic features. The students are also asked to take the ReLDI course *Introduction to Corpus-Based Methods in Linguistics.*

In all but two last classes all students present the results of their small-size empirical research and their presentations get discussed. In the two last classes, the students hold their final presentations in which they present the research on which they chose to submit their final report.

At the end of the course, the students submit their final report using the template created by Tanja Samardžić, Maja Miličević Petrović and Genoveva Puskas within the course "Introduction to research in linguistics: theory, logic, method" (https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info), which was eventually which was eventually adapted into the UPSKILLS research report template.

vii. Research related subject

Morphology and semantics of the secondary imperfectivization of the Slavic type.

viii. Data the students work with

BCMS corpora hrWac, srWac, bsWac, the databases of the project Hyperspacing the verb.

ix. Overview of topics

A: Research design
A1: General research design
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Adapting the general research design to the specific topic of interest
Identifying the patterns of secondary imperfectivization and both their predictors (thematic vowel, prosodic pattern, final segment, aspectual properties of the base, other possible factors) and the predictors of the very possibility of secondary imperfectivization of a verb (aspectual class, other possible factors)
A2.1: Formulation of questions and hypothesis in terms of variables
A2.2: Formulation of predictions of H0 and H1
A2.3: Selection of optimal research techniques, selection and creation of corresponding data sources ● Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading) ● Developing and exploiting databases and corpora (e.g. manual data annotation)
A2.4: Identifying the optimal data analysis method
A2.5: Inferring theoretical consequences from the specific data analysis results
A3: Adapting the research design to the available research infrastructures
Familiarising with the type of data extractable from the available corpora; Familiarising with the type of data extractable from online surveys; Choosing the optimal way of annotating the properties
A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3) ● data compilation, data analysis;

- understanding, selecting and performing optimal statistical tests and models,
- identifying the optimal data analysis method

A4: Research reporting

Developing and formulating formalizations of the relevant semantic regularities in terms of the lambda calculus. Selecting the optimal formal morphological framework and using it to present the observed morphological regularities

A4.1 Presentation modes for research reporting (short oral presentation, report)

A4.2 Established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/squib/report,
- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions

B: Infrastructures & techniques

B1: For obtaining literature

[GENERAL-PURPOSE REPOSITORY] Research Gate, Google Scholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, semantic scholar

B2: For obtaining, sharing and managing data

Advanced use of the available corpora for the target language, advance use of online surveys (use of CQLs, use of additional options)

B2.1: Definition of research infrastructures, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**

B2.2: Platforms and repositories

B2.3:Identifying, collecting, creating and/or using relevant data for research projects

- Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them
- Citing linguistic data sets as appropriate.
- Familiarity with online survey tools

B3: For analysing data

B3.1: Softwares for statistical analysis (R)

B3.2: Corpus management and text analysis software (NoSketch Engine, AntConc)

C: Subject-specific topics
C1: What is verbal aspect?
C2: How does the verbal aspect relate to verbal morphology in Slavic and beyond?
C3: Why and when is there optionality?
C4: What are characteristic properties of Slavic verbal morphology?
C5: Basic concepts of Formal Semantics

x. Learning outcomes

A: Research design
A1: Students will be able to make an overview of the general research design.
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Students will be able to create a suitable research design for the specific topic of interest.
Students will be able to identify the patterns of secondary imperfectivization, their predictors and the predictors of the possibility of secondary imperfectivization of a verb (aspectual class, other possible factors).
A2.1: Students will be able to formulate questions and hypotheses in terms of variables.
A2.2: Students will be able to formulate H0 and H1.
A2.3: Students will be able to select optimal research techniques, and create corresponding data sources. • Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading) • Developing and exploiting databases and corpora (e.g. manual data annotation).
A2.4: Students will be able to select and implement the optimal data analysis method.

A2.5: Students will be able to infer theoretical consequences from the specific data analysis results.
A3: Students will be able to adapt a research design to the available research infrastructures.
Students will be familiar with the type of data extractable from the available corpora; Students will be familiar with the type of data extractable from online surveys; Students will be able to choose the optimal way to annotate the relevant semantic and morphological properties.
A3.1 Students will be able to select of optimal research techniques, select and create corresponding data sources (see also A2.3) ● data compilation, data analysis; ● understanding, selecting and performing optimal statistical tests and models.
A4: Students will be able to report on their performed research in accordance with standards and conventions in the field.
Students will be able to represent the relevant semantic regularities in terms of the lambda calculus. Students will be able to select the optimal formal morphological framework for their data and use it to represent the observed morphological regularities.
A4.1 Students will be able to select and implement different presentation modes for research reporting (short oral presentation, report, article etc.).
A4.2 Students will be able to implement established procedures and conventions in research reporting, such as: ● the ordering of thematic units in an article/report, ● organization of the presentation, ● amount of text and graphical items on a slide/handout, ● terminology, ● citing conventions.
B: Infrastructures & techniques
B1: Students will be able to identify and apply suitable infrastructures & techniques for obtaining literature.
[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, semanticscholar.
B2: Students will be able to identify and apply suitable infrastructures &

techniques for obtaining, sharing and managing data.
Students will be able to extract data from the available corpora for the target language, Students will be able to obtain data in online surveys.
B2.1: Students will understand what research infrastructures are, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**.
B2.2: Students will be able to identify suitable platforms and repositories.
B2.3: Students will be able to identify, collect, create and/or use relevant data for their research projects • Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them • Citing linguistic data sets as appropriate. • Familiarity with online survey tools.
B3: Students will be able to identify and apply suitable infrastructures & techniques for analysing data.
B3.1: Students will be able to select and use softwares for statistical tests.
B3.4: Students will be able to select and use concordancers for the analysis of corpora.
C: Subject-specific learning outcomes
C1: Students will be able to discuss the core issues of verbal aspect.
C2: Students will be able to discuss the relation between verbal aspect and verbal morphology in Slavic.
C3: Students will be able to identify the potential loci of optionality regarding the pattern of secondary imperfectivization in Slavic.
C4: Students will be able to apply morphological and semantic models onto linguistic data.
C5: Students will be able to represent semantic insights in terms of lambda calculus formulae and to interpret such representations.

xi. Overview of evaluation	
Rubric	Weighing
Participation in classes (initiative, forward thinking, problem solving, critical thinking, organisation, time	20%

management)	
Homework (data collecting, annotation, analysis)	20%
Outputs based on the final research report ○ oral presentation ○ final written report	60%

xii. Career paths

a. Academia
b. Natural language processing
c. Market analysis

and all careers involving data analysis

xiii. Reading materials

Arsenijević, B. (2013). The for-phrase. Philologia Mediana, 515–538.

Łazorczyk, A. A. (2010). Decomposing Slavic Aspect: the Role of Aspectual Morphology in Polish and Other Slavic Languages [University of Southern California]: Chapter 4: Secondary imperfectives, pp. 104-161.

Markman, V. G. (2008). On Slavic semelfactives and secondary imperfectives: Implications for the split 'AspP.' Proceedings of the 31st Annual Penn Linguistics, Article 20.

Polančec, J. (2018). Osamostaljeni izvedeni nesvršeni glagoli u hrvatskom jeziku. Suvremena Lingvistika, 85, 113–138.

Simonović, Marko and Petra Mišmaš. 2020. √ov is in the air: The extreme multifunctionality of the Slovenian affix ov. Linguistica 60.1. DOI: 10.4312/linguistica.60.1.83-102.

Tatevosov, S. (2015). Severing imperfectivity from the verb. In G. Zybatow, P. Biskup, M. Guhl, C. Hurtig, O. Mueller-Reichau, & M. Yastrebova (Eds.), Slavic Grammar from a Formal Perspective (pp. 465–494). Peter Lang.

Tatevosov, S. (2018). On the aspectual architecture of Russian. Ms. Mosscow: esp. Sect. 3: Imperfectivity and "imperfective" morphology.

I.II. Course Description 2: Correlates of thematic roles in Bosnian/Croatian/Montenegrin/Serbian

i.	**Name of the course** Correlates of thematic roles in Bosnian/Croatian/Montenegrin/Serbian
ii.	**Level of the course** BA advanced
iii.	**Workload** 6 ECTS
iv.	**Institution** University of Graz
v.	**Course instructor(s)** Boban Arsenijević
vi.	**Brief course description**

This course focuses on the notion of thematic roles assigned by the verb, and brings it in connection with the much debated topic of the syntactic/semantic contribution of theme vowels.
The course enhances the problem-solving and data-analysis skills of the students, thus preparing the students for a wide range of possible careers. It also provides the students with first-hand scientific research experience, including the development of a database, the annotation and handling of the disagreement between annotators, quantitative analysis.
The course starts (the first three weeks) with a very basic sketch of general research design and a general introduction into the thematic and participant roles. Simultaneously, some preliminary discussion is held about testing the predictions of various theories. After this, students are assigned 1000 verbs from the database of the ongoing research project Hyperspacing the verb. The students first excerpt a sample of 200 randomly selected concordances of each of the verbs assigned to them from the hrWac corpus. Then the students build a database in which each of the verbs they have been assigned is annotated for the number of times it takes each of the following thematic roles:

- agent,
- patient,
- theme,
- recipient and
- experiencer.

The thematic roles are identified using two tests for each role. Particular attention is paid to the cases where the two tests give different results for the same concordance. The students are then expected to discuss the reason why such mismatches emerge, and what they tell us about the verb and about the thematic role.
Halfway through the course, the students submit their final annotations, on which inter-annotator agreement is computed. One session is devoted to the discussion of the results. In the second half of the course, the students compute and discuss correlations between morphosyntactic properties of the verb (such as theme vowels, lexical and grammatical aspect) and the likelihood that the verb selects each of the targeted thematic roles.

At the end of the course, students are expected to submit a final report in which they discuss the established correlations or lack thereof.

vii. Research related subject

Correlations between theme vowels and thematic roles

viii. Data the students work with

Data obtained from corpora, data obtained from human subjects

ix. Topics

A: Research design

A1: General research design

[Teaching materials]
UPSKILLS Moodle course First steps into scientific research
https://upskillsproject.eu/project/scientific_research/
Movetia/ReLDI courses:
PHIL: Movetia101 Introduction to research in linguistics: theory, logic, method
https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English)
ReLDI-Project: ReLDI101 Introduction to Research Methodology in Linguistics
https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)

A2: Adapting the general research design to the specific topic of interest

Identifying correlations between thematic roles and other properties of the verbs as the relevant measure; identifying potential relations between the predictors and the thematic roles

A2.1: Formulation of questions and hypothesis in terms of variables

A2.2: Formulation of predictions of H0 and H1

A2.3: Selection of optimal research techniques, selection and creation of corresponding data sources

- Developing and exploiting databases and corpora (e.g. manual data annotation)

A2.4: Identifying the optimal data analysis method

A2.5: Inferring theoretical consequences from the specific data analysis results

A3: Adapting the research design to the available research infrastructures

Comparing the available corpora for their suitability for the research of thematic roles and their predictors, choosing the best one, planning the research considering the resources that it offers; learning about the databases of the project Hyperspacing the verb and the properties annotated

A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3)

- data compilation, data analysis;
- understanding, selecting and performing optimal statistical tests and models

A4: Research reporting

Identifying optimal formats for the representation of the correlations between thematic roles and other verbal properties as well as of the research results

A4.1 Presentation modes for research reporting (short oral presentation, report, article etc.)

A4.2 Established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/squib/report,
- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions

B: Infrastructures & techniques

B1: For obtaining literature

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz

B2: For obtaining, sharing and managing data

B2.1: Definition of research infrastructures, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**
B2.2: Platforms and repositories
B2.3: Identifying, collecting, creating and/or using relevant data for research projects • Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them
B3: For analysing data
Understanding statistical tests of correlation and their implications
B3.1: Softwares for statistical tests
B3.2: Concordancers for the analysis of corpora
C: Subject-specific topics
C1: Are there interactions between different grammatical phenomena and why?
C2: Identification of thematic roles
C3: Assessing verb meaning in terms of argument structure
C4: Explaining interactions between different properties of the verb

x. Learning outcomes

A: Research design
A1: Students will be able to make an overview of the general research design.
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: PHIL: Movetia101 Introduction to research in linguistics: theory, logic, method https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) ReLDI-Project: ReLDI101 Introduction to Research Methodology in Linguistics https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Students will be able to create a suitable research design for the specific topic of interest.

Students will be able to identify correlations between thematic roles and other properties of the verbs as the relevant measure; Students will be able to identify potential relations between the predictors and the thematic roles.
A2.1: Students will be able to formulate questions and hypothesis in terms of variables.
A2.2: Students will be able to formulate H0 and H1.
A2.3: Students will be able to select optimal research techniques, and create corresponding data sources ● Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading) ● Developing and exploiting databases and corpora (e.g., manual data annotation).
A2.4: Students will be able to select and implement the optimal data analysis method.
A2.5: Students will be able to infer theoretical consequences from the specific data analysis results.
A3: Students will be able to adapt a research design to the available research infrastructures.
Students will be able to compare the available corpora for their suitability for the research of thematic roles and their predictors, Students will be able to select the best corpus for their research, Students will be able to plan research considering the resources that the selected data source offers; Students will be familiar with the databases of the project Hyperspacing the verb and the properties annotated.
A3.1 Students will be able to select of optimal research techniques, select and create corresponding data sources (see also A2.3) ● data compilation, data analysis; ● understanding, selecting and performing optimal statistical tests and models.
A4: Students will be able to report on their performed research in accordance with standards and conventions in the field.
Students will be able to select optimal formats for the representation of the correlations between thematic roles and other verbal properties.
A4.1 Students will be able to select and implement different presentation modes for research reporting (short oral presentation, report, article etc.).

A4.2 Students will be able to implement established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/report,
- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions.

B: Infrastructures & techniques

B1: Students will be able to identify and apply suitable infrastructures & techniques for obtaining literature

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, ROA.

B2: Students will be able to identify and apply suitable infrastructures & techniques for obtaining, sharing and managing data

B2.1: Students will understand what research infrastructures are, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**.

B2.2: Students will be able to identify suitable platforms and repositories.

B2.3: Students will be able to identify, collect, create and/or use relevant data for their research projects

- Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them
- Citing linguistic data sets as appropriate.
- Familiarity with online survey tools.

B3: Students will be able to identify and apply suitable infrastructures & techniques for analysing data.

Students will be able to discuss statistical tests of correlation and its significance and their implications.

B3.4: Students will be able to select and use concordancers for the analysis of corpora.

C: Subject-specific learning outcomes

C1: Students will be able to discuss interactions between different grammatical phenomena.

C2: Students will be able to identify thematic roles.

C3: Students will be able to assess verb meaning in terms of argument structure.

C4: Students will be able to explain interactions between different properties of the verb.

xi. Overview of evaluation	
Rubric	Weighing
Participation incl. homework (initiative, forward-thinking, problem solving, critical thinking, organisation, time management)	30%
Outputs based on the final research report ● oral presentation ● final written report	70%

I.III. Course Description 3: Deverbal derivations

i.	**Name of the course** Deverbal derivations
ii.	**Level of the course** BA
iii.	**Workload** 5 ECTS
iv.	**Institution** University of Graz
v.	**Course instructor(s)** Boban Arsenijević
vi.	**Brief course description** The course focuses on the derivation of new words from verbal bases (i.e. on deverbal derivations). The main questions targeted are what morphological operations apply in deverbal derivations, what restrictions can be observed that are imposed by verbal properties such as aspect, argument structure or morphological make up, which among these operations are productive and how is the prosody of

the derived words determined. The third step begins with the concrete research design: its adaptation to the topic of the course and to the available infrastructure. In the rest of the course, the students will pursue the research as designed, report the results and discuss them in the class, and finally write a seminar paper.
The course begins with a general introduction to the scientific method in linguistics in the first two weeks, and is then followed by three larger blocks. During the introductory block, the basic notions of lexical categories, properties of verbs and derivation are presented and discussed. They are brought in connection with empirical research methods.
The second block discusses the available infrastructure - in particular the database of Bosnian/Croatian/Serbian and Slovenian verbs that serves as the base or the course. The students get familiar with the retrieval of the databases, with the use of corpora to pursue further annotation of the databases and with rudimentary automatic retrieval and annotation. Statistical tests and tools available for their implementation are introduced, discussed and tried hands-on. The students begin conducting their empirical investigations, discuss their difficulties and experiences
The third part is entirely dedicated to the students' research projects. They continue with the empirical investigations, develop their databases including their own decisions about certain open aspects of their design, apply statistical analyses and discuss the resulting tendencies of certain classes of verbs regarding undergoing certain derivational operations. The students submit their preliminary drafts which are read before class and discussed in class. The lecturer also provides written feedback. After this, the students hold oral presentations of their projects. At the end of the course, the students submit their final research reports.

vii. Research related subject

Morphological operations, verbal features

viii. Data the students work with

Data obtained from corpora, data obtained from human subjects

ix. Topics

A: Research design

A1: General research design

[Teaching materials]
UPSKILLS Moodle course First steps into scientific research
https://upskillsproject.eu/project/scientific_research/
Movetia/ReLDI courses:
https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English)

https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Adapting the general research design to the specific topic of interest
Decomposing derivations into the components of morphological operations (affixation, allomorphy, prosodic changes…), the semantic shifts, and the category change, where available; already annotated properties of verbs in the databases that the students work on Representing these in terms of variables and levels, distinguishing between predictors and 'predictees' Recognizing mutually dependent variables and discussing possible designs when no such combinations are included at the same time
A2.1: Formulation of questions and hypothesis in terms of variables
A2.2: Formulation of predictions of H0 and H1
A2.3: Selection of optimal research techniques, selection and creation of corresponding data sources • Experimental paradigms (e.g., elicitation, judgements, forced-choice) • Developing and exploiting databases and corpora (e.g. manual data annotation)
A2.4: Identifying the optimal data analysis method
A2.5: Inferring theoretical consequences from the specific data analysis results
A3: Adapting the research design to the available research infrastructures
Familiarisation with the databases, corpora and the script for retrieval
A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3) • data compilation, data analysis; • understanding, selecting and performing optimal statistical tests and models
A4: Research reporting
Identifying optimal formats for the representation of correlations between annotated properties of verbs and the derivational patterns they undergo
A4.1 Presentation modes for research reporting (short oral presentation, report, article etc.)
A4.2 Established procedures and conventions in research reporting, such as: • the ordering of thematic units in an article/squib/report, • organization of the presentation,

- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions

B: Infrastructures & techniques

B1: For obtaining literature

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz

B2: For obtaining, sharing and managing data

Annotating data in a database, both manually and automatically

B2.1: Definition of research infrastructures, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**

B2.2: Platforms and repositories

B2.3: Identifying, collecting, creating and/or using relevant data for research projects

- Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them
- Citing linguistic data sets as appropriate
- Familiarity with online survey tools

B3: For analysing data

Linear Mixed Effects Model and its fitting to the data. Issues with non-binary categorical predictor variables, using ANOVA instead

B3.1: Softwares for statistical tests

B3.2: Concordancers for the analysis of corpora

C: Subject-specific topics

C1: How properties of the base influence its derivational potential?

C2: What properties of the base can survive through morphological operations?

C3: What are the quantitative signatures of semantic and morphological relations?

x. Learning outcomes

A: Research design
A1: Students will be able to make an overview of the general research design.
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Students will be able to create a suitable research design for the specific topic of interest.
Students will be able to identify potentially relevant properties of the linguistic objects under study and formalize them as variables, listing their levels exhaustively and parsimoniously. Students will be able to calculate the required amount of observations for reliable results. Students will be able to retrieve the required amount of observations from the corpus and databases, and select and apply the appropriate statistical tests.
A2.1: Students will be able to formulate questions and hypothesis in terms of variables.
A2.2: Students will be able to formulate H0 and H1.
A2.3: Students will be able to select optimal research techniques, and create corresponding data sources ● Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading) ● Developing and exploiting databases and corpora (e.g., manual data annotation).
A2.4: Students will be able to select and implement the optimal data analysis method.
A2.5: Students will be able to infer theoretical consequences from the specific data analysis results.
A3: Students will be able to adapt a research design to the available research infrastructures.
Students will be familiar with the basic corpus linguistic tools, Students will be able to assess the relevant properties of the available corpora / databases, Students will be able to adapt the RD to the limitations of these infrastructures.

A3.1 Students will be able to select of optimal research techniques, select and create corresponding data sources (see also A2.3)

- data compilation, data analysis;
- understanding, selecting and performing optimal statistical tests and models.

A4: Students will be able to report on their performed research in accordance with standards and conventions in the field.

Students will be able to select the optimal format for the representation of the attestd interactions between verbal properties and each of the considered derivational patterns.

A4.1 Students will be able to select and implement different presentation modes for research reporting (short oral presentation, report, article etc.).

A4.2 Students will be able to implement established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/report,
- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions.

B: Infrastructures & techniques

B1: Students will be able to identify and apply suitable infrastructures & techniques for obtaining literature.

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, ROA.

B2: Students will be able to identify and apply suitable infrastructures & techniques for obtaining, sharing and managing data.

Students will be able to extract data from the available databases.
Students will be able to perform moderately complex searches of the available corpora.

B2.1: Students will understand what research infrastructures are, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**.

B2.2: Students will be able to identify suitable platforms and repositories.

B2.3: Students will be able to identify, collect, create and/or use relevant data for their research projects

• Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them • Citing linguistic data sets as appropriate. • Familiarity with online survey tools.
B3: Students will be able to identify and apply suitable infrastructures & techniques for analysing data.
Students will be able to use ANOVA and linear mixed model tests.
B3.4: Students will be able to select and use concordancers for the analysis of corpora
C: Subject-specific learning outcomes
C1: Students will be able to discuss how various properties of the base influence its derivational potential.
C2: Students will be able to predict whether a semantic property of the base will also be observable on the derived word.
C3: Students will be able to calculate the basic quantitative properties of morphological operations such as productivity or relative size of the base.

xi. Overview of evaluation	
Rubric	Weighing
Participation incl. homework (initiative, forward-thinking, problem solving, critical thinking, organisation, time management)	40%
Outputs based on the final research report • oral presentation • final written report	60%

I.IV. Course Description 4: Theoretical and methodological basics: Multipurpose Suffixes

i.	**Name of the course** Theoretical and methodological basics: Multipurpose Suffixes
ii.	**Level of the course** MA
iii.	**Workload** 5 ECTS
iv.	**Institution** University of Graz
v.	**Course instructor(s)** Boban Arsenijević
vi.	**Brief course description** The course introduces the student to the scientific methodology: what is a scientific question, a hypothesis, an analysis, a model, a theory, a prediction, how predictions can be tested, what happens when they are confirmed, and what when they are rejected. Discussed topics include the experimental design, continuous and discrete variables, the role of statistics, basic statistical models - fitting the models to the experiment or corpus research, as well as fitting the experiment or corpus research to the available models. All these notions are discussed on linguistic material (mostly from Slavic languages), which changes each year. In this semester (spring 2021/22), illustrative empirical phenomena and theoretical models are sought for in the area of phonology, morphology and semantics of multipurpose suffixes: identical suffixes which occur in different environments and yield different morphological effects. The data is mainly from BKS, but other (Slavic) languages are included too (in particular Slovenian, but also Russian, as well as Hungarian, Dutch etc.). The course begins with a theoretical introduction in the general scientific methodology, including the two Movetia courses listed below as homework, followed by one class of general overview of morphology, and two classes discussing processing signatures of grammatical and semantic relations and operations. We then begin with the readings on multipurpose suffixes - first three readings discussing various languages, and then at least three, possibly up to seven readings (depending on the pace) on multipurpose suffixes in South Slavic. Each article is first discussed for clarification and regarding its methodological setting

(identifying the theoretical background, the research questions, the hypotheses, the tests / evidence for hypotheses as given in the paper, but also potential counterevidence and inconsistencies), and then the students are encourage to come up with possible experimental designs to test the theoretical models proposed.
vii. Research related subject Morphological operations, types of morphological units
viii. Data the students work with Data obtained from corpora, data obtained from human subjects
ix. Topics

A: Research design

A1: General research design

[Teaching materials]
UPSKILLS Moodle course First steps into scientific research
https://upskillsproject.eu/project/scientific_research/
Movetia/ReLDI courses:
https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English)
https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)

A2: Adapting the general research design to the specific topic of interest

The current course has a somewhat different approach. As the main topic of the course is the general research design, the course takes both directions: generalizing from the concrete research designs encountered in the course readings and from own investigations towards the more general notions, formulating concretizations of these general notions for the given article (identifying its research questions, hypotheses, predictions, tests). An additional difficulty comes from the highly theoretical nature of this year's topic, which is to be targeted by a rather empirically oriented final discussion.

A2.1: Formulation of questions and hypothesis in terms of variables

A2.2: Formulation of predictions of H0 and H1

A2.3: Selection of optimal research techniques, selection and creation of corresponding data sources

- Experimental paradigms (e.g., elicitation, judgements, forced-choice)
- Developing and exploiting databases and corpora (e.g. manual data annotation)

A2.4: Identifying the optimal data analysis method

A2.5: Inferring theoretical consequences from the specific data analysis results
A3: Adapting the research design to the available research infrastructures
Familiarization with the software and online platforms for the development and administration of linguistic experiments
A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3) • data compilation, data analysis; • understanding, selecting and performing optimal statistical tests and models
A4: Research reporting
Identifying optimal formats for the representation of the experimental insights in the processing aspects of multipurpose suffixes
A4.1 Presentation modes for research reporting (short oral presentation, report, article etc.)
A4.2 Established procedures and conventions in research reporting, such as: • the ordering of thematic units in an article/squib/report, • organization of the presentation, • amount of text and graphical items on a slide/handout, • terminology, • citing conventions
B: Infrastructures & techniques
B1: For obtaining literature
[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz
B2: For obtaining, sharing and managing data
Learning to use Ibex Farm to administer experiments. Advancing the use of MSExcel-equivalent software
B2.1: Definition of research infrastructures, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**
B2.2: Platforms and repositories
B2.3: Identifying, collecting, creating and/or using relevant data for research projects

• Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them • Citing linguistic data sets as appropriate • Familiarity with online survey tools
B3: For analysing data
Linear Mixed Effects Model and its fitting to the data and alternative statistical tools
B3.1: Softwares for statistical tests
B3.2: Concordancers for the analysis of corpora
C: Subject-specific topics
C1: Are there multipurpose suffixes, or are they all different homophonous suffixes?
C2: If they exist, are multipurpose suffixes roots or functional items? In the former case, are other derivational suffixes roots? What about inflection?
C3: What are the processing signatures of identity, roots, homonymy?

x. Learning outcomes

A: Research design
A1: Students will be able to make an overview of the general research design.
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Students will be able to create a suitable research design for the specific topic of interest.
Students will be able to summarize a theoretical linguistic article in terms of the basic methodological notions: assumptions, research questions, hypotheses, their predictions and the matching between predictions and empirical data.

Students will be able to identify potentially relevant properties of the linguistic objects under study and formalize them as variables, listing their levels exhaustively and parsimoniously. Students will be able to design appropriate experiments to test the empirical predictions of models in the intersection of morphology with phonology, i.e. with semantics.
A2.1: Students will be able to formulate questions and hypothesis in terms of variables.
A2.2: Students will be able to formulate H0 and H1.
A2.3: Students will be able to select optimal research techniques, and create corresponding data sources ● Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading) ● Developing and exploiting databases and corpora (e.g., manual data annotation).
A2.4: Students will be able to select and implement the optimal data analysis method.
A2.5: Students will be able to infer theoretical consequences from the specific data analysis results.
A3: Students will be able to adapt a research design to the available research infrastructures.
Students will be familiar with one platform for the administration of linguistic experiments, Students will be able to adapt the RD to the limitations of this platform, and maximally use its features to get as precise and informative data as possible.
A3.1 Students will be able to select of optimal research techniques, select and create corresponding data sources (see also A2.3) ● data compilation, data analysis; ● understanding, selecting and performing optimal statistical tests and models.
A4: Students will be able to report on their performed research in accordance with standards and conventions in the field.
Students will be able to select the optimal format for the presentation of the critical assessment of a theory as well as of the experimental design they come up with.
A4.1 Students will be able to select and implement different presentation modes for research reporting (short oral presentation, report, article etc.).
A4.2 Students will be able to implement established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/report,
- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions.

B: Infrastructures & techniques

B1: Students will be able to identify and apply suitable infrastructures & techniques for obtaining literature.

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, ROA.

B2: Students will be able to identify and apply suitable infrastructures & techniques for obtaining, sharing and managing data.

Students will be able to administer an experiment on the platform of choice (planned: Ibex Farm).
Students will be able to extract data from the platform.

B2.1: Students will understand what research infrastructures are, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**.

B2.2: Students will be able to identify suitable platforms and repositories.

B2.3: Students will be able to identify, collect, create and/or use relevant data for their research projects

- Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them
- Citing linguistic data sets as appropriate.
- Familiarity with online survey tools.

B3: Students will be able to identify and apply suitable infrastructures & techniques for analysing data.

Students will be able to use the linear mixed effects model.

B3.4: Students will be able to select and use concordancers for the analysis of corpora.

C: Subject-specific learning outcomes

C1: Students will be able to weigh arguments for and against the existence of multipurpose suffixes.

C2: Students will be able to compare models of multipurpose suffixes in terms of roots and

in terms of functional morphemes..
C3: Students will be able to plan an experimental investigation of multipurpose suffixes.

xi. Overview of evaluation	
Rubric	Weighing
Participation incl. homework (initiative, forward-thinking, problem solving, critical thinking, organisation, time management)	40%
Outputs based on the final research report • oral presentation • final written report	60%
xii. Reading Materials	Arsenijević, B. 2020. Deverbal nouns in -ie and their variation across the South Slavic area. Linguistica, 60(1), 7–29. https://doi.org/10.4312/linguistica.60.1.7-29 Creemers, A., Don, J. & Fenger, P. 2018. Some affixes are roots, others are heads. Nat Lang Linguist Theory 36, 45–84. https://doi.org/10.1007/s11049-017-9372-1 Kenesei, István. 2014. On a multifunctional derivational affix : Its use in relational adjectives or nominal modification and phrasal affixation in Hungarian. Word Structure, 7 (2), 214-239. Simonović, Marko. 2022. Derivational affixes as roots across categories. To appear in Journal of Slavic linguistics. Simonović, Marko and Petra Mišmaš. 2022. Lowest theme vowels or highest roots? An 'unaccusative' theme-vowel class in Slovenian. To appear in Glossa. DOI: https://doi.org/10.16995/glossa.5809

II. Beyond Hyperspacing: Individual UPSKILLS RBT courses

II.I. Course Description 5: Current trends in Phonology

i.	**Name of the course** Current trends in Phonology (Morphologically conditioned phonological alternations)
ii.	**Level of the course** MA, PhD (can also be taught to advanced BA students)
iii.	**Workload** 6 ECTS
iv.	**Institution** University of Nova Gorica/University of Ljubljana
v.	**Course instructor(s)** Marko Simonovic
vi.	**Brief course description** The course provides a state-of-the-art overview of the literature and ongoing research of a phenomenon at the interface between phonology and morphology/syntax. The course enhances the problem-solving and data-analysis skills, thus preparing the students for a wide range of possible careers. It also provides the students with first-hand scientific research experience. The course consists of five parts. During the first, introductory part, the students get familiarised with morphologically conditioned phonological alternations and their formalisation. The example used to illustrate the phenomenon is the velar/strident alternation in Bosnian/Croatian/Montenegrin/Serbian. An algorithm for the various types of conditioning and an Optimality Theory account are presented.

In the second part the students identify morphologically conditioned phonological alternations in their native language. The starting point are the existing descriptions including the standard grammars. Special attention is devoted to cases of optionality, variation and gaps. All students submit a descriptive summary of the conditioning (operationalised as an algorithm) based on available descriptive sources.

The third part focuses on general research design. The students define their research question and start sketching their research report based on a provided research-report template.

The fourth part focuses on obtaining data from corpora. After getting familiarised with the available corpora for their target language, students design corpus studies aimed at answering the question defined in the previous part. All students submit a summary of their findings in the form of a short intermediate report.

The fifth part focuses on obtaining data from human subjects. The students create surveys and obtain data from subjects aimed at answering the question defined in the previous part. All students present a summary of their findings in the form of a short oral presentation.

At the end of the course, the students submit a final research report.

vii. Research related subject

Morphologically conditioned phonological alternations.

viii. Data the students work with

Data obtained from corpora, data obtained from human subjects

ix. Topics

A: Research design

A1: General research design

[Teaching materials]
UPSKILLS Moodle course First steps into scientific research
https://upskillsproject.eu/project/scientific_research/
Movetia/ReLDI courses:
PHIL: Movetia101 Introduction to research in linguistics: theory, logic, method
https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English)
ReLDI-Project: ReLDI101 Introduction to Research Methodology in Linguistics

https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Adapting the general research design to the specific topic of interest
Identifying the predictors of variation and avoidance: position in the paradigm, borrowed vs. native, word frequency, morphological-context frequency, phonological-environment frequency
A2.1: Formulation of questions and hypothesis in terms of variables
A2.2: Formulation of predictions of H0 and H1
A2.3: Selection of optimal research techniques, selection and creation of corresponding data sources • Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading) • Developing and exploiting databases and corpora (e.g. manual data annotation)
A2.4: Identifying the optimal data analysis method
A2.5: Inferring theoretical consequences from the specific data analysis results
A3: Adapting the research design to the available research infrastructures
Familiarising with the type of data extractable from the available corpora; Familiarising with the type of data extractable from online surveys; Choosing the optimal level of transcription, phonemic vs. phonetic
A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3) • data compilation, data analysis; • understanding, selecting and performing optimal statistical tests and models
A4: Research reporting
Identifying the optimal format for representing paradigms with gaps, avoidance, results of frequency counts
A4.1 Presentation modes for research reporting (short oral presentation, report, article etc.)
A4.2 Established procedures and conventions in research reporting, such as: • the ordering of thematic units in an article/squib/report, • organization of the presentation,

- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions

B: Infrastructures & techniques

B1: For obtaining literature

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, ROA

B2: For obtaining, sharing and managing data

Advanced use of the available corpora for the target language, advance use of online surveys

B2.1: Definition of research infrastructures, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**

B2.2: Platforms and repositories

B2.3: Identifying, collecting, creating and/or using relevant data for research projects

- Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them
- Citing linguistic data sets as appropriate.
- Familiarity with online survey tools

B3: For analysing data

B3.1: Softwares for statistical tests

B3.2: Concordancers for the analysis of corpora

C: Subject-specific topics

C1: Why are there variation, paradigm gaps and avoidance?

C2: Does variation always lead to avoidance?

C3: Basic concept of descriptive phonology applied to the target languages

C4: Basic concept of descriptive morphology applied to the target languages

C5: Basic concepts of Optimality Theory

x. Learning outcomes

A: Research design

A1: Students will be able to make an overview of the general research design.

[Teaching materials]
UPSKILLS Moodle course First steps into scientific research
https://upskillsproject.eu/project/scientific_research/
Movetia/ReLDI courses:
PHIL: Movetia101 Introduction to research in linguistics: theory, logic, method
https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English)
ReLDI-Project: ReLDI101 Introduction to Research Methodology in Linguistics
https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)

A2: Students will be able to create a suitable research design for the specific topic of interest.

Students will be able to identify the predictors of variation and avoidance: position in the paradigm, borrowed vs. native, word frequency, morphological-context frequency, phonological-environment frequency.

A2.1: Students will be able to formulate questions and hypothesis in terms of variables.

A2.2: Students will be able to formulate H0 and H1.

A2.3: Students will be able to select optimal research techniques, and create corresponding data sources

- Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading)
- Developing and exploiting databases and corpora (e.g., manual data annotation).

A2.4: Students will be able to select and implement the optimal data analysis method.

A2.5: Students will be able to infer theoretical consequences from the specific data analysis results.

A3: Students will be able to adapt a research design to the available research infrastructures.

Students will be familiar with the type of data extractable from the available corpora;
Students will be familiar with the type of data extractable from online surveys;
Students will be able to choose the optimal level of transcription, phonemic vs. phonetic.

A3.1 Students will be able to select of optimal research techniques, select and create corresponding data sources (see also A2.3)

- data compilation, data analysis;
- understanding, selecting and performing optimal statistical tests and models.

A4: Students will be able to report on their performed research in accordance with standards and conventions in the field.

Students will be able to select the optimal format for representing paradigms with gaps, avoidance, and results of frequency counts.

A4.1 Students will be able to select and implement different presentation modes for research reporting (short oral presentation, report, article etc.).

A4.2 Students will be able to implement established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/report,
- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions.

B: Infrastructures & techniques

B1: Students will be able to identify and apply suitable infrastructures & techniques for obtaining literature.

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, ROA.

B2: Students will be able to identify and apply suitable infrastructures & techniques for obtaining, sharing and managing data.

Students will be able to extract data from the available corpora for the target language, Students will be able to obtain data in online surveys.

B2.1: Students will understand what research infrastructures are, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**.

B2.2: Students will be able to identify suitable platforms and repositories.

B2.3: Students will be able to identify, collect, create and/or use relevant data for their research projects.

• Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them, • Citing linguistic data sets as appropriate, • Familiarity with online survey tools.
B3: Students will be able to identify and apply suitable infrastructures & techniques for analysing data.
B3.1: Students will be able to select and use software for statistical tests.
B3.4: Students will be able to select and use concordancers for the analysis of corpora.
C: Subject-specific learning outcomes
C1: Students will be able to discuss sources of variation, paradigm gaps and avoidance.
C2: Students will be able to discuss the relation between variation and avoidance.
C3: Students will be able to apply the basic concepts of descriptive phonology to the target language.
C4: Students will be able to apply the basic concepts of descriptive morphology to the target language.
C5: Students will be able to apply the basic concepts of Optimality Theory to the targeted dataset.

xi. Overview of evaluation	
Rubric	Weighing
Participation incl. homework (initiative, forward-thinking, problem solving, critical thinking, organisation, time management)	30%
Outputs based on the final research report • oral presentation • final written report	70%

II.II. Course Description 6: Multilingualism

i.	**Name of the course** Multilingualism: Obtaining, representing and analysing empirical data in linguistics (Field-work, building corpora, running experiments)
ii.	**Level of the course** MA, PhD (can also be taught to advanced BA students)
iii.	**Workload** 6 ECTS
iv.	**Institution** University of Graz
v.	**Course instructor(s)** Boban Arsenijević, Marko Simonović
vi.	**Brief course description** This course focuses on the hotly debated topics (both in Slavic studies and in the Austrian public arena) of Slavic heritage varieties, language contact and multilingualism. The course enhances the problem-solving and data-analysis skills, thus preparing the students for a wide range of possible careers. It also provides the students with first-hand scientific research experience. The course consists of three parts. During the introductory part, the central concepts of research in heritage varieties, contact linguistics and multilingualism are introduced (heritage vs. baseline variety, code-switch, borrowing, interlanguage, additive vs. subtractive bilingualism), mostly through the discussion of handbook chapters and articles (app. 80% of each class). Simultaneously, a very basic sketch or general research design is introduced and some preliminary discussion is held about testing the predictions of various theories (app. 20% of each class). The second part focuses on the same two topics as the first part (multilingualism and research). Now 80% is dedicated to research design. During this part, most preparatory work is done by attending the Movetia course *Introduction to research in linguistics: theory, logic, method* and coming up with an own research project. In addition, students start sketching their research report based on the research-report template (created by Tanja Samadžić and Maja Miličević Petrović). The remaining 20% are reserved for the discussion of research articles (and, where necessary handbook chapters) which are deemed instrumental given the specific projects selected by the students. The third part is entirely dedicated to the students' research projects. The students submit their preliminary drafts which are read before class and discussed in class. The lecturer also

provides written feedback. After this, the students hold oral presentations of their projects. At the end of the course, the students submit their final research reports.
vii. Research related subject Slavic heritage varieties, language contact and multilingualism
viii. Data the students work with Data obtained from corpora, data obtained from human subjects
ix. Topics

A: Research design
A1: General research design
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Adapting the general research design to the specific topic of interest
Different groups of speakers, heritage vs. baseline (2 baselines: monolinguals and L2 learners of the contact language) or heritage vs. heritage (where the influence of the contact language is to be explored); Defining measurable phenomena, borrowing vs. code-switch, Identifying correlations between contact phenomena and other factors. Distinguishing between mere consequences of a lower degree of exposure (more errors generally, slower overall processing, smaller vocabulary) and concrete signatures of the heritage nature of the language in particular aspects of grammar (verbal prefixation, telicity, gender, pronouns, lexical prosody), including the potential influence of the contact language
A2.1: Formulation of questions and hypothesis in terms of variables
A2.2: Formulation of predictions of H0 and H1
A2.3: Selection of optimal research techniques, selection and creation of corresponding data sources ● Experimental paradigms (e.g., elicitation, judgements, forced-choice) ● Developing and exploiting databases and corpora (e.g. manual data annotation)
A2.4: Identifying the optimal data analysis method
A2.5: Inferring theoretical consequences from the specific data analysis results

A3: Adapting the research design to the available research infrastructures

Familiarisation with common bilingual and L2 corpora, identifying available L2 research networks, adapting the RD to the limitations of these infrastructures.

A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3)

- data compilation, data analysis;
- understanding, selecting and performing optimal statistical tests and models

A4: Research reporting

Identifying optimal formats for the representation of correlations between contact phenomena and linguistic and non-linguistic factors

A4.1 Presentation modes for research reporting (short oral presentation, report, article etc.)

A4.2 Established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/squib/report,
- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions

B: Infrastructures & techniques

B1: For obtaining literature

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz

B2: For obtaining, sharing and managing data

Familiarity with L2 and bilingual corpora

B2.1: Definition of research infrastructures, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**

B2.2: Platforms and repositories

B2.3: Identifying, collecting, creating and/or using relevant data for research projects

- Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them
- Citing linguistic data sets as appropriate.
- Familiarity with online survey tools

B3: For analysing data
ANOVA and linear mixed model tests Tests for categorical non-binary variables
B3.1: Softwares for statistical tests
B3.2: Concordancers for the analysis of corpora
C: Subject-specific topics
C1: How do languages interact within speakers and communities?
C2: Are heritage languages full-fledged languages?
C3: Heritage vs baseline variety
C4: Code-switching vs borrowing

x. Learning outcomes

A: Research design
A1: Students will be able to make an overview of the general research design.
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Students will be able to create a suitable research design for the specific topic of interest.
Students will be able to define different groups of speakers, heritage vs. baseline; Students will be able to define measurable phenomena, borrowing vs. code-switch, Students will be able to identify correlations between contact phenomena and other factors.
A2.1: Students will be able to formulate questions and hypothesis in terms of variables.
A2.2: Students will be able to formulate H0 and H1.

A2.3: Students will be able to select optimal research techniques, and create corresponding data sources

- Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading)
- Developing and exploiting databases and corpora (e.g., manual data annotation).

A2.4: Students will be able to select and implement the optimal data analysis method.

A2.5: Students will be able to infer theoretical consequences from the specific data analysis results.

A3: Students will be able to adapt a research design to the available research infrastructures.

Students will be familiar with common bilingual and L2 corpora,
Students will be able to identify available L2 research networks,
Students will be able to adapt the RD to the limitations of these infrastructures.

A3.1 Students will be able to select of optimal research techniques, select and create corresponding data sources (see also A2.3)

- data compilation, data analysis;
- understanding, selecting and performing optimal statistical tests and models.

A4: Students will be able to report on their performed research in accordance with standards and conventions in the field.

Students will be able to select the optimal format for the representation of correlations and interactions between contact phenomena and linguistic and non-linguistic factors.

A4.1 Students will be able to select and implement different presentation modes for research reporting (short oral presentation, report, article etc.).

A4.2 Students will be able to implement established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/report,
- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions.

B: Infrastructures & techniques

B1: Students will be able to identify and apply suitable infrastructures &

techniques for obtaining literature.
[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, ROA.
B2: Students will be able to identify and apply suitable infrastructures & techniques for obtaining, sharing and managing data.
Students will be able to extract data from the available L2 and bilingual corpora.
B2.1: Students will understand what research infrastructures are, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**.
B2.2: Students will be able to identify suitable platforms and repositories.
B2.3: Students will be able to identify, collect, create and/or use relevant data for their research projects • Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them • Citing linguistic data sets as appropriate. • Familiarity with online survey tools.
B3: Students will be able to identify and apply suitable infrastructures & techniques for analysing data.
Students will be able to use ANOVA and linear mixed model tests.
B3.4: Students will be able to select and use concordancers for the analysis of corpora.
C: Subject-specific learning outcomes
C1: Students will be able to discuss how languages interact within speakers and communities.
C2: Students will be able to evaluate arguments for and against considering heritage languages full-fledged languages.
C3: Students will be able to distinguish between heritage and baseline varieties in different contexts.
C4: Students will be able to use standard tools for distinguishing between code-switching and borrowing.

xi. Overview of evaluation	

Rubric	Weighing
Participation incl. homework (initiative, forward-thinking, problem solving, critical thinking, organisation, time management)	30%
Outputs based on the final research report • oral presentation • final written report	70%

II.III. Course Description 7: Acquisition of English as a Second Language

i.	**Name of the course** Acquisition of English as a Second Language
ii.	**Level of the course** MA
iii.	**Workload** 3 ECTS
iv.	**Institution** University of Rijeka - Faculty of Humanities and Social Sciences
v.	**Course instructor(s)** Tihana Kraš
vi.	**Brief course description** The course deals with the acquisition of English as a second language and the way this phenomenon can be studied experimentally collecting data from human participants. The students are first provided with an introduction into research design and research infrastructures and techniques with a particular focus on experimental research into second language acquisition. After this, they do two things in parallel: they read critically and discuss in class published research studies on the acquisition of English as a second language, and design and conduct their own small-scale experimental study focusing on an aspect of the acquisition of English as a second language of their choice. For the purposes of the latter task, they work in pairs and produce a final research report in pairs. The submission of the report is preceded by two oral presentations, in which pairs of students present their work in progress. Students receive feedback from their peers and the course instructor with the aim of improving their study before its completion.

	The course enhances the students' analytical, problem-solving and data-analysis skills, thus preparing them for a wide range of possible careers. It also provides the students with first-hand scientific research experience.
vii.	**Research related subject** Acquisition of English as a Second Language
viii.	**Data the students work with** Data obtained from human participants by means of experimental methods
ix.	**Topics**

A: Research design
A1: General research design
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English)
A2: Adapting the general research design to the specific topic of interest
Explaining how different factors affect different aspects of the acquisition of English as a second language. Reading critically published research studies on the acquisition of English as a second language. Identifying a research problem concerning an aspect of this phenomenon. Designing and conducting an experimental study to address this problem. Analysing the data statistically and interpreting the results. Inferring theoretical and practical consequences from the research findings.
A2.1: Identification and formulation of a research problem
A2.2: Formulation of aims, research questions and hypotheses
A2.3: Selection of optimal research techniques, selection and creation of corresponding data sources relating to experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading)
A2.4: Identifying the optimal data analysis method
A2.5: Inferring theoretical consequences from the specific data analysis results
A2.6: Inferring practical (pedagogical) consequences from the specific data analysis results

A3: Adapting the research design to the available research infrastructures

Familiarisation with IRIS, a digital repository of instruments and materials for research into second languages (https://www.iris-database.org/iris/app/home/index;jsessionid=FE234F9CB66D8604FC97BB8D7D3C1A36)
Familiarisation with the ReLDI repository of data collection instruments (https://reldi.spur.uzh.ch/data-collection-instruments/)

A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3), data compilation, data analysis;

- understanding, selecting and performing optimal statistical tests and models.

A4: Research reporting

Compiling a research report and preparing and giving an oral presentation using conventions specific for the field of second language acquisition

A4.1 Presentation modes and formats for research reporting (short oral presentation, report, article etc.)

A4.2 Established procedures and conventions in research reporting, such as:

- the ordering of thematic units in a report,
- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions

B: Infrastructures & techniques

B1: For obtaining literature

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, IRIS

B2: For obtaining, sharing and managing data

B2.1: Definition of research infrastructures, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**

B2.2: Platforms and repositories.

- **General-purpose repositories** and **disciplinary repositories**
  - [GENERAL-PURPOSE REPOSITORY] Zenodo, FigShare
  - [DISCIPLINARY REPOSITORY] CLARIN, The Language Archive

B2.3:Identifying, collecting, creating and/or using relevant data for research projects

- Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them
  - [DISCIPLINARY REPOSITORY] CLARIN, ELRC-SHARE, the Language Archive.
- Citing linguistic data sets as appropriate.
- Depositing research data in a **certified repository** and selecting an appropriate licence for sharing their data
- The **versioning** policy of repositories
- Familiarity with online survey tools

B2.4: **Data management plan**

- Understanding the **data lifecycle**
- Understanding how to generate data, analyse and handle it
- Understanding the **legal and ethical issues** around data generation and use (e.g. licensing, GDPR compliance, anonymisation, the importance of FAIR principles and Open Access).
- Secure storage and backup of research data
- Documenting workflows and what metadata to use to describe the nature of the data based on existing standards.
- What data needs to be destroyed, preserved in a data repository and made available for reuse.

B3: For analysing data

Conducting suitable descriptive and inferential statistical analyses of the data. Explaining and choosing appropriate statistical tests. Reporting and interpreting the results of the data analysis

B3.1: Softwares for statistical tests

C: Subject-specific topics

C1: How do different factors affect different aspects of the acquisition of English as a second language?

C2: Reading critically published empirical studies on the acquisition of English as a second language

C3: Designing and conducting a small-scale experimental study to explore one aspect of the acquisition of English as a second language

C4: Explaining in what way the results of an experimental study are relevant for the theory of second language acquisition

C5: Explaining in what way the results of an experimental study are relevant for learning

and/or teaching English as a second language acquisition.

x. Learning outcomes

A: Research design
A1: Students will be able to make an overview of the general research design.
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English)
A2: Students will be able to create a suitable research design for the specific topic of interest.
Students will be able to explain how different factors affect different aspects of the acquisition of English as a second language. Students will be able to read critically published research studies on the acquisition of English as a second language. Students will be able to identify a research problem concerning an aspect of the acquisition of English as a second language. Students will be able to design and conduct an experimental study to address the research problem they have identified. Students will be able to analyse the data statistically and interpret the results. Students will be able to infer theoretical and practical consequences from their research findings.
A2.1: Students will be able to identify and formulate a research problem.
A2.2: Students will be able to formulate aims, research questions and hypotheses.
A2.3: Students will be able to select optimal research techniques, and create corresponding data sources concerning experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading).
A2.4: Students will be able to select and implement the optimal data analysis method.

A2.5: Students will be able to infer theoretical consequences from the specific data analysis results.
A2.6: Students will be able to infer practical (pedagogical) consequences from the specific data analysis results.
A3: Students will be able to adapt a research design to the available research infrastructures.
Students will be familiar with IRIS, a digital repository of instruments and materials for research into second languages (https://www.iris-database.org/iris/app/home/index;jsessionid=FE234F9CB66D8604FC97BB8D7D3C1A36). Students will be familiar with the ReLDI repository of data collection instruments (https://reldi.spur.uzh.ch/data-collection-instruments/).
A3.1 Students will be able to select optimal research techniques, and to select and create corresponding data sources (see also A2.3) • perform data compilation, data analysis; • understand, select and perform optimal statistical tests and models.
A4: Students will be able to report on their performed research in accordance with standards and conventions in the field.
Students will be able to compile a research report, and prepare and give an oral presentation using conventions specific for the field of second language acquisition.
A4.1 Students will be able to select and implement different presentation modes and formats for research reporting (short oral presentation, report, article etc.)
A4.2 Students will be able to implement established procedures and conventions in research reporting, such as: • ordering of thematic units in a report, • organization of the presentation, • amount of text and graphical items on a slide/handout, • terminology, • citing conventions.
B: Infrastructures & techniques
B1: Students will be able to identify and apply suitable infrastructures and techniques for obtaining literature

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, ROA.
B2: Students will be able to identify and apply suitable infrastructures & techniques for obtaining, sharing and managing data.
B2.1: Students will understand what research infrastructures are, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**.
B2.2: Students will be able to identify suitable platforms and repositories.
B2.3: Students will be able to identify, collect, create and/or use relevant data for their research projects • Citing linguistic data sets as appropriate. • Familiarity with online survey tools.
B3: Students will be able to identify and apply suitable infrastructures & techniques for analysing data.
Students will be able to conduct suitable descriptive and inferential statistical analyses of the data. Students will be able to explain and choose appropriate statistical tests. Students will be able to report and interpret the results of the data analysis.
B3.1: Students will be able to select and use softwares for statistical tests.
C: Subject-specific learning outcomes
C1: Students will be able to explain how different factors affect different aspects of the acquisition of English as a second language.
C2: Students will be able to read critically published empirical studies on the acquisition of English as a second language.
C3: Students will be able to design and conduct a small-scale experimental study to explore one aspect of the acquisition of English as a second language.
C4: Students will be able to explain in what way the results of their study are relevant for the theory of second language acquisition.
C5: Students will be able to explain in what way the results of their study are relevant for learning and/or teaching English as a second language.

xi. Overview of evaluation	

Rubric	Weighing
Participation incl. homework (initiative, forward-thinking, problem solving, critical thinking, organisation, time management) or a test (based on the research papers read in class)	30%
Outputs based on a research report • two oral presentations • the written report	70% (10% + 10% + 50%)

Oral presentation marking scale

An oral presentation may be awarded a maximum of 10 and a minimum of 5 points, according to the following criteria:

	Criterion	Grade points
1.	Ideas in the presentation are presented clearly and logically (i.e. the presentation is clear and coherent).	0–2
2.	The slides are well-organised and easy to read, and do not contain typos or language mistakes. In-text citations and full references are provided. A chosen reference style is followed consistently and accurately.	0–2
3.	The presentation contains data and/or examples, which are presented in a clear and effective way.	0–2
4.	The language used to give the presentation is fluent and accurate with appropriate lexical choices.	0–2
5.	The presentation is given in a clear, engaging and confident manner, providing strong evidence of preparation and organisation.	0–2

Research report marking scale

A research report may be awarded a maximum of 50 and a minimum of 26 grade points, according to the following criteria:

Criterion	Grade points
General	
1. Comprehensibility Is the output comprehensible for audiences who did not participate in its creation (clarity of thought and expression, quality of language, etc.)?	0–4
2. Coherence Is the output internally coherent (logical flow of ideas, no contradictions, consistent terminology, consistent referencing style, consistent formatting, etc.)?	0–4
3. Field-specific conventions Does the output use the conventions typically used in the field (structure, terminology, artwork, a referencing style, etc.)?	0–3
Research-related	
4. Identification of the research problem Is the research problem clearly identified? Is the research problem sufficiently justified?	0–3
5. Literature review Are the relevant theories described clearly and with sufficient detail? Is previous relevant research described clearly and with sufficient detail? Is previous relevant research reviewed critically?	0–4

6. Formulation and suitability of research aims, research questions and hypotheses Are aims, research questions and hypotheses clearly formulated? Are research aims, research questions and hypotheses suitable?	0–3
7. Description and suitability of the study design Is the study design clearly described? Is the study design suitable?	0–3
8. Description and suitability of the participant sample Is the participant sample clearly described? Is the participant sample suitable?	0–3
9. Application of the ethical principles concerning data collection and use Have the ethical principles concerning data collection and use been applied correctly?	0–2
10. Familiarity with the relevant resources, tools and infrastructures presented in class Has sufficient familiarity with the relevant resources, tools and infrastructures presented in class been demonstrated?	0–2
11. Description and suitability of experimental techniques Are the data collection instruments clearly described? Are the data collection instruments suitable?	0–3
12. Description and suitability of the experimental procedure Is the experimental procedure clearly described? Is the experimental procedure suitable?	0–2

13. Description and suitability of the data analysis method Is the data analysis method clearly described? Is the data analysis method suitable?	0–3
14. Presentation of the data analysis results Are the data analysis results clearly presented?	0–3
15. Interpretation and discussion of the data analysis results Are the data analysis results plausibly interpreted? Are the data analysis results discussed with reference to the relevant theories and previous research findings? Are the limitations of the study reasonably identified? Are suitable directions for future research given?	0–4
16. Inferring theoretical consequences from the specific data analysis results Are reasonable broader theoretical consequences inferred from the data analysis results?	0–2
17. Inferring practical consequences from the specific data analysis results Are reasonable practical consequences (i.e. pedagogical implications) inferred from the data analysis results?	0–2

xiv. Career paths

a. Academia
b. Natural language processing
c. Market analysis
d. and all careers involving data analysis

xv. Reading materials

OBLIGATORY SOURCES

Kraš, T., & Miličević, M. (2015). *Eksperimentalne metode u istraživanjima usvajanja drugoga jezika*. Rijeka: Filozofski fakultet Sveučilišta u Rijeci.

Litosseliti, L. (2018). *Research methods in linguistics*. London/New Delhi/New York/ Sydney: Bloomsbury Academic.

Phakiti, A. (2014). *Experimental research methods in language learning*. London/New Delhi/New York/ Sydney: Bloomsbury Academic.

OPTIONAL SOURCES

Blom, E., & Unsworth, S. (Eds.) (2010). *Experimental methods in language acquisition research*. Amsterdam/Philadelphia: John Benjamins.

Bhatia, T. K., & Ritchie, W. C. (Eds.) (2012). *The handbook of bilingualism and multilingualism*. Oxford: Blackwell.

de Bot, K., Lowie, W., & Verspoor, M. (2005). *Second language acquisition: An advanced resource book*. London/New York: Routledge.

Brown, J. D., & Rodgers, T. S. (2002). *Doing second language research*. Oxford: Oxford University Press.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Dörnyei, Z. (2003). *Questionnaires in second language research: Construction, administration and processing*. London/New York: Routledge.

Doughty, C. J., & Long, M. H. (Eds.) (2003). *The handbook of second language acquisition*. Oxford: Blackwell.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.

Gass, M. S., & Mackey, A. (2011). *The Routledge handbook of second language acquisition*. London/New York: Routledge.

Gass, M. S., & Mackey, A. (2007). *Data elicitation for second and foreign language research*. Mahwah, NJ/London: Lawrence Erlbaum Associates.

Kroll, J. F., & De Groot, A. M. B. (Eds.) (2005). *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press.

Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research*. Longman.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ/London: Lawrence Erlbaum Associates.

Mackey, A., & Gass, S. M. (2012). Research *methods in second language acquisition: A practical guide*. Malden, MA: Wiley-Blackwell.

Ritchie, W. C., & Bhatia, T. K. (Eds.) (2009). *The new handbook of second language acquisition*. Bingley: Emerald.

Robinson, P., & Ellis, N. C. (Eds.) (2008). *Handbook of cognitive linguistics and second language acquisition*. New York/London: Routledge.

White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

II.IV. Course Description 8: Syntax of the DP

i. Title of the course
The Syntax of the DP

ii. Level of the course
BA for Linguistics students, MA for language students

iii. Workload
6 ECTS

iv. Institution
University of Geneva

v. Course instructor(s)
Margherita Pallottino, Genoveva Puskas

vi. Brief course description
The course explores the complex morpho-syntactic properties of nominal expressions (DPs). The course consists of five parts.

The first part presents the syntactic properties of nominal structures. The students are presented with a set of examples which illustrate some properties of DPs. The students are asked to provide comparable data in other languages, in order to open a discussion on variation at the level of the DP.

The second part focuses on the theoretical accounts of variation in the domain of the DP. The students read a golden standard paper (Cinque 2005), and the key theoretical points are discussed in class.

The third part focuses on general research design. The students are divided into small groups (max three students) and propose a research topic. The research question is refined in class. Students also investigate various survey designs.

The fourth part is devoted to data collection tools. The students work on their survey design and on the various elicitation methods.

Part five is the data collection itself from human subjects. All students present a short report of their findings in the form of an oral report.

The final assessment consists of either a written project report or an exam focusing on their research and findings.

vii. Research related subjects

The distribution of adjectives and classifiers in the DP

viii. Tools and data the students work with

Data: judgement data and data elicited data from human subjects
Tools: Google FormsSurvey

ix. Topics

A: Research design

A1: General research design

[Teaching materials]
UPSKILLS Moodle course First steps into scientific research
https://upskillsproject.eu/project/scientific_research/
Movetia/ReLDI courses:
https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English)
https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)

A2: Adapting the general research design to the specific topic of interest

A2.1: Formulation of questions and hypotheses in terms of variables

A2.2: Formulation of predictions of H0 and H1

A2.3: Selection of appropriate research techniques, selection and creation of corresponding data sources

- Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading)

A2.4: Inferring theoretical consequences from the specific data analysis results

A3: Adapting the research design to the available research infrastructures

A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3)

- data compilation, data analysis, data archiving (e.g., XML, XLS), data reuse

A4: Research reporting

A4.1 Presentation modes for research reporting (short oral presentation, poster, squib, report, article etc.)

A4.2 Established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/squib/report,
- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions

B: Infrastructures & techniques

B1: For obtaining literature

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu,

[DISCIPLINARY REPOSITORY] lingbuzz, ROA

B2: For obtaining, sharing and managing data

B2.31: Identifying, collecting, creating and/or using relevant data for research projects

- Familiarity with online survey tools

B2.24: **Data management plan**

- Understanding how to generate data, analyse and handle it
- Understanding the **legal and ethical issues** around data generation and use (e.g. licensing, GDPR compliance, anonymisation, the importance of FAIR principles and Open Access)

C: Subject-specific topics

C1: Familiarity with the nominal system and its internal complexity (the DP structure)
C2: Cross-linguistic variation within the nominal system

x. Learning outcomes

A: Research design
A1: Students will be able to make an overview of the general research design.
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Students will be able to create a suitable research design for the specific topic of interest.
A2.1: Students will be able to formulate questions and hypotheses in terms of variables.
A2.2: Students will be able to formulate H0 and H1.
A2.3: Students will be able to select optimal research techniques, and create corresponding data sources • Experimental paradigms (e.g., elicitation, judgements, forced-choice).
A2.4: Students will be able to infer theoretical consequences from the specific data analysis results.
A3: Students will be able to adapt a research design to the available research infrastructures.

A3.1 Students will be able to select of optimal research techniques, select and create corresponding data sources (see also A2.3)

- data compilation, data analysis, data archiving (e.g., XML, XLS), data reuse.

A4: Students will be able to report on their performed research in accordance with standards and conventions in the field.

A4.1 Students will be able to select and implement different presentation modes for research reporting (short oral presentation, poster, squib, report, article etc.).

A4.2 Students will be able to implement established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/squib/report,
- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions.

B: Infrastructures & techniques

B1: Students will be able to identify and apply suitable infrastructures & techniques for obtaining literature.

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu,

[DISCIPLINARY REPOSITORY] lingbuzz, Rutgers Optimality Archive.

B2: Students will be able to identify and apply suitable infrastructures & techniques for obtaining, sharing and managing data.

B2.3: Students will be able to identify, collect, create and/or use relevant data for their research projects

- Cite linguistic data sets as appropriate.
- Use online survey tools.

B2.4: Students will be able to create a **data management plan**

- Understand how to generate data, analyse and handle it

- Understand the **legal and ethical issues** around data generation and use (e.g. licensing, GDPR compliance, anonymisation, the importance of FAIR principles and Open Access).

C: Subject-specific topics

C1: Students will be able to provide a overview ofFamiliarity with the nominal system and its internal complexity (the DP structure)

C2: Students will be able to discuss cross-linguistic Variation within the nominal system

xi. Evaluation

Rubric	Weighing
○ oral presentation	20%
○ final written report	80%

xii. Further information on the course that the instructor considers relevant (assessment, background, reading materials, detailed weekly plan, career paths etc.)

- The piloted course consisted of 12in 12 weeks (2 hours/week) of teaching, of which 5 weeks were devoted to theoretical questions. Student evaluations revealed that the proposed ratio theoretical introduction/research was not optimal and that at least 2/3 of the time need to be allotted to the research design and its implementation.The theoretical goal should be adapted to the available time keeping in mind this principle.
- The part of the theory that cannot be taught is compensated by the gain that the students have in terms of research skills, in anticipation of the research paper they have to produce for their MA thesis, and of the confidence that they are able to launch into research-based projects in their potential future career.
- It is advisable to give the students an attestation of the acquired competences which can be produced in the process of a job interview.

xiii. Reading Materials

Cinque, G. (2005). Deriving Greenberg's Universal 20 and its exceptions. *Linguistic inquiry*, *36*(3), 315-332.

II.V. Course Description 9: Collecting and analyzing corpus data in hypothesis-driven linguistic research: The /k, g, x/ → /t͡s, z, s/ alternation in srWaC

i.	**Name of the course** Collecting and analyzing corpus data in hypothesis-driven linguistic research: The /k, g, x/ → /t͡s, z, s/ alternation in srWaC
ii.	**Level of the course** Advanced BA, MA
iii.	**Workload** 3 ECTS
iv.	**Institution** University of Novi Sad
v.	**Course instructor(s)** Marko Simonovic
vi.	**Brief course description** The course provides hands-on experience of corpus data extraction and analysis, targeting a phenomenon which is well known from prescriptive sources, but insufficiently described and virtually unaccounted for in formal approaches. Unlike most cases, where individual students or groups work on a representative sample, in this course the group works as a whole and targets all words attested in the corpus (above a certain frequency threshold). The course enhances the problem-solving and data-analysis skills, thus preparing the students for a wide range of possible careers. It also provides the students with first-hand scientific research experience. The course consists of six parts. During the first, introductory part (10% of the time available), the velar/strident alternation in BCMS is described and the environments in which it occurs are

identified. The starting point are the standard grammars. Special attention is devoted to cases of optionality, variation and gaps and the native speakers are encouraged to share their intuitions which diverge from the standard grammars. All students submit a descriptive summary of the velar/strident alternation in BCMS based on available descriptive sources.
The second part (10% of the time available) focuses on general research design and its application to the alternation in focus. Specifically, the various environments in which the alternation applies are translated into independent variables, whereas the application of the alternation is conceptualised as the dependent variable.
In the third part (10% of the time available) the focus is on obtaining data from corpora, specifically, from the Serbian web corpus (srWaC). The students get a quick introduction to CQLs and learn about what kind of data can be obtained.
In the fourth part (20% of the time available), a common project is set up targeting one of the morphological contexts for the alternation (the most probable candidate being the dative/locative singular context). A common document is created and shared with all participants, in which the data collection procedure is described. A Google sheet is created where data get collected. The specifics of the data collection are agreed upon:

- the specific CQLs to be used (+ whether different CQLs are used for triangulation),
- inclusion criteria for lemmas,
    - frequency threshold,
    - word status,
    - unresolvable homonymy etc.
- splitting lemmas (cases of resolvable homonymy),
- merging lemmas (different spellings)
- variables for which the lemmas will be annotated.

The initial division of labour is agreed upon.
Part 5 (40% of the time available) is the central part of the course. In between classes all students do a portion of data collection. The classes serve for discussions of issues and agreeing on changes in the data collection procedures (which get 'registered' in the relevant document). The teacher informs of the descriptive statistics of the data collected up to the point
Part 6 (10% of the time available) is used for a wrap-up of the empirical results.

vii. Research related subject

Conditioning of phonological alternations.

viii. Data the students work with

Data obtained from corpora, descriptive statistics.

ix. Topics

A: Research design

A1: General research design

[Teaching materials]
UPSKILLS Moodle course First steps into scientific research
https://upskillsproject.eu/project/scientific_research/
Movetia/ReLDI courses:
PHIL: Movetia101 Introduction to research in linguistics: theory, logic, method
https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English)
ReLDI-Project: ReLDI101 Introduction to Research Methodology in Linguistics
https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)

A2: Adapting the general research design to the specific topic of interest

Identifying the predictors of alternation: position in the paradigm, borrowed vs. native, word frequency, phonological environment, animacy etc.

A2.1: Formulation of questions and hypothesis in terms of variables

A2.2: Formulation of predictions of H0 and H1

A2.3: Selection of optimal research techniques, selection and creation of corresponding data sources

- Developing and exploiting databases and corpora (e.g. manual data annotation)

A2.4: Identifying the optimal data analysis method

A2.5: Inferring theoretical consequences from the specific data analysis results

A3: Adapting the research design to the available research infrastructures

Familiarising with the type of data extractable from the available corpora

A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3)

- data compilation, data analysis;
- understanding, selecting and performing optimal statistical tests and models

B: Infrastructures & techniques

B1: For obtaining, sharing and managing data

Advanced use of srWaC
B1.1: Identifying, collecting, creating and/or using relevant data for research projects ● Searching relevant corpora, ● Citing linguistic data sets as appropriate
B1.2: Document a research process
B2: For analysing data
B2.1: Concordancers for the analysis of corpora
C: Subject-specific topics
C1: Basic concept of descriptive phonology of BCMS
C2: Basic concept of descriptive morphology of BCMS
C3: Basic concepts of Optimality Theory

x. Learning outcomes

A: Research design
A1: Students will be able to make an overview of the general research design.
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: PHIL: Movetia101 Introduction to research in linguistics: theory, logic, method https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) ReLDI-Project: ReLDI101 Introduction to Research Methodology in Linguistics https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Students will be able to create a suitable research design for the specific topic of interest.
Students will be able to identify the predictors of variation and avoidance: position in the paradigm, borrowed vs. native, word frequency, phonological environment, animacy etc.
A2.1: Students will be able to formulate questions and hypothesis in terms of variables.

A2.2: Students will be able to formulate H0 and H1.

A2.3: Students will be able to select optimal research techniques, and create corresponding data sources.

- Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading)
- Developing and exploiting databases and corpora (e.g., manual data annotation).

A2.4: Students will be able to select and implement the optimal data analysis method.

A2.5: Students will be able to infer theoretical consequences from the specific data analysis results.

A3: Students will be able to adapt a research design to the available research infrastructures.

Students will be familiar with the type of data extractable from the available corpora.

A3.1 Students will be able to select of optimal research techniques, select and create corresponding data sources (see also A2.3)

- data compilation, data analysis;
- understanding, selecting and performing optimal statistical tests and models.

B: Infrastructures & techniques

B1: Students will be able to identify and apply suitable infrastructures & techniques for obtaining, sharing and managing data.

Students will be able to use srWaC in order to extract alternation data.

B1.1: Students will be able to identify, collect, create and/or use relevant data for their research projects

- Searching relevant corpora
- Citing linguistic data sets as appropriate.

B1.2: Students will be able to document a research process.

B3: Students will be able to identify and apply suitable infrastructures & techniques for analysing data.

B3.1: Students will be able to select and use concordancers for the analysis of corpora.

C: Subject-specific learning outcomes

C1: Students will be able to apply the basic concepts of descriptive phonology to BCMS.

C2: Students will be able to apply the basic concepts of descriptive morphology to BCMS.
C3: Students will be able to apply the basic concepts of Optimality Theory to the alternation in focus.

xi. Overview of evaluation

This course is suitable for pass/fail grading, where all students who regularly contribute to the data collection and in-class discussions pass the course.

Rubric	Weighing
Participation incl. homework (initiative, forward-thinking, problem solving, critical thinking, organisation, time management)	100%

II.VI. Course Description 10: From computational linguistics via clinical linguistics to forensic linguistics

i. Name of the course
Applications of Linguistics: From computational linguistics via clinical linguistics to forensic linguistics

ii. Level of the course
MA (can also be taught to advanced BA students)

iii. Workload
5 ECTS

iv. Institution
University of Graz

v. Course instructor(s)
Stefan Milosavljević

vi. Brief course description
In this semester (winter, 2021/2022), the interdisciplinary course *Applications of Linguistics: From computational linguistics via clinical linguistics to forensic*

linguistics focuses on the applications of the software-based Social Network Analysis (SNA) in a variety of fields linking linguistics to both natural and social sciences.
The course is organized in six main units.
In the first part of the course, network theory and SNA are introduced and their application in different areas are discussed: in exploring social relationships (e.g. online social networks such as *Facebook* or *Twitter*), sports (e.g., networks of players based on mutual interactions), psychology (e.g., the mental lexicon as a network); epidemiology (e.g., how viruses spread through a network); movies (networks of actors), etc. This unit is meant to introduce the main concepts and methodology by using familiar examples.
The second part of the course is devoted to the application of SNA in exploring narratives. The students learn how to use the software-based SNA to analyze the basic properties of (literary) texts, such as identifying the main characters, extracting the most important segments of a given text, comparing 'fictional' and historical ('realistic') texts, and identifying the authorship. SNA is conducted in the programming language *R*. For concreteness, the SNA of the *Dictionary of the Khazars* by Milorad Pavić is used as a case-study.
The identification of authorship, which is introduced in this part of the course, anticipates the role of Forensic Linguistics, which is a topic of the fifth part of the course.
In the third part of the course, the focus is on the software-based SNA of language networks: the mental lexicon, semantic networks of the creative mind, and/or networks based on grammatical and/or derivational relations. A practical part consists in analysing a network of Serbo-Croatian prefixes, which is extracted from the database of Western South Slavic languages (WeSoSlaV), being developed within the project *Hyperspacing the verb* at the University of Graz and the University of Nova Gorica.
The fourth part of the course brings together Network Science, Computational Linguistics and Clinical Linguistics, by focusing on the software-based analysis of networks extracted from the clinical language data (autism, schizophrenia, aphasia, late-talking children, children with Down syndrome). In this part of the course, the focus is on discussing the published papers on the topic. The students learn how computational and network science tools and metrics (familiar from the previous classes) can be applied in analysing clinical cognitive networks and how these sets of tools can be used for the diagnosis and classification of language-related diseases, the quantification of condition severity, as well as for preparing treatment protocols.
In the fifth part of the course, Forensic Linguistics is introduced from the interdisciplinary perspective: the focus is on using the computational network tools

and metrics in the field of author identification and plagiarism, as prominent topics in Forensic Linguistics.
Each of the above described five parts of the course consists of three main building blocks: i) the instructor's brief introduction of the phenomenon, ii) reading an article related to a given topic, and iii) practical exercises (modelling the relevant networks, preparing the material using online corpora, data analysis in the programming language *R*).
At the transition point from the first to the second unit, the students are introduced with a form of a final report that they handle over at the end of the course, in which they describe a computational analysis of a network based on some language/textual data. They choose (in consultation with the instructor) some narrative and/or language phenomenon that can be analyzed from the SNA perspective. During the entire course, a part of each class is devoted to discussing their projects.
In the final, sixth part of the course, the students work exclusively on finalising their projects and writing up the final report, which they submit at the end of the course (or after the course, depending on the established deadline).

vii. Research related subject

Language and narrative networks

viii. Data the students work with

Data obtained from corpora: Stanford Large Network Dataset Collection (http://snap.stanford.edu/data/index.html); Project Gutenberg (www.gutenberg.org); BCMS corpora srWaC, hrWac, bsWac, meWac (https://www.clarin.si/noske/index.html); the databases of the project Hyperspacing the verb (in preparation).

ix. Topics

A: Research design

A1: General research design

[Teaching materials]
quanteda: Quantitative Analysis of Textual Data (tutorial in English): http://quanteda.io/
Network Analysis and Visualization with *R* and *igraph* (tutorial in English): https://kateto.net/networks-r-igraph
Network science (an interactive online book in English): http://networksciencebook.com/

A2: Adapting the general research design to the specific topic of interest

Identifying the key properties of textual/narrative networks: the main characters, the most important segments of texts, authorship, etc.; analyzing the structure of mental lexicon by

extracting the most prominent types of words and their relations; learning about the possibilities of diagnosing different types of language impairments based on network properties; detecting plagiarism and identifying authors based on network properties

A2.1: Formulation of questions and hypotheses in terms of variables

A2.2: Selection of appropriate research techniques, selection and creation of corresponding data sources

- Computational tools for textual analysis (the programming language *R*, especially the library *quanteda*)
- Computational tools for SNA (especially the *igraph* package in the programming language *R*)
- Developing and exploiting databases and corpora (e.g. extraction of language networks from the BCMS corpora *srWac* and *hrWac*, as well as from the databases within the project *Hyperspacing the verb*)

A2.3: Identifying the optimal data analysis method

A2.4: Inferring theoretical consequences from the specific data analysis results

A3: Adapting the research design to the available research infrastructures

Familiarising with the available computational tools relevant for SNA of narratives and language
Familiarising with the type of data extractable from the available datasets

A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3)

- computational analysis of textual data, Social Network Analysis (SNA);
- modelling textual and language networks, selecting optimal computational tools (packages, libraries) for extracting the relevant material; choosing appropriate SNA metrics for performing the analysis and visualising the results

A4: Research reporting

Identifying optimal formats for presenting different types of research outcomes and making them available to a wider audience: using the repositories for the public storing of the relevant code (e.g., github.com), using appropriate visualisation techniques (graphs, charts, tables, etc.) depending on the type of a written form (reports, presentations, posters, articles)

A4.1 Presentation modes for research reporting (short oral presentation, poster, squib, report, article etc.)

A4.2 Established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/squib/report,
- organization of the presentation,
- amount of text and graphical items on a poster (including text size),
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions

B: Infrastructures & techniques

B1: For obtaining literature

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Google Scholar, Academia.edu, slideshare.net
[DISCIPLINARY REPOSITORY] lingbuzz, Stanford Large Network Dataset Collection (http://snap.stanford.edu/data/index.html), Project Gutenberg (www.gutenberg.org)

B2: For obtaining, sharing and managing data

Use of the repositories for sharing the relevant code (e.g. guthub.com), advanced use of the available corpora for extracting narrative and language networks

B2.1: Definition of research infrastructures and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**

B2.2: Platforms and repositories

- **General-purpose repositories** and **disciplinary repositories**
  - [GENERAL-PURPOSE REPOSITORY] FigShare, github.com, pixabay.com;
  - [DISCIPLINARY REPOSITORY] CLARIN; Stanford Large Network Dataset Collection; Project Gutenberg

B2.3:Identifying, collecting, creating and/or using relevant data for research projects

- Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them
- Citing linguistic data sets as appropriate.
- Depositing research data in a **certified repository** and selecting an appropriate licence for sharing their data
- The **versioning** policy of repositories
- Familiarity with online survey tools

B2.4: **Data management plan**

- Understanding the **data lifecycle**

- Understanding how to generate data, analyse and handle it
- Understanding the **legal and ethical issues** around data generation and use (e.g. licensing, GDPR compliance, anonymisation, the importance of FAIR principles and Open Access).
- Secure storage and backup of research data
- Documenting workflows and what metadata to use to describe the nature of the data based on existing standards.
- What data needs to be destroyed, preserved in a data repository and made available for reuse

B3: For analysing data
B3.1: Software for computational text analysis (the programming language *R*, library *quanteda*)
B3.2: Software for SNA (the programming language *R*, package *igraph*)
C: Subject-specific topics
C1: Basic concepts, terms and methods in Computational Linguistics and their application
C2: Basic concepts, terms and methods in Clinical Linguistics and their application
C3: Basic concepts, terms and methods in Forensic Linguistics and their application
C4: Basic concepts, terms and methods in Network Science and their application
C5: Basic concepts, terms and methods in Social Network Analysis and their application

x. Learning outcomes

A: Research design
A1: Students will be able to make an overview of the general research design.
[Teaching materials] *quanteda*: Quantitative Analysis of Textual Data (tutorial in English): http://quanteda.io/ Network Analysis and Visualization with *R* and *igraph* (tutorial in English): https://kateto.net/networks-r-igraph Network science (an interactive online book in English): http://networksciencebook.com/
A2: Students will be able to create a suitable research design for the specific topic of interest.

Students will be able to identify the key properties of textual networks (the main characters, the important segments of texts, autorship), to analyze the structure of mental lexicon by extracting the most prominent types of words and their relations, to recognize the network properties of textual/language data that may be relevant in diagnosing different types of language impairments, in detecting plagiarism, and in identifying the authorship.

A2.1: Students will be able to formulate questions and hypothesis in terms of variables.

A2.2: Students will be able to select optimal research techniques, and create corresponding data sources

- Computational tools for textual analysis (the programming language *R*, especially the library *quanteda*)
- Computational tools for SNA (especially the *igraph* package in the programming language *R*)
- Developing and exploiting databases and corpora (e.g. extraction of language networks from the BCMS corpora *srWaC* and *hrWaC*, and from the databases within the project Hyperspacing the verb).

A2.3: Students will be able to select and implement the optimal data analysis method.

A2.4: Students will be able to infer theoretical consequences from the specific data analysis results.

A3: Students will be able to adapt a research design to the available research infrastructures.

Students will be familiar with the available computational tools relevant for the language and texts analysis from the perspective of network theory;
Students will be familiar with the type of data extractable from the available datasets.

A3.1 Students will be able to select optimal research techniques and data sources, in particular:

- computational analysis of textual/language data, Social Network Analysis (SNA);
- modelling textual and language networks, selecting optimal computational tools (packages, libraries) for extracting the relevant material; choosing appropriate SNA metrics for performing the analysis and visualising the results.

A4: Students will be able to report on their performed research in accordance with standards and conventions in the field.

Students will be able to identify optimal formats for presenting different types of research outcomes and make them available to a wider audience: to use the repositories for the public storing of the relevant code (e.g. github.com), and to apply appropriate visualisation

techniques (graphs, charts, tables, etc.) depending on the type of a written form (reports, presentations, posters).

A4.1 Students will be able to select and implement different presentation modes for research reporting (short oral presentation, poster, squib, report, article etc.).

A4.2 Students will be able to implement established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/squib/report,
- organization of the presentation,
- amount of text and graphical items on a poster (including text size),
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions.

B: Infrastructures & techniques

B1: Students will be able to identify and apply suitable infrastructures & techniques for obtaining literature.

- [GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu,
- [DISCIPLINARY REPOSITORY] lingbuzz, Stanford Large Network Dataset Collection (http://snap.stanford.edu/data/index.html), Project Gutenberg (www.gutenberg.org).

B2: Students will be able to identify and apply suitable infrastructures & techniques for obtaining, sharing and managing data.

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Google Scholar, Academia.edu, slideshare.net, github.com
[DISCIPLINARY REPOSITORY] lingbuzz.

B2.1: Students will understand what research infrastructures are, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**.

B2.2: Students will be able to identify suitable platforms and repositories, and to understand the difference between **general-purpose repositories** and **disciplinary repositories**.

B2.3: Students will be able to identify, collect, create and/or use relevant data for their research projects, and

- to cite linguistic data sets as appropriate,
- to deposit their research data in a **certified repository** of their choice and select an appropriate licence for sharing their data,
- to understand the **versioning** policy of the repository,
- to use online survey tools.

B2.4: Students will be able to create a **data management plan** • Understand the **data lifecycle** • Understand how to generate data, analyse and handle it • Understand the **legal and ethical issues** around data generation and use (e.g. licensing, GDPR compliance, anonymisation, the importance of FAIR principles and Open Access). • Know how to securely store and backup their research data • Know how to document their workflows and what metadata to use to describe the nature of the data based on existing standards. • Know what data needs to be destroyed, preserved in a data repository and made available for reuse.
B3: Students will be able to identify and apply suitable infrastructures & techniques for analysing data.
B3.1: Software for computational text analysis (the programming language *R*, library *quanteda*).
B3.2: Software for SNA (the programming language *R*, package *igraph*).
C: Subject-specific learning outcomes
C1: Students are familiar with the basic concepts, terms and methods in Computational Linguistics and are able to apply them in specific problem-solving.
C2: Students are familiar with the basic concepts, terms and methods in Clinical Linguistics and are able to apply them in specific problem-solving.
C3: Students are familiar with the basic concepts, terms and methods in Forensic Linguistics and are able to apply them in specific problem-solving.
C4: Students are familiar with the basic concepts, terms and methods in Network Science and are able to apply them in specific problem-solving.
C5: Students are familiar with the basic concepts, terms and methods in Social Network Analysis and are able to apply them in specific problem-solving.

xi. Overview of evaluation	
Rubric	Weighing
Participation in classes incl. homework (mini-projects related to each of the topics, mini-presentations, discussing the reading material)	40 %

Final written report	60 %

xii. Reading materials

Beckage, N. M., & Colunga, E. (2016). Language Networks as Models of Cognition: Understanding Cognition through Language. In A. Mehler, L. Andy, B. Sven, P. Blanchard, & B. Job (Eds.), *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks* (pp. 3–28). Springer.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774.

Castro, N., Stella, M., & Siew, C. S. Q. (2020). Quantifying the Interplay of Semantics and Phonology During Failures of Word Retrieval by People With Aphasia Using a Multiplex Lexical Network. *Cognitive Science*, *44*(9).

Cummings, L. (2013). Clinical linguistics: state of the art. *International Journal of Language Studies*, *7*(3), 1–32.

Jockers, M. L. (2014). *Text Analysis with R for Students of Literature*. Springer.

Kenett, Yoed N., and Miriam Faust (2019). Clinical Cognitive Networks: A Graph Theory Approach. In M. S. Vitevitch (ed.), *Network Science in Cognitive Psychology*, 136–165. Routledge.

Olsson, J. (2008). *Forensic Linguistics: Second Edition*. Continuum.

Raj S., Kannan B., Jagathy Raj V. P. (2021) Significance of Network Properties of Function Words in Author Attribution. In: Satapathy S., Zhang YD., Bhateja V., Majhi R. (eds.) Intelligent Data Engineering and Analytics. Advances in Intelligent Systems and Computing, vol 1177. Springer, Singapore.

Stewart, L. L. (2006). Computational Stylistics. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics*. Elsevier.

II.VII. Course Description 11: Automatic Speech Recognition/Forced Alignment

i.	**Name of the course** Automatic Speech Recognition/Forced Alignment
ii.	**Level of the course** BA Linguistics, but can be scaled up towards MA, research MA, PhD
iii.	**Workload** 6 ECTS
iv.	**Institution** Radboud University, Nijmegen, NL
v.	**Course instructor(s)** Louis ten Bosch c.s.
vi.	**Brief course description** This ASR/FA course is based on research-based teaching (RBT), i.e. will be based on the integration of ongoing research into ongoing teaching. The course's topic, ASR and FA, is quite technical in nature and progresses rapidly over recent years. It is therefore not at all straightforward to have bachelor students involved in ongoing research projects or even have them apply the same research tools and infrastructures that the researcher uses. However, it is possible for them to acquire and actively apply problem-solving and analytical skills using methods based on actual research if their background matches sufficiently with the background of the course. In this way, students gain experience from actual real-life research as part of their training. Since ASR has matured into a broad research domain with over 10 specialized subdomains, short-term student involvement in projects is necessarily very specific, both in terms of the topic and in terms of method. In order to guarantee that students get a sufficiently wide perspective on ASR/FA research, this course provides a general introduction to ASR/FA principles. **The course is built around the following structure** (a) general framework and theoretical design (b) analytic skills, tools, use of infrastructure, and use of literature (c) thesis/research report The course consists of 10 modules. Each of these modules addresses a specific topic within ASR/FA. The last module (number 10) consists in the writing of a thesis. The topic of this thesis is up to the student (e.g.,can be chosen from a predefined list).
vii.	**Research related subject** Speech decoding, signal analysis

viii. Data the students work with Open source audio data (mainly English)
ix. Topics

A: Research design
A1: General research design
[TEACHING MATERIAL] https://docs.google.com/document/d/1nSP-QeJJHWCfjDQ_0IUmAfF61kw7D2kf/edit?usp=sharing&ouid=111755079485368776002&rtpof=true&sd=true https://web.stanford.edu/~jurafsky
A2: Adapting the general research design to the specific topic of interest
A2: Adapting the general research design to the specific topic of interest
A2.1: Formulation of a specific theoretical or practical issue in the general framework of signal-to-word ASR pipeline
A2.2: Formulation of detailed question
A2.3: Selection of appropriate research techniques, selection and creation of corresponding data sources • Experimental paradigms (e.g., feature selection, acoustic modelling, interpretation of non-semantical ASR output) • AM/LM-based architectures versus end-to-end • Developing and exploiting databases and corpora for improving ASR performance
A2.4: Identifying an ASR performance error analysis method
A2.5: Inferring theoretical and practical consequences of chosen approach
A3: Adapting the research design to the available research infrastructures
A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3) • literature search in ejournals • connection literature findings to own results
A4: Research reporting
A4.1 Presentation modes for research reporting: thesis (3EC)

A4.2 Established procedures and conventions in research reporting, such as: • the ordering of thematic units in the thesis, • organization and clarity of themes addressed, • balance text and graphical items in thesis • clarity terminology, • citing conventions
• B: Infrastructures & techniques
B1: For obtaining literature
[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu
B2: For obtaining, sharing and managing data
B2.1: Definition of research infrastructures are, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**.
B2.2: Platforms and repositories. • **General-purpose repositories** and **disciplinary repositories**
B2.3:Identifying, collecting, creating and/or using relevant data for research projects • Searching and selecting relevant speech corpora • Citing data sets as appropriate. • Awareness about research data, certified repository and selection of an appropriate licence for sharing data
• B3: For analysing data
B3.1: Software for investigating audio signals
B3.2: Software platform for computations (python)
B3.3: Familiarity with existing ASR platforms

x. Learning outcomes

A: Research design
A1: Students will be able to make an overview of the general research design.

A2: Students will be able to create a suitable research design for the specific topic of interest.
A2.1: Students will be able to formulate general questions and hypothesis in terms of ASR/FA pipeline.
A2.2: Students will be able to formulate a specific question about an aspect of the ASR/FA pipeline.
A2.3: Students will be able to select optimal research techniques, and search/create corresponding data sources.
A2.4: Students will be able to select and implement the optimal speech decoding method.
A2.5: Students will be able to infer (speech-related, linguistic, theoretical) consequences from the specific ASR/FA outcomes.
A3: Students will be able to adapt a research design to the available research infrastructures.
A3.1 Students will be able to select of optimal research techniques, select and create corresponding data sources (see also A2.3) • students will be able to link research outcomes to data.
• A4: Students will be able to report on their performed research in accordance with standards and conventions in the field.
A4.1 Students will be able to cast research outcomes in terms of a thesis.
A4.2 Students will be able to implement established procedures and conventions in research reporting, such as: • the ordering of thematic units • organization of the presentation, • balance text/graphical items • proper citation of references • citing conventions.
B: Infrastructures & techniques
B1: Students will be able to identify and apply suitable infrastructures & techniques for obtaining literature.
B2: Students will be able to identify and apply suitable infrastructures & techniques for obtaining, sharing and managing data.

B2.1: Students will understand what ASR/FA pipelines are.
B2.2: Students will be able to identify suitable platforms ASR/FA. ● Understand the difference between ASR and FA.
B2.3: Students will be able to identify, collect, create and/or use relevant data for their research projects ● Cite audio data sets as appropriate.
B2.4: Students will be aware of the relevance of **data management plan** ● Understand the **data lifecycle** ● Understand how to generate data, analyse and handle it ● Understand the **legal and ethical issues** around speech data generation and use.
B3: Students will be able to identify and apply suitable infrastructures & techniques for analysing data.
B3.1: Students will be able to select and use software for speech analyses.
B3.2: Students will be able to select and use software for ASr/FA.
B3.3: Students will be able to select and use software for interpreting ASR output.

xi. Overview of evaluation	
Rubric	Weighing
Participation incl. homework (initiative, forward-thinking, problem solving, critical thinking, organisation, time management)	30%
On (individual) written final research report in line with the local university standards	70%

II.VIII. Course Description 12: Research-based Course in Multilingual NLP

i. Name of the course
Research-based Course in Multilingual NLP

ii. Level of the course
BA, MA

iii. Workload
6 ECTS

iv. Institution
University of Zurich

v. Course instructor(s)
Tanja Samardžić

vi. Brief course description

The students enrolled in this course will pursue small individual research projects, each tackling a separate problem as one part of a bigger project. They will be shortly introduced to the main opportunities and challenges of multilingual NLP and then select the problem to work on during the course. For example, the question of what tokenization method is optimal for which script requires studying several scripts and several tokenization methods. Each student can work on one script comparing several tokenization methods.
At the end of the course, each student will have written a 10-page research report following the same template. During the course, the students will regularly submit in-progress reports and meet with the lecturer in weekly interactive sessions.
The main output of the course activities will be students' final reports. In addition to this, the lecturer will compile a progress report summarising observations on how the course advanced each week. The lecturer's report will contain an overview of the students' projects and how they relate to each other.

vii. Research related subject
Multilingual NLP

viii. Tools and data the students work with
Own Python programs, data from highly multilingual data sets (our own TeDDi sample, other data sets derived from Wikipedia)

ix. Topics
See "Learning outcomes".

x. Learning outcomes
The learning outcomes of this course are divided into two parts, topic-specific and transversal. Regarding the topic-specific part, the students who complete this course will have an overview of the issues related to multilingual processing such as the impact of various writing systems on text encoding, subword tokenization and basic

text statistics. They will be able to apply relevant steps in the current processing pipeline (using large pre-trained models and cross-lingual transfer). Regarding the transversal skills, the students will gain experience with problem solving, forward thinking and independent work. They will activate their own curiosity as a driver of research activities.
xi. Evaluation Research report evaluation, following the criteria outlined in Figure 4 above.

II.IX. Course Description 13: Directed Study in Linguistics

i. Name of the course Directed Study in Linguistics
ii. Level of the course BA
iii. Workload 4 ECTS
iv. Institution University of Malta
v. Course instructor(s) Marc Tanti
vi. Brief course description This course is aimed at students who wish to deepen their knowledge of some area of linguistics, by carrying out an assigned research or analysis task under supervision. It is on offer to 2nd and 3rd year undergraduate students, as well students in the MA Preparatory Course, which is designed to allow graduates of non-Linguistics-related degrees to undertake an MA in Linguistics. In the specific iteration run for UPSKILLS, the directed study is in the area of Computational Linguistics. Students will be scraping and parsing recipe webpages in order to automatically extract a CookLang formatted recipe. With the help of an assigned tutor, the students will:

- Write a Python script that downloads the HTML of a handful of assigned recipe webpages.
- Use BeautifulSoup and regular expressions in order to find the title, ingredients, and procedure of the recipe.
- Parse the ingredients to separate the item, quantity, and unit of each ingredient.
- Determine the different ways that the ingredients are referred to in the recipe procedure. An important distinction of the CookLang recipe format is that ingredients are highlighted within the recipe procedure rather than given as a list at the top, so some minor natural language processing needs to be applied in order to determine different ways how an ingredient is mentioned, such as olive oil being referred to as 'the oil'.
- Output a text file containing the recipe formatted in CookLang.
- Compare the automatically generated text files with manually written gold text files in order to evaluate the program.
- Write a report about how the program works, how it performed, what was learned, and how the program can be improved.

vii. Research related subject

Natural Language Processing, Information extraction

viii. Tools and data the students work with

Python, HTML, regular expressions, shallow parsers

ix. Topics

The topics to be covered are:

- Natural language processing (without machine learning).
- Information extraction.
- Shallow parsing.
- Text distance measures (for comparing the automatically generated output to the gold output).

x. Learning outcomes

The students will learn:

- How to scrape websites.
- How to identify patterns that can be exploited to make a small program work on different websites.
- How to extract reliable bits of information from text.
- How to convert data between different formats.
- How to evaluate text generation.

xi. Evaluation

Research report evaluation, following the criteria outlined in Figure 4 above.

xii. Further information on the course that the instructor considers relevant (assessment, background, reading materials, detailed weekly plan, career paths etc.)

Given that this course comprises an extensive practical component, the following sources will be used to familiarise the students with the task they need to complete:

- https://www.geeksforgeeks.org/python-web-scraping-tutorial/
- https://beautiful-soup-4.readthedocs.io/en/latest/
- https://docs.python.org/3/howto/regex.html
- https://www.nltk.org/api/nltk.chunk.html

II.X. Course Description 14: Research-Based Course in Experimental Linguistics

i. Name of the course

Advanced Research Methods in Linguistics 2

ii. Level of the course

MA

iii. Workload

6 ECTS

iv. Institution

University of Malta

v. Course instructor(s)

Paul Marty

vi. Brief course description

The aim of this course is to provide MA students with a solid grounding in behavioural research methods for linguistics, to equip them with practical skills for conducting experimental linguistic research and to provide them with first-hand research experience.
The lectures cover various topics in experimental linguistics, from core experimental techniques to ethical considerations, each of which is approached both from a theoretical and practical perspective. The main focus of the course, however, is to guide students to pursue

small individual projects related to their MA dissertation topic. Depending on the student's topic and the progress of their research, this project may take any one of the following forms:

- Creating the materials and design for a pilot study
- Programming an online experiment (e.g., a web-based survey)
- Analysing original data or data of interest from published studies, or reanalysing data using other analysis techniques (e.g., Bayesian statistics)

The purpose of the project is primarily to enhance problem-solving skills and provide further motivation for the students to pursue their research independently, while also complementing the work leading up to their MA dissertation. Students may also take this opportunity to conduct a pilot study prior to conducting a larger-scale study for their MA dissertation.

In the last class, students present the results of their project in front of fellow students and faculty staff and each presentation gets discussed (short conference format: 10/15-minute presentation followed by 5-minute Q&A session). At the end of the course, students submit a research report (about 3,000 words). The report takes the form of a short experimental paper similar in essence to those published in scientific journals and includes a discussion of the comments and potential concerns raised during the presentation.

vii. Research related subject

Quantitative research methods, experimental research designs, language data science

viii. Tools and data the students work with

- The tools that the students work with are determined by considering their domain of inquiry, the progress of their research and their programming skills. In general, the students will be introduced to tools for managing, annotating and analysing data (Excel, Jamovi, R, Praat, ELAN), for building online tasks/surveys (Qualtrics, Gorilla task builder, PennController for IBEX), for recruiting participants (Prolific) and for making their data and results publicly available (OSF, lingbuzz).
- The students are encouraged to work with their own data. If a student has not collected any data yet, open source data related to the student's chosen topic or research method is provided to exemplify certain aspects of data collection and analysis that are of primary interest for the student.

ix. Topics

The course covers a range of classical topics in experimental linguistics including experimental measurement techniques, experimental design and design sensitivity, sampling methods, data collection and management, statistical inferences and hypothesis testing, ethical issues, among others. The list of topics is usually finalised at the start of the semester and tailored to the more immediate research goals of the students enrolled in the course. Tools for managing and analysing data, creating online studies, etc. are introduced in the course of the semester based on the students' self-reported needs.

x. Learning outcomes

The learning outcomes of this course are both general and student-specific.

1. General learning outcomes:

   By the end of this course, the student will be able to:

- appreciate how knowledge is created in experimental linguistics
- create small-scale psycholinguistics experiments
- present research results to an audience/reader

2. Student-specific learning outcomes:
   By the end of this course, the student will be able to:

- select or devise an experimental method suitable for their research question
- formulate the rationale for their choice of design and data analyses
- fine-tune their experimental studies for their MA dissertation

xi. Overview of evaluation

Rubric	Weighing
Oral presentation	25 %
Final written research report	75%

xii. Career paths

1. Academia
2. Language data science and, more generally, all careers involving the collection and analysis of quantitative data (e.g., marketers, government agencies, economists, research groups)

xiii. Readings

Readings partly depend on the student's research focus and chosen topic. The references listed below are relevant to various topics in experimental linguistics and they are provided to the students to ensure a common base of knowledge and practical skills. Supplementary readings are determined in the course of the class typically through discussions with the students and their MA supervisor(s).

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press.

Desagulier, G., Desagulier, G., & Amboy. (2017). *Corpus linguistics and statistics with R*. Springer International Publishing.

Fink, A. (1995). *How to ask survey questions*. London: Sage.

Gillioz, C., & Zufferey, S. (2020). *Introduction to experimental linguistics*. John Wiley & Sons.

Goodall, G. (2021). *Theory and Experiment in Syntax*. Routledge.

Gries, S. T. (2021). *Statistics for Linguistics with R*. De Gruyter Mouton.

Johnson, K. (2008). *Quantitative methods in linguistics*. John Wiley & Sons.

Kumar, Ranjit (2005). *Research Methodology: A Step-by-Step Guide for Beginners*. London: Sage.

Woods, A., P. Fletcher and A. Hughes (1986). *Statistics in language studies*. Cambridge: Cambridge University Press.

II.XI. Course Description 15: Language acquisition in Slavic

i.	**Name of the course** Language acquisition in Slavic: Obtaining, representing and analysing empirical data in linguistics (Field-work, building corpora, running experiments)
ii.	**Level of the course** MA, PhD (can also be taught to advanced BA students)
iii.	**Workload** 6 ECTS
iv.	**Institution** University of Graz
v.	**Course instructor(s)** Boban Arsenijević
vi.	**Brief course description** The course enhances the problem-solving and data-analysis skills, thus preparing the students for a wide range of possible careers. It also provides the students with first-hand scientific research experience. The course consists of three parts. During the introductory part, the central concepts of research in language acquisition are introduced (stages in language acquisition, aspects of language acquisition, central questions), mostly through the discussion of handbook chapters and articles (app. 80% of each class). Simultaneously, a very basic sketch or general research design is introduced and some preliminary discussion is held about testing the predictions of various theories (app. 20% of each class). The second part focuses on the same two topics as the first part (language acquisition and research). Now 80% is dedicated to research design. During this part, most preparatory work is done by attending the Movetia course *Introduction to research in linguistics: theory, logic, method* and coming up with an own research project. Students are also introduced to the Slavic collections of CHILDES (a CLARIN K-Centre), the corpus of child and child-directed language. In addition, students start sketching their research report based on the research-report template. The remaining 20% are reserved for the discussion of research articles (and, where necessary handbook chapters) which are deemed instrumental given the specific projects selected by the students.

The third part is entirely dedicated to the students' research projects. The students submit their preliminary drafts which are read before class and discussed in class. The lecturer also provides written feedback. After this, the students hold oral presentations of their projects. At the end of the course, the students submit their final research reports.
vii. Research related subject Language acquisition, specifics of language acquisition f Slavic languages
viii. Data the students work with Data obtained from corpora, data obtained from human subjects
ix. Topics

A: Research design
A1: General research design
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Adapting the general research design to the specific topic of interest
• Different stages in acquisition, ages, populations, various types of controls / baselines • Selecting the aspect of acquisition, defining measurable phenomena • Specifying the methods of empirical access, i.e. distributing between corpus and experimental methods • Distinguishing between demographic, social and linguistic data and measures, in the design of the research.
A2.1: Formulation of questions and hypothesis in terms of variables
A2.2: Formulation of predictions of H0 and H1
A2.3: Selection of optimal research techniques, selection and creation of corresponding data sources • Experimental paradigms (e.g., elicitation, judgements, forced-choice) • Developing and exploiting databases and corpora (e.g. manual data annotation)

A2.4: Identifying the optimal data analysis method
A2.5: Inferring theoretical consequences from the specific data analysis results
A3: Adapting the research design to the available research infrastructures
Familiarization with CHILDES, identifying available research networks, adapting the RD to the limitations of these infrastructures
A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3) • data compilation, data analysis; • understanding, selecting and performing optimal statistical tests and models
A4: Research reporting
Identifying optimal formats for the representation of the acquired data in terms of factors and effects
A4.1 Presentation modes for research reporting (short oral presentation, report, article etc.)
A4.2 Established procedures and conventions in research reporting, such as: • the ordering of thematic units in an article/squib/report, • organization of the presentation, • amount of text and graphical items on a slide/handout, • terminology, • citing conventions
B: Infrastructures & techniques
B1: For obtaining literature
[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz
B2: For obtaining, sharing and managing data
Familiarity with CHILDES and other acquisitional corpora
B2.1: Definition of research infrastructures, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**
B2.2: Platforms and repositories
B2.3: Identifying, collecting, creating and/or using relevant data for research projects

• Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them • Citing linguistic data sets as appropriate. • Familiarity with online survey tools
B3: For analysing data
ANOVA and linear mixed model tests. Tests for categorical non-binary variables
B3.1: Softwares for statistical tests
B3.2: Concordancers for the analysis of corpora
C: Subject-specific topics
C1: How does language develop in an individual?
C2: What factors determine this process?
C3: Innate vs. acquired language competences
C4: Aspects of atypical language development

x. Learning outcomes

A: Research design
A1: Students will be able to make an overview of the general research design.
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: PHIL: Movetia101 Introduction to research in linguistics: theory, logic, method https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) ReLDI-Project: ReLDI101 Introduction to Research Methodology in Linguistics https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Students will be able to create a suitable research design for the specific topic of interest.
Students will be able to describe and recognize different stages in language development and match them with approximate age;

Students will be able to define measurable phenomena for a particular aspect of language acquisition; Students will be able to design studies of the development of a particular phenomenon in language acquisition (e.g. verbal aspect, palatalization, clitic second).
A2.1: Students will be able to formulate questions and hypotheses in terms of variables.
A2.2: Students will be able to formulate H0 and H1.
A2.3: Students will be able to select optimal research techniques, and create corresponding data sources ● Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading) ● Developing and exploiting databases and corpora (e.g., manual data annotation).
A2.4: Students will be able to select and implement the optimal data analysis method.
A2.5: Students will be able to infer theoretical consequences from the specific data analysis results.
A3: Students will be able to adapt a research design to the available research infrastructures.
Students will be familiar with CHILDES, Students will be able to identify available language acquisition research networks, Students will be able to adapt the RD to the limitations of these infrastructures.
A3.1 Students will be able to select of optimal research techniques, select and create corresponding data sources (see also A2.3) ● data compilation, data analysis; ● understanding, selecting and performing optimal statistical tests and models.
A4: Students will be able to report on their performed research in accordance with standards and conventions in the field.
Students will be able to select the optimal format for the representation of the development of a particular language acquisition phenomenon.
A4.1 Students will be able to select and implement different presentation modes for research reporting (short oral presentation, report, article etc.).
A4.2 Students will be able to implement established procedures and conventions in research reporting, such as: ● the ordering of thematic units in an article/report,

- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions.

B: Infrastructures & techniques

B1: Students will be able to identify and apply suitable infrastructures & techniques for obtaining literature.

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, ROA.

B2: Students will be able to identify and apply suitable infrastructures & techniques for obtaining, sharing and managing data.

Students will be able to extract data from CHILDES.

B2.1: Students will understand what research infrastructures are, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**.

B2.2: Students will be able to identify suitable platforms and repositories.

B2.3: Students will be able to identify, collect, create and/or use relevant data for their research projects

- Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them
- Citing linguistic data sets as appropriate.
- Familiarity with online survey tools.

B3: Students will be able to identify and apply suitable infrastructures & techniques for analysing data.

Students will be able to use ANOVA and linear mixed model tests.

B3.4: Students will be able to select and use concordancers for the analysis of corpora.

C: Subject-specific learning outcomes

C1: Students will be able to discuss how language develops in typically developing individuals.

C2: Students will be able to evaluate arguments for and against treating certain linguistic capacities innate.

C3: Students will be able to distinguish between typical and atypical language development

and to identify the type of the atypical ones.
C4: Students will be able to identify typical traits of the acquisition of (case-rich) Slavic languages.

xi. Overview of evaluation	
Rubric	Weighing
Participation incl. homework (initiative, forward-thinking, problem solving, critical thinking, organisation, time management)	30%
Outputs based on the final research report • oral presentation • final written report	70%

xii. Reading materials

1. Sauerland, U., Grohmann, K., Guasti, M., Andelković, D., Argus, R., Armon Lotem, S., et al. (2016). How do 5-year-olds understand questions? Differences in languages across Europe. FIRST LANGUAGE, 36(3), 169-202 [10.1177/0142723716640236].
2. Anđelković, D. and M. Mirić. Distribution of verbal overgeneralizations in the Serbian Corpus of Early Child Language. Psihološka istraživanja 20 (2), 291–310.
3. Savić, Maja, Maša Popović and Darinka Anđelković. 2017. Psihologija 50 (4), 427-444. 10.2298/PSI160921007S.
4. Pavlinušić Vilus, Eva and Gordana Hržica.2022. Omissions and Overgeneralizations of Reflexive Clitic in the Acquisition of Reflexive Constructions in Croatian as L1. Časopis Instituta za hrvatski jezik i jezikoslovlje, 48 (1), 151-169. 10.31724/rihjj.48.1.7
5. Ilić, Tatjana. 2008. What does the acquisition of the involuntary state construction in Serbo-Croatian have to tell us about the ability to represent A-Chains in 3-year-olds? MS, University of Hawaii at Manoa
6. Helen, Goodluck., Danijela, Stojanovic. 1996. The Structure and Acquisition of Relative Clauses in Serbo-Croatian. Language Acquisition, 5(4):285-315. 10.1207/S15327817LA0504_2

II.XII. Course Description 16: Tense in Bosnian / Croatian / Serbian

i. Name of the course
Tense in Bosnian/Croatian/Serbian (BCS)

ii. Level of the course
MA, PhD

iii. Workload
5 ECTS

iv. Institution
University of Graz

v. Course instructor(s)
Boban Arsenijevic

vi. Brief course description
The course provides a state-of-the-art overview of the literature and ongoing research of one of the most hotly debated issues in Slavic linguistics. While the traditional view is that Slavic languages have tense inflection, and hence also tense as a grammatical category, recent advances signal that this might not be the case, and that Slavic languages might be coding tense information in terms of aspect and its pragmatic inferences. The course enhances the problem-solving and data-analysis skills, thus preparing the students for a wide range of possible careers. It also provides the students with first-hand scientific research experience.
A collection of state-of-the-art articles on the topic are read and discussed in each class. This discussion amounts to approximately 50% of the class.
For the second half of the class, the students are asked to focus on one of the claims made in the article read for the respective class (if the article is not about BCS, the students verify whether the claim about the other Slavic language applies to BCS) and present the results of their small-size empirical research testing this claim. The students are also asked to take the ReLDI course *Introduction to Corpus-Based Methods in Linguistics*.
In all but two last classes all students present the results of their small-size empirical research and their presentations get discussed.
In the two last classes, the students hold their final presentations in which they present the research on which they chose to submit their final report.
At the end of the course, the students submit their final report using the provided template.

vii. Research related subject
Syntax, morphology and semantics of the subjects and subject agreement. The syntactic height of the verb. Semantics of BCS verb forms.

viii. Data the students work with
BCMS corpora hrWaC, srWaC, bsWaC, the databases of the project Hyperspacing

the verb.
ix. Overview of topics

A: Research design
A1: General research design
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Adapting the general research design to the specific topic of interest
Identifying the syntactic and semantic indicators of tense, formulating semantic and syntactic tests. Establishing parallels with the nominal domain
A2.1: Formulation of questions and hypothesis in terms of variables
A2.2: Formulation of predictions of H0 and H1
A2.3: Selection of optimal research techniques, selection and creation of corresponding data sources • Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading) • Developing and exploiting databases and corpora (e.g. manual data annotation)
A2.4: Identifying the optimal data analysis method
A2.5: Inferring theoretical consequences from the specific data analysis results
A3: Adapting the research design to the available research infrastructures
Familiarising with the type of data extractable from the available corpora; Familiarising with the type of data extractable from online surveys; Choosing the optimal way of annotating the properties
A3.1 Selection of optimal research techniques, selection and creation of corresponding data sources (see also A2.3) • data compilation, data analysis; • understanding, selecting and performing optimal statistical tests and models,

- identifying the optimal data analysis method

A4: Research reporting

- Modeling the data in terms of the generative syntactic formalisms.
- Developing and formulating formalizations of the relevant semantic regularities in terms of the lambda calculus.
- Selecting the optimal formal morphological framework and using it to present the observed morphological regularities

A4.1 Presentation modes for research reporting (short oral presentation, report)

A4.2 Established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/squib/report,
- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions.

B: Infrastructures & techniques

B1: For obtaining literature

[GENERAL-PURPOSE REPOSITORY] Research Gate, Google Scholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, semantic scholar

B2: For obtaining, sharing and managing data

- Advanced use of the available corpora for the target language,
- Advance use of online surveys (use of CQLs, use of additional options).

B2.1: Definition of research infrastructures, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**

B2.2: Platforms and repositories

B2.3:Identifying, collecting, creating and/or using relevant data for research projects

- Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them
- Citing linguistic data sets as appropriate.
- Familiarity with online survey tools

B3: For analysing data

B3.1: Softwares for statistical analysis (R)

B3.2: Corpus management and text analysis software (NoSketch Engine, AntConc)
C: Subject-specific topics
C1: What is verbal tense?
C2: How does the verbal tense relate to clausal syntax and verbal morphology in Slavic and beyond?
C3: What is the semantics of tense and is it invariant or are there Slvic specifics?
C4: What are characteristic properties of Slavic verbal morphology?
C5: Basic concepts of Formal Semantics

x. Learning outcomes

A: Research design
A1: Students will be able to make an overview of the general research design.
[Teaching materials] UPSKILLS Moodle course First steps into scientific research https://upskillsproject.eu/project/scientific_research/ Movetia/ReLDI courses: https://phil.openedx.uzh.ch/courses/course-v1:PHIL+Movetia101+2046/info (in English) https://phil.openedx.uzh.ch/courses/course-v1:PHIL+ReLDI101+2018/info (in BCMS)
A2: Students will be able to create a suitable research design for the specific topic of interest.
Students will be able to identify the tense semantics, as well as the syntactic presence of tense, in a clause in an arbitrary language.
A2.1: Students will be able to formulate questions and hypotheses in terms of variables.
A2.2: Students will be able to formulate H0 and H1.
A2.3: Students will be able to select optimal research techniques, and create corresponding data sources • Experimental paradigms (e.g., elicitation, judgements, forced-choice, self-paced reading)

- Developing and exploiting databases and corpora (e.g. manual data annotation).

A2.4: Students will be able to select and implement the optimal data analysis method.

A2.5: Students will be able to infer theoretical consequences from the specific data analysis results.

A3: Students will be able to adapt a research design to the available research infrastructures.

Students will be familiar with the type of data extractable from the available corpora;
Students will be familiar with the type of data extractable from online surveys;
Students will be able to choose the optimal way to annotate the relevant syntactic, semantic and morphological properties.

A3.1 Students will be able to select of optimal research techniques, select and create corresponding data sources (see also A2.3)

- data compilation, data analysis;
- understanding, selecting and performing optimal statistical tests and models.

A4: Students will be able to report on their performed research in accordance with standards and conventions in the field.

Students will be able to represent the relevant semantic regularities in terms of the lambda calculus.
Students will be able to select the optimal formal syntactic framework for their data and use it to represent and explain the observed patterns.

A4.1 Students will be able to select and implement different presentation modes for research reporting (short oral presentation, report, article etc.).

A4.2 Students will be able to implement established procedures and conventions in research reporting, such as:

- the ordering of thematic units in an article/report,
- organization of the presentation,
- amount of text and graphical items on a slide/handout,
- terminology,
- citing conventions.

B: Infrastructures & techniques

B1: Students will be able to identify and apply suitable infrastructures & techniques for obtaining literature.

[GENERAL-PURPOSE REPOSITORY] ResearchGate, Googlescholar, Academia.edu, [DISCIPLINARY REPOSITORY] lingbuzz, semanticscholar.
B2: Students will be able to identify and apply suitable infrastructures & techniques for obtaining, sharing and managing data.
Students will be able to extract data from the available corpora for the target language, Students will be able to obtain data in online surveys.
B2.1: Students will understand what research infrastructures are, and the main concepts around **data interoperability**, such as **data**, **metadata** and **standards**.
B2.2: Students will be able to identify suitable platforms and repositories.
B2.3: Students will be able to identify, collect, create and/or use relevant data for their research projects • Searching, identifying and selecting relevant corpora from language resources platforms and repositories hosting them • Citing linguistic data sets as appropriate. • Familiarity with online survey tools.
B3: Students will be able to identify and apply suitable infrastructures & techniques for analysing data.
B3.1: Students will be able to select and use softwares for statistical tests.
B3.4: Students will be able to select and use concordancers for the analysis of corpora.
C: Subject-specific learning outcomes
C1: Students will be able to discuss the core issues of verbal tense.
C2: Students will be able to discuss the relation between tense and verbal morphology in Slavic.
C3: Students will be able to identify the potential mismatches between syntax, semantics and morphology regarding tense..
C4: Students will be able to apply syntactic, morphological and semantic models onto linguistic data.
C5: Students will be able to represent semantic insights in terms of lambda calculus formulae and to interpret such representations.

xi. Overview of evaluation	

Rubric	Weighing
Participation in classes (initiative, forward thinking, problem solving, critical thinking, organisation, time management)	20%
Homework (data collecting, annotation, analysis)	20%
Outputs based on the final research report ○ oral presentation ○ final written report	60%

xii. Career paths

a. Academia
b. Natural language processing
c. Market analysis

and all careers involving data analysis

xiii. Reading materials

Todorović, Neda. 2016. On the Presence/Absence of TP: Syntactic Properties and Temporal Interpretation. PhD Dissertation, University of Connecticut.

Арсенијевић, Бобан. 2013. Временско и аспектуално значење аориста. Српски језик 18, 253–263.

Seth Cable's handout on the theoretical background of tense: http://people.umass.edu/scable/PNWSeminar/handouts/Tense/Tense-Background.pdf https://www.youtube.com/watch?v=hvVMnlVuo9c