Open Call Collection OC-2020-1

Proposal Reference OC-2020-1-24356

Title: **Strategic Targeting Of Rich Morphology: Tools for language learning**

Acronym: **STORM**

Summary

Strategic Targeting Of Rich Morphology (STORM): Tools for language learning

Instead of requiring foreign and second language (L2) learners to absorb myriads of inflected forms, the Action uses technology to facilitate strategic focus on the most frequent forms, reducing the learning burden.

Many languages have “rich inflectional morphology”, meaning that words can have many different forms to signal grammatical categories such as case, number, tense, etc. Rich morphology presents a challenge for L2 learners because even a basic vocabulary of a few thousand words can entail mastery of over a hundred thousand word forms. However, only a handful of the potential forms of a given word occur frequently, while the remainder are rare.

Now that linguists have digital collections of language samples known as “corpora”, it is possible to scientifically determine which forms of any given word are of highest frequency, as well as what grammatical and collocational contexts motivate those few frequent forms. This makes it possible to strategically focus language learning in a new way and make new tools available to the public.

The Action envisions the creation of Tools for L2 language learning, a specific, measurable, tangible, and innovative product that will be delivered free and open-source. International coordination will scale up the creation of Tools with guidelines and standards portable across varieties of languages and users. Tools are relevant to the linguistic needs of Europe and timely given the recent advent of big language data and trends in Europe due to migration and economic expansion.

Key Expertise needed for evaluation

Languages and literature
Second language teaching and learning
Languages and literature
Linguistics: formal, cognitive, functional and computational linguistics
Other engineering and technologies
Databases, data mining, data curation, computational modelling for other engineering and technologies

Keywords

second language learning

linguistics

morphological complexity of languages

web-based language learning tools

digital humanities

TECHNICAL ANNEX

1 S&T EXCELLENCE

1.1 SOUNDNESS OF THE CHALLENGE

1.1.1 DESCRIPTION OF THE STATE-OF-THE-ART

Many languages have “rich inflectional morphology”, meaning that words can have many different forms to signal grammatical categories such as case, number, person, tense, etc. In Czech, for example, the paradigm for nouns has up to fourteen slots, and the paradigms for adjectives and verbs have over thirty slots each. This complexity is illustrated in Table 1 by a grammatical analysis of the word forms in a Czech proverb extolling the benefits of multilingualism. With the exception of the adverb *tolikrát* ‘that many times’, every word form belongs to a large paradigm and signals multiple grammatical categories.

Kolik řečí znáš, tolikrát jsi člověkem ‘How many languages you know – that many times you are a person’		
Word	**Gloss**	**Grammatical categories marked by morphology**
Kolik	How many	Indefinite numeral in Accusative case
řečí	language	Feminine noun in Genitive case and Plural number
znáš	know	Imperfective verb in Present tense, Second person Singular
tolikrát	that many times	Adverb
jsi	be	Imperfective verb in Present tense, Second person Singular
člověkem	person	Masculine noun in Instrumental case and Singular number

Table 1: The grammatical categories expressed by each word in a Czech proverb.

Rich inflectional morphology presents a special challenge for second and foreign language (L2) learners because even a basic vocabulary of a few thousand words can entail mastery of over one hundred thousand word forms. Morphology is considered both to be essential to L2 acquisition and to be a “bottleneck”, more difficult than both syntax and semantics [1, 2, 3]. The frequency of word forms follows a highly skewed Zipfian distribution [4], meaning that only a handful of the potential forms of any given word occur frequently, while the remainder are rare (some vanishingly so). This is particularly the case with Slavic languages (Croatian, Czech, Polish, Russian, Slovak, etc., cf. [5, 6]); Baltic languages (Lithuanian and Latvian); and Uralic languages (Estonian, Finnish, Hungarian, etc., cf. [7]). Moreover, rich morphology is not irrelevant for Romance languages (Spanish, Portuguese, etc.) that have large paradigms for verbs. Rich morphology is likewise frequent among minority, indigenous, and immigrant languages (Basque, Saami languages, Scottish Gaelic, Welsh, Turkish etc., cf. [8]). From the perspective of typology, many languages that are high in inflectional morphology belong to the Slavic (a sub-family of Indo-European) and Uralic families. In Europe, the distribution of majority languages with relatively rich inflectional morphology largely corresponds to COST’s list of Inclusiveness Target Countries, as displayed in Table 2.

It is not entirely clear how first (native) language (L1) learners acquire and navigate rich morphologies. A linguistic conundrum termed the “Paradigm Cell Filling Problem” [9] highlights the fact that native speakers of languages with rich inflectional morphology routinely recognize and produce forms that they have never been exposed to. For example, the gerund *nedokarmlivaja* ‘while underfeeding’ and the participle *nedokarmlivaemyj* ‘being underfed’ have no attestations in the Russian National Corpus (RNC, http://ruscorpora.ru/). Given that the RNC contains over 360 million words (as of March 2020), a quantity

roughly equivalent to the lifetime exposure of a human being of between 40 and 70 years old to their native language, this fact indicates that many native speakers have probably never encountered these word forms. However, all native speakers of Russian can be expected to readily understand and to produce these forms in appropriate contexts, as evidenced by rare occurrences that turn up in Google searches. A computational learning experiment [6] gives evidence that exclusive focus on the most frequent word forms can yield better results in terms of ability to produce forms of newly introduced words than expecting L2 learners to memorize entire paradigms. Irregular word forms that are harder to learn also tend to be highly frequent, making the exploitation of frequency data additionally strategic.

	Approximate Richness of Inflectional Morphology		
	Low	**Medium**	**High**
Languages of COST Inclusiveness Target Countries		Bulgarian(S), Luxembourgish(G), Macedonian(S), Portuguese(R)	Albanian(IE), Bosnian(S), Croatian(S), Montenegrin(S), Serbian(S), Czech(S), Estonian(U), Greek(IE), Hungarian(U), Latvian(B), Lithuanian(B), Maltese(SEM), Polish(S), Romanian(R), Slovak(S), Slovenian(S), Turkish(T)
Languages of other COST Member Countries	Danish(G), Dutch(G), English(G), Norwegian(G), Swedish(G)	French(R), German(G), Italian(R), Spanish(R)	Finnish(U), Irish(C)
Minority languages of Europe (sample)		Catalan(R), Yiddish(G)	Basque(I), Kven(U), Lower Sorbian(S), Romani(IE), Rusyn(S), Saami languages(U), Scottish Gaelic(C), Welsh(C)

Table 2: Approximate richness of inflectional morphology found in languages of COST Inclusiveness Target Countries, other COST Member Countries, and a sample of minority languages of Europe. Low = small verbal and nominal paradigms; Medium = large verbal and small nominal paradigms or medium-sized paradigms for both; High = large verbal and nominal paradigms. Language families are marked in parentheses. Within the Indo-European family, sub-families are B=Baltic, C=Celtic, G=Germanic, R=Romance, S=Slavic. Other Indo-European languages are marked IE. Non-Indo-European languages are marked I=Isolate, SEM=Semitic,T=Turkic, U=Uralic.

There exist large digital collections of language samples known as “corpora”. However, with some notable exceptions [10], corpus resources have been aimed primarily at linguists, not L2 learners, and it has been difficult to find ways to connect L2 learners to the powerful benefits of using corpus language data. Corpora make it possible to determine which forms of any given word are of highest frequency, as well as to discover what grammatical and collocational contexts motivate the most frequent forms. Contemporary textbooks to some extent represent the skewed distribution of forms by relying on the intuitions of their authors, but now a language technology solution can provide objective empirical evidence for these choices. Big language data and the tools to analyze this data make it possible to strategically focus L2 language learning in a new way and make new tools available.

The Action envisions a strategic leap forward in the use of technology to facilitate a bridge between evidence-based empiricism and the teaching and learning of languages: an online interactive Tool will offer strategic input for learning rich morphology. The Action is accordingly named Strategic Targeting Of Rich Morphology: Tools for language learning, with the acronym “STORM”. Each Tool will connect a “Target Language” (the L2 that a learner is acquiring) to a “User Language” (the language of the interface for the Tool, a language that the learner is already proficient in).

At present there exists one proof-of-concept, namely a fully functional Tool for one Target Language with rich inflectional morphology completed in early 2020 (the name of the Target Language and link to this webpage have been omitted for purposes of anonymity). This Tool is a free, publicly available interactive internet webpage that presents inflectional morphology for 3000 words in the Target Language stratified according to the Common European Framework of Reference for Languages (CEFR) Levels A1, A2, B1, and B2. In addition to filtering words according to Levels, users can filter for

Topics (such as “food” and “transport”), Analysis (grammatical categories such as case, number, tense, etc.), or select words from a Dictionary. Corpus research shows that three word forms account for over 80% of the uses of most words, so for each word, the user is shown the three most frequent word forms for that word (or only one or two forms if those account for over 90% of the uses of that word). Each word form appears in a sentence that illustrates the grammatical and lexical context that motivates the use of that form (relevant grammatical constructions and collocations). The user can click to hear audio of each sentence in the Target Language and to receive translations of each sentence into the User Language. All of the code and source files for the Tool are open-source and are housed on GitHub. A generic version of the programming code has been drafted to facilitate the creation of a Tool for any combination of Target Language and User Language.

While the existing proof-of-concept Tool is a breakthrough, it raises a raft of research questions that must be addressed in order to take this innovation to the next level. These questions are posed in Section 1.1.2.

1.1.2 DESCRIPTION OF THE CHALLENGE (MAIN AIM)

The main aim of STORM is to streamline and coordinate the creation and implementation of multiple Tools optimized to meet Europe’s needs for second and foreign language learning. Instead of requiring L2 learners to memorize myriads of inflected forms, the STORM Action uses technology to facilitate strategic focus on the most frequent forms, greatly reducing the burden of forms and the burden on processing efficiency in L2 inflection acquisition [11].

The challenge is to use language data to make it possible for L2 learners of languages with rich inflectional morphology to approach the proficiency of native speakers in terms of recognition and production of word forms. STORM will facilitate this by creating Tools for European languages with rich inflectional morphology (High and Medium levels in Table 2). This challenge evokes several sets of research questions (RQs) that address **Users**, **Linguistic facts**, **Technology**, and **Theories**. These research questions motivate Working Groups 1-4 (Section 4).

While some research relevant to these RQs is underway, at present this research is fragmented and scattered across languages and countries. The STORM Action is needed to bring this patchwork together into a coherent and efficient whole.

Users and their needs will determine the scope and design of Tools, which in turn need to be evaluated for efficacy. The Action thus researches not only the Tool, but also its use.

RQ1: Language Selection	What is the optimal set of Target Language and User Language combinations that serve Europe’s needs?
RQ2: L2 Learner Variation	How can the Tools be used most effectively for different kinds of L2 learners, including non-traditional learners and special needs learners?
RQ3: Evaluation	What are the optimal language teaching applications of the Tools, and how can the Tools be empirically evaluated (e.g., back-end data, experiments)?

Linguistic facts will require adaptations both to the overall structures of Tools and to questions of how the most strategic word forms and their contexts are identified for inclusion in Tools.

RQ4: Typology	How can the Tools account for linguistic facts specific to given languages, for example the differences between languages with agglutinative (e.g., Turkish) vs. fusional (e.g., Polish) morphology?
RQ5: Lexicon	What is the optimal way to select the lexicon for Tools, taking into account factors such as user needs, language proficiency levels, slang, jargon, taboo language, polysemy, diglossia, dialects?
RQ6: Gender	How can the Action create Tools that are gender inclusive, given the strong statistical bias in corpora for masculine over feminine word forms [12, 13]?
RQ7: Constructions and Collocations	What is the best way to determine the most typical grammatical constructions and lexical collocations that motivate the use of each of the most frequent word forms?

Technology relevant to the Tools is both specific to the language sciences and more general for the production of apps for users.

RQ8: Corpus Linguistics	How can the Action streamline extraction of the most frequent word forms from language corpora? How can the Action overcome challenges such as

	the lack of large corpora and other language technology for some languages, and overrepresentation of written genres in existing corpora?
RQ9: Computational Technology	What kinds of technological solutions such as crowdsourcing and gamification can be implemented to accelerate and optimize the implementation of the Tools?
RQ10: Programming	What are the best programming solutions for extending generic open-source code to create Tools for new Target Language and User Language combinations and for various types of L2 learners (see RQ2)?

Theories about language, how it is modeled and how language is acquired.

RQ11: Modelling morphology	The highly skewed distributions of word forms show that word forms are not equiprobable. What is a realistic cognitive model of inflectional morphology?
RQ12: Acquisition	How does frequency of word forms impact L1 acquisition (i.e., children learning their mother tongue) as compared with L2 acquisition?
RQ13: Informativity and Distinctiveness	How can the importance of factors other than frequency, such as informativity and distinctiveness of word forms be compared (cf. varying points of view on this issue [14, 15, 16])?

The STORM Action is timely both for global and for regional reasons. Two global science and technology factors that facilitate creation of these new Tools are the advent of large language corpora (often with linguistic annotation) and the development of powerful software for statistical analysis (such as R). Both the data and the means to analyze them have reached a point where they can be harnessed for creative and useful purposes. It is now possible both to scientifically determine which word forms are most frequent and to deliver a technological Tool that can bring about a sea change in how languages with rich morphology are taught and learned. Minority languages such as Lower Sorbian, which tend to be rich in inflectional morphology, are threatened with extinction: the UNESCO Atlas of the World's Languages in Danger [17] finds that there are endangered languages in every country in Europe, and indeed the world. Saving these languages is a race against time. Furthermore, in the period of 1990-2018 the economies of most COST Inclusiveness Target Countries have surged by several hundred percent [18]. This surge has spurred the immigration of workers who need to learn languages with rich morphology. Building Tools for these languages can have a positive effect on the European economy, by facilitating mobility within Europe and making these countries more accessible to international trade.

1.2 PROGRESS BEYOND THE STATE-OF-THE-ART

1.2.1 APPROACH TO THE CHALLENGE AND PROGRESS BEYOND THE STATE-OF-THE-ART

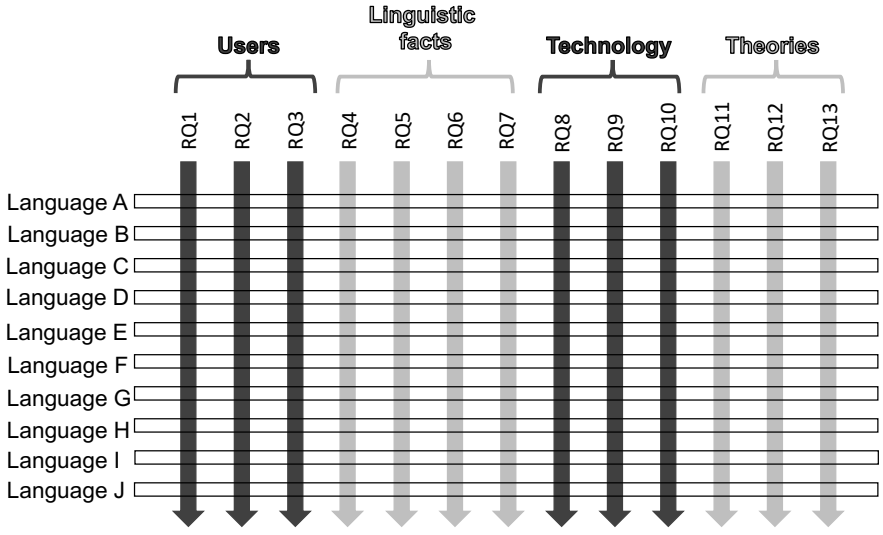

Figure 1: The STORM Action in terms of weaving, with Research Questions as the warp and a potentially unlimited number of Target Languages as the weft.

An apt analogy for the approach of the STORM Action is weaving, as depicted in Figure 1. The lengthwise or longitudinal warp yarns are Research Questions that form the structure of the loom. The transverse weft yarns are the Target Languages for the creation of Tools. The number of Target Languages is in principle unlimited.

The problem is that current research is mostly limited to individual “nodes”, or intersections of Target Languages and RQs as in Figure 1. The STORM Action is needed to stitch the entire cloth together and ensure that each achievement at a given node can spread along the warp lines of Research Questions. Thus the Action will support the development of Tools in all of the languages in the weft lines. A tapestry will emerge from the collective coordinated efforts of the researchers, whose partially overlapping subsets of intersections strengthen the fabric of the whole.

The integration of Research Questions across Target Languages in the Action creates a rich dynamic environment that advances both the field of linguistics and the production of Tools for Europe. The Action will make significant progress beyond the state of the art in these aspects:

1. Encourage novel collaborations across research domains and languages by threading together researchers in a new network that has not previously existed.

2. Set the standards for best practices in research on acquisition of rich morphology.

3. Devise a shared pipeline for the development of Tools that optimizes efficiency and allows for individual variation (since each Target Language will present specific challenges).

4. Devise standards for implementation of international conventions, such as ISO (International Organization for Standardization) language codes, Leipzig glossing rules, and the norms developed for the CLARIN European Research Infrastructure for Language Resources and Technology.

5. Ensure that Tools are not commercialized, but instead remain freely available and open-source in the public domain, and that both the research and the Tools themselves achieve the goals of FAIR (findable, accessible, interoperable, reusable) standards.

1.2.2 OBJECTIVES

1.2.2.1 Research Coordination Objectives

Figure 1 highlights the need to coordinate efforts a) across Target Languages, and b) across groups of Research Questions. The STORM Action will pursue the following sets of Research Coordination Objectives (RCOs) to ensure that insights and results from the Action maximize synergy.

Research Coordination Objectives across Target Languages:

RCO1: Overall Research Coordination across Target Languages. In order to maximize efficiency and prevent duplication of efforts, the Action will set up a Tool production pipeline, with coordinated steps for research and development. Coordination will facilitate production for multiple Target and User Languages, ensuring that each achievement is then easily spread to multiple user audiences. In many cases machine translation can speed deployment to additional User Languages, by producing first drafts of translations for subsequent manual editing. Coordination across Target Languages will facilitate dissemination both among stakeholders and to the general public.

RCO2: Coordination across Majority and Minority Languages. Most majority languages, particularly those belonging to larger countries, have language technology resources such as corpora and morphological annotation tools that can support Tool development, and coordinating their activities will increase their efficiency. It is essential that the needs of minority languages be addressed and that achievements for majority languages serve to bolster development for minority languages. The Action will prioritize sharing of solutions from majority languages to minority languages and across minority languages.

RCO3: Languages across Families and Regions. Closely related languages share features, facilitating parallel development of Tools thanks to intercomprehension and receptive multilingualism. Some minority languages are related to majority languages, and by coordinating across language families, the Action will maximize the ways in which work on majority languages benefits minority languages as well. Geographic contiguity also plays an important role, since it is particularly important

for countries to communicate well with their neighbors, but neighboring languages are often missing from educational curricula.

Research Coordination Objectives across groups of Research Questions:

RCO4: Coordination across Users and Technology. Coordination is essential to both development and evaluation of the efficacy of Tools. Gamification can aid development and engage L2 learners. Technological enhancements to Tools can collect and analyze back-end click-stream data to better understand user behavior. Technological and design accommodations will be implemented to serve special needs learners such as older learners and the visually impaired.

RCO5: Coordination across Linguistic facts and Technology. Each Target Language will present a unique set of technological challenges imposed by its linguistic structure, but from a broader perspective some of those challenges will be similar. For example, in some Slavic languages the reflexive marker on verbs is a clitic particle that "floats" toward the beginning of a sentence, with the result that it may be far away from the rest of the verb. Some Germanic languages have separable verbal prefixes that "float" in the opposite direction, toward the end of a sentence. And it is not uncommon for verb forms in many languages to have an auxiliary that can be separated from the rest of the verb form by several other words. While at first glance these three situations appear quite different, they present a similar problem for natural language processing, namely: how to identify non-contiguous word forms. Coordination will make it possible to find common ground among disparate challenges and share technical solutions.

1.2.2.2 Capacity-building Objectives

Capacity-building Objectives (COs) maximize the reach and efficacy of the STORM Action.

CO1: Inclusiveness. The Action creates a network of researchers with expertise in a majority of the languages found in COST Member Countries, Near Neighbor Countries, and International Partner Countries for the building and researching of Tools. Because the distribution of languages with rich inflectional morphology in Europe happens to align well with the list of COST Inclusiveness Target Countries (Table 2), the Action will emphasize those countries and their languages. The Action explicitly includes minority languages, which are typically under-resourced for language technology. The Action will remain open to admit additional researchers.

CO2: Training Schools and STSMs for Early Career Investigators (ECIs). The Action will support Training Schools and Short-Term Scientific Missions for ECIs to build their professional networks and promote collaboration.

CO3: International Dissemination and Collaboration. Proposers of the Action will participate in international conferences in disciplines relevant to the Action where relevant researchers and practitioners who are stakeholders convene (e.g., International Cognitive Linguistics Conference, Conference on Natural Language Learning, European Second Language Association, American Association of Applied Linguistics) and collaborate with existing organizations (e.g., CLARIN: Common Language Resources and Technology Infrastructure, and the European Language Grid), new projects and initiatives, and existing online learning platforms. In addition to university curricula, the Action will use a website, social media, webinars, and podcasts to disseminate to users, and proposers will present project results to local public and media.

2 NETWORKING EXCELLENCE

2.1 ADDED VALUE OF NETWORKING IN S&T EXCELLENCE

2.1.1 ADDED VALUE IN RELATION TO EXISTING EFFORTS AT EUROPEAN AND/OR INTERNATIONAL LEVEL

Technological advances in second and foreign language pedagogy are currently not adequately shared across Target Languages, countries, and educational institutions. Second and foreign language teaching programs tend to be siloed, instead of accessing and building on each other's achievements.

The STORM Action specifically aims to maximize added value by facilitating the sharing of solutions in developing Tools for multiple Target Languages.

No other COST Action previously or currently underway focuses on the use of linguistic data science and technology to enhance learning of languages with rich inflectional morphology. CA18209 “European network for Web-centred linguistic data science” (NexusLinguarum) strives to construct a system of multilingual linguistic data and support language resources such as corpora, dictionaries and natural language processing. NexusLinguarum strengthens the infrastructure upon which Tools can be built. These two Actions are thus compatible and complementary. NexusLinguarum is primarily about digital links across texts and machines, whereas the STORM Action is about language learning to facilitate human-to-human communication. Past COST Actions that have developed language technology resources, such as the European Network of e-Lexicography (ENeL) have likewise contributed achievements that make the tasks of the STORM Action more feasible. In a similar vein, EU-funded projects on Language Technologies [19] have focused mainly on big data and machine translation rather than on resources that strategically improve L2 learning outcomes. These projects have helped to create the environment in which it is possible to envision the STORM Action.

A COST Action that began in 2013, IS1306 “New Speakers in a Multilingual Europe: Opportunities and Challenges” addressed some of the stakeholders relevant to the STORM Action, namely “new speakers” of minority languages, immigrants, and transnational workers, focusing particularly on the social challenges of adapting to new linguistic spaces. By contrast, the STORM Action proposes to create Tools to help all of these stakeholders and others achieve language proficiency.

2.2 ADDED VALUE OF NETWORKING IN IMPACT

2.2.1 SECURING THE CRITICAL MASS AND EXPERTISE

NB: Figures in this section reflect the Action’s network as of 23.09.2020.

The STORM Action will create an open pan-European network to promote research on the acquisition of languages with rich inflectional morphology and develop Tools to enhance such acquisition. At the time of submission, 62 proposers from 21 countries represent expertise in 80% of European Target Languages with inflectional morphology marked as “High” or “Medium” in Table 2. Proposers cover the wide spectrum of fields addressed by the Research Questions, among them theoretical linguistics, morphology, construction grammar, language acquisition, second and foreign language pedagogy, computational linguistics, lexicography, language typology, and computer programming. While 80.6% of proposers list their primary expertise as “Language and Linguistics”, many of them also have significant expertise in computer science and technological applications, and 8% of the proposers have primary expertise in computer and information science. The Action remains open to include other researchers who may choose to join in the future.

The Action’s network has a significant number of Early Career Investigators (32%), and 48% of the proposers are female.

The majority of countries represented in this proposal are Cost Inclusiveness Target Countries (63%), with the aim of including all of these countries.

The Action will use Workshops, Training Sessions, meetings, and conference participation to consolidate the network and integrate all proposers into the research and technological development of Tools.

2.2.2 INVOLVEMENT OF STAKEHOLDERS

The spectrum of stakeholders includes language teachers, L2 learners and others. Online webinar workshops with unlimited enrolment will be held to recruit all types of stakeholders.

Language teachers include:

- Instructors at schools, high schools, colleges, universities, and private language schools where languages with rich inflectional morphology are taught, in particular instructors in charge of developing syllabi, degree programs, and courses;
- Independent instructors and speech therapists;

- Developers of MOOCs (Massive Open Online Courses) and other online courses for Target Languages, developers of language tests, language teaching materials, textbooks;
- Online platforms that support second language acquisition and the organizations that create and maintain them.

Most of the proposers represent this group of stakeholders. Given the fact that the languages of COST Inclusiveness Target countries constitute a majority of European languages with High or Medium levels of inflectional morphology (Table 2), it is natural that the majority of stakeholders represent these countries, and that proposers from these countries will have leading roles in the STORM Action.

L2 learners include:

- Pupils and students at schools, high schools, colleges, universities, and private language schools where languages with rich inflectional morphology are taught;
- Independent learners and other non-traditional learners among the general public;
- Distance learners;
- Heritage speakers who have acquired a language in the home;
- Asylum seekers and other vulnerable groups of L2 learners.

Stakeholders representing L2 learners will be recruited through the network of proposers, website, social media, webinars, and podcasts. Learners will be directly integrated into the development of Tools by using gamified crowdsourcing to create and edit content for Tools, in collaboration with the Action's network and other stakeholders. The fact that the Tool is easy to access and use and free of charge makes it an attractive option for asylum seekers and other vulnerable L2 learners.

Other stakeholders include:

- Businesses and organizations that depend on the existence of proficient L2 speakers of languages with rich inflectional morphology;
- Policy-makers;
- Other European projects and initiatives.

These stakeholders will be able to integrate Tools into their operations and assist with their evaluation. The Action will collaborate with the European Commission for Education and Training.

The institutions of higher education that the Action's proposers are affiliated with, along with their own professional networks will inform local communities and the public about the Action, and L2 learners will be directly involved in the development of Tools. A website, social media, webinars, and podcasts will be used to further engage stakeholders in the Action. Workshops will educate language instructors on how to implement Tools in L2 teaching.

2.2.3 MUTUAL BENEFITS OF THE INVOLVEMENT OF SECONDARY PROPOSERS FROM NEAR NEIGHBOUR OR INTERNATIONAL PARTNER COUNTRIES OR INTERNATIONAL ORGANISATIONS

The STORM Action includes proposers from Russia and the USA. Russian is a language with a high degree of rich inflectional morphology, and is one of the largest languages in Europe's immediate neighborhood. Russian is a Slavic language, and solutions developed for Russian will be portable to over a dozen other genetically related Slavic languages spoken in Europe. Russian is well resourced with regard to language technology. The Action will benefit from Russian institutional and personal expertise in the fields of corpus linguistics, e-learning, computational technologies, data curation, and cross-language frequency studies. Collaboration with Russian and US colleagues will expand the scope of experimentation to show the efficacy of Tools and facilitate dissemination beyond Europe. Given ongoing political tensions, it is to the advantage of all European countries to nurture domestic expertise in Russian, motivating the building of Tools for Russian as a Target Language paired with a multitude of European User Languages.

Europe recruits large numbers of highly skilled professionals and international students from the USA and Russia, and Russia likewise recruits similar talent from Europe and other countries, creating a huge demand in terms of L2 learners of Russian. Technological solutions such as the Action's Tools are particularly advantageous in this situation.

3 IMPACT

3.1 IMPACT TO SCIENCE, SOCIETY AND COMPETITIVENESS, AND POTENTIAL FOR INNOVATION/BREAK-THROUGHS

3.1.1 SCIENTIFIC, TECHNOLOGICAL, AND/OR SOCIOECONOMIC IMPACTS (INCLUDING POTENTIAL INNOVATIONS AND/OR BREAKTHROUGHS)

The Tools created by the STORM Action are a specific, measurable, tangible, and innovative product. International coordination in the Action will scale up the creation of Tools with guidelines and standards that are portable across varieties of languages and users. The existing proof-of-concept Tool proves that building Tools is an achievable objective. Tools are relevant to the linguistic needs of Europe and timely given the recent advent of big language data. The needs of Europe for multilingual workers due to migration and economic expansion also contribute to the timeliness of the Action.

Impact to Science: Second and Foreign Language Acquisition
The STORM Action will push forward the frontiers of knowledge in L2 acquisition, with particular emphasis on morphology, the integration of L2 learner data into research, and the solution of corpus challenges.

Morphology: The STORM Action will advance linguistic understanding of how word forms are learned, comprehended, and produced by both L1 (native) speakers and L2 learners, yielding a more realistic and applicable cognitive model of inflectional morphology. Some scholars (e.g., [15]) claim that one form is the most informative or distinctive for predicting other forms of a word and there have been attempts to measure informativity (a.k.a. “conditional entropy” [16]). Others (e.g., [14]) counter that no word form has privileged status. It is also possible that both informativity and distinctiveness play a role, and their relative importance can be tested. It is theorized that words form neighborhoods (groups of words with similar paradigmatic patterns and phonological shape) that enable extrapolation from partial paradigms, but this needs to be tested out on a larger sample of data. The volume of data generated by the Action on the distributions of word form frequencies and their function in language acquisition will shed crucial light on these controversial issues of morphology. It is largely agreed that (functional) morphology imposes a learning bottleneck for adult L2 learners [1, 2, 3, 11]. Developing and implementing Tools will help us to unpack why this is and find ways to bypass the bottleneck in the acquisition process.

L2 Learner Data: Anonymized back-end data on the behavior of Tool users will track the exposure of L2 learners and provide an empirical foundation for understanding and improving learner outcomes. Experiments will compare the outcomes of Tool users as opposed to L2 learners in environments that do not implement Tools. Both back-end data and data from classroom experiments will be compared with data from machine learning experiments (such as [6]) in the acquisition of inflectional morphology. The integration of back-end data with experiments on both human and machine learning will give new insights into second language acquisition. The Action will thus coordinate a variety of measurements of language acquisition and of the effectiveness of Tools.

Corpus Linguistics: While it might appear straightforward to harvest high-frequency words and word forms from corpora, a host of challenges arise, and by addressing these challenges the STORM Action will make important contributions to corpus linguistics. These challenges include: user needs, polysemy, morphological ambiguity, paucity of data, language variety/register, and gender. The order in which words are ideally learned does not directly correspond to corpus frequency [20]. Priorities in the lexicon that a L2 learner needs to acquire are influenced by numerous other factors, including the types of situations in which the learner is likely to use the Target Language and specific facts about the Target Language and its culture. While the lexicon will vary from one Target Language to another due to language- and culture-specific facts, topics that are relevant to L2 learner needs, such as “health”, “time”, “weather”, “politics” will be shared across languages. A method for combining topics with corpus frequency can establish cross-linguistic standards for lexicon across proficiency levels. Many words in all languages have multiple meanings, and such polysemy is particularly prevalent among the high-frequency words that L2 learners must prioritize, however distinguishing between meanings remains a challenge for corpus linguistics. Likewise, morphological ambiguity (homonymy of word forms within and across paradigms) is a ubiquitous challenge. While much progress has been made in automated morphological disambiguation, it pertains mostly to a limited number of languages and challenges persist. The Action can contribute by strategically identifying the most frequent meaning and parse for ambiguous forms along with their most typical contexts. For example, the paradigm of the Russian word

radost' 'joy' contains the form *radosti* in five of its paradigm slots, but only the Genitive Singular is of high frequency, and a typical context follows the preposition *ot* 'from'. For Russian there exists a one million word "gold standard" corpus that has been manually checked to resolve morphological ambiguities. However, word form frequencies from this resource are reliable only for a small set of the most frequent words. Furthermore, most of the world's languages do not have such a resource; for example, neither Latvian nor North Saami have a sufficient gold standard corpus. Corpora tend to overrepresent written genres at the expense of spoken language, and many features of language variety are marginalized or absent, such as slang, jargon, taboo language, dialects, and diglossia (where there are two distinct language varieties, as in modern Czech). The Action will investigate the use of various kinds of corpora, such as CHILDES (https://childes.talkbank.org/access/) to compensate for underrepresentation of spoken language. Languages with grammatical gender show a strong statistical tendency for gender bias. For example, past tense verb forms in Slavic languages express the gender of the subject, and verbs that have human subjects have on average several times as many masculine forms as feminine forms [12, 13]. Some languages have separate masculine vs. feminine nouns to represent persons from various ethnic groups and professions, cf. Czech *politik* 'politician (male)' with a frequency of 18.42 per million words in the Czech National Corpus (korpus.cz) vs. *politička* 'politician (female)' with a frequency of 1.58 per million words. Selecting words and word forms according to corpus frequency alone will underrepresent women, requiring adjustments to correct for gender bias. The Action will seek scientific systematic ways to solve or work around these problems.

Impact to Society:
The Action will optimize L2 learner outcomes by empowering learners to benefit from the scientific insights from e-learning and language corpora. There has been much research on e-learning, but this has not delivered a corresponding impact on language pedagogy. Likewise, the advent of large language corpora and evolution of corpus linguistics have not yet had a major impact on language learning. The Action closes both of these gaps between research and impact. The Tool is a scientifically-motivated technological application that exploits corpus data, and there is a proof-of-concept Tool that actually works and has users. For example, when L2 learners are grappling with mastering a specific grammatical category such as the imperative verb form or the dative case, a Tool can provide instant access to words that are most frequently used precisely in those forms, along with examples of typical grammatical constructions and collocations. Filters in the Tool for grammatical analysis and proficiency level make it possible for users to access maximally strategic input. Frequency-based Tools are expected to help L2 learners develop their language comprehension and production similar to the way native speakers do. The Action thus addresses the United Nations' Sustainable Development Goal #4: Quality Education, by improving the quality of language pedagogy. The Action further contributes to Goal #17: Partnerships, by promoting multilingualism that facilitates cross-border cooperation. Because the Tool is delivered as an interactive webpage, it is compatible with distant and digital learning in situations such as the recent pandemic in which traditional classroom instruction had to be curtailed.

Impact to Competitiveness:
Language skills are essential on the international job market. The European Commission has an ambitious goal for its citizens to learn at least two foreign languages [21], a goal that has proven elusive [22] and usually limited to a narrow set of languages (English, Spanish, French, German). The Tools proposed in the Action will contribute to raising language skills across and beyond Europe. Tools across a wide range of Target Languages will facilitate interpersonal communication in business, professional, academic, public, and private arenas. Tools will maximize inclusiveness by catering to special needs, minority, and heritage language learners [23, 24], so that the European economy can benefit from their participation.

3.2 MEASURES TO MAXIMISE IMPACT

3.2.1 KNOWLEDGE CREATION, TRANSFER OF KNOWLEDGE AND CAREER DEVELOPMENT

Knowledge Creation
The STORM Action will create new scientific knowledge (see 3.1.1) and new technological knowledge about the design and development of strategic second and foreign language learning Tools. Neither kind of knowledge could be created without dedicated collaboration of researchers with expertise spanning a multitude of languages, various subdisciplines of linguistics, and information technology.

Transfer of Knowledge

Primary venues for transfer of the Action's knowledge are through STSMs, Training Schools, and Workshops. Knowledge transfer will facilitate scaling up the building of Tools to cover a wide range of Target and User Language combinations both during and beyond the period of the Action.

Career Development
ECIs have already played a formative role in the drafting of the Technical Annex for the Action, by proposing both Target Languages and Research Questions (especially RQ2, RQ3, RQ5, RQ15). Each Working Group will have a Leader and a Co-Leader. Working Groups 1 and 3 will have ECIs as Leaders. Working Groups 2, 4, and 5 will have ECIs as Co-Leaders. At least 40% of the Management Committee will be comprised of ECIs. The Action will give ECIs top priority for Short Term Scientific Missions and showcase the achievements of ECIs in the Action's Workshops, Training Schools, and other dissemination activities.

3.2.2 PLAN FOR DISSEMINATION AND/OR EXPLOITATION AND DIALOGUE WITH THE GENERAL PUBLIC OR POLICY

Dissemination to Academic Community. The Action will disseminate both Tools and research findings to the scholarly community. Proposers will present findings at prestigious conferences and in scholarly journals devoted to language pedagogy, corpus linguistics, theoretical linguistics, and language technology. Conferences may include: Applied Linguistics and Foreign Language Teaching Conference; Linguistics, Language Learning and Teaching Conference; Language and Applied Corpus Linguistics Conference; International Cognitive Linguistics Conference; Conference on Computational Natural Language Learning; Teaching and Language Corpora Conference. Open-access top-ranked journals accessible to stakeholders will be prioritized and may include: *Language Teaching Research*, *International Journal of Corpus Linguistics*, *Corpus Linguistics and Linguistic Theory*, *Studies in Language*, *Natural Language Engineering*, *Language Learning*, *Studies in Second Language Acquisition*, *Applied Psycholinguistics*, *Language*, *Morphology*. ECIs will particularly be encouraged to participate and invited to co-publish with more established proposers. The Action's three Training Schools will be announced through professional networks and on international platforms, such as the World Academy of Science, Engineering and Technology (waset.org).

Dissemination to Second and Foreign Language Instructors, L2 Learners, Businesses, Policy Makers, and the Public. The Action will host four Workshops open to all stakeholders. The Workshops will disseminate Tools as well as the technology to develop new Tools. The Workshops will foster bi-directional dialog in which stakeholders participate in crowdsourcing and evaluation of Tools. The Action's dedicated website, social media, webinars, and podcasts will further disseminate results.

4 IMPLEMENTATION

4.1 COHERENCE AND EFFECTIVENESS OF THE WORK PLAN

4.1.1 DESCRIPTION OF WORKING GROUPS, TASKS AND ACTIVITIES

Four Working Groups (WG1-WG4) are directly linked to the Research Questions presented in 1.1.2. A fifth Working Group (WG5) is devoted to administration and dissemination. Membership will overlap across Working Groups.

WG1: Users. This Working Group will direct the building of Tools. This is the largest Working Group in the Action, and will coordinate the efforts of subgroups, each of which will develop one or more Tools.

Task 1.1: Language selection. This task will develop a stratified list of prioritized Target Languages. The first generation of Tools will include at minimum the following Target Languages: one Slavic language, one Uralic language, one language that is either Romance or Germanic, and one minority language. One User Language will be paired with each Target Language. Subsequent generations will both expand the list of Target Languages and extend first-generation Tools to new User Languages. Within the period of the Action, two generations of Tools will be delivered, so that the production of Tools will scale up and be self-perpetuating beyond the end of the Action.

Task 1.2: Traditional and non-traditional L2 learners. "Traditional learners" applies to L2 learners who receive instruction, most often in schools and universities. Most other learners who acquire a second or foreign language online or through self-study are "non-traditional learners". This task will

devise a menu of exercises to implement Tools in the classroom and in independent study and propose gamification enhancements to boost user interaction for both groups of learners. This task will also conduct classroom and other experiments as well as analysis of back-end data to evaluate and fine tune Tools.

Task 1.3: Heritage learners and related languages. Heritage speakers have grown up exposed to a minority language (either immigrant or indigenous) in the home [23, 24], and have a head start toward acquiring the standard variety of that language. Similarly, L2 learners acquiring a language closely related to their native tongue or another language they know will recognize much of the vocabulary and grammatical structure of their new L2. This task will design and implement Tools that capitalize on what these learners already know so that their language acquisition proceeds at an advanced pace. Experiments and back-end data analysis will be conducted.

Task 1.4: Special needs learners. Learners with aphasia, other learning impairments, visual and/or hearing impairment, and older learners will benefit from targeted accommodations in Tool design. This task will propose such accommodations and evaluate their efficacy through experiments and/or back-end data analysis. In the case of aphasia, Tools can aid in acquisition and/or retention of a native language (L1). The needs of asylum seekers and other vulnerable L2 learners will also be addressed.

WG2: Linguistic facts. Various features that are shared across some or all languages will be investigated with relation to their implications for Tool design.

Task 2.1: Typology. Rich inflectional morphology exists along a range of types, characterized as primarily “agglutinative” (e.g., Uralic and Turkic languages) or “fusional” (e.g., Indo-European languages like Slavic, Germanic, and Romance). In the agglutinative languages there is a relatively transparent composition of markers, each of which corresponds to a grammatical category. For example, a case marker is added to a number marker. In some languages this system leads to very large potential paradigms, with hundreds and even thousands of possible word forms for a single word. In fusional languages, the indication of each combination of multiple grammatical categories is fused into a single non-transparent marker. Furthermore, some languages have large paradigms only for their verbs (languages tagged “Medium” in Table 2), whereas others have large paradigms for both verbs and nouns (languages tagged “High” in Table 2). This task will examine these and other typological factors to arrive at guidelines to aid in the development of relevant Tools.

Task 2.2: Lexicon and Gender. This task will set a standard list of topics for selection of lexicon, as well as a means to implement corpus frequency in this selection. This task will devise strategies for representation vs. avoidance of various marginalized lexical items such as slang, jargon, and taboo, taking into account user needs, proficiency levels, and culturally appropriate norms (since, for example, taboo words can be much more unacceptable in one language than in another language). Language variation due to dialect and diglossia will also be considered. Means to compensate for the underrepresentation of feminine gender will be proposed where relevant.

Task 2.3: Constructions and collocations. Discovery procedures will be developed to arrive at best practices for determining what grammatical constructions and collocations are most typical for given word forms. This information is needed in order to make the examples of usage presented in Tools maximally strategic. For most larger languages, existing open-source corpus resources make it possible to search for collocates to the left or right of a given word form, providing some empirical basis for this work. However, best practices need to be carefully designed, along with workarounds for languages lacking sufficient corpus resources.

WG3: Technology. This Working Group will secure empirical data on word form frequency and design the software architecture for Tools.

Task 3.1: Corpus linguistics. This task will address morphological and semantic disambiguation specifically needed for Tool development, maximizing use of existing open-source corpus resources, and devising workarounds for under-resourced languages. The underrepresentation of spoken language in most corpora will be addressed, yielding proposals to correct for this imbalance.

Task 3.2: Computational technology. Computational enhancements for Tools are the focus of this task, among them gamification, integration of text-to-speech (using existing resources) so that L2 learners can listen to examples, and crowdsourcing (especially in development of Tools). This task will collaborate with tasks 1.2 and 1.4.

Task 3.3: Programming and pipeline design. A step-by-step pipeline for building Tools will be created. Generic open-source code for the building of Tools will be extended and refined, leaving room for necessary language-specific adjustments. Guidelines for how to implement the code will be drafted and shared using an open platform (Github).

WG4: Theories. Theoretical assumptions about morphology and how it is acquired will be tested with respect to Tools.

Task 4.1: Modelling morphology. Detailed frequency distribution data will be used to confront existing theories and to suggest a realistic cognitive model of inflectional morphology.

Task 4.2: Acquisition. The acquisition patterns for rich inflectional morphology will be compared across L1 and L2, the latter with and without Tools.

Task 4.3: Informativity and distinctiveness. This task will devise and undertake experiments and studies to assess the relative import of informativity and distinctiveness (see 3.1.1 Impact to Science).

WG5: Administration and dissemination. This Working Group will ensure that WGs 1-4 stay on task to produce their deliverables, that events (Training Schools, Workshops, etc.) are organized, and will maximize dissemination of results across all relevant audiences.

Task 5.1: Administration. The overall administration and coordination of the Action, including organizing Management Committee (MC) meetings, will be the remit of this task.

Task 5.2: Collaboration across Working Groups. This task will ensure that the Working Groups and their tasks are smoothly integrated into the function of the Action.

Task 5.3: Transfer of knowledge and career development. This task will organize and coordinate Training Schools and STSMs to disseminate the achievements of the Working Groups and will maximize the visibility of ECIs in leadership roles in the Action. Workshops held at universities and schools and webinars will directly involve stakeholders in the Action.

Task 5.4: Dissemination. This task will manage a website publicizing the progress of the Action, social media streams, podcasts, and press releases. This task will also coordinate the dissemination of results to the scholarly community through conferences and publications.

4.1.2 DESCRIPTION OF DELIVERABLES AND TIMEFRAME

Deliverables are allocated across the 48-month period (see Gantt chart in Figure 2) according to the remit of each Working Group. Training Schools and Workshops showcase the deliverables achieved in the months that precede each of those events.

Deliverables	**Month**
WG1: Users	
Prioritized list of Target and User Languages	3
First generation of Tools	24
Second generation of Tools	48
Exercises for implementation of Tools in classroom and self-study	9
Guidelines for heritage and related language learners	9
Guidelines for special needs users	9
Classroom experiments	33
Back-end data analysis	39
WG2: Linguistic facts	
Typological adjustments	6
Lexicon guidelines	9
Guidelines to correct for corpus gender imbalance	9
Discovery procedures for grammatical constructions and collocations	15
WG3: Technology	
Frequency distribution of word forms for first generation Tools, including morphological and semantic disambiguation	6
Frequency distribution of word forms for second generation Tools, including morphological and semantic disambiguation	27

Corrective measures to increase representation of spoken language	9
Workarounds for under-resourced languages	9
Gamification, crowdsourcing, and text-to-speech implementation	21
Pipeline with open-source code and implementation guidelines	21
WG4: Theories	
Confrontation of theories with data and proposal of realistic cognitive model	21
Comparison of L1 vs. L2 acquisition of rich morphology with and without Tools	30
Experiments and studies to assess relative import of informativity and distinctiveness	33
WG5: Administration and dissemination	
Website, webinars, podcasts, social media streams (maintained through end of Action)	throughout
Organize meetings: Kick-Off (KO), Management Committee (MC), Working Groups (WG)	see Figure 2
Training School 1: Data and technology for Tools. Optimizing use of data, accommodations for special needs users	15
Training School 2: Tool pipeline and enhancements. Optimizing the pipeline, integration of gamification and crowdsourcing	27
Training School 3: Experiments and analysis of back-end data	39
Workshop 1: Integrating Tools into classrooms and self-study	12
Workshop 2: Launch of first generation Tools	24
Workshop 3: How to build your own Tool	36
Workshop 4: Launch of second generation Tools	48
Scientific presentations/publications on linguistic facts in relation to Tools	24
Scientific presentations/publications on technological implementation of Tools	30
Scientific presentations/publications on pedagogical implementations of Tools	48
Scientific presentations/publications on theoretical issues	48

Table 3: Schedule of deliverables for each Working Group

4.1.3 RISK ANALYSIS AND CONTINGENCY PLANS

Table 4 lists risks associated with the STORM Action, along with mitigating factors and measures.

Working Group	Risk	Level	Mitigating factors and measures
WG1	User data privacy	Low	Well-established procedures and ethical guidelines for anonymizing data and protecting the privacy of users will be followed.
WG1	Recruitment of participants for experiments	Mid	The need to recruit participants, particularly for classroom experiments, will be consistently broadcast on the website, webinars, Workshops, Training Schools, etc.
WG2, WG3	Restricted access to corpus data and data sparsity	Mid	There are freely available corpora for most majority languages of Europe. The nature of the research does not require extraction of more than 1 sentence at a time, thus not conflicting with copyright. Workarounds will have to be devised for some smaller and minority languages.
WG3	Varying technological expertise across proposers	Mid	All proposers have at least minimal knowledge of corpus linguistics, most have advanced expertise in one or more relevant technical fields. All Working Groups will include some members with advanced computational expertise and the Action will remain open to admit additional collaborators.
WG3	Interoperability of computer systems and access to sufficient tools	Low	Only open-source code and software will be used. The proof-of-concept Tool proves that this Action requires only computers of the type that most European researchers already have.
WG3	Code, data, and software maintenance	Mid	While the use of open-source code reduces this risk, maintenance is always a technical challenge. A sufficient number of proposers with the necessary expertise has been recruited, yet the numbers leave

			little room for attrition. The Action will recruit more technical expertise.
WG4	It is unknown whether a realistic cognitive model can be achieved	High	The results of theoretical inquiries can never be predicted in advance. The Action will provide more empirical grounds for such inquiries.
WG5	Low productivity or quality of results by proposers	Low	Established methods for project planning and control will be used. WG chairs will be responsible for timely and accurate communication of internal procedures and will monitor quality of deliverables. All deliverables will undergo prior internal peer-review.
WG5	Geographical dispersion and travel restrictions	High	The pandemic has occasioned widespread use of teleconferencing; most researchers are now adept and can use these means to communicate in groups and one-on-one.
WG5	Strength of infrastructure and internet access across stakeholders	Mid	Content will be delivered in multiple formats that can be accessed asynchronously.
WG5	Critical mass of stakeholders needs to be attracted	Low	The number and diversity of the initial network of proposers reduces this risk and the Action will remain open to include additional collaborators. Professional networks will be exploited to announce workshops and attract stakeholders.

Table 4: Risks, mitigating factors and measures.

4.1.4 GANTT DIAGRAM

	Year 1				Year 2				Year 3				Year 4			
Months	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48
Activities																
Meetings	KO	MC	WG			MC	WG			MC	WG			MC	WG	
Tool Workshops				1				2				3				4
Training Schools					1				2				3			
STSMs																
Working Groups																
WG1: Users																
WG2: Linguistic facts																
WG3: Technology																
WG4: Theories																
WG5: Admin & diss																
Tools																
First generation																
Second generation																

Figure 2: Gantt Diagram (KO = Kick-Off, MC = Management Committee, WG = Working Groups)

References

[1] Slabakova, Roumyana. 2009. How is inflectional morphology learned? *Eurosla Yearbook*, *9*, 56-75. DOI: 10.1075/eurosla.9.05sla

[2] Jensen, I. N., Slabakova, R., Westergaard, M., & Lundquist, B. 2019. The Bottleneck Hypothesis in L2 acquisition: L1 Norwegian learners' knowledge of syntax and morphology in L2 English. *Second Language Research*. DOI: 10.1177/0267658318825067

[3] Slabakova, Roumyana. 2014. The bottleneck of second language acquisition. *Foreign Language Teaching and Research*, 46 (4), 543-559.

[4] Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort*. Reading, MA: Addison-Wesley.

[5] Schöne, Karin. 2015. *Zkoumání hierarchizace pádů českého substantiva v sémantických (kolokačních) třídách.* PhD dissertation, Charles University. https://is.cuni.cz/webapps/zzp/detail/104275

[6] Janda, Laura A. & Francis M. Tyers. 2018. Less is More: Why All Paradigms are Defective, and Why that is a Good Thing. *Corpus Linguistics and Linguistic Theory* 14(2), 33pp. doi.org/10.1515/cllt-2018-0031.

[7] Karlsson, Fred 1986. Frequency Considerations in Morphology. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*. Berlin 39/1, pp. 19-28.

[8] Miestamo, M. 2008. Grammatical complexity in a cross-linguistic perspective. In M. Miestamo, K. Sinnemäki, and F. Karlsson (eds.), *Language complexity: Typology, contact, change*, 23-41. Amsterdam: John Benjamins.

[9] Ackerman, Farrell, James P Blevins & Robert Malouf. 2009. "Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter." In James P. Blevins and Juliette Blevins (eds.), *Analogy in Grammar: Form and Acquisition*, 54–82. Oxford: Oxford University Press.

[10] Sinclair, John McH. (ed.) 2004. How to use corpora in language teaching (= *Studies in Corpus Linguistics*, 12).

[11] Hopp, Holger. 2010. Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua* 120, 901-931.

[12] Kuznetsova, Julia. 2015. Linguistic Profiles: Going from Form to Meaning via Statistics. Berlin: De Gruyter.

[13] Elmerot, Irene. 2017. *These women's verbs. A combined corpus and discourse analysis on reporting verbs about women and men in Czech media 1989–2015.* MA thesis, Stockholm University. http://www.diva-portal.org/smash/get/diva2:1161582/FULLTEXT01.pdf.

[14] Bochner, Harry. 1993. *Simplicity in Generative Morphology.* Mouton de Gruyter.

[15] Albright, Adam. 2002. *The Identification of Bases in Morphological Paradigms.* PhD Dissertation,

Funded by the Horizon 2020 Framework Programme of the European Union

University of California, Los Angeles.

[16] Ackerman, F. and R. Malouf. 2016. Implicative relations in word-based morphological systems. In Hippisley A, Stump G (eds.), *Cambridge Handbook of Morphology*, Cambridge: Cambridge University Press, 272– 296.

[17] UNESCO Atlas of the World's Languages in Danger. http://www.unesco.org/languages-atlas/index.php Accessed 03.09.2020.

[18] Walker, Shaun. 2019. 'This is the golden age': Eastern Europe's extraordinary 30-year revival. *The Guardian* 26.10.2019. https://www.theguardian.com/world/2019/oct/26/this-is-the-golden-age-eastern-europes-extraordinary-30-year-revival Accessed 03.09.2020.

[19] European Commission. Shaping Europe's digital future. EU-funded projects on Language Technologies. https://ec.europa.eu/digital-single-market/en/programme-and-projects/eu-funded-projects-language-technologies Accessed 03.09.2020.

[20] Comer, William. 2019. Measured words: Quantifying vocabulary exposure in beginning Russian. *Slavic and East European Journal* 63(1), 92–114.

[21] European Commission. Education and Training. Multilingualism Policy. https://ec.europa.eu/education/policies/multilingualism/about-multilingualism-policy_en Accessed 03.09.2020.

[22] European Commission. Education and Training. Council Recommendation on a comprehensive approach to the teaching and learning of languages. https://ec.europa.eu/education/education-in-the-eu/council-recommendation-improving-teaching-and-learning-languages_en Accessed 03.09.2020.

[23] Montrul, Silvina. 2015. *The acquisition of heritage languages*. Cambridge: Cambridge University Press.

[24] Polinsky, Maria. 2018. *Heritage languages and their speakers*. Cambridge: Cambridge University Press.

COST Mission and Policies

Strategic Targeting Of Rich Morphology (STORM): Tools for language learning

Relevance to COST Mission and Policy

This document addresses eight points in the COST Mission and Policy statement, followed by the provisions in the STORM Action that target each point.

Point 1: The STORM Action will deliver an innovative technological product, a “Tool” for language learning that strengthens European research and innovation capacities by promoting human-to-human multilingual communication. This product will be provided as a free and open-source resource in the public domain.

Provisions:

- Working Group 1 and its Deliverables, especially:
    - Prioritized list of Target and User Languages for Tools
    - First generation of Tools
    - Second generation of Tools
- Working Group 3 and its Deliverables, especially:
    - Pipeline with open-source code and implementation guidelines

Point 2: The STORM Action will integrate all stakeholders, including second and foreign language instructors, language learners, enterprises that develop online materials for language learning, businesses and organizations that depend on proficient L2 speakers, policy-makers, other European projects and initiatives.

Provisions:

- Working Group 1 and its Deliverables, especially:
    - Exercises for implementation of Tools in classroom and self-study
    - Guidelines for heritage and related language learners
    - Guidelines for special-needs users
    - Classroom experiments
    - Back-end data analysis
- Working Group 3 and its Deliverables, especially:
    - Gamification, crowdsourcing, and text-to-speech implementation

Point 3: The STORM Action will disseminate research results to the scientific community, to stakeholders, and to the general public.

Provisions:

- Working Group 5 and its Deliverables, especially:
    - Website, webinars, podcasts, and social media streams
    - Training School 1: Data and technology for Tools. Optimizing use of data, accommodations for special-needs users
    - Training School 2: Tool pipeline and enhancements. Optimizing the pipeline, integration of gamification and crowdsourcing
    - Training School 3: Experiments and analysis of back-end data
    - Workshop 1: Integrating Tools into classrooms and self-study
    - Workshop 2: Launch of first generation Tools
    - Workshop 3: How to build your own Tool
    - Workshop 4: Launch of second generation Tools
    - Scientific presentations/publications on pedagogical implementations of Tools
    - Scientific presentations/publications on linguistic facts in relation to Tools
    - Scientific presentations/publications on technological implementation of Tools
    - Scientific presentations/publications on theoretical issues

Point 4: The STORM Action will remain open to include all researchers, encouraging participation across geographic location, age groups, and gender. The Action increases the visibility of researchers and promotes acquisition of leadership skills regardless of location, age or gender.

Provisions:

- Because the Action focuses on languages with rich inflectional morphology and in Europe these languages are primarily found in COST Inclusiveness Target Countries (see Table 2 in section 1.1.1), it is natural that researchers from these countries are prioritized in recruitment and will be given leadership roles in the Action.
- Early Career Invstigators (ECIs) have played a formative role in the drafting of the proposal and are thus poised to take on leadership roles. All Working Groups will have an ECI as either Leader or Co-Leader and at least 40% of the Management Commitee will be ECIs. The Action will give ECIs top priority for Short Term Scientific Missions and showcase the achievements of ECIs in the Action's dissemination activities.
- ECIs will particularly be encouraged to participate in scholarly dissemination at conferences and invited to co-publish with more established proposers.

Point 5: The STORM Action's product, the Tool, is designed to be portable across a range of Target and User Languages both in Europe and globally. International collaboration on shared guidelines and technological solutions promotes efficiency and effectiveness.

Provisions:

- Working Group 2 and its Deliverables, especially:
    - Typological adjustments
    - Lexicon guidelines
    - Guidelines to correct for corpus gender imbalance
    - Discovery procedures for grammatical constructions and collocations
- Working Group 3 and its Deliverables, especially:
    - Corrective measures to increase representation of spoken language
    - Workarounds for under-resourced languages
    - Pipeline with open-source code and implementation guidelines

Point 6: The STORM Action counterbalances research communities' unequal access to knowledge, infrastructures, funding and resources, particularly with respect to language technology for small and minority languages.

Provisions:

- Small and minority languages tend to be underresourced in terms of language technology. These languages also tend to be "High" on the scale of rich inflectional morphology (see Table 2 in Section 1.1.1), and are thus specifically targeted in the Action.
- The first generation of Tools will include one with a minority language as its Target Language.

Point 7: The STORM Action contributes to triggering structural changes in the research agendas of COST members by making contributions to the scientific understanding of inflectional morphology and its acquisition.

Provisions:

- Working Group 4 and its Deliverables, especially:
    - Confrontation of theories with data and proposal of realistic cognitive model
    - Comparison of L1 vs. L2 acquisition of rich morphology with and without Tools
    - Experiments and studies to assess relative import of informativity and distinctiveness

Point 8: The STORM Action supports the participation of both Near Neighbor and Non-COST Countries.

Provisions:

- The Action includes proposers from Russia as a Near Neighbor Country and from the USA as a Non-COST Country.

Network of Proposers - Features

COST Inclusiveness target countries
63.16 %

Number of Proposers
62

Geographic Distribution of Proposers

Country	ITC/ non ITC/ other	Number of institutions from that country	Number of researchers from that country	Percentage of the proposing network
Bulgaria	ITC	2	2	3.23 %
Croatia	ITC	3	3	4.84 %
Czech Republic	ITC	5	5	8.06 %
Estonia	ITC	2	2	3.23 %
Finland	non ITC	6	6	9.68 %
France	non ITC	1	1	1.61 %
Hungary	ITC	1	1	1.61 %
Italy	non ITC	3	3	4.84 %
Latvia	ITC	1	1	1.61 %
Lithuania	ITC	1	1	1.61 %
North Macedonia	ITC	2	2	3.23 %
Norway	non ITC	9	12	19.35 %
Poland	ITC	5	5	8.06 %
Russian Federation	other	4	4	6.45 %
Slovakia	ITC	2	2	3.23 %
Slovenia	ITC	3	3	4.84 %
Spain	non ITC	1	1	1.61 %
Switzerland	non ITC	1	1	1.61 %
Turkey	ITC	2	2	3.23 %
United Kingdom	non ITC	2	2	3.23 %
United States	other	3	3	4.84 %

Gender Distribution of Proposers
51.6% Males
48.4% Females

Average Number of years elapsed since PhD graduation of Proposers with a doctoral degree
11.6

Number of Early Career Investigators
20

Core Expertise of Proposers: Distribution by Sub-Field of Science
80.6% Languages and literature

8.1% Computer and Information Sciences
3.2% Other humanities
1.6% Chemical sciences
1.6% Educational sciences
4.8% Other

Institutional distribution of Network of Proposers

98.4% Higher Education & Associated Organisations
1.6% Business enterprise

Higher Education & Associated Organisations:61

- Number by Field of Science of Department/Faculty of Affiliation
  Languages and literature:47
  Computer and Information Sciences:6
  Educational sciences:2
  Other humanities:1
  Other social sciences:2
  Interdisciplinary:2
  Electrical engineering, electronic engineering, Information engineering:1
- Number by Type
  Education Oriented:41
  Research Oriented:20
- Number by Ownership
  Fully or mostly public:57
  Fully or mostly private:3
  50-50 Public and Private:1

Business enterprise:1

- Number by Market sector of unit of affiliation
  Information And Communication:1
- Number by Type
  Private enterprises:1
- Number by Ownership and International Status
  Independent Enterprise:1
- Number by Size
  SME (EU Definition provided underneath after selection):1

COST Country(19) : Bulgaria , Croatia , Czech Republic , Estonia , Finland , France , Hungary , Italy , Latvia , Lithuania , North Macedonia , Norway , Poland , Slovakia , Slovenia , Spain , Switzerland , Turkey , United Kingdom
International Partner Country(1) : United States
Near Neighbour Country(1)
European Commission or EU Agency(0)
European RTD Organisation(0)
International Organisation(0)

Network of Proposers - Details

Main Proposer's Details

Title:	Prof		
First Name:	Laura A.	**Gender:**	F
Last Name:	Janda	**Years from PhD:**	36
Institution:	UiT The Arctic University of Norway	**Type of Institution:**	COST Country
Sub-field of Science of Department:	Languages and literature	**Core Area of Expertise:**	Languages and literature (Linguistics: formal, cognitive, functional and computational linguistics)

Secondary Proposers' Details

Bulgaria

Dr Svetlana Nedelcheva (Konstantin Preslavsky University of Shumen [Department of English Studies])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: No PhD

Prof Stanimir Zhelezov (Konstantin Preslavsky University of Shumen [FACULTY OF MATHEMATICS AND COMPUTER SCIENCE, DEPARTMENT OF COMPUTER SYSTEMS AND TECHNOLOGIES])

Participating as Secondary Proposer
Core Expertise: Computer and Information Sciences: Cryptology, security, privacy
Gender: M
Years from PhD: 4

Croatia

Dr Sanda Lucija Udier (Faculty of Humanities and Social Sciences [Croaticum - Centre for Croatian as Foreign and Second Language])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning
Gender: F
Years from PhD: 12

Dr Darko Matovac (Faculty of Humanities and Social Sciences, University of Zagreb [Croaticum - Centre for Croatian as a Second and Foreign Language])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning
Gender: M
Years from PhD: 7

Dr Petra Bago (Faculty of Humanities and Social Sciences)

Participating as Secondary Proposer
Core Expertise: Other social sciences: Information Sciences
Gender: F
Years from PhD: 6

Czech Republic

Dr Vaclav Cvrcek (Charles University - Faculty of Arts [Institute of the Czech National Corpus])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Databases, data mining, data curation, computational modelling
Gender: M
Years from PhD: 16

Mr Michal Láznička (Charles University [Department of Linguistics])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: No PhD

Dr Adrian Jan Zasina (Univerzita Karlova [Ústav Českého národního korpusu])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 1

Dr Jiří Januška (Charles University [Department of Central European Studies])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: typological, historical and comparative linguistics
Gender: M
Years from PhD: 3

Dr Jaroslava Hlaváčová (Charles University - Faculty of Mathematics and Physics [Institute of Formal and Applied Linguistics])
Participating as Secondary Proposer
Core Expertise: Computer and Information Sciences: Artificial intelligence, intelligent systems, multi agent systems
Gender: F
Years from PhD: No PhD

Estonia

Prof Ilona Tragel (University of Tartu [Department of General Linguistics])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 17

Dr Heiki-Jaan Kaalep (University of Tartu - Institute of Computer Science)
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 21

Finland

Dr Ilmari Ivaska (University of Turku [Scool of Languages and Translation Studies])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 5

Prof Marja-Liisa Helasvuo (University of Turku)
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 23

Dr Mikhail Kopotev (U of Helsinki - Mikhail Kopotev [Dept of languages])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 11

Mr Aleksandr Klimov (University of Helsinki [Faculty of Arts])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 0

Prof Johanna Viimaranta (University of Helsinki [Department of languages])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 14

Dr Helen Plado (University of Helsinki [Department of Finno-Ugrian and Scandinavian Studies])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 7

France

Dr Rafael Marín (CNRS - CNRS / Université de Lille [UMR 8163])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 19

Hungary

Dr Veronika Vincze (MTA-SZTE Research Group on Artificial Intelligence)
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 8

Italy

Dr Jacopo Saturno (Università di Bergamo [Dipartimento di Lingue])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning
Gender: M
Years from PhD: 3

Dr Marco Magnani (University of Trento [Department of Humanities])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning
Gender: M
Years from PhD: 4

Dr Daniele Artoni (Università di Verona - Department of Foreign Languages and Literatures)
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 5

Latvia

Dr Inga Kaija (Riga Stradiņš University [Language Center])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 2

Lithuania

Ms Inga Daraškienė (Vilniaus universitetas [Filologijos fakultetas])

Participating as Secondary Proposer
Core Expertise: Other humanities: Sociolinguistics
Gender: F
Years from PhD: No PhD

North Macedonia

Prof Eleni Buzarovska (Ss. Cyril and Methodius University in Skopje ["Blaze Koneski" Faculty of Philology])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 23

Dr Liljana Mitkovska (AUE-FON University [Faculty of Humanities])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 15

Norway

Prof Antonio Fábregas (UiT-Norway's Arctic University [HSL-Fakultet])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 15

Prof Ljiljana Saric (University of Oslo [ILOS])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 21

Prof Tore Nesset (UiT The Arctic University of Norway)

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 23

Dr Radovan Bast (UiT The Arctic University of Norway [Department of information technology])

Participating as Secondary Proposer
Core Expertise: Chemical sciences: Theoretical and computational chemistry
Gender: M

Years from PhD: 12

Dr Svetlana Sokolova (UiT The Arctic University of Norway [Department of Language and Culture])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 8

Mr Mikhail Voronov (UiT Arctic University of Norway [Department of Language and Culture])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: typological, historical and comparative linguistics
Gender: M
Years from PhD: No PhD

Prof Trond Trosterud (UiT Noregs arktiske universitet [Institutt for språk og kultur])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 16

Dr Lene Antonsen (UiT The Arctic University of Norway [Department of Language and Culture])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Use of language: form, pragmatics, sociolinguistics, discourse analysis, lexicography, terminology
Gender: F
Years from PhD: 2

Ms Chiara Argese (UiT The Arctic University of Norway [Department of Language and Culture])
Participating as Secondary Proposer
Core Expertise: Computer and Information Sciences: Algorithms, data processing and visualisation in language technology
Gender: F
Years from PhD: No PhD

Dr Gustavo Guajardo (Uit The Arctic University of Norway)
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 3

Prof Jason Rothman (UiT the Arctic University of Norway - UiT the Arctic University of Norway [Language and Culture (ISK)])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: 15

Poland

Prof Przemysław Gębal (Silesian University of Technology [Institute of Education and Communication Research])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning

Gender: M
Years from PhD: 15

Ms Slawomira Kolsut (Politechnika Śląska [Instytut Badań nad Edukacją i Komunikacją (Institute of Education and Communication Research)])
Participating as Secondary Proposer
Core Expertise: Educational sciences: Education: training, pedagogy, didactics
Gender: F
Years from PhD: No PhD

Ms Elena Savina (Uniwersytet Warszawski - Institute of Western and Southern Slavistics)
Participating as Secondary Proposer
Core Expertise: Languages and literature: Use of language: form, pragmatics, sociolinguistics, discourse analysis, lexicography, terminology
Gender: F
Years from PhD: No PhD

Ms Olga Popova (Warsaw University)
Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning
Gender: F
Years from PhD: No PhD

Mr Andrii Baran (Uniwersytet Warszawski - Wydział Polonistyki [Instytut Polonistyki Stosowanej])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning
Gender: M
Years from PhD: No PhD

Russian Federation

Ms Valentina Zhukova (National Research University Higher School of Economics [School of Linguistics])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: No PhD

Prof Olga Lyashevskaya (National Research University Higher School of Economics [School of linguistics])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 21

Ms Olga Eremina (National Research University - Higher School of Economics [School of Linguistics, Faculty of Humanities])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 8

Mr Andres Felipe Cuellar Rojas (National Research University: Higher School of Economics [Faculty of linguistics])
Participating as Secondary Proposer

Core Expertise: Languages and literature: Second language teaching and learning
Gender: M
Years from PhD: No PhD

Slovakia

Mr Vladimir Benko (Slovak Academy of Sciences - L. Stur Institute of Linguistics)
Participating as Secondary Proposer
Core Expertise: Languages and literature: Use of language: form, pragmatics, sociolinguistics, discourse analysis, lexicography, terminology
Gender: M
Years from PhD: No PhD

Dr Martina Ivanova (Prešov University - Faculty of Arts [Department of Slovak Language])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 13

Slovenia

Dr Nikola Ljubesic (Jožef Stefan Institute [Department of Knowledge Technologies])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Databases, data mining, data curation, computational modelling
Gender: M
Years from PhD: 11

Ms Kaja Dobrovoljc (University of Ljubljana [Centre for Language Resources and Technologies])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Use of language: form, pragmatics, sociolinguistics, discourse analysis, lexicography, terminology
Gender: F
Years from PhD: 2

Mr Mladen Uhlik (University of Ljubljana - Faculty of Arts University of Ljubljana [Oddelek za slavistiko])
Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: typological, historical and comparative linguistics
Gender: M
Years from PhD: No PhD

Spain

Dr Benamí Barros García (University of Granada)
Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning
Gender: M
Years from PhD: 9

Switzerland

Dr Tanja Samardzic (University of Zurich [URPP Language and Space])
Participating as Secondary Proposer
Core Expertise: Other humanities: Computational linguistics
Gender: F
Years from PhD: 7

Turkey

Mr Taner Sezer (Mersin University [Department of Linguistics])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: M
Years from PhD: No PhD

Dr GULSEN ERYIGIT (Istanbul Technical University)

Participating as Secondary Proposer
Core Expertise: Electrical engineering, electronic engineering, Information engineering: Human computer interaction and interface, visualization and natural language processing
Gender: F
Years from PhD: 13

United Kingdom

Ms Natalia Parker (The University of Leeds [School of Languages, Cultures and Societies])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Second language teaching and learning
Gender: F
Years from PhD: No PhD

Mr Turker Sezer (TS DESIGN INFORMATICS LTD. - Turker Sezer)

Participating as Secondary Proposer
Core Expertise: Computer and Information Sciences: Natural Language Processing
Gender: M
Years from PhD: No PhD

United States

Dr Francis Tyers (Indiana University [Department of Linguistics])

Participating as Secondary Proposer
Core Expertise: Computer and Information Sciences: Artificial intelligence, intelligent systems, multi agent systems
Gender: M
Years from PhD: 7

Prof Vsevolod Kapatsinski (University of Oregon [Linguistics])

Participating as Secondary Proposer
Core Expertise: Psychology: Psycholinguistics: acquisition and knowledge of language, language pathologies
Gender: M
Years from PhD: 11

Prof MASAKO FIDLER (Brown University [Department of Slavic Studies])

Participating as Secondary Proposer
Core Expertise: Languages and literature: Linguistics: formal, cognitive, functional and computational linguistics
Gender: F
Years from PhD: 29