THE ROLE OF TEXT-TYPE AND GENRE IN THE ANALYSIS OF SUBTITLES: TOWARDS MEASURING HOW GREEN THE MEDIA ARE

Pilar ORERO
Autonomous University of Barcelona

Alexander SHVETS
TALN, Pompeu Fabra University

Abstract

The United Nations agency UNESCO issued in 2009 the Paris Declaration on Broadcast Media and Climate Change: A Public Service Remit where the role of communication was underlined as vital in informing and educating the public about the realities of climate change and the costs of inaction. Citizens are informed of the environmental crisis through media channels. By integrating the topic of sustainability into all television programming, from children’s programmes to reality TV shows, we can elevate this once-peripheral issue to a prominent position in the public consciousness.

Understanding, measuring, and effectively reporting the degree to which the topic of sustainability is featured in broadcast media is challenging (McDonagh and Orero forthcoming). While some analyses and methodologies exist, these can be difficult to implement and time-consuming. Currently, there are no automatic tools to measure large language-based broadcast corpus data. Some companies have developed their own private methodology. Seeking to remedy this critical oversight, the goal of this paper is to propose the use of subtitles as a tool and explain the singularity of subtitles as a text type. At present, subtitles are already used to mine information, which is the second part of the article. As a text type, subtitles features need to be taken into consideration. The final part defines the objective of our research: to use subtitles as a metric to measure the frequency, topic coverage, and accessibility of sustainability-related content in television. By drawing on subtitling data, we can measure the degree to which sustainability is discussed and presented in the wider broadcast mediascape. These data are expected to be instrumental in drafting reports,
guidelines, and recommendations on sustainability for broadcasting in the region with the possibility of influencing national and international policy. By using subtitles as measurable raw data to develop metrics to gauge the quantity of sustainability-related content in broadcast television, we can better understand how climate change is currently covered. Armed with this knowledge, we can develop strategies and benchmarks to promote sustainability as a concept and contribute towards Net Zero targets.

**Keywords:** text-type, genre, subtitles, green media, accessibility

## INTRODUCTION

Subtitles are a complex text-type due to their formal components and the many features and formats leading to a fluid terminology. Subtitles have been defined as an audiovisual translation modality—with a large body of academic research (Gregory and Caroll; Titford; Agost; Chaume; Gottlieb, “Subtitles-Reading dialogue?”). One subtitle type is included in the category of Media Accessibility, i.e., subtitles for the deaf and hard of hearing (Neves; Matamala and Orero). The subtitle shift from Translation Studies to Media Accessibility is due to two main issues: they are in the same language, and were originally defined as a service for people with loss of hearing, or deaf people.

In 2006 the United Nations approved the Convention of the Rights of People with Disabilities¹ (CRPD). Its signature and ratification by almost all countries made subtitles for the deaf and hard of hearing a default accessibility service worldwide. The subtitles’ compulsory media ubiquity and the rapid development of digital technology led to a text-type which is difficult to classify in the 21st century. Today, information and communication are screen based and digital, usually within the context of the Internet. Subtitles are binary coded as 1 and 0, and there is no agreed format for their distribution. The European Broadcasting Union² defines subtitles as follows:

---


² [https://tech.ebu.ch/docs/tech/tech3350v1-0.pdf](https://tech.ebu.ch/docs/tech/tech3350v1-0.pdf)
Subtitles are created, edited, exchanged, and archived in many ways. At one extreme subtitles may be closely linked to the video, e.g. as burned-in (so-called open, or forced) subtitles in the video, at the other extreme they may be loosely coupled to the audio/video essence (e.g. stored on an external storage medium and associated with the video at the moment of playout/viewing).

In the next section some subtitle features are analysed to determine their text type.

THE SUBTITLE TEXT TYPE

Subtitles have their own text type, and its definition is elusive. Subtitles are a form of written text that appears in one or two lines on the screen, usually at the bottom, to provide a translation or transcription of the spoken dialogue or audio content in a video or film. They convey the spoken content of the video or film to viewers who may not understand the language being spoken or who may have hearing impairments. In this second case, they are also called subtitles for the deaf or captions (Neves). Still, subtitles are more complex, because in turn they represent in text the original narrative, a dialogue in a movie, a narration in a documentary, sounds, or song lyrics.

The subtitle text type, as part of audiovisual translation, was classified first as an audiovisual text according to Reiss’ categories:

1. **Informative Texts**: These texts aim to convey information or knowledge to the audience. They are factual, objective, and often used in scientific, technical, or academic contexts. Informative texts focus on providing accurate and reliable information, presenting facts, data, and explanations in a clear and logical manner.

2. **Expressive Texts**: These texts express the author's emotions, opinions, thoughts, or feelings. They are subjective and often used in creative writing, literature, or personal communication. Expressive texts aim to evoke emotions, create a specific atmosphere, or convey the author's perspective, and may use literary devices, such as imagery, metaphors, and tone, to achieve their purpose.

3. **Operative Texts**: These texts aim to persuade, instruct, or influence the audience to take action. They are typically used in advertising, marketing,
legal, or administrative contexts. Operative texts use persuasive language, call-to-action statements, and specific instructions to motivate the audience to perform a particular task, make a decision, or adopt a certain behaviour.

4. Audio-visual Texts: These texts refer to texts that are presented in visual or audio form, such as films, videos, or multimedia presentations. Audio-visual texts combine different elements, such as spoken language, images, sound effects, and music, to create a cohesive communication experience. They can serve various purposes, including entertainment, education, information, or persuasion.

This classification mixes the audiovisual format with the function, leading to a hybrid category. Audiovisual is added to any of the three categories: informative, expressive, or operational. Subtitles can also be all the four categories at once, since in the same text (a TV programme) they may inform, persuade, and engage. A good example applied to the topic of sustainability would be the subtitles for any of David Attenborough’s programmes dealing with climate change, to name three: State of the Planet (BBC 2000), The Truth about Climate Change (BBC 2006), or The death of the oceans? (BBC 2010).

Subtitles, as a text type, have specific characteristics that make them different from other types of text: time, space, language, and syntax.

a) timing constraints: Time constraints are a critical aspect of subtitles, as they determine the duration for which each subtitle is displayed on screen. Subtitles need to be timed to match the spoken dialogue or audio content in the video or film, ensuring that they appear and disappear at the appropriate moments. Subtitles that are not properly timed can result in a mismatch between the audio and the text, which can confuse viewers and diminish the effectiveness of the subtitles (Szarkowska and Gerber-Morón, “Two or three lines”). Subtitle display can vary depending on the complexity and length of the text (Perego et al.), the pacing of the video, and the target audience’s reading speed (d’Ydewalle et al., “Reading a message”, “Watching subtitled television”; De Bruycker & d’Ydewalle; d’Ydewalle and De Bruycker), and the device where the subtitle is displayed (Gerber-Morón et al.). It is important to strike a balance between providing enough time for viewers to read the subtitles comfortably without them lingering on the screen for too long and disrupting the flow of the video.
(Romero-Fresco, “More haste less speed”, “Final thoughts”; Szarkowska and Gerber-Morón, “Viewers can keep up”).

In addition to the duration per line, subtitles also need to account for other time-related factors, such as the time it takes for a subtitle to fade in and out, any transitions or effects applied to the subtitles, and any pauses or gaps in the spoken dialogue. Properly timing subtitles requires careful consideration of these factors to ensure that the subtitles are synchronised with the audiovisual content. It is important to note that time constraints in subtitles can pose challenges for the translation, often leading to different strategies, such as condensing or paraphrasing the original dialogue to fit the allotted time (Gottlieb, “Subtitling”).

b) limited space: the number of characters per line in subtitles can vary depending on various factors, such as the font size, typeface, screen size, and language being used. However, as a general guideline, subtitles are typically kept within a range of 30 to 40 characters per line, including spaces.

c) the language and syntax (Gottlieb, “Subtitling”; Perego). Since subtitles have to be cut in lines—chunked—their syntax is altered, leading to their own characteristic micro- and macro-syntactic order (Gottlieb, “Subtitles-Reading dialogue?”). Also the vocabulary of translated subtitles has its own features, closer to the source language from which it has been translated (Gottlieb, “The Impact of English”, “Screen Translation”; Dhevi et al.).

All these considerations need to be contemplated when using subtitles as a tool in any analysis that has semantic implications.

**SUBTITLES AS A TOOL FOR SEMANTIC ANALYSIS**

Texts, as a format, are classified, stored, retrieved, analysed, etc. Audiovisual texts, as a format, also have tools and methodologies for their storage, retrieval, analysis, etc. In Media Studies and Translation Studies the most popular audiovisual analysis methodology is manual. To continue with examples from Climate Change terminology, let us consider the four reports published by ECODES\(^5\) (the Spanish Observatory of Climate Change Communication). The

\(^5\) [https://ecodes.org/](https://ecodes.org/)

---

213
latest report issued by the Observatory (76) describes the search methodology as manual: establishing 5 terms "climate change", "global warming", "climate crisis", "greenhouse effect" or "climate emergency", and then defining the analysis sample as subtitles from all television broadcasts (DTT) from January to July 2021 in Spain. They further scaled the search to four full weeks in six months, chosen at random. The analysis measured the frequency of these five terms and the percentage applied to time of day (seven sections), genre of programme (four genres), with a discourse analysis of seven sectors (politics, economy, scientific, extreme weather, technology, socio-cultural, activism, health, and environment). Other information was also derived from the search regarding the geographic dimension (local, state, European, worldwide), who the speaker was, etc. And finally, the last step was the extraction of semantics from subtitles, using tools from the fields of Computational Linguistics, Natural Language Processing and Data Retrieval (Reidsma et al.; Declerck et al.).

**Scientific research**

Subtitles are a popular source of data to perform analyses. Classifying media content is an important task in the media industry and an active research area. Understanding media content is basic towards building recommendation systems for the audience, or for understanding theme frequency, etc. Subtitles have been used for example to predict word frequency (New et al.; Cai and Brysbaert). As early as 2004 the project POLYSEMA\(^4\) began to develop an end-to-end platform for interactive TV services (Katsiouli et al.). This included a novel residential gateway architecture that can provide intelligent iTV services by exploiting the metadata of the broadcast transmission. Media content has many elements, from the obvious, such as sound, to some less noticeable, such as colour, to name a couple. Each element might be classified with different labels, for example sound could be labelled as music, but also as sentiment. Media classification techniques use mono labelling, generally based on audio-visual features. Recently, multilabel movie classification models have been built using supervised machine learning techniques. The clear benefit is the robustness of the results of these hybrid classification systems, but still much work and data

are needed. Using subtitles as a label we find the work of Bougiatiotis and Giannakopoulos to understand media content representation, which in itself allows to match similarity in the video content. Saumya et al. and Hasan et al. used subtitles towards movie genre detection. Subtitles can be used to analyse the genre of media content, such as movies, TV shows, or online videos. Researchers can extract textual features from subtitles, such as keywords, dialogue patterns, or visual cues, to classify the media content into different genres, such as comedy, drama, action, or romance (Carvalho et al.). This can provide insights into genre conventions, audience preferences, and cultural differences in media content (Armstrong et al.). Subtitles can be used to classify the emotional tone of media content, such as positive, negative, or neutral sentiment (Hughes and Armstrong). Researchers can apply natural language processing techniques to analyse the sentiment expressed in the subtitles, such as sentiment lexicons, machine learning algorithms, or deep learning models (Oliveira et al.). This can help understand the emotional impact of media content on audiences and explore sentiment patterns across different genres, languages, or cultural contexts. Subtitles can be used to categorise media content based on specific themes, topics, or subject matter (Langlois et al.). Researchers can use keywords, topic modelling, or text clustering techniques to identify and classify the main themes or topics discussed in the media content. This can be useful for studying patterns of representation, discourse, or ideology in media content, as well as for tracking changes in content over time or across different media sources (Bethard et al.). Subtitles can be used to identify the language or languages used in media content towards understanding language use in different contexts, or language policies in media production and distribution. Subtitles are also used to analyse cultural elements, such as references to cultural norms, values, or practices. Researchers can examine the translation and adaptation of cultural references in subtitles to investigate how media content is localised for different audiences or cultural contexts, like in video games (O’Hagan and Mangiron). This can provide insights into cross-cultural communication, media globalisation, or cultural representation in media content (Pym). Subtitles are also used towards the automatic generation of trailers, or video summarisation (Hesham). Subtitles may guide in the choice of descriptive frames for the video. Prediction of movie genres is an intriguing
problem that has several applications in designing recommendation systems for the audiences, analysing movie box office performance, and understanding the theme of the movie, to list just some.

**Subtitle analysis in the media production industry**

The media industry, as any other industrial sector, needs urgently to design CO2 industry related calculators to report their action against climate change. Different variables have been used to build CO2 calculators related to: transport, lights, catering, etc. As far as we know, the first organised initiative to deal with the issues of sustainability in the film industry began as part of CineRegio⁵, as a network of regional film funds in Europe. The network began with 15 regional film funds from 11 European countries and today represents 52 regional film funds from 12 EU Member States, Norway, Switzerland, and the UK. CineRegio’s mission is to support schemes and services related to film culture, encourage social cohesion, and build regional infrastructure. Its main objectives are:

- To exchange views, perspectives, good practices, and information for the benefit of the European film industry, including integrating new regional film funds.
- To raise awareness, represent and promote regional audiovisual interests across Europe, including members' interests towards European institutions and other organisations which play a role in determining the set of rules and conditions for regional film funds.
- To strengthen the co-development and co-production of audiovisual products, fusing talents, and resources in different regions for a wider market - including stimulating artistic, technical, and creative exchange and know-how throughout Europe.

In 2012, Green-Regio started as a branch of CineRegio and in 2013, at Cannes Film Festival, it issued the first interim report⁶ which was followed by reports in

---

⁵ [https://www.cineregio.org/about_cineregio/the_story_so_far/2005/](https://www.cineregio.org/about_cineregio/the_story_so_far/2005/)
⁶ [https://www.screensouth.org/wp-content/uploads/2020/02/Screen-South-Sustainable-Environment-Policy-FINAL.pdf?_ga=2.127782477.1924804755.1680338427-81341374.1680338427&_gl=1*1npjznzk*_ga*ODEzNDEzNzQuMTY4MDMzODQyNw..*_ga_ZBES4QNZTV*MTY4MDMzODQyNi4xLjEuMTY4MDMzODQsMy4wLjAuMA.](https://www.screensouth.org/wp-content/uploads/2020/02/Screen-South-Sustainable-Environment-Policy-FINAL.pdf?_ga=2.127782477.1924804755.1680338427-81341374.1680338427&_gl=1*1npjznzk*_ga*ODEzNDEzNzQuMTY4MDMzODQyNw..*_ga_ZBES4QNZTV*MTY4MDMzODQyNi4xLjEuMTY4MDMzODQsMy4wLjAuMA.)
2014, 2015, 2017, and 2020. From the early days, the “Sustainability in Vision” was drafted from experiences on sustainability production in different countries. The report looks at collaborative ways to support the industry to be more sustainable in an accessible and cost-effective way. It touches on setting the “Green Production” strategies and practices in the wider context of the audio-visual industries’ “Corporate and Social Responsibility” agenda. Green-Regio reports are built from use cases, and some of them flag up approaches such as: a simple, robust, and systematic methodological approach to sustainability, partnerships for awareness projects, standardising international certification, education and training, and promoting the use of various accessible and compliant carbon tracking software for production—but none assesses the media content. Media industry carbon calculators offer the possibility of understanding the sustainability of movie production and distribution. The industry should also analyse the media content: do they deliver engaging sustainable narratives; how much attention is paid to climate change as a discourse. In 2019, ALBerti started using subtitles to “analyse how the broadcast industry is covering climate change on our screens.” Every year they publish a report, based on the research performed by the private auditing company Deloitte. The research started with data from BBC, ITV, Channel 4 and Sky, representing 40 channels and 128,719 distinct programmes from September 2017 to September 2018. From an academic point of view, it is fundamental to understand the different elements for the set up: the choice of the corpus gathered, the tools for analysis, and the methodology used in the study, towards replication. This is not possible in the case of ALBERT, since the information given regarding the methodology followed to build the glossary (or the themes, as they are called) is quite vague: “The themes for analysis were chosen based on those

---

11 [https://wearealbert.org/](https://wearealbert.org/)
12 [https://www.deloitte.com/](https://www.deloitte.com/)
which fell within the UN’s sustainable development goals that relate to the environment.” The themes shown are:

- **Food**: food waste, vegan, meat free, vegetarian, meat substitute, food miles.
- **Resources**: upcycle, single use, recycle, environmental impact
- **Travel**: electric vehicle, hybrid car, clean air, carbon offset, electric car
- **Energy**: green energy, solar power, wind power, renewables, clean energy
- **Climate Knowledge**: carbon emissions, climate change, carbon footprint, global warming, eco-friendly

This clashes with the type of language used in TV content where programmes are for diverse audiences: for children to news or sport. Also, the justification for the choice of words in each theme is not very clear: “Words were chosen to populate each theme using a combination of manual selection and machine learning.” Also problematic is the lack of information regarding the tools used in the selection. “The word list was then expanded using language association tools; however not all new terms were added as some were considered false positives (e.g., “vegetable” strongly related to “vegan” but in most instances was not relevant to environmental impact).” Again, it is not possible to replicate this, since there is no further information regarding words, nor tools.

While **ALBERT** does not fulfil scientific principles, it is a good first step towards the use of subtitles to analyse sustainability in media content. **ALBERT** (2019) would also like to “capture full sentiment analysis to indicate whether our television coverage of the topic is neutral, positively or negatively biased.” While they reported that, their aim was not entirely achieved. In particular, it was not possible to attribute sentiment to every instance of the search tewrms we analysed, nor was it possible to quantify whether mentions were supportive, impartial, or dismissive. However, spot checks were carried out to indicate relevance and to begin to build a picture of sentiment. General spot checks confirmed that across all the findings, 90% of instances were used in the context expected, i.e., using ‘beef’ to refer to food rather than climate arguments. Exploring ‘beef’ in more detail, analysing 100 of the 14,984 total mentions of ‘beef’ revealed that 95% of these mentions were in the context of ‘beef’ as a food source,
5% of mentions were out of context and 0% were in relation to reducing beef consumption to reduce our carbon footprint.

Though the subtitle analysis performed by Deloitte for ALBERT should be taken as any industry-led study, the objective is very interesting and can be extrapolated to other languages and improved with scientific robustness, since it seems to report on major trends. For 2021 ALBERT concluded that “the word frequency is associated with five key sustainability topics: food, travel, resources, energy, and climate knowledge. Interestingly the green discourse on UK media is not increasing. The 2021 report finds 14,540 mentions of climate change, while climate justice, climate action and climate solution only received 296 mentions collectively.” This data can be interpreted as anything you want, since the tools for analysis are not disclosed.

Another example of subtitle analysis is that from the Spanish public broadcaster RTVE. Though no publication has been found, the information is from the presentation at the 2023 EBU summit “SDGs key programming”\textsuperscript{13}. On the RTVE webpage\textsuperscript{14} the subtitle analysis tool “Parliament 2030” has the following definition: Parliament 2030 is an innovative tool that tracks, gathers, and offers information on the activity of the RTVE programmes related to the Sustainable Development Goals (SDGs). Designed to overcome the challenges posed by the cross-cutting nature of the 2030 Agenda, SDG Programming classifies information related to the SDGs thanks to an advanced automatic mass labelling system. This innovative technology allows users to browse TV content related to the SDGs through a free and open online search engine. The information provided is essential for the monitoring and accountability of the implementation of the 2030 Agenda at the national level. It works by creating a database through an automatic mass labelling process that classifies content according to its relationship with any SDG or goal. It combines advanced computer science - necessary to scrape 520 subtitling hours in which information is published and transform it into a structured database - with the design of SDG dictionaries that automatically relate folksonomies to objectives and targets. Then, through an intuitive, attractive, and accessible search engine, it offers all the relevant information in reusable formats. Users can select their

\textsuperscript{13} https://tech.ebu.ch/publications/demo-rtve-2030-sdgs-key-programming
\textsuperscript{14} https://rtve.politicalwatch.es
own search terms to tailor searches to their needs, as can be seen in figure 1 below.

![Los 17 ODS en la programación](image)

**Figure 1.** User interface of RTVE SDG app

Again, as in the case of ALBERT, no information is given related to the corpus, tools, or methodology beyond what is published on the website. And as in the case of ALBERT, it might not be scientific, but it offers a good example for ways to go about gathering indicators for sustainable media content.

**CONCLUSION**

Sir David Attenborough recently remarked that “saving our planet is now a communications challenge.” There is a growing interest in engaging all citizens from all walks of life in discussions about the climate crisis. This intensification of communication strategies as an action for helping with climate change has also been promoted by the calls for funding issued in 2023 by the European Commission as part of the Green Deal. It is important to give citizens clear and scientifically robust information to fully comprehend and value the magnitude of the ongoing climate crisis, allowing individuals to better respond to it.

Public broadcasters have the social responsibility to inform all their viewers about environmental issues and the climate crisis through their programming. However, understanding and measuring the degree to which the topic of sustainability is featured in broadcast media is still an open issue. For years, public broadcasters have been legally obligated to provide subtitles as per
the terms of the Convention on the Rights of People with Disabilities (CRPD). With nearly 100% of their content subtitled, this large corpus could be used to extract information on the following: 1) Words related to environmental issues (such as climate change) in different programmes. 2) Number of hours allotted for discussing various topics related to environmental issues. 3) The types of programmes that make the most reference to environmental issues.

To date, subtitles have been used to perform visual data indexing, but the methodology can still be improved. The subtitle text type has to date not been taken into consideration, hence the results of any data analysis are challenging. The use of subtitles as a metric for automatic linguistic analysis and structuring of media content in relation to the topic of the climate crisis has been proposed in this paper. This first article will be followed by the development of a methodology for measuring the prevalence of environmental issues in broadcast television based on the resulting structured data. These measurements will enable academic, societal, and industrial partners to draw data-driven conclusions regarding the content coverage in broadcast television that can aid in the creation and implementation of sustainability-focused policies, and ultimately contribute towards creating a local green media agenda in which environmental issues are front and centre in most programming. The language analysis will provide an understanding of media content from various perspectives by detecting keywords, creating their specific definitions in the climate change context, and identifying covered sources of environmental issues, actions, and impacts with respect to a geographic dimension and a style in which the information is communicated to citizens.

Acknowledgements
This article has been partially funded by the H2020 project GreenSCENT: Smart Citizen Education for a Green Future with grant no. 101036480, the EC-funded research and innovation programme Horizon Europe under the grant agreement no. 101070278 (the ReSilence project), and the EC funded MSCA CLEARCLIM with grant no. 101131220. Pilar Orero is part of the Catalan Government funded research group TransMediaCatalonia 2021SGR000077 and part of the Network ACCESSCAT 2021 XARDI 000007.

Works Cited

Agost, Rosa. “Diversitat tipològica i traducció audiovisual.” La diversitat discursiva, edited by M.D.
https://doi.org/10.1371/journal.pone.0010779


Katsioulis, Polyxeni, Vassileios Tsetsos, Stathes Hadjieftymiades. “Semantic video classification based on subtitles and domain terminologies.” *SAMT’07 Workshop on Knowledge Acquisition from Multimedia Content (KAMC)*, Genova, Italy, December 2007 ftp.informatik.rwth-aachen.de


Papadimitriou, Antonis, Christos Anagnostopoulos, Vassileios Tsetsos, Sarantis Paskalis, Stathes


Szarkowska, Agnieszka and Olivia Gerber-Morón. “Viewers can keep up with fast subtitles: Evidence from eye movements.” Plos One, vol. 13, no. 6, 2018, 10.1371/journal.pone.0199331

Szarkowska, Agnieszka and Olivia Gerber-Morón. “Two or three lines: a mixed-methods study on subtitle processing and preferences.” Perspectives, vol. 27, no. 1, 2019, pp. 144-164. 10.1080/0907676X.2018.1520267


BIONOTES

Pilar ORERO, PhD (UMIST, UK), works at Universitat Autònoma de Barcelona (Spain) in the TransMedia Catalonia Lab. She has written and edited many books, near 100 academic papers and almost the same number of book chapters – all on Media Accessibility. Leader and participant on numerous EU funded research projects focusing on media accessibility. She works in standardisation and participates in the
UN ITU IRG-AVA - Intersector Rapporteur Group Audiovisual Media Accessibility, ISO, SAGA, and ANEC. She has been working on Immersive Accessibility for the past 4 years first in a project called ImAc, which results are now further developed in TRACTION, MEDIAVERSE, MILE, and has just started to work on green accessibility in GREENSCENT. She led until December 2022 the EU network LEADME on Media Accessibility. She is the Co-Chair of Study Group on Accessibility and Inclusion in the ITU Metaverse Focus Group.

E-mail: Pilar.Orero@uab.cat

Dr. Alexander SHVETS is a Lead Researcher at the Natural Language Processing Group (TALN) at UPF. He earned his Diploma in Applied Mathematics and Computer Science from the Siberian Federal University and his Ph.D. in Computer Science from the Russian Academy of Sciences. His research interests lie in the areas of multilingual content extraction, with application to text understanding, natural language generation, and dialogue management. His recent papers include studies on computational lexicography and hate speech analysis. In the past thirteen years, Dr. Shvets has been actively participating in various European and national research projects in dialogue management, opinion-driven design, disaster management, and scientometrics, being the UPF team coordinator in several of them.

E-mail: alexander.shvets@upf.edu