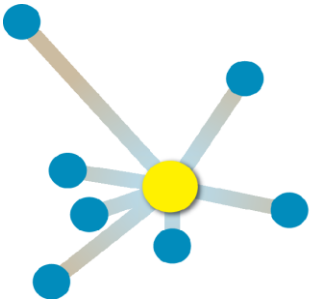
European Language Resource Coordination 3.0 (ELRC3.0) is a service contract operating under the EU's Connecting Europe Facility SMART 2019/1083 programme.

SMART 2019/1083
ACTION ON CEF AUTOMATED TRANSLATION
CORE SERVICE PLATFORM

D5.2.3 Report of Workshop 3

Author(s): **Tom Vanallemeersch (CrossLang)**

Version: **1**
Date: **2022-02-28**



European Language Resource Coordination
Connecting Europe Facility

Table of Contents

1. Introduction **1**
2. Setup **2**
2.1 Preparation 2
2.2 Agenda 3
2.3 Participants 4
2.4 Practical organisation 5
3. Presentations **6**
3.1 Welcome by project representative 6
3.2 Introduction by Head of Unit at DG CNECT 6
3.3 Easy-to-understand language: practice, training and standards 6
3.4 The H2020 Easy Reading software framework 7
3.5 Norm-breaking research: creating understandable content for migrants 8
3.6 Easy Language in Belgium - prospects for automatic simplification 9
3.7 Natural language processing can support clear writing: AMesure 10
3.8 Challenges in evaluation of automatic text simplification 10
3.9 Less is more: keeping it short and simple with ASD-STE100 11
3.10 Final discussion 12
4. Conclusions **14**

1. Introduction

The workshops organised in the SMART 2019/1083 programme support the continued development of the EC's eTranslation system and a wider deployment of DG CNECT's services in terms of language resources and tools. The third of these workshops, *Simplify Language – Capture Audience*, focused on software, techniques and procedures for text simplification, for measuring readability of texts, and for implementing controlled language by restricting vocabulary and grammar. It covered objectives concerning social inclusion and better accessibility of technical information to a large audience.

The workshop included 8 speakers. Besides the organisers and the speakers (5 from academia, 1 from a company, 1 freelancer and 1 from a public administration), the 100 participants included staff of EU administrations and Member State administrations, representatives of companies and academia, and freelancers.

The following subtopics were dealt with in the presentation:

- easy language (make text readable and comprehensible for people with cognitive disabilities);
- plain language/clear writing (make information accessible to the general public, for instance administrative texts);
- multimodal integration (for instance in the context of audiovisual information);
- support for migrants;
- guidelines for clear writing;
- resources for supporting text simplification;
- tools for automated text simplification and readability assessment;
- evaluation of text simplification results;
- controlled language;
- standardisation.

This document is structured as follows. Section 2 provides information on the setup of the workshop (preparation steps, agenda, participants, and practical organisation of the workshop day). Section 3 provides the abstracts of the presentations, questions of participants, and a description of the final discussion between the speakers and other participants. Section 4 provides conclusions.

2. Setup

This section provides information on the setup of the workshop, i.e. the preparation, the agenda, the participants, and the practical organisation of the workshop day.

2.1 Preparation

CrossLang settled a date for the workshop in agreement with the Contracting Authority and set up an agenda including 8 speakers (5 from academia, 1 from a company, 1 from a public administration, and 1 freelancer). The agenda set up by CrossLang was submitted to and approved by the Contracting Authority. The speakers submitted their presentation abstracts in order to complete the agenda.

A page with the workshop agenda was set up in the ELRC website (https://lr-coordination.eu/tw), as well as a page with the abstracts of the speakers and a registration page. CrossLang and DG CNECT created awareness of the workshop among potential participants inside and outside the EU administrations by providing them with a link to the workshop agenda. The workshop was promoted through social media activities (Facebook, LinkedIn, Twitter). While the audience of the workshop consisted of staff of EU and Member State administrations in the first place, the workshop was also open to a wider audience of people with an interest in techniques and research relating to the topic. Therefore, the speakers were invited to contact their colleagues. A total of 165 people registered: representatives from administrations, companies and academia, and freelancers. A Zoom session was set up and the people who registered were provided with the Zoom link.

Speakers sent their slides beforehand in order to anticipate on potential Internet connection problems during the workshop; for instance, in case of screen sharing problems, navigation through the slides can be performed by the organisers based on instructions by a speaker.

Two poll questions and potential topics for the final discussion were set up by consortium members, i.e. by the organisers from CrossLang and by Khalid Choukri, who offered to act as a moderator during the final discussion; he chairs the committee *ISO/IEC JTC 1/SC 35* (User accessibility and ICT products and services).[1]

[1] https://www.iso.org/committee/45382.html

2.2 Agenda

AGENDA

10:00	Welcome by **Tom Vanallemeersch** (SMART 2019/0183 Project representative)
10:05	**Introduction** by June Lowery Kingston, Head of Unit "Accessibility, Multilingualism, Safer Internet" (DG CNECT)
10:10	**Easy-to-understand language: practice, training and standards** **Anna Matamala** (Universitat Autònoma de Barcelona)
10:50	**The H2020 Easy Reading software framework** **Klaus Miesenberger** (Johannes Kepler University Linz, JKU)
11:20	**Norm-breaking research: creating understandable content for migrants** **Susanna Laurin** (company Funka, Sweden)
11:50	**Easy Language in Belgium - prospects for automatic simplification** **Vincent Vandeghinste** (Instituut voor de Nederlandse Taal / Dutch Language Institute)
12:20	**Discussion**
12:35	**Lunch**
13:45	**Natural language processing can support clear writing: the example of the AMesure platform** **Thomas François** (CENTAL, IL&C, UCLouvain)
14:15	**Challenges in evaluation of automatic text simplification*** **Fernando Alva-Manchego** (University of Sheffield)
14:45	**Less is more: keeping it short and simple with ASD-STE100** **Daniela Zambrini** (Freelance technical translator)
15:15	**Discussion**
15:45	**Conclusion**
16:00	**End**

2.3 Participants

165 people registered (including the speakers and organisers). The registered people have the following profiles:

- staff of EU administrations: 50 (30%);
- staff of Member State administrations: 48 (29%);
- representatives of companies: 35 (21%);
- representatives of academia: 25 (15%);
- freelancers: 7 (4%).

Of those who registered, 98 (59%) people logged in to the workshop.

2.4 Practical organisation

From the part of CrossLang, three people participated in the workshop: one for chairing the workshop, one for providing technical support to the chair and the attendees, and one for supporting the chair in following up the chat questions.

The speakers were given the *co-host* status in order to be able to share their screen. They were asked to log in 30 minutes before the start of the workshop in order to test whether screen sharing worked without problems and, in case they had audiovisual material (which was the case for one speaker), in order to test whether the rendering of video and audio worked well.

When attendees logged in, the workshop organisers admitted them manually. Attendees used the chat function of Zoom for questions. After each presentation, the chair read questions from the chat window. The attendees could also use the *raise hand* button after a presentation in order to request the floor.

In order to give the workshop participants the possibility to interact in an Informal way during the lunch break, CrossLang set up break-out rooms.

After the presentation of the second speaker, the organisers asked a poll question using Zoom functionality:

What is/are your objective(s) when producing plain or easy language?

Possible answers (one or more per participant):

1. Making the target audience aware of specific information
2. Preventing the audience from interpreting a text incorrectly
3. Increasing the efficiency of the documentation workflow
4. Other objective(s)

The first answer was the one provided by the largest number of poll participants, the second and third objectives being considered as less important overall.

After the lunch break, another poll question was presented to the audience:

Is it important for you to define a specific terminology for the target audience?

Possible answers:

1. Yes, it is highly important
2. Yes, but it is not the focus of our attention
3. No, not at all

The first option was the one provided by the great majority of poll participants. Only one poll participant selected the third option. It is clear that terminology needs to be taken into account when it comes to plain, easy and controlled language.

The slides of the presentations were uploaded to the workshop page on the ELRC website. The link to the recordings of the presentations was also made available to the workshop participants.

3. Presentations

This section provides the abstracts of the presentations and the questions following presentations. The section also reports on the final discussion, which involved speakers as well as other attendees.

3.1 Welcome by project representative

The project representative, Tom Vanallemeersch (CrossLang) welcomes the audience.

3.2 Introduction by Head of Unit at DG CNECT

June Lowery-Kingston, Head of Unit "Accessibility, Multilingualism, Safer Internet" at DG CNECT, explained the potential of language technologies for Persons with Disability (PwD), as well as the Web Accessibility Directive (WAD), related initiatives from W3C, and an ETSI[2] standard. She indicated that the average literacy level across the EU (around 15% at or below level 1 on a 5-point scale) has implications for all public sector communications, not just the one targeted towards PwD. She further indicated that easy to read material is difficult for humans to produce but at the same time essential for the public and private sector. The purpose of the workshop is to bring together different stakeholders from across the different elements of language technology and accessibility. Ms Lowery-Kingston issued a challenge for the workshop participants: can you produce an easy version of the WAD?

3.3 Easy-to-understand language: practice, training and standards

Anna Matamala, associate professor at the Universitat Autònoma de Barcelona, provided an overview of easy-to-understand language, ranging from easy language (or easy-to-read) to plain language. She presented some of the main challenges in terms of terminological choices and end user definition, and provided an overview of the current situation of easy-to-understand language training and practice across Europe. To this end, she used the results of a 3-year Erasmus + project funded by the European Commission (2018-2021): EASIT, Easy Access for Social Inclusion Project.[3] It is an innovative educational project that stresses the need to go beyond written documents and considers how easy-to-understand language could be integrated in audiovisual media and audiovisual access services. The presentation also highlighted the need for standardisation and further research that supports current standardisation efforts. More specifically, it referred to the Spanish standard on Easy-to-Read and to the ISO standard on making written text easy to read and easy to understand (ISO/IEC CD 23859-1), currently under development.

[2] European Telecommunications Standards Institute: globally applicable standards for ICT-enabled systems, applications and services.
[3] pagines.uab.cat/easit

The introductory presentation by prof. Matamala set the ground for other presentations in the workshop, which adopted a more technological approach, considering tools that support writing processes in easy-to-understand language.

The following questions were asked to the presenter:

Do you use (or would it be a good idea to use) machine translation to make the EASIT materials more broadly available in other languages?

We have the materials on Youtube and they are visualised on the platform. So our users could activate automatic subtitles. The result would of course not be the same, but we are open to discussion and to any collaboration.

But machine translation goes against personalisation?

Machine translation can help you customise your content as a user. The question is whether it can achieve personalisation or on the contrary has a limited, more general result. There is a hot debate on human translation vs. machine translation. We have done research on machine translation in audiodescription and voice-over. Our personal view is that both human and automatic translation have a place and are good choices for users to personalise. When organising a public, face-to-face event, we will employ a subtitler producing text in English and Catalan. But at the same time, participants will have a wider choice and can personalise through an application on their mobile phone for reading the subtitles as well as choosing to translate them in many languages.

Would translated "easy content" be also easy content?

If machine translation takes place without post-editing, the situation is unclear.

3.4 The H2020 Easy Reading software framework

Klaus Miesenberger, head of the Institut Integriert Studieren[4] at Johannes Kepler University Linz, presented the Easy Reading software framework, a browser extension for users with cognitive disabilities and a backend for care providers, supporting the personalisation of web content in real time. This is achieved by (a) adaption of layout and structure of web pages, (b) annotation of web content with symbols, videos, etc., and (c) conversion and translation of content into a different language level, e.g. plain language or easy-to-read language. The Easy Reading project is funded by the European Union‘s Horizon 2020 research and innovation programme. The personalisation in the software framework is only done on demand, allowing the user to remain and work within the original content. This fosters independent access, supports learning to cope with original content and keeps the user in the inclusive discourse to enable participation and contribution. The backend for care providers and web owners allows for customising and controlling how web content is displayed for individual users or for groups of users. The Easy Reading tool, developed in cooperation with end users as co-researchers, allows for addressing cognitive accessibility at the highest level without separating the user from the original content.

[4] www.jku.at/iis

The following questions were asked to the presenter:

Does the auto symbol translation support multiple symbol sets?

Yes. Accessibility and assistive support is provided a service. It integrates specialists, trainers, therapists, parents, relatives, etc. Together with them you can integrate and import the symbol set you want to use. For instance, you click on a word and get a proposal of different symbols. This supports the service provider in selecting the right symbol for the user. We are still working on these backend functionalities and try to support as many symbol systems as possible.

Could a one-click easy-to-read converter be potentially created for the PDF format?

In terms of research, this would require two to three developers for a year. It is not on the agenda because the tricky issue is to get to the PDF standard. We would be happy to find partners to cooperate on this. Our actual goal is to bring our functionality from the web browser to the operating system so it works across all applications that run on your computer, for instance when reading PDFs. We are confident that this integration into an operating system will take place sooner or later. It needs a huge investment in terms of programming, but the benefit is obvious according to the end users.

Could your strategy of supporting users also serve a more general audience, e.g in the context of plain language?

Yes, the field of assistive technology often works this way. For instance, predictive typing or text to speech were technologies developed by and for people with disabilities. They entered in almost every system. When technology works for the “extreme user”, we have a very high confidence that is beneficial for everybody.

3.5 Norm-breaking research: creating understandable content for migrants

Susanna Laurin, Chief Research and Innovation Officer of Swedish company Funka, discussed the company's projects related to simplified language. She focused on a project carried out together with a local government involving a population covering a large number of different languages and a high percentage of recently arrived migrants. In this project, Funka investigated, tested and created guidelines on how to best create information that reaches these target audiences when translating into all native languages is not feasible. Other projects of Funka concerning simplified language involve the conversion of ordinary text from public sector websites into plain language using machine learning, and the communication of information and instructions from public sector agencies to disabled people, using multimodal means (text combined with video, audio and images).

The following questions were asked to the presenter:

Have you ever worked with the Italian public sector?

Yes, but not on the level of content. Our team includes a person that speaks Italian, but we do not have easy to read or plain language knowledge for this specific language. We have clients in most European member states, including the Italian public sector.

Do you take the cultural background of migrants into account, or do you go for a more generic approach?

We do not only translate text but also provide background. For instance, migrants from Syria need more information than German migrants on how Swedish society works. There is a need for contextual information and explanation. Digital provision of information is easier because you can provide links. There is a misunderstanding on easy-to-read language when it comes to cognitive disabilities: you do not take things away, but, on the contrary, provide more information.

3.6 Easy Language in Belgium - prospects for automatic simplification

Vincent Vandeghinste, researcher at the Dutch Language Institute (Instituut voor de Nederlandse Taal) in Leiden, the Netherlands, and at the Centre for Computational Linguistics and Leuven.AI (the Leuven Institute for Artificial Intelligence) at KU Leuven, presented the chapter about Belgium of the Handbook of Easy Languages in Europe, released in the summer of 2021. The chapter was co-written with Adeline Müller, Thomas François and Orphée De Clercq, and zooms in on the needs for easy language, the past and current status of easy language and plain language, and their target groups. According to Lindhom & Vanhatalo (2021), "the notion of Easy Language refers to modified forms of standard languages that aim to facilitate reading and language comprehension. Easy Languages have been adapted in terms of content, vocabulary, and structure to make them more readable and comprehensible. They are intended for people who have difficulties understanding standard language." The second part of the presentation focused on a number of writing aid tools for Dutch and some easy language resources and their linguistic properties, and how they could be used to train automatic simplification systems.

The following questions were asked to the presenter:

How much training material is needed to train the simplification language model? How does this quantity compare to the one needed for machine translation?

The required quantity will become apparent once we train systems. In the unsupervised stage, we have a lot of general training data: the BERT[5] model is based on 6 billion words of Dutch. If we finetune it, we should not need too much data. Moreover, there should be sufficient specific material, as Wablieft contains 2 million words, which is not huge in machine learning terms but is still quite substantial. The corpus may get larger in the future if we or Wablieft find the time to update it. It now contains data up till 2018.

You mentioned different associations in the Dutch-speaking and French-speaking part of Belgium. Is there cooperation between them or are their strategies different?

For plain language, the federal level issues bilingual guidelines and organises bilingual courses, activities in this context are pretty parallel. Regional governments have their own emphasis and marketing campaigns. Some organisations have a pendant in the other country part. One the level of easy language, there is more separation: in the Dutch-speaking part, there is a newspaper, while the French part has a book publisher. However, disability organisations have sister partners because health regulations are federal matter. It also depends on the specific user group.

[5] Bidirectional Encoder Representations from Transformers. A BERT model is highly rich in contextual information: it takes into account the context of each occurrence of a word in a text rather than keeping a single representation per unique word in the text.

3.7 Natural language processing can support clear writing: AMesure

Thomas François, assistant professor in Applied Linguistics at UCLouvain in Belgium presented the AMesure platform.

Recent research (e.g. Nord, 2018) reveals that, although several plain language guides are available to assist writers of administrative texts in their work, the guidelines provided are not always followed by writers, mainly because they are too vague and too numerous. To support the practice of clear writing and the application of all these guidelines, prof. François and his team designed a web platform, AMesure, which aims at automatically identifying clear writing issues in administrative texts and provides simple writing advice that is contextually relevant (François et al., 2020). In its current state, the platform offers four main functionalities: (1) providing an overall readability score based on a formula tailored to administrative texts, (2) identifying, in a text, the linguistic phenomena that are assumed to have a negative effect on the comprehension of the text, (3) for these phenomena, proposing simplification advice from clear language guides, and (4) automatically generating simpler synonyms of difficult words. The presentation briefly sketched various opportunities offered by the field of natural language processing to clear writing studies. Based on a recent survey of the practice of professional writers (Müller et al., 2021), the speaker reported some current challenges and opportunities for clear writing and discussed how a platform such as AMesure could assist writers of French administrative texts to simplify their writing.

The following questions were asked to the presenter:

How language-specific are the linguistic features that you detect? Would it be hard to adapt the system to another language?

We use handcrafted rules and lists, which you could probably adapt to other languages. For the synonym generation we apply a more standard approach, which is very easy to adapt to other languages.

Does your approach take into account the typical length of texts of specific types in number of words?

As AMesure is a prototype, we have so far decided not to make it adaptable for various aspects. These aspects include the genre of the text and, maybe even more important, the type of public. We need to have an idea of the type of reader we are targeting (general public or reader with difficulties) in order to set a threshold for deciding whether a word or sentence is too rare, too long, etc.

3.8 Challenges in evaluation of automatic text simplification

Fernando Alva-Manchego, Postdoctoral Research Associate at the University of Sheffield and a member of the Natural Language Processing Group, explained that most current models for automatic text simplification are data-driven: given a large dataset of parallel original-simplified sentences, models are trained to implicitly learn how to perform a variety of editing operations that aim to make a text easier to read and understand. However, how can one verify whether an automatic output is actually 'simpler' than its original version? As is the case for many natural language processing tasks, this should be done using both automatic and manual assessments. The speaker first presented the results of a meta-evaluation of automatic metrics for automatic sentence simplification and showed how much the correlation between metrics and human judgements is affected by factors such as the perceived simplicity of the outputs, the system type, and the set of references used for computation. After that, he presented some preliminary results on a study of joint translation and simplification, and showed how difficult it can be for lay users to

manually assess simplicity. He concluded with some recommendations and ideas for future work in evaluation of automatic simplifications.

The following questions were asked to the presenter:

As there is no such thing as an average user of a simplification system, should one collect references from a specific group?

In this particular case, for all of these three datasets, all references were collected through crowdsourcing. Annotators included people with high proficiency in English and people with a low proficiency. The annotators were not split up in further ways. From a natural language processing perspective, we suffer a lot from the fact that we cannot check additional considerations (circumstances) affecting the target audience; this limitation is due to budget constraints.

Evaluations are performed using the BLEU measure – how useful is this measure given the fact that experiments show that it does not bring any added value beyond 4 references?

There have been some studies on BLEU for other natural language generation tasks such as machine translation. They show that not only the number of references is important, but also the quality and the variability. There are other types of research that do not look into comparing one output to references but generate a diversity of outputs and compare the latter against a diversity of references. Therefore, you can compare all the characteristics of the model as well. Variation in natural language generation, methodology, and human and automatic evaluation are actively being explored.

3.9 Less is more: keeping it short and simple with ASD-STE100

Daniela Zambrini, who has been working in the airline industry since 1987 and as a freelance technical translator for over twenty years, presented the purposes of Simplified Technical English (STE) and illustrated the structure of the ASD-STE100 specification and its advantages for translators and technical authors. ASD-STE100 is a controlled natural language which was developed in the early eighties (as AECMA Simplified English) to help the users of English-language aircraft maintenance documentation understand what they read. STE addresses difficulties in English comprehension that are related to complex sentence structures, confusing word forms, and ambiguous vocabulary. Although STE was created to improve aircraft maintenance documentation, its principles can dramatically improve the reading quality of technical documentation in any industry. Its advantages in the translation process are directly related to the reduced chance of risk and ambiguity.

The following questions were asked to the presenter:

To what extent are/can the writing rules (be) automated? Or is it feasible to automatically detect "issues"?

Yes, to assist potential users of STE there are quite a few companies that market software supporting STE. These efforts are greatly appreciated by ASD (Association of Aerospace Industries) and STEMG (Simplified Technical English Maintenance Group). Our approach is mutual. We do not endorse any kind of software as our principle is that software does not think in your place. But as technology is fast developing and we want to merge the present and the future, our idea is to organise as wide as possible a forum with all the

stakeholders, i.e. users, software developers, trainers, etc., and to see how we can work together to make something feasible.

Are there equivalents for other languages?

Yes, Italiano Tecnico Semplificato, Español Técnico Simplificado, Français rationalisé (Groupement des industries françaises aéronautiques et spatiales, GIFAS), equivalents for Japanese, Korean and other languages.

3.10 Final discussion

Khalid Choukri acted as a moderator during the final discussion. He started by summarising the various topics that were discussed in the workshop presentations: easy to read text, monomodalities, multimodalities (text, images, audio, text to speech, sign language), multilinguality (understanding text in other languages), definitions (easy language, plain language, etc.), automation, evaluation (benchmarking), cultural adaptation (e.g. Swedish), initiatives in research and in standardisation. The latter initiatives include W3C, ETSI and ISO SC 35 (user interfaces, accessibility for ICT products and services). Most of these standardisation bodies are there to make sure we come up with best practices both in designing easy to read language and integrating it in workflows.

The moderator then asked the presenters the following questions:

- *How much cultural/linguistic/community/group adaptation is needed in the area of simplification?*

  Answer from Vincent Vandeghinste: There are different kinds of user groups, but they are not easy to access. An option is getting corpora for a specific user group in order to model the group, instead of going to the user groups directly. For instance, Wablieft are experts in writing easy language news for low-literate people but not necessarily for language learners, migrants or children. If you have a corpus for such a group, you can model that culture. If not, you should go to the group directly, e.g. some groups with cognitive disabilities. There is also the need for cultural adaptation, for instance in case of refugees, or in case of sign language translation for the hard-of-hearing. The latter have reading difficulties because they do not have the sound analog. They want to be treated as a subculture.

- *There are a lot of initiatives to include easy-to-read and easy-to-understand language in regulatory requirements or legislative systems. For instance, in Canada, subtitling is mandatory by law for all the broadcasters since the last ten years. How can the research community bring language technologies like speech-to-text, machine translation, automatic subtitling and captioning to target communities?*

  Answer from Vincent Vandeghinste: Researchers need data from the target communities to train systems, but it is a very difficult task for them to get the license for the data. GDPR does not make it easier, especially when data cannot be anonymised, e.g. video or audio. Obtaining textual data is less of a problem but there is a huge fear from legal departments e.g. of broadcasting companies. The people that work with the data may be happy to share it, but the lawyers are very hesitant. Even when you get the data, the question is what you are allowed to do with it. You need the consent from all the people that contributed (e.g. the ones who created the data) and to set up a list of things you want to use the data for. This is an underestimated problem. Another issue related to language technologies is that they are not perfect. They may for instance work for 90%, but errors of combined systems add up, so accuracy goes down, which may lead output to become less instead of more understandable. It

is risky to fully automate things: not so much for subtitling a TV show, but in case of official documents, text needs to be to the point and accurate.

- *Evaluation takes place in various ways, such as performing a linguistic analysis or a task. How can we involve a group of potential users in the evaluation?*

  Answer from Daniela Zambrini: In case of STE, the evaluation is practically oriented. STE was created for aviation mechanics who do not have English as their native language. The effectiveness and simplicity of a text is measured based on the lack of serious incidents, near-accidents, or missed accidents. For any kind of accident occurring, it needs to be thoroughly analysed whether improper writing of the manual may have caused confusion. Measurement also takes place based on feedback of users to our group on potential changes to rules. Based on this feedback, we bring things forwards in a next issue.

One of the attendees added the following comment to the discussion:

We are focusing on computer models for how to make information clear, but it is also important to develop communities of people who only speak a dialect or a simplified language version. It is interesting to hear what happens in Sweden in terms of simplifying information, but what are governments doing to help migrants learn the language of the country where they will be living? There are two things to be developed, (1) governments have to simplify language, but this does not necessarily mean the information will come across (you lose information in the translation) and (2) people should be empowered, i.e. should be given the IT tools that they need to help themselves, not only to understand the information but also to be able to educate themselves and to be able to improve their language skills. Two elements are at stake here: improving the skills of the people who are going to be communicating and improving the ability of people to clearly communicate.

4. Conclusions

The workshop *Simplify Language – Capture Audience* focused on software, techniques and procedures for text simplification, for measuring readability of texts, and for implementing controlled language by restricting vocabulary and grammar. It covered objectives concerning social inclusion and better accessibility of technical information to a large audience. A very large number of people registered. A total of around 100 people participated. The audience included staff of EU administrations and Member State administrations, representatives of companies and academia, and freelancers.

The speakers had various backgrounds (academia and industry). An introduction to the workshop was provided by June Lowery-Kingston, Head of Unit "Accessibility, Multilingualism, Safer Internet" at DG CNECT. She pointed out the need for the workshop given the implications of the average literacy level across the EU on public sector communications, and the need for language technologies in this context. The speakers dealt with topics like easy language (for people with cognitive disabilities), plain language (clear writing), multimodal integration (text, images, audio, text to speech, sign language), multilinguality (understanding text in another language, support for migrants), automation of text simplification and readability assessment, evaluation of text simplification results, cultural adaptation, controlled language, and initiatives in research and standardisation.

The discussions between speakers and audience related to the use of machine translation, personalisation of systems, reusability of approaches (for other target audiences, languages and text types), cultural adaptation, the potential of automation, the characteristics of training and evaluation data, the involvement of potential users in the evaluation process, and the role of the research community.