Intellingo: An Intelligible Translation Environment

Peer-reviewed author version

COPPERS, Sven; VAN DEN BERGH, Jan; LUYTEN, Kris; CONINX, Karin; van der Lek-Ciudin, Iulianna; Vanallemeersch, Tom & Vandeghinste, Vincent (2018) Intellingo: An Intelligible Translation Environment. In: Proceedings of the ACM conference on Human Factors in Computing Systems, CHI 2018, ACM, (ART N° 524).

DOI: 10.1145/3173574.3174098
Handle: http://hdl.handle.net/1942/25605

DOI: 10.1145/3173574.3174098
Handle: http://hdl.handle.net/1942/25605

Intellingo: An Intelligible Translation Environment

Sven Coppers[1], Jan Van den Bergh[1], Kris Luyten[2], Karin Coninx[2], Iulianna van der Lek-Ciudin[3], Tom Vanallemeersch[4], Vincent Vandeghinste[4]
[1]Hasselt University – tUL – imec, Expertise Centre for Digital Media, Diepenbeek, Belgium
[2]Hasselt University – tUL – Flanders Make, Expertise Centre for Digital Media, Diepenbeek, Belgium
[3]KU Leuven, Faculty of Arts, Campus Sint-Andries, Antwerp, Belgium
[4]KU Leuven, Centre for Computational Linguistics, Leuven, Belgium
[1,2]firstname.lastname@uhasselt.be, [3,4]firstname.lastname@kuleuven.be

ABSTRACT
Translation environments offer various translation aids to support professional translators. However, translation aids typically provide only limited justification for the translation suggestions they propose. In this paper we present Intellingo, a translation environment that explores intelligibility for translation aids, to enable more sensible usage of translation suggestions. We performed a comparative study between an intelligible version and a non-intelligible version of Intellingo. The results show that although adding intelligibility does not necessarily result in significant changes to the user experience, translators can better assess translation suggestions without a negative impact on their performance. Intelligibility is preferred by translators when the additional information it conveys benefits the translation process and when this information is not part of the translator's readily available knowledge.

ACM Classification Keywords
H.5.2. User Interfaces: Graphical user interfaces (GUI); J.5 ARTS AND HUMANITIES: Linguistics

INTRODUCTION
Most professional translators nowadays use computer-assisted translation (CAT) tools, also known as translation environments [11, 35]. CAT tools integrate various translation aids that present different kinds of automated translation suggestions. We can discern three main types of translation suggestions [8]. First, translations for the whole *source segment* (sentence being translated) are provided by means of *machine translation*. In some cases, the machine translation engine can even adapt its output during translation based on the corrections that the user makes [19]. Second, lists of translations per word or word group are shown, based on a *term base* that stores terms and their meta-data. Finally, a *translation memory* stores sentences and their human translation from a specific domain. Given a source segment, the translation memory provides the user sentences that have the same or a similar vocabulary and/or grammar. Sentence similarity is expressed in a matching score, varying from 100% for exact matches to lower scores for matches which have one or more differences in grammar/vocabulary (i.e. the matches are *fuzzy*). The machine translation engine is automatically constructed from existing repositories of bi-lingual sentences within a given domain (e.g. medical terminology). Such a repository is for instance provided by a translation memory. The term base may equally be constructed (semi-)automatically from such a set. The relations between different translation aids are, however, not made visible within current CAT tools, because they are presented in multiple secluded panels. This may affect the understanding of users on why the computer makes certain suggestions, or in other words, limit the intelligibility of the tool. Intelligibility, or understandability, has been addressed in context-aware systems, which undertake some actions based on the context. These systems become much more reliable and trusted by their users when their behavior is externalized through the user interface [4, 39]. Improved intelligibility of the translation aids may enable translators to estimate the value of suggestions in a particular context more accurately.

In this paper, we present Intellingo, an intelligible translation environment. It reveals why and how suggestions are provided by the translation aids. Figure 1 provides an overview of the intelligibility features for the algorithms of Intellingo. It shows suggested words or word groups and alternatives, as well as additional metadata about where they originate from: term base, translation memory, the machine translation engine, or a combination of these resources. Examples of these words or word groups are highlighted in the fuzzy matches and/or machine translation. In contrast to existing systems, Intellingo presents all this information compactly just below the edit box. Intellingo was developed using a user-centered approach in which we designed and developed our user interface in close collaboration with professional translators and translation researchers. This enabled us to define the type of activities to be supported by a translation environment, as well as the various tools that are useful to improve both the efficiency and quality of the translation work. Before the start of the design, we conducted a survey and performed interviews and contextual inquiries. Both early and more advanced design iterations

CHI 2018, April 21–26, 2018, Montréal, QC, Canada.

ACM ISBN 978-1-4503-5620-6/18/04 ...$15.00.
https://doi.org/10.1145/3173574.3174098

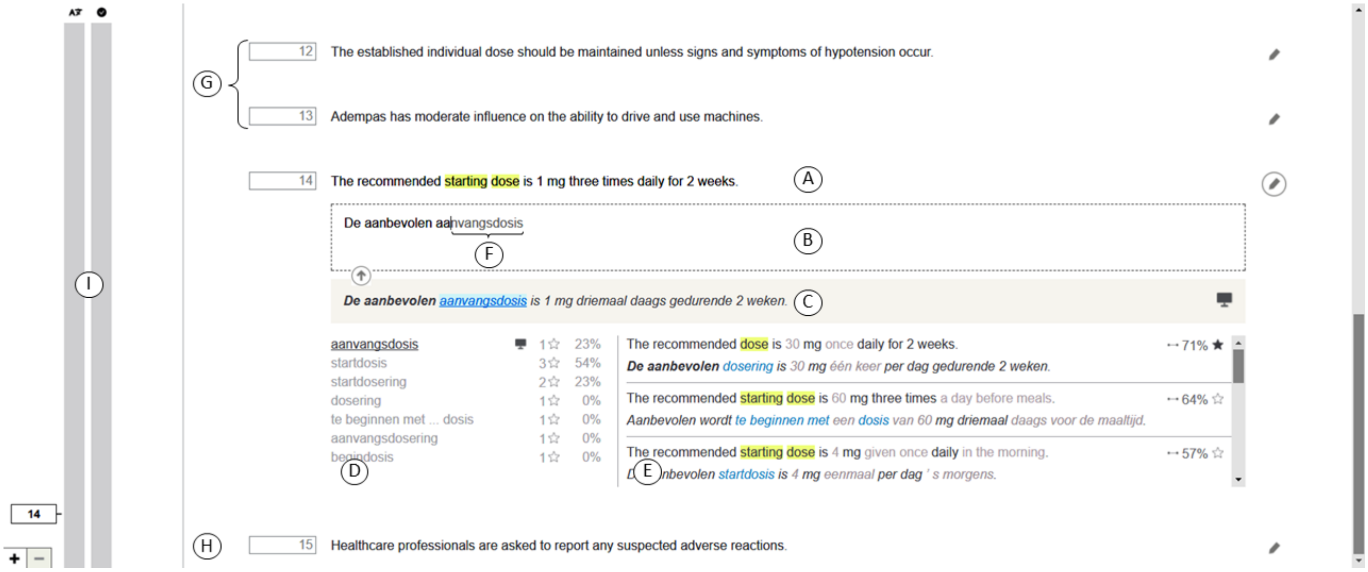

Figure 1. An overview of the Intellingo user interface when translating: Ⓐ the source segment, Ⓑ a text area for the user translation, Ⓒ machine translation, Ⓓ alternative word groups, Ⓔ related earlier translations (translation memory), Ⓕ auto-completion, Ⓖ two preceding segments, Ⓗ the next segment and Ⓘ a progress bar indication of translation and revision. Upon hovering over the translation suggestions, the related phrases are highlighted in yellow (source language) and blue (target language) across all suggestions.

were evaluated with translation experts and/or professional translators. Furthermore, we compared two versions of the interface of Intellingo that provide exactly the same functionality, with and without intelligibility. The results revealed the impact of the intelligibility features on the user experience.

RELATED WORK

We cover relevant literature for two important facets of our research. We first review related work in the domain of intelligibility. Secondly, we describe studies related to existing translation environments.

Intelligibility

Intelligent systems such as context-aware applications and recommender systems are widely used. They outperform humans on specific tasks [38] and often guide processes of human understanding and decisions [14]. However, their behaviour is rarely clear to users [1, 32]. They involve smart algorithms to reason about data that might be incomplete [9, 14]. They act like *black boxes* in the sense that they do not offer any insight into the system's behavior [39, 49]. When users have a low understanding of the system's behavior, they lose trust [20, 36, 39] and can become frustrated [3].

For decades, researchers have been aware that explaining system behavior is important for good interaction between users and complex systems [23]. In literature, this concept is inconsistently referred to as intelligibility [4, 45], transparency [16, 39, 46], explainability [20, 27, 42] and interpretability [14]. In this paper, we use the term *intelligibility*. Intelligibility allows users to verify whether the system behavior is sound [14] and to judge the appropriateness of the results [39]. Consequently, these systems become more reliable, predictable and trusted by their users [4, 14, 42], thereby giving users confidence to act on its results [27]. In the context of recommender systems, users also like recommendations more if they are intelligible [39]. Doshi-Velez et al. [14] warn that intelligibility might not be useful in applications in which wrong recommendations are acceptable (e.g. ad servers) or when trust in the system is already high (e.g. air craft collision avoidance systems). In the context of professional translation environments, mistakes in the final translation result are not acceptable and trust in machine translation without post-editing is low.

Honan et al. report that some users might try to reverse engineer the algorithm to understand it [21]. For systems to properly support intelligibility, they must make additional information available [39]. Belloti et al. state, for example, that these systems need to present what they know, how they know it and what they will do with this information [4]. The Intelligibility Toolkit considers a wider range of explanations, including certainty, what information was used as input, why a system comes up with a certain result or why it did not [12, 31]. Other authors even argue that algorithms themselves should become part of the user experience [27, 49].

Evaluating the user experience requires a broader set of measures than have been commonly used [27]. Recently, Doshi-Velez et al. [14] identify two methods to evaluate intelligibility. The first method is to compare the application with a *quantifiable proxy*, another system we know is intelligible. The second method, which we used in our research, is to evaluate the system in the intended scenario and to consider context-specific metrics. Konstan et al. highlight the lack of understanding about intelligibility in live, deployed systems and user-contributed content [27]. Our work focuses on understanding intelligibility in the context of translation environments and covers user-contributed content such as a translation memory.

Translation Environments

In the past, research in CAT tools was mainly driven by technical improvements in algorithms behind translation aids, and not from a user-centered perspective [28, 35]. More recently, several surveys and contextual inquiries have investigated the features of current translation environments as well as work practices and needs of translators. Studies revealed that experienced translators have low trust in machine translation technology [28]. Similarly, in a recent study involving 403 respondents (231 complete) by Moorkens et al. [35], only 18% of translators like to use machine translations, while 40% think machine translation technology is still problematic due to the number of errors it produces. 80% say they would like to see subsegment machine translation suggestions provided as a dropdown list. Translation memory technology, which can be useful for collaboration [24], is among the most frequently used features in the survey of Van den Bergh et al. [11]. 56% of the respondents like to use translation memory technology, of which 75% believe it helps them preserve consistency in their work and increase productivity [35]. Even when a machine translation suggestion received a higher confidence score than any fuzzy match available, 88% still want to consult the fuzzy matches. Furthermore, 90% feel it would be useful to combine subsegments from the machine translation and fuzzy matches. 62% prefer the translation environment to automatically generate such hybrid translations, provided that it is made clear where each subsegment is coming from. The same study [11] reveals that ease of use and time to learn are among the most important aspects when choosing a translation environment. Other studies reveal the importance of the translation memory metadata. Information about the source database and the translation author increases the translator's trust in the suggestions [24, 35].

We briefly review the interfaces of existing translation environments. Matecat [34, 17] puts all suggestions close to the edit field in a tabbed view. Machine translation and translation memory are combined in the same tab, only discerning the suggestions based on a textual and color-coded label. Glossary and concordance are displayed in different tabs, making it impossible to view these features at the same time. Matecat offers some form of intelligibility by showing an estimation of machine translation quality. Casmacat [7] shows similar confidence measures and features a word alignment visualization. Lilt [22], which builds on the research by Green et al. [18], focuses on providing the best translation option based on their adaptive machine translation or on a translation memory. The translation suggestion is displayed immediately below the edit field. The concordance and glossary are combined into the lexicon feature that is offered in a separate part of the interface. In contrast to these online tools, SDL Trados [37], one of the most frequently used translation environments [35], and several other commercial and open source tools, have a flexible layout. Although this allows translators to juxtapose all information, it omits the details of this information. Translators want all information to be displayed immediately when typing, without having to navigate or shift between areas [15, 29]. This is an important challenge, especially on lower screen resolutions. In summary, translation environments usually offer some form of intelligibility, by showing matching scores, for example. However, the inner logic of the algorithms is rarely shown, and there is plenty of potential to enhance their intelligibility.

DESIGN PROCESS

The user interface of Intellingo was designed and developed using a user-centered approach for which we collaborated closely with professional translators and translation experts. In preparation for this process, we conducted a literature study concerning translation practices, translator's needs and state-of-the-art translation tools from both industry and academia, which we reported in the related work section.

First, we identified the needs of professional translators using a web-based survey among language professionals (181 respondents, 72.38% freelance translators, 24.31% in-house translators) to get an update on the usage of CAT tools, what training is involved to use them, and what features are considered important. Translators most commonly used SDL Trados [37], followed by memoQ [25], CafeTran [10] and XTM International [48], and were mostly self-trained. Almost all respondents used tools with support for translation memories. To obtain a better understanding, we complemented this survey with semi-structured interviews and contextual inquiries with translators (5 in-house and 4 freelancers). Ease of use was the most important motivation for choosing a CAT tool, closely followed by speed of performance and features such as management of translation memories and terminology. Translators heavily rely on keyboard shortcuts and rarely customize the user interface of their CAT tool, which highlights the importance of sensible defaults. A thorough overview of the survey and interview is reported in [11], including a comparison with the results of other surveys [2, 13, 29].

Next, high-fidelity mockups were designed for the user interface of Intellingo in roughly four iterations. The multi-disciplinary team driving this design consisted of a professional graphical designer, HCI researchers, computer scientists, and translation researchers. An analysis of the underlying mechanisms and translation algorithms was performed, after which the HCI and translation researchers selected features that would potentially benefit from additional information provisioning (intelligibility). Each iteration was presented by means of demonstrations and walkthroughs to a user group, composed form various companies involved in translation work. Their feedback was used to improve the mockups and to refine the focus of the project. Multiple layouts were designed in parallel, ranging from Figure 2.Ⓐ to Figure 2.Ⓑ).

Finally, we implemented a functional prototype and performed two rounds of evaluation with end-users. The design of the interface further evolved significantly throughout this phase, as illustrated by the difference between the two mockups (Figure 2.Ⓐ and Figure 2.Ⓑ) and a screenshot of the prototype (Figure 2.Ⓒ). The first iteration of the prototype focused on ease of use and integration of various translation technologies such as machine translation, translation memories and terminology. To enhance the intelligibility of the prototype, we designed and implemented visualizations that present the

Figure 2. Artefacts created during the user-centered design process show substantial improvements in the Intellingo user interface: Ⓐ + Ⓑ static, detailed designs that are the result of a strong collaboration between experts in a multidisciplinary team. Ⓒ Screenshot of the high-fidelity prototype that was evaluated in the first round of evaluations by 8 professional translators.

relationships between different kinds of translation suggestions, by including context and statistical information and by highlighting related information (also know as brushing [6]). From a formative evaluation with 8 professional translators, we learned that intelligibility can influence the perceived usefulness of translation suggestions. For example, matches that would have been ignored due to their low matching scores, were perceived as useful because the interface emphasized parts that could be reused in the final translation (P5: "*I could clearly see for which parts the match could be useful*", P7: "*You can see right away if a fuzzy match is close to the sentence you need to translate, and whether it is useful to you or not*"). P3 and P8 participants even explicitly expressed the need to combine these parts with parts of the machine translation. However, not all intelligible visualizations were appreciated. The morphological function of the suggested alternatives (e.g. V-PAPA[1]) is considered as distracting and annoying (P3). We found this is an essential aspect of deciding whether to add intelligibility or not: an intelligible visualization is only perceived as useful when the information it conveys benefits the translation process, and when this information is not part of the translator's readily available knowledge. We used these insights to improve the implementation. 26 other professional translators were recruited for the evaluation of the second iteration, which is reported in the remainder of this paper.

INTELLINGO

Intellingo is an online translation environment targeted towards professional translators and translation experts. Four different translation aids are included in Intellingo, as well as visual explanations of the algorithms behind them: (1) a list of matches from a translation memory, (2) suggestions from a hybrid machine translation engine, (3) a list of possible alternatives for the selected term and (4) an auto-completion feature that predicts the rest of a word (or word group).

Fuzzy Matches

Fuzzy matches are segments in a translation memory that are similar to the source segment. We developed a web service that allows for requesting fuzzy matches in the publicly available EMEA translation memory [41], given one or more similarity metrics. Many metrics for similarity exist, for instance based on edit distance [30] or on the automatic linguistic analysis of a sentence (syntax tree) [44]. In Intellingo, the metric being applied is shown through an appropriate icon: ↔ or ⅄ respectively, alongside the match itself (Figure 3.Ⓖ). The parts that are potentially useful, according to the matching algorithms, are emphasized in Intellingo (Figure 3.Ⓓ). Note that some other interfaces, such as Matecat [34], emphasize differences instead of similarities (Figure 3.Ⓗ). These visualizations are designed to help the translator to quickly understand *why* a fuzzy match was found and *what* parts are potentially useful.

Hybrid Machine Translation

In a machine translation system, the source segment is translated from scratch. An established toolkit for building machine translation systems is Moses [26]. Based on a large set of sentences and their translation (a parallel corpus), Moses automatically extracts word groups and their potential translations, and builds a *language model* in order to be able to create fluent translations. We integrated the Moses machine translation system as a web service. Similar to Lilt [18], the machine translation is shown very close to the source segment and directly available to the translator at any time (Figure 1.Ⓒ).

Consistent with a study of Moorkens et al., our testers of an older prototype expressed the need for a feature that combines words from machine translation and matches from the translation memory into combined suggestions [35]. We have implemented this feature in Intellingo by automatically detecting parts in the source segment that are already translated in the fuzzy matches and treating these parts as *pre-translations*. To make this functionality clear to the translator, pre-translations are printed in bold in both the machine translation and the fuzzy matches they originate from. Icons are used to discriminate matches with pre-translations (★) from matches without any (☆). This feature explains *how* the machine translation was constructed, as requested by translators [35].

Translation Alternatives

For the word or word group being translated, a list of translation alternatives is shown (Figure 3.Ⓐ). The alternatives are aggregated from various sources, such as the machine translation (🖥), the fuzzy matches (☆ or ★) and from a term base generated that was generated from EMEA using TExSIS [33]. For each source, the associated metric is shown to clarify how often a word or word group occurred in that source (Figure

[1]Past participle of verb

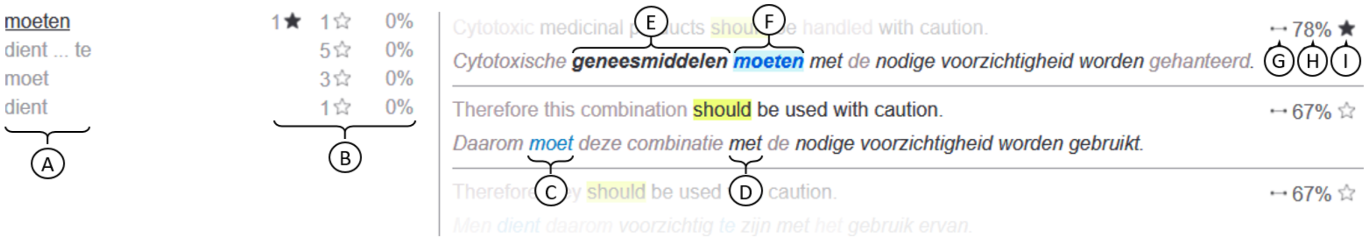

Figure 3. For the word (group) being translated: Ⓐ a list of alternatives and Ⓑ an overview of its occurrences in machine translation, translation memory and term base to explain the ranking. Words in the fuzzy matches are shown Ⓓ darker when they match with the source segment. Occurrences of all alternatives are highlighted in blue Ⓒ, and the corresponding source words are highlighted in yellow. The occurrences of the translation option selected by the user are emphasized Ⓕ. When parts of a match were used by the machine translation algorithm, these parts are shown in bold Ⓔ and Ⓘ a filled star is shown. Other intelligible details include Ⓖ the matching metric and Ⓗ the match score.

3.Ⓑ). For alternatives originating from the machine translation or the translation memory, an absolute value is shown. For alternatives from the term base, a relative frequency is included. Note that this summary also clarifies the order in which the alternatives are sorted: alternatives from the hybrid machine translation are more important than alternatives in the translation memory. Alternatives in the term base are the least important. This order focuses on intelligibility: a term that is found in the term base may also occur in matches from the translation memory. When a term occurs often within the matches, it can become a pre-translation in the hybrid machine translation. This order assumes the machine translation engine, the match engine and the term base are all using the same parallel-corpus. The occurrences of these alternatives are highlighted in blue in the fuzzy matches (Figure 3.Ⓒ). The occurrences of the selected translation option are additionally emphasized (Figure 3.Ⓕ). When the first emphasized occurrence is not within the viewport, the overview of matches automatically scrolls.

Intellingo explains *where* translation alternatives come from, in *what* context(s) they have been used before and *how often* they have been used before. As a result, translators can make quick and well-informed decisions on the suitability of multiple alternatives in a particular translation context.

Auto-completion

The need to efficiently combine subsegments from both the machine translation and fuzzy matches arose from both our user-centered approach and from the work of Moorkens et al. [35]. It is addressed in Intellingo by offering a hybrid machine translation and by implementing a version of auto-completion that considers both the hybrid machine translation and the fuzzy matches. To improve the typing speed, Intellingo suggests the remainder of a word or word group (Figure 1.Ⓕ). By pressing `ENTER`, the translator can add this suggestion to the translation, which will subsequently cause a new suggestion for the next word to be predicted.

The auto-completion algorithm will suggest terms from the current translation context, including the machine translation, fuzzy matches and corresponding alternatives from the term base. The algorithm justifies its prediction in order to be more intelligible, as recommended by Sinha et al. [39]. The origin of the prediction is made clear by programmatically selecting the alternative in the list of alternatives (Figure 3.Ⓐ). Since this list uses the same metrics and priorities as the ones used in the auto-completion algorithm, it effectively explains what alternatives were considered and why the prediction was chosen, without requiring new UI elements. In addition, the overview of fuzzy matches will automatically scroll to the first match that contains this suggestion in order to illustrate how it was used previously. The algorithm prefers terms that are already shown in the list of alternatives for more stable predictions and to decrease distraction. Alternatives can be selected by clicking them or by "navigating" through them with the keyboard shortcuts (`SHIFT+ARROWS`). This improves the user control over the auto-completion algorithm, as recommended by Tintarev et al. [43].

Our intelligible implementation of auto-completion shows *where* a suggestion comes from, *how* that suggestion was generated, *what* the alternatives were and *why* a suggestion was predicted instead of the alternatives.

Relationships between translation suggestions

The Intellingo user interface clarifies relationships between the different types of recommendations. The relationships are highlighted with visual annotations: Figure 1 includes these highlights in yellow, blue and gray. A translator can explore up to four different relationships simultaneously: (1) the relationship between words and word groups in the source segment, (2) words that are translations for each other, (3) synonyms, and (4) usage examples of these synonyms within the matches. All translation aids require only limited space and can be combined into a compact recommendation overview. This way, all relevant information is available to the translator close to the source segment. This minimizes the visual focus shifts required in traditional translation environments, as desired by translators [15, 29]. It ensures the translator can use translation aids as part of the ongoing translation activities, instead of separating navigation and exploration of recommendations from doing the actual translation. Multiple preceding (Figure 1.Ⓖ) and following source segments (Figure 1.Ⓗ) remain visible to provide additional context to the translator.

WALKTHROUGH

This section contains an example scenario of how the features described in the previous section fit into the translation workflow of professional translators and illustrates how our contributions can be useful to them.

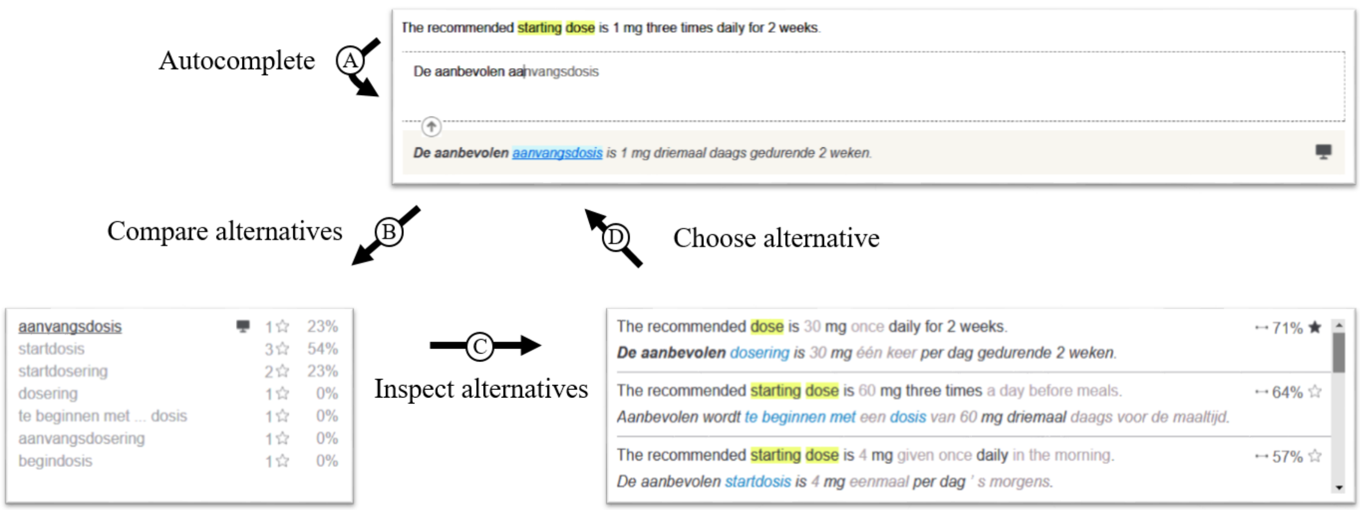

Figure 4. A possible workflow in which several translation aids can be combined. Ⓐ The translator types and auto-completion generates new predictions. Ⓑ The translator consults the list of translation alternatives to understand why the prediction was generated. Ⓒ The translator selects an alternative and inspects its occurrences in the matches. Ⓓ The translator decides which alternative to use and adds it to the translation.

Eva is a professional freelance translator who gets paid per translated word. A client sends her a request to translate a medical document of 10 pages from English to Dutch. Eva opens the source text with Intellingo (Figure 1), loads a medical term base and translation memory and starts translating.

When translating the first source segment, Eva analyses a few matches from the translation memory in which potentially relevant parts are emphasized in dark. The low number of emphasized words together with the match percentages help Eva to take a quick decision about the translation strategy: she decides to translate this source segment from scratch instead of post-editing one of the matches. She can see that the auto-completion suggestion (Figure 4.Ⓐ) is part of the machine translation and it also occurs in three matches (Figure 4.Ⓑ). From this information, she infers that her colleagues often use the same translation and she decides to confirm this prediction by pressing `ENTER`. Thereupon the auto-completion algorithm predicts a suggestion for the next word. Eva believes this prediction fits very well, since the highlighting shows her that the prediction often appears with the previous word in the matches.

The next prediction comes from the machine translation, but it does not completely fit the rest of the sentence. In the list of alternatives for this prediction, Eva considers using the second alternative because it occurs in two of the matches (Figure 4.Ⓑ). As soon as she clicks on the second alternative, the overview of matches automatically scrolls to the first match in which the alternative occurs (Figure 4.Ⓒ). This match has a relatively high score and Eva decides to add the alternative to her translation (Figure 4.Ⓓ). By pressing `ENTER`, she also adds the next three predictions because they occurred together in the match.

Next she reaches a prediction that occurs both in the machine translation and in two other matches. All occurrences are shown in bold (Figure 3.Ⓔ). Eva understands that these parts in the fuzzy matches were considered as pre-translations in the machine translation engine. This helps her to quickly determine that this part of the machine translation is correct and can be added to her translation. When the sentence is finished, Eva confirms her translation by pressing `CTRL+ENTER`.

EVALUATION

In our study, we compare the intelligible version of Intellingo with a version that does not contain the intelligible visualizations. Two groups of thirteen (13) professional translators were asked to translate the same texts with the two different versions of the interface in a within-subject experimental design. The aim of the study is to measure the impact of intelligibility on the perceived value, trust, usage, performance and user satisfaction for different translation aids. We want to learn whether intelligibility can improve the understanding of translations aids and the trust professional translators have in them. Three hypotheses are postulated and drive the analysis of the study results:
H1. In Intellingo, the intelligible translation interface improves the user experience of both the overall interface and the translation suggestions;
H2. In Intellingo, the intelligible translation interface ensures more appropriate and conscious use of suggestions provided by translation aids;
H3. In Intellingo, translators prefer intelligible translation aids above their *simple* counterparts.

Participants

Twenty-six (26) professional translators and translation experts were recruited to participate in our study. Nine participants use SDL Trados [37] for their daily translation activities and four use MemoQ [25]. Other translation environments that are often used include Across [40], Cafetran [10] and Wordfast [47]. On average, participants have 7.7 years of professional experience ($SD = 9.4$). They are mostly hired for translating technical, political, medical, administrative or legal texts. Translators focus on speed, because they are usually

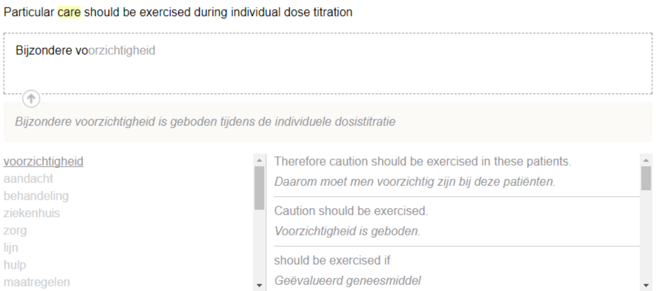

Figure 5. The simple version of the translation environment has all intelligibility features disabled.

	Simple	Intelligible	Trend
Time	71.3 (s)	72.5 (s)	+1.7%
Clicking words in translation memory	0.46	0.64	+39.4%
Clicking alternatives	0.08	0.09	+12.5%
Navigation shortcuts for translation memory	6.08	6.37	+4.8%
Navigation shortcuts for alternatives	0.22	0.19	-11.7%
Accept auto-completion/selected word	6.36	6.46	+1.6%
Hovering words in source segment	1.68	1.42	-15.6%
Hovering words in machine translation	4.60	4.42	-3.8%
Hovering source words in translation memory	2.98	2.31	(*) -22.5%
Hovering target words in translation memory	3.60	3.45	-4.1%
Copy machine translation (ALT+M)	0.19	0.21	+8.1%
Backspaces and characters typed	48.05	47.97	-0.18%

Table 1. Average usage of the Intellingo UI elements per source segment. Only trends indicated with a star (*) are statistically significant.

paid per word. Participants received a voucher worth €20 as a token of our gratitude for participating in the study. This is well below the rate they charge for their translation work, ensuring participation in the study is voluntary and unbiased.

Apparatus

Participants use two different versions of the Intellingo translation environment: a "simple" version stripped from the intelligible visualizations (**S**), depicted in Figure 5, and the intelligible version (**I**) as described in the previous section. The overall layout and the behaviour of both versions is identical to eliminate the influence of differences in functionality or structure of the user interface. The suggestions coming from various translation aids are exactly the same in both versions. The difference between both versions is that all intelligible visualizations we discussed in the previous section are disabled in the simple version. Both versions run in a browser.

We use an online remote usability testing tool[2] for our study, which allows us to capture the screen of the user, record a video of the user during the test and survey the user after each session of the study. Furthermore, the tool can be used in the home environment, where most of the professional translators also perform their daily work activities. All interaction events are logged when they use Intellingo. These add up to 24 different types of interaction, such as hovering words in the translation memory, typing, and using keyboard shortcuts.

In total, three English texts needed to be translated to Dutch by each participant. The texts are prepared by selecting sentences containing between 50 and 150 characters from the Adempas leaflet[3]. Next, the order of the sentences is randomized to ensure the resulting texts have similar characteristics. The first five sentences are used as a *training text*. The following fifteen sentences end up in *text 1* and another group of fifteen sentences are used for *text 2*.

Procedure

A schematic overview of the procedure is shown in Figure 6. Participants took part in the experiment through a web page. First they are briefed about the purpose of the experiment and the reward they will receive upon completing the study. Next, participants are guided through the necessary steps to set up the recording of videos of the experiment. These steps mainly concern requesting permission to use the webcam and selecting a screen to share. The experiment starts with a short survey about the demography (DEM). Next, participants are instructed on how to use the simple version of the translation environment for translating 5 sentences.

During the actual study that follows, the participants need to translate two texts, each containing fifteen sentences (source segments). Both texts are translated with a different version of the interface. The order of the texts is the same for all participants. The order of the interface versions is randomized to eliminate training effects. After translating each text, the System Usability Scale (SUS) [5] is used to rate the usability of each interface. Next, the participants fill in a short questionnaire about the user experience (FEA). At the end of the experiment, they are asked to compare both interfaces in terms of translation time, quality and user experience (COM).

Results

The median of the SUS score increased from $SUS = 60.0$ (fair-good) in the simple version to $SUS = 67.5$ (good) for the intelligible version. Figure 7 reveals that participants tend to agree more to positive statements about the intelligible version than to such statements about the simple version, although the difference is not significant. Similarly, they agree less to negative statements. Displaying more information to enhance intelligibility had a slightly positive influence on the general user experience and does not require more training. Participants remain mostly neutral when asked which interface was better regarding criteria such as enjoyability, trustworthiness and efficiency (Figure 8). Surprisingly, participants believe the simple version helped them the most to improve quality, even though the suggestions of the translation aids were identical and both interfaces were perceived as non-distracting.

Table 1 shows a summary of statistics about the average usage of the UI elements in Intellingo when translating a single source segment. A Wilcoxon Signed-rank test shows that intelligibility significantly lowered the amount of mouse hovers over words in the translation memory in the source language ($Z = -2.12, p < 0.05, r = 0.29$). There is a similar trend for hovering words in the source segment, in the machine translation and in the translation memory in the target language, although not significant. This suggests that participants needed less manual effort to explore these words, and instead relied on the highlighting visualization to discover potentially useful words. After using both versions of the

[2] www.uxpro.be
[3] www.medicines.org.uk/emcmobile/PIL.28743.latest.pdf

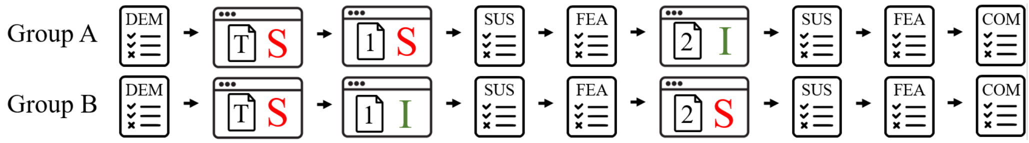

Figure 6. All participants start with a short demography survey (DEM) and a training phase in which they translate the training text using the simple version of the interface. The actual study has a split-plot design: participants in group A used the simple version (S) of the interface to translate the first text and used the intelligible version (I) for the second text. Participants in group B translated the texts in the same order, with the opposite versions of the interface. After translating each text, a SUS test and an intermediate survey (FEA) are taken. A Closing survey (COM) finishes the study.

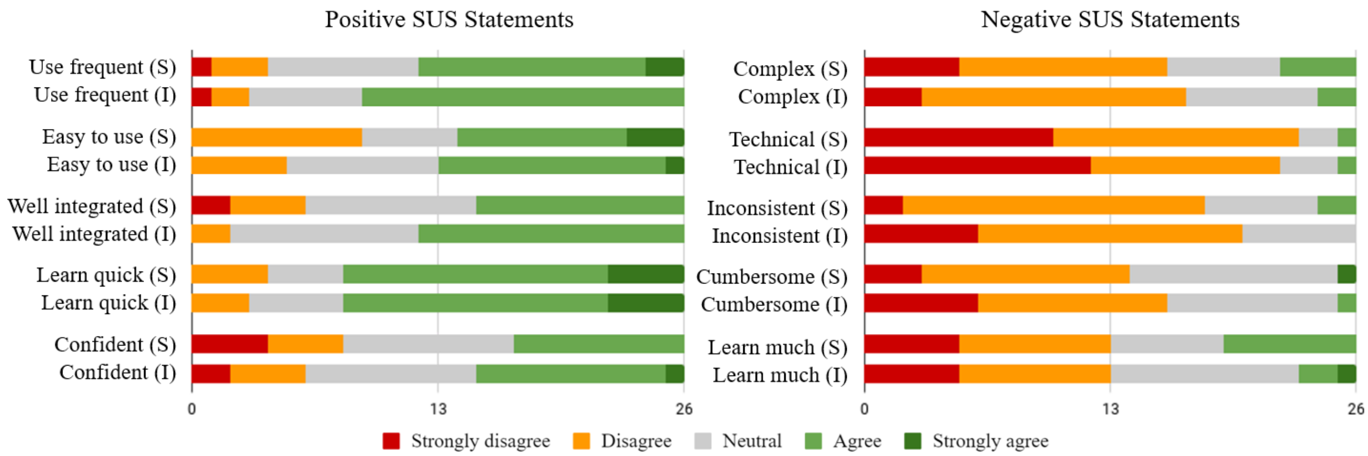

Figure 7. Level of agreement with statements about the simple interface (S) and the intelligible interface (I) in the SUS test. Even though there are no significant differences, the agreement leans more towards increments in the intelligible version for positive statements (left) and decrements in the intelligible version for negative statements (right).

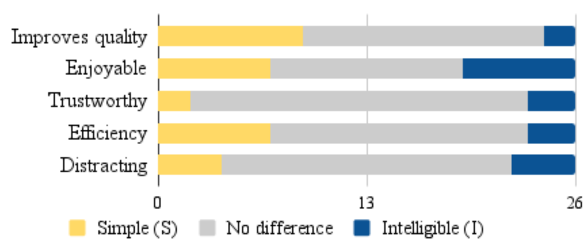

Figure 8. Answers when asked which version of the interface was better regarding criteria such as enjoyability, trustworthiness and efficiency.

interface, P16 remarked "*the [matches] from the translation memory in the [intelligible] interface appeared to be of lesser quality and thus less helpful*", while in fact these matches have the same quality. In this case, the low matching percentages may have negatively impacted the perception of quality. The correlation between clicking a word in a match and clicking again to add it to the translation increased from moderate (Pearson's $r(24) = 0.47, p < 0.05$) to strong (Pearson's $r(24) = 0.85, p < 0.001$). This data suggests intelligibility helps participants to be more efficient and confident in discovering useful words in matches. The time to translate did not differ significantly. It is remarkable to see how often translators used our new keyboard shortcuts. On average, translators used the `ALT+M` keyboard shortcut every five source segments to copy the suggestion of the machine translation engine to the editing field. More than six times per source segment, translators used the `ALT+ARROWS` shortcuts to navigate between words in the translation memory.

After translating each text, the user experience of the translation aids was evaluated using Likert scales. All answers are presented in Figure 9. It can be seen that none of the translation aids was perceived as distracting, even when more information is shown to enhance intelligibility. Auto-completion was the most appreciated feature. After having used only the simple version (**S**), P15 suggested a feature: "*It might be interesting to see the prediction automatically highlighted in a sentence / translation memory, if present*". She expressed the need for an intelligible visualization that is actually present in the intelligible version (**I**) of the interface. For both auto-completion and machine translation, participants were noticeably less neutral about the impact of these features on quality when the intelligible version is presented. This suggests intelligibility can improve the awareness of the quality, which is useful for both high and low quality suggestions. Most participants expressed trust in the translation aids. Machine translation technology is the least trusted, which is consistent with other literature [28]. The intelligible version gave them more reasons to accept or reject the proposed translation suggestions. Auto-completion makes the machine translation suggestions more quickly accessible.

Figure 10 presents the level of agreement with Likert scale questions about features that were only present in the intelligible version. None of the intelligible features was perceived as distracting. Enhancements to the representation of matches were the most appreciated intelligible feature. P2 and P3 especially liked the matching percentage, which is already available in some existing translation environments. P6, who used the intelligible interface (**I**) first, became frustrated when us-

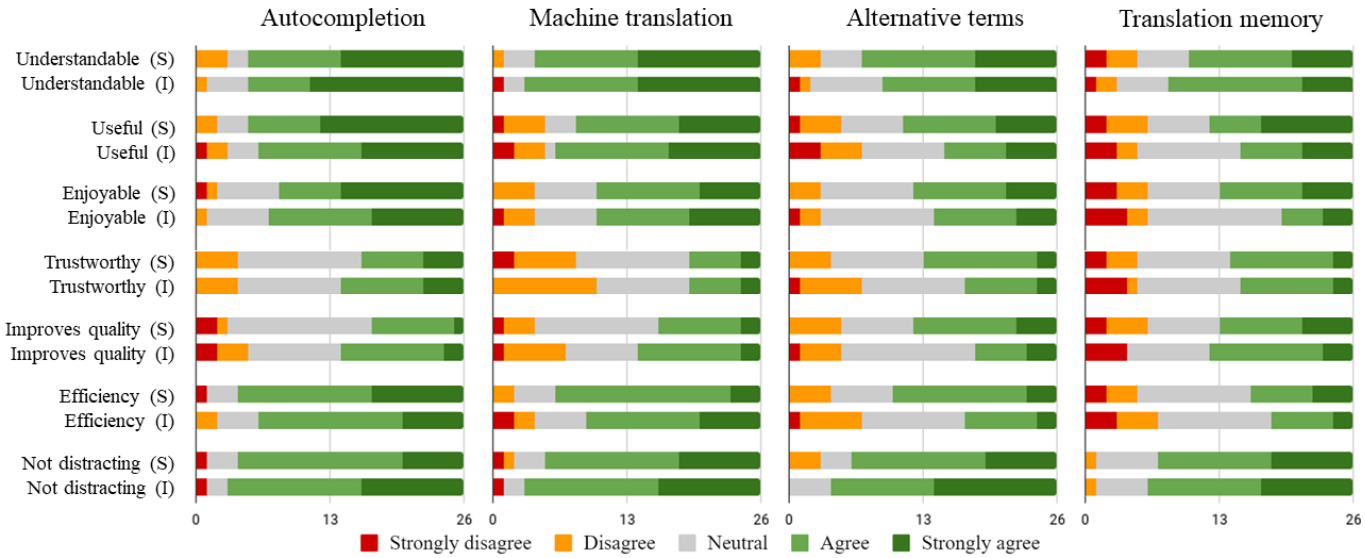

Figure 9. Level of agreement with Likert scale questions about the perception of features that were available in both the simple version (S) and the intelligible version (I) of the interface. Intelligibility helps translators to determine whether the results of the machine translation engine and auto-completion can improve the quality of their translation.

ing the simple version (**S**) and explicitly complained about the lack of background information in the simple version. There were mixed reactions to displaying metrics next to alternative terms and highlighting them in the matches. One translator pointed out that "*too much information becomes an obstacle*" (P10). P3, P15 and P20 stated that the feature can be distracting for basic vocabulary, but that it is particularly useful for difficult terms and sentences. P5 explained his usage pattern: "*I just translated and let the system inspire me for translations, while judging about the quality of suggested words myself*". In order to inform the translator about how a suggested term can be used, the occurrences of that term are highlighted in the different matches in the translation memory. P14 explicitly remarked he enjoyed using this feature. Participants were the least positive about visualizing pre-translations. For instance, P10 commented "*I don't think it is useful to know exactly how the machine translation was made; a translator can judge the quality of a machine translation by him/herself*".

DISCUSSION

Our study on displaying more contextual information to enhance the intelligibility of existing translation aids shows a slight improvement in the appreciation of these translation aids in terms of perceived usefulness and general usability as measured by the SUS scoring method. However, our data does not show significant improvements in other parameters of the user experience and we cannot accept hypothesis **H1** (*In Intellingo, the intelligible translation interface improves the user experience of both the overall interface and the translation suggestions*). Nor does intelligibility have a negative impact: translators reported similar amounts of training required to use both interfaces, even though more information is presented in the intelligible version. Some translators remarked they were not aware of any differences in the interfaces while others indicated they did not always understand how to use the intelligibility features, which were never explained to them. From this we deduce that intelligibility does not necessarily increase the amount of training required to use a specific feature. Translators can choose to ignore the additional information available. When interested, they can spend more time to learn how to interpret that information. Strong correlations suggest a proper understanding of the interface will in turn result in a more enjoyable and confident user experience. We can not make claims on whether translators are more likely to interact with more intelligible suggestions over *simple* suggestions.

Intelligibility triggered less neutral scores on the perceived contribution to quality of the suggestions provided by the machine translation engine and by the auto-completion feature. Translators were equally critical about both the high and low quality suggestions. Even matches with lower matching scores become worth exploring in the intelligible interface, since the matching words are revealed. Such lower matches would otherwise be ignored completely. We conclude that the translators gained a better insight into the usefulness of the suggestions. We confirm hypothesis **H2** (*In Intellingo, the intelligible translation interface ensures more appropriate and conscious use of suggestions provided by translation aids*). The quality of the suggestions is of primary importance. Therefore, translators appreciate being better informed on why a suggestion was made, and what the origin of the suggestion was.

We see conflicting results regarding the preferences of the translators themselves. On the one hand, translators remain mainly neutral about which interface was the most enjoyable and trustworthy. They indicate the simple version helps them to improve the quality of the translations. We believe this is because professional translators develop habits over time to dealing with translation suggestions of varying quality. Some participants have a negative attitude towards translation aids and intelligibility in general: "*[Machine translation] is always slower than a translator's brain...*" (P19) and "*we don't think the exact numerical value is very useful to a translator, who*

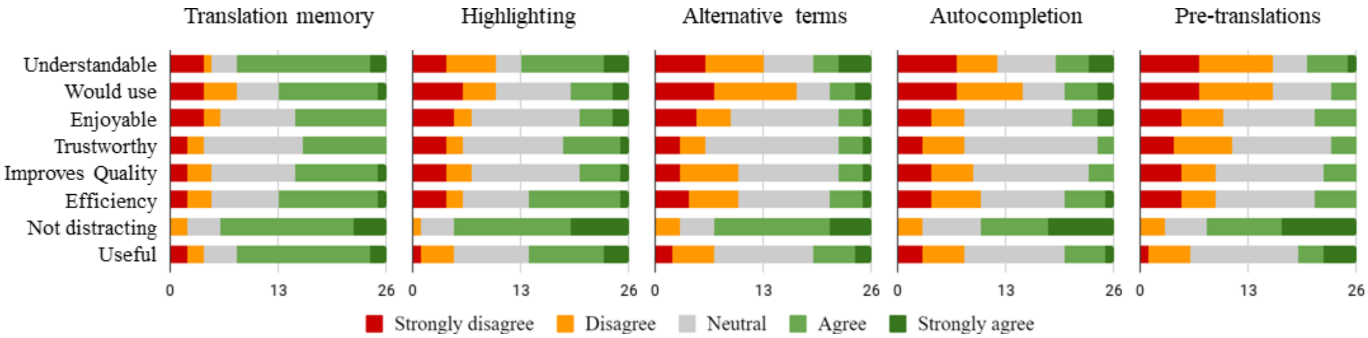

Figure 10. Level of agreement with Likert scale questions about features that were only present in the intelligible version (I) to improve the user experience. Visualizing which words from the translation memory match was perceived as the most useful. Translators were neutral about highlighting and negative about the explanations of pre-translations and the auto-completion algorithm. None of the intelligible features was perceived as distracting.

has to trust her/his own judgment anyway. That is precisely the added value of a human translator" (P7). On the other hand, participants react very positively to intelligibility features such as highlighting usage examples of a term and emphasizing matching similarities, which resembles highlighting differences, a feature found in some other translation environments as well. Some translators mention they were already familiar with matching scores. Many participants indicate they would need explanations and/or more time to get used to intelligible translation suggestions. Some participants who used the intelligible version first, complain about missing intelligibility in the simple version, such as the widely known matching scores. As a result, we can not confirm **H3** (*In Intellingo, translators prefer intelligible translation aids above their simple counterparts*). Translators only prefer intelligible translation aids when the additional information that is conveyed benefits the translation process, and when this information is not part of the translator's readily available knowledge. Highlighting is a great example of how intelligibility can be achieved by using existing UI elements instead of creating new ones that require more space and increase complexity.

One limitation of the experimental design is that participants were not informed about the intelligibility features in order to prevent bias. Not all visualizations were self-explaining enough to the point where they are readily understandable. It is possible that preferences change when users understand better how all intelligibility visualizations can be interpreted. The duration of this study was too short for several translators to include new intelligibility features in their workflow. In that sense, the neutral to positive levels of appreciation reported in this study are a good start. In our tests, we considered two extremes: one version with all intelligibility features enabled, and one without. The results show that not all of the additional information shown is useful to all translators. Because of the diverse opinions on this topic, we advise developers of professional translation environments to make the visibility of intelligible information configurable.

CONCLUSIONS

In this paper we presented Intellingo, an intelligible translation environment for professional translation work that includes various translation aids, such as term bases, translation memory and machine translation. In contrast with traditional translation environments, the user interface of Intellingo presents translation aids in a more intelligible manner in order to allow translators to use them in a more efficient way. It provides additional context about the quality and source of the suggestions, as well as subtle explanations why suggestions were made. Furthermore, Intellingo highlights how translation suggestions are related to the context of the ongoing translation.

To investigate the impact of intelligible translation aids on the translation process, a user study was performed with professional translators. Participants were positive to very positive in their comments about both the *simple* version and the intelligible version of Intellingo. The results show that intelligibility does help professional translators to assess the quality of the generated suggestions and to understand how these suggestions can be used in translation, without distracting them or negatively impacting their efficiency. Surprisingly, the study showed that adding intelligibility in the user interface design did not result in significant changes to the user experience.

Intelligible design of translation aids does not affect the quality of the translation suggestions themselves. However, the intelligible features inform translators quickly about the quality and context of the suggestions to support better decision making. Translators only prefer intelligible translation aids when the additional information they convey benefits the translation process, and when this information is not part of the translators readily available knowledge. Usage of intelligibility in design needs to be carefully balanced: providing more information and context might lead to a decrease in efficiency and a potential information overload. We showed various possibilities to add intelligibility to professional translation tools, and explored their impact on the user experience.

ACKNOWLEDGEMENTS

The SCATE (Smart Computer-Aided Translation Environment) project IWT 130041 is funded by the Flemish Institute for the Promotion of the Scientific-Technological Research in the Industry (IWT Vlaanderen). We would like to thank all participants who took part in our study for their time and for the feedback they provided. Special thanks to Karel Robert for the graphical design of the Intellingo user interface, and to Gustavo Rovelo Ruiz for his help in analyzing the data of the user studies.

REFERENCES

1. Stavros Antifakos, Nicky Kern, Bernt Schiele, and Adrian Schwaninger. 2005. Towards Improving Trust in Context-aware Systems by Displaying System Confidence. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices &Amp; Services (MobileHCI '05)*. ACM, New York, NY, USA, 9–14. DOI: http://dx.doi.org/10.1145/1085777.1085780
2. Edmund K. Asare. 2011. *An Ethnographic Study of the Use of Translation Tools in a Translation Agency: Implications for Translation Tool Design*. Ph.D. Dissertation. Kent State University.
3. Louise Barkhuus and Anind Dey. 2003. Is Context-Aware Computing Taking Control away from the User? Three Levels of Interactivity Examined. In *Proceedings of UbiComp 2003: Ubiquitous Computing: 5th International Conference, Seattle, WA, USA, October 12-15, 2003.* Springer Berlin Heidelberg, Berlin, Heidelberg, 149–156. https://doi.org/10.1007/978-3-540-39653-6_12 DOI: 10.1007/978-3-540-39653-6_12.
4. Victoria Bellotti and Keith Edwards. 2001. Intelligibility and Accountability: Human Considerations in Context-Aware Systems. *Human-Computer Interaction* 16, 2 (Dec. 2001), 193–212. DOI: http://dx.doi.org/10.1207/S15327051HCI16234_05
5. John Brooke. 1996. SUS: A quick and dirty usability scale. (1996).
6. Andreas Buja, John Alan McDonald, John Michalak, and Werner Stuetzle. 1991. Interactive Data Visualization Using Focusing and Linking. In *Proceedings of the 2Nd Conference on Visualization '91 (VIS '91)*. IEEE Computer Society Press, Los Alamitos, CA, USA, 156–163. http://dl.acm.org/citation.cfm?id=949607.949633
7. Casmacat. 2016. Casmacat. (11 November 2016). Retrieved January 3, 2018 from http://www.casmacat.eu/.
8. Chan Sin-wai. 2004. *A dictionary of Translation and Technology*. Hong Kong: Chinese University Press.
9. Keith Cheverst, Nigel Davies, Keith Mitchell, and Christos Efstratiou. 2001. Using Context as a Crystal Ball: Rewards and Pitfalls. *Personal and Ubiquitous Computing* 5, 1 (Feb. 2001), 8–11. DOI: http://dx.doi.org/10.1007/s007790170020
10. Collaborative Translation Networks, LLC. 2018. CafeTran Espresso. (2018). Retrieved January 3, 2018 from https://www.cafetran.com/.
11. Jan Van den Bergh, Eva Geurts, Donald Degraen, Mieke Haesen, Iulianna van der Lek-Ciudin, and Karin Coninx. 2015. Recommendations for Translation Environments to Improve Translators' Workflows. In *Translating and the Computer 37*. Asling.
12. Anind K. Dey and Alan Newberger. 2009. Support for Context-aware Intelligibility and Control. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 859–868. DOI: http://dx.doi.org/10.1145/1518701.1518832
13. Gavin Doherty, Nikiforos Karamanis, and Saturnino Luz. 2012. Collaboration in Translation: The Impact of Increased Reach on Cross-organisational Work. *Computer Supported Cooperative Work* 21, 6 (2012), 525–554. DOI: http://dx.doi.org/10.1007/s10606-012-9175-1
14. F. Doshi-Velez and B. Kim. 2017. Towards A Rigorous Science of Interpretable Machine Learning. *ArXiv e-prints* (Feb. 2017).
15. Maureen Ehrensberger-Dow and Gary Massey. 2014. Cognitive ergonomic issues in professional translation. *The development of translation competence: Theories and methodologies from psycholinguistics and cognitive science* (2014), 58–86.
16. Michael D. Ekstrand, F. Maxwell Harper, Martijn C. Willemsen, and Joseph A. Konstan. 2014. User Perception of Differences in Recommender Algorithms. In *Proceedings of the 8th ACM Conference on Recommender Systems (RecSys '14)*. ACM, New York, NY, USA, 161–168. DOI: http://dx.doi.org/10.1145/2645710.2645737
17. Marcello Federico, Nicola Bertoldi, Mauro Cettolo, Matteo Negri, Marco Turchi, Marco Trombetti, Alessandro Cattelan, Antonio Farina, Domenico Lupinetti, Andrea Martines, and others. 2014. The MateCat tool. In *Proceedings of COLING*. 129–132. Extended abstract on free online tool for analysis of effect of MT on post-editing.
18. Spence Green, Jeffrey Heer, and Christopher D. Manning. 2013. The Efficacy of Human Post-editing for Language Translation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 439–448. DOI: http://dx.doi.org/10.1145/2470654.2470718
19. Spence Green, Jeffrey Heer, and Christopher D. Manning. 2015. Natural Language Translation at the Intersection of AI and HCI. *Commun. ACM* 58, 9 (Aug. 2015), 46–53. DOI:http://dx.doi.org/10.1145/2767151
20. Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining Collaborative Filtering Recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW '00)*. ACM, New York, NY, USA, 241–250. DOI: http://dx.doi.org/10.1145/358916.358995
21. Mat Honan. 2014. I liked everything I saw on Facebook for two days. Here‘s what it did to me. *I Wired, publisert* 8 (2014), 2014.
22. Lilt Inc. 2017. Lilt. (2017). Retrieved January 3, 2018 from https://www.lilt.com.

23. Hilary Johnson and Peter Johnson. 1993. Explanation facilities and interactive systems. ACM Press, 159–166. DOI:http://dx.doi.org/10.1145/169891.169951

24. Nikiforos Karamanis, Saturnino Luz, and Gavin Doherty. 2011. Translation practice in the workplace: contextual analysis and implications for machine translation. *Machine Translation* 25, 1 (2011), 35–52.

25. Kilgray. 2018. MemoQ Translation software. (2018). Retrieved January 3, 2018 from https://www.memoq.com/en/.

26. Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions (ACL '07)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 177–180. http://dl.acm.org/citation.cfm?id=1557769.1557821

27. Joseph A Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction* 22, 1-2 (2012), 101–123.

28. Elina Lagoudaki. 2008. The value of machine translation for the professional translator. In *Proceedings of the 8th Conference of the Association for Machine Translation in the Americas*. 262–269.

29. Elina Lagoudaki. 2009. Translation editing environments. In *MT Summit XII: Workshop on Beyond Translation Memories*.

30. Vladimir Iosifovich Levenshtein. 1966. Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics-Doklady* 10, 8 (1966). https://nymity.ch/sybilhunting/pdf/Levenshtein1966a.pdf

31. Brian Y. Lim and Anind K. Dey. 2010. Toolkit to support intelligibility in context-aware applications. ACM Press, 13. DOI:http://dx.doi.org/10.1145/1864349.1864353

32. Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. 2009. Why and Why Not Explanations Improve the Intelligibility of Context-aware Intelligent Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 2119–2128. DOI: http://dx.doi.org/10.1145/1518701.1519023

33. Lieve Macken, Els Lefever, and Veronique Hoste. 2013. Texsis: bilingual terminology extraction from parallel corpora using chunk-based alignment. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 19, 1 (2013), 1–30.

34. MateCat. 2018. MateCat. (2018). Retrieved January 3, 2018 from https://www.matecat.com/.

35. Joss Moorkens and Sharon O'Brien. 2016. Assessing User Interface Needs of Post-Editors of Machine Translation. *Human Issues in Translation Technology: The IATIS Yearbook. Abingdon: Routledge (to appear)* (2016).

36. Bonnie M. Muir. 1994. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* 37, 11 (Nov. 1994), 1905–1922. DOI: http://dx.doi.org/10.1080/00140139408964957

37. SDL. 2018. SDL Trados Studio. (2018). Retrieved January 3, 2018 from http://www.sdl.com/software-and-services/translation-software/sdl-trados-studio/.

38. David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 7587 (Jan. 2016), 484–489. DOI: http://dx.doi.org/10.1038/nature16961

39. Rashmi Sinha and Kirsten Swearingen. 2002. The Role of Transparency in Recommender Systems. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. ACM, New York, NY, USA, 830–831. DOI:http://dx.doi.org/10.1145/506443.506619

40. Across Systems. 2018. Across Translator Edition. (2018). Retrieved January 3, 2018 from https://www.my-across.net/en/.

41. Jörg Tiedemann. 2009. News from OPUS-A collection of multilingual parallel corpora with tools and interfaces. In *Recent advances in natural language processing*, Vol. 5. 237–248.

42. Nava Tintarev and Judith Masthoff. 2007a. Effective Explanations of Recommendations: User-centered Design. In *Proceedings of the 2007 ACM Conference on Recommender Systems (RecSys '07)*. ACM, New York, NY, USA, 153–156. DOI: http://dx.doi.org/10.1145/1297231.1297259

43. N. Tintarev and J. Masthoff. 2007b. A Survey of Explanations in Recommender Systems. In *2007 IEEE 23rd International Conference on Data Engineering Workshop*. 801–810. DOI: http://dx.doi.org/10.1109/ICDEW.2007.4401070

44. Tom Vanallemeersch and Vincent Vandeghinste. 2015. Assessing linguistically aware fuzzy matching in translation memories. In *Proceedings of the 18th Annual Conference of EAMT*.

45. Jo Vermeulen, Kris Luyten, and Karin Coninx. 2013. Intelligibility Required: How to Make Us Look Smart Again. In *Proceedings of the 10th Romanian Conference on Human-Computer Interaction, ROCHI*, Vol. 13. 36.

46. Jesse Vig, Shilad Sen, and John Riedl. 2009. Tagsplanations: Explaining Recommendations Using Tags. In *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI '09)*. ACM, New York, NY, USA, 47–56. DOI: http://dx.doi.org/10.1145/1502650.1502661

47. Wordfast, LLC. 2018. Wordfast. (2018). Retrieved January 3, 2018 from http://www.wordfast.com/.

48. XTM International Ltd. 2018. XTM - Better Translation Technology. (2018). Retrieved January 3, 2018 from https://xtm-intl.com/.

49. Jianlong Zhou, M. Asif Khawaja, Zhidong Li, Jinjun Sun, Yang Wang, and Fang Chen. 2016. Making machine learning useable by revealing internal states update - a transparent approach. *International Journal of Computational Science and Engineering* 13, 4 (2016), 378. DOI: http://dx.doi.org/10.1504/IJCSE.2016.080214