International Journal of Interpreter Education

Volume 14 | Issue 1 Article 5

7-2022

CAI Tool-Supported SI of Numbers: A Theoretical and Methodological Contribution

Francesca Maria Frittella
Shanghai International Studies University

Recommended Citation
Frittella, Francesca Maria (2022) "CAI Tool-Supported SI of Numbers: A Theoretical and Methodological Contribution," *International Journal of Interpreter Education*: Vol. 14: Iss. 1, Article 5.
DOI: https://doi.org/10.34068/ijie.14.01.05
Available at: https://tigerprints.clemson.edu/ijie/vol14/iss1/5

CAI Tool-Supported SI of Numbers: A Theoretical and Methodological Contribution

Francesca Maria Frittella
Shanghai International Studies University

Abstract

Numbers are an area of interpreting that is particularly prone to human error. Thanks to recent advancements in automatic speech recognition (ASR) and artificial intelligence (AI) technology, computer-assisted interpreting (CAI) tools may soon be used to enhance delivery accuracy for numbers during simultaneous interpreting (SI).

Given the novelty of the topic, the impact of in-booth CAI tool support on the SI of numbers is still largely under-researched. First, only a few studies have addressed the topic. Second, due to a number of methodological limitations, their findings yield only a partial understanding of the issue. The present work aims to make a theoretical and methodological contribution to this new area of inquiry. It identifies different research approaches to the interpretation of numbers. It then presents an empirical study on the computer-assisted SI of numbers with five Italian conference interpreters and the AI-powered CAI tool SmarTerp. The analysis contrasts two different research approaches and reveals the impact of speech design and evaluation methods on results. Implications and methodological recommendations for future studies are discussed.

Keywords: numbers, simultaneous interpreting, computer-assisted interpreting, automatic speech recognition, methodology

1. Introduction

Numbers are a recurring stumbling block for interpreters, which "make[s] simultaneous interpreting (SI) a 'finite and fallible function'" (Mazza, 2001, p. 103). In other words, they are "among the source speech elements that are particularly vulnerable to incompleteness and inaccuracy in an interpretation" (Mead, 2015, p. 286).

Through the years, several studies have explored the SI of these highly challenging elements and quantified their impact on the interpreter's output (i.e., the "delivery"; see, for example, Braun & Clarici, 1996a; Cheung, 2008, 2009; Frittella, 2017, 2019a; Kajzer-Wietrzny et al., 2021; Korpal & Stachowiak-Szymczak, 2018; Mazza, 2001; Pinochi, 2009). As reported by Desmet et al. (2018), the average error rate across multiple studies lies between 45% and 55% for students and 30% and 40% for professionals in experimental settings.

Although most studies report "accuracy rates" and "error rates," they differ sensibly in key methodological aspects, such as the definition of the research issue (i.e., what "interpreting numbers" entails) and how to adequately explore the issue empirically. Consequently, methods used to evaluate the delivery and the design of the test speech differ across studies. The error rates of some studies refer to the rendition of the bare numeral only (e.g., Braun & Clarici, 1996; Mazza, 2001), while other studies consider the number's "context" too (Korpal & Stachowiak-Szymczak, 2018). Some studies present participants with sentences that contain one numeral only, others work with number-dense passages, and yet others do not specify the characteristics of the speech units in which numerals are embedded. All this variety limits the comparability and reproducibility of findings. It may even be conjectured that these methodological issues may introduce biases in the results and their interpretation. Concerning the unit of analysis, for instance, Cheung (2009) points out:

> The validity of research findings based solely on [the accuracy of interpreted numerals], however, is questionable. For instance, how should renditions that contain gibberish but correctly translated numbers be graded? . . . Therefore, examining correctly translated numbers would only be a valid criterion if one is interested in one particular aspect of CI [here: conference interpreting]. However, this may not be the case for most research, quantitative or qualitative, which tends to focus on messages being interpreted rather than on numbers or words out of context. (p. 66)

In recent years, researchers' interest in the SI of numbers has been fueled by the integration of automatic speech recognition (ASR) and artificial intelligence (AI) technology into computer-assisted interpreting (CAI) tools, which now make it conceivable to enhance human performance through technology.[1] Early explorations report that the use of CAI tools may improve accuracy rates by more than 30%, even when subjects are master's degree students who have not received any previous training on CAI (Defrancq & Fantinuoli, 2020; Desmet et al., 2018). However, the methodological issues pointed out above with regard to the "traditional" SI of numbers may be found in current research on the CAI tool–supported SI of numbers too. Given these limitations, it is questionable whether reported findings are actually reflective of broadly conceived delivery accuracy and may be transferred from the context of the experiment to real-life assignments.

The present paper aims to make a theoretical and methodological contribution to research on the interpretation of numbers, in general, and CAI tool–supported SI of numbers, in particular, by identifying different approaches to the empirical exploration of this issue, analyzing their methodological components, and shedding light on their impact on results. Thereby, the study aims to contribute to the development of a solid empirical foundation for this area of increasing scientific and practical interest. At the same time, it aims to provide a starting point for a more holistic approach to the analysis of the CAI tool–supported SI of numbers. Deeper knowledge of the impact of CAI tool use on the SI of numbers and the challenges to the successful execution of this task is fundamental to tackle the paramount pedagogical issue of our time: ensuring that professionals and students may leverage technological innovation.

The aim of the paper is accomplished through a literature review and an empirical study. The literature review identifies three main approaches to the empirical investigation of the SI of numbers—*cognitive*, *syntactic*, and *communicative*—highlighting their conceptual premises, methodological implications, and limitations.

The empirical part draws on data gathered during a pilot study conducted within the framework of the Innovation Activity *SmarTerp*,[2] funded by EIT Digital. Five Italian conference interpreters interpreted a speech simultaneously from English into Italian with the support of a mock-up of the AI-powered CAI tool SmarTerp. The data set is analyzed by two distinct methods to ascertain whether the choice of approach and methodology has an impact on results.

The body of the paper comprises four sections: literature review, study design, results, and discussion. The conclusion highlights the relevance of the present discussion for the research milieu, points to future study directions, and summarizes the methodological recommendations for future research on the interpretation of numbers. Detailed study materials are provided in the appendix of the paper.

Literature Review: Research Approaches to the Interpretation of Numbers

Studies on the interpretation of numbers differ on a series of key methodological aspects. The present paper proposes that they may be divided into three groups based on some common theoretical and methodological denominators[3] : the *cognitive*, the *syntactic*, and the *communicative* approaches. For each approach, these key methodological issues are analyzed in the following review:

(1) *Conceptualization of the research issue:* How does the study define the issue *interpretation of numbers*?
(2) *Research question:* What type of research question does this approach answer?
(3) *Unit of analysis:* What is the unit of analysis corresponding to the conceptualization of the research issue in the study?
(4) *Speech design:* What speech variables are manipulated to explore the research issue and answer the research question?
(5) *Evaluation methods:* How are errors in the interpreter's delivery defined? How are *omissions* evaluated?
(6) *Limitations:* How are research findings (e.g., accuracy rates) obtained by this method limited?

The Cognitive Approach

A first approach, arguably the most represented, defines the SI of numbers as the conversion of numerals from one coded representation (the source language, SL) to another (the target language, TL, or a graphic code, e.g., Arabic numerals). This approach is little concerned with the semantic aspects of the SI of numbers, and the transcoding task is even postulated to be an asemantic process (Braun & Clarici, 1996; Mazza, 2001; Pinochi, 2009). We may call this approach *cognitive*, as it is mostly concerned with cognitive aspects of the interpretation.

The unit of analysis corresponding to this approach is the bare numeral. This means that data analysis is restricted to the evaluation of how the numeral is interpreted, whereas other aspects of the delivery (for instance, how the whole sentence is rendered) are not considered.

As already mentioned, this *cognitive* approach is most suitable to answer research questions concerning cognitive aspects of the interpretation of numbers. For instance, Braun and Clarici (1996) adopt a cognitive standpoint to define the nature of the mental processes that underlie the SI of numbers, and Korpal and Stachowiak-Szymczak (2019) explore the cognitive load involved in the task. Other studies have adopted this approach to explore the impact of specific variables on the rendition of the bare numeral, such as the use of note-taking (Mazza, 2001), the verbal properties of the source-language numerals (Pinochi, 2009), and speed (Korpal & Stachowiak-Szymczak, 2020).

Because the focus of studies adopting a cognitive approach is on the numeral, the salient feature of the test speech design is the introduction of numerals of different sizes and types. For instance, Mazza (2001) crafts her test speech by including (a) whole numbers above the order of magnitude "thousand" (with four or more digits),

(b) whole numbers below "thousand" (with fewer than four digits), (c) decimals, (d) ranges, and (e) dates. A similar choice is adopted by Pinochi (2009) and Desmet et al. (2018).

Consistent with the conceptualization of the research issue and the corresponding unit of analysis, the evaluation of the delivery focuses only on the bare numeral. The analysis typically yields an error rate describing how many numerals are correctly rendered (if the numeral corresponds exactly to the original) and how many are misinterpreted (if an error occurs in the transcoding process), approximated (if another numeral that is close in value is rendered), or omitted (if the numeral is not interpreted). The analysis may further qualify the phenomena identified, such as in the error classification proposed by Braun and Clarici (1996) based on studies in numerical cognition (McCloskey, 1992) and later adopted by Mazza (2001), Pinochi (2009), and Desmet et al. (2018), among others.

As already stated, studies adopting a cognitive stance are particularly suitable to "zoom in" on the mental processes underlying the rendition of numbers and shed light on the variables that may influence the processing operations as well as the interpreter's output. It may, hence, be regarded as a microanalysis concerned with the smallest unit of analysis in the interpretation of numbers. However, this approach has two major limitations related to its speech design features and evaluation methods. The first limitation is the lack of focus on the linguistic context in which numerals occur in the source speech: Because the focus is on the numeral, further variables inherent to the source speech are not taken into account, such as the sentence structure, the density of numerals, and other problem triggers in the speech passage. The second limitation is that the evaluation does comprise delivery aspects other than the bare numeral. Consequently, this approach fails to describe the impact of the interpreted numeral on the overall delivery quality. For instance, this method does not capture whether a misinterpreted numeral corresponds to a minor inaccuracy or a major semantic error (Frittella, 2019a). In the same way, every SL numeral that is not rendered as the TL numeral is regarded as an omission, irrespective of whether the interpreter expresses the numeral with other words (e.g., 2021 → this year) or decides to leave out a numeral repeated multiple times by the speaker. As a consequence of these two limitations, accuracy rates obtained by the cognitive approach may not be regarded as a reliable measure of the overall delivery accuracy, only as a description of a particular aspect of the interpretation of numbers and only under the conditions specified by the particular design of the speech.

The Syntactic Approach

A second research approach to the interpretation of numbers has emerged as authors have pointed out the need to expand the unit of analysis to the number and its referent (Gotri, 2003, cited in Moratto, 2011, p. 214; Korpal & Stachowiak-Szymczak, 2018; Pellatt, 2006). Embracing the views of Jones (1998, p. 130), these authors stress that numbers do not occur in isolation in a speech and do not convey meaning out of context. To reliably evaluate the accuracy of the SI of numbers, one should analyze not just the delivery of the bare numeral but also its "context." We may call this approach *syntactic* because the unit of analysis is defined by syntactic criteria (i.e., by the elements constituting the "context" of the numeral).

It must be noted that different authors provide different definitions of such "context," and so the unit of analysis varies across studies aligned to a syntactic approach. Although Jones (1998) stresses that several elements contribute to constituting such context, Korpal and Stachowiak-Szymczak choose the numeral and referent combination (i.e., the numeral together with the entity it refers to) as their unit of analysis—a choice that has also appeared in recent master's degree theses (e.g., Canali, 2018). Frittella (2017, 2019a) elaborates on Jones's definition and proposes to call the unit of meaning constituted by the numeral and its "context" the *numerical information unit* (NIU). The components of the NIU are as follows (Frittella, 2019a, p. 80): (a) the *numeral* itself, (b) the *referent* (i.e., the entity that the numeral refers to, such as export value), (c) the *unit of measurement* (e.g., U.S. dollars), (d) the *relative value* (e.g., increase, decrease, or leveling off), (e) the *time reference* (e.g., in 2019), and (f) the *geographical location* (e.g., in China).

This approach is most suitable to answer research questions concerning the impact of different syntactic variables on the rendition of the numeral as well as the NIU. Some research questions could ask "How does the sentence structure affect the rendition of the numeral and the NIU?" or "Does the nature of the referent (e.g., whether it is an acronym, a specialized term, or another numeral) affect the rendition of the numeral and the NIU?" Pellatt (2006) stresses that patterns emerging in the interpretation of numbers may not be fully understood without considering the crucial impact of "the linguistic environment" in which numerals occur in the speech (i.e., the sentence structure, the number of components in the NIU, and the density of numbers in a speech passage). Korpal and Stachowiak-Szymczak (2018) use a syntactic approach to investigate whether a difference may be identified in practicing conference interpreters' and trainees' renditions of numerals and their referents.

To be aligned to research questions formulated from a syntactic stance, the speech design should make it possible to observe the impact of the specific syntactic variables considered. One method could, for instance, present interpreters with NIUs of various lengths or with a specified number of numerals. If the aim is to explore the impact of different types of referents on the delivery, the speech design should comprise these elements.

Evaluation under the syntactic approach stresses that to ascertain whether the delivery is correct, the researcher must look beyond the rendition of the bare numeral. From this standpoint, the delivery is evaluated as accurate only if the numeral and the other elements constituting the unit of analysis are rendered correctly. In Korpal and Stachowiak-Szymczak (2018), for instance, the correct rendition of the numeral and the referent are conditional to the deliveries being evaluated as accurate. Based on Frittella's (2019a) definition of NIU, all constituents should be rendered accurately for the delivery to be evaluated as accurate.

Although the syntactic approach expands the unit of analysis of the cognitive approach, its scope continues to present limitations—especially if only the numeral and referent combination is considered rather than the whole NIU. In the speech design, this approach fails to account for influencing variables beyond the sentence level—for instance, are numerals repeated in the speech section? As for the evaluation, this approach may fail to unveil that interpreted numerals in the speech are mutually contradictory or represent a plausibility error or, again, that the omission of a redundant numeral may represent a strategy rather than a problem (Frittella, 2019a).

The Communicative Approach

A third approach, underrepresented compared to the previous two (particularly compared to the cognitive approach), sees the SI of numbers as an act of comprehension and interpretation of a message. With the exception of Frittella (2017, 2019a), scholars have endorsed this approach without in-depth theoretical justifications (e.g., Alessandrini, 1990; Cheung, 2009), but rather following the pretheoretical intuition that the whole of interpreting is "meant to reproduce ideas not words" (Alessandrini, 1990, p. 78) and that numbers should not be regarded as an exception. By this approach, "interpreting numbers" is conceptualized as the interpretation of a message, which contributes to constituting the logical chain of reasoning of a text, carries an extralinguistic semantic meaning, and is purposefully used by a speaker in a defined communicative context to achieve a specific goal. We may define this approach as *communicative* because it considers the SI of numbers to be an act of communication, like the whole of interpreting.

The communicative approach and its corresponding units of analysis are best formalized in the *processing ladder model for the interpretation of numbers* (Frittella, 2017, 2019a), which is inspired by Chernov's (2004) *probability prediction model*. The processing ladder model comprises the units identified by the cognitive and syntactic approaches and further expands the scope of analysis to consider the whole *numerical information*, as shown in Figure 1.

Examples of research questions that may be answered by this approach include the following: What is the impact of numbers on the broader accuracy of the interpreter's delivery? What factors beyond the characteristics of the numeral and the NIU may influence the interpreting process and product? What strategies make it possible to overcome difficulties inherent to the interpretation of numbers without severely compromising the transmission of the message? Adopting this approach, Alessandrini (1990, p. 78) investigates "what exactly happens when an interpreter comes across numbers," Cheung (2008) analyses the impact of different types of exercises (numerals and numeral and referent drills) on students' interpretation of numbers, and Frittella (2017, 2019a) explores the impact of different possible causes of error in the SI of numbers.

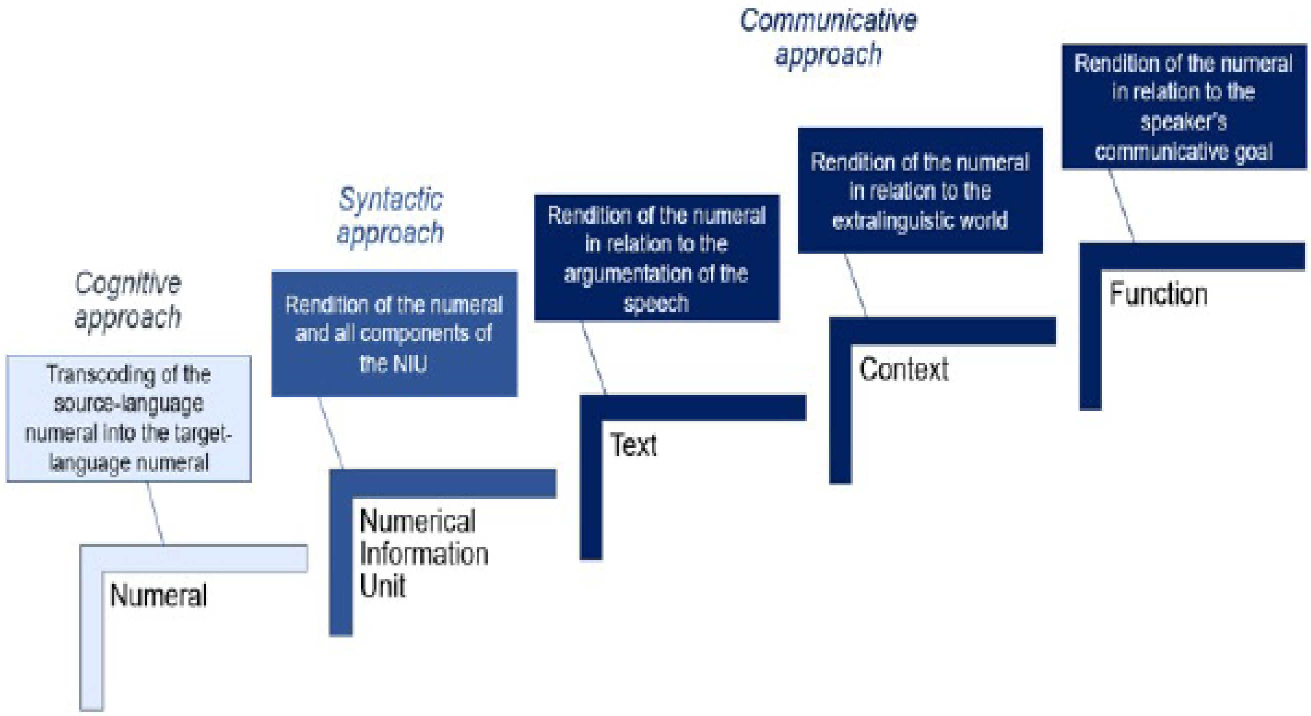

Figure 1. Processing ladder model for the interpretation of numbers; adapted from Frittella (2019a)

As this approach strives to holistically capture delivery accuracy, the speech design should comprise different types of challenges inherent to the interpretation of numbers. For instance, several authors highlight that number-dense speech passages may be associated with a higher error rate than isolated numbers (e.g., Mazza, 2001; Pellatt, 2006), which points to the need to take this variable into account. Frittella (2017, 2019a) proposes to include in the test speech numerical information of different complexities defined by the notion of *objective redundancy* (Chernov, 2004)—which may be broadly defined as the part of the message that is repeated or predictable based on the linguistic properties of discourse. Frittella (2017, 2019a) defines the objective redundancy of a speech unit (from a single sentence to a whole passage) containing numbers by (a) the numeral size and the number of digits, (b) the syntactic structure of the speech unit, (c) the number of components in each NIU constituting the speech unit, (d) the number of problem triggers within the unit, (e) the number of numerals within the unit, and (f) the number of repetitions of numerals and other NIU components within the unit.

Evaluation from a communicative stance considers the delivery to be accurate if all numerical information is interpreted correctly. In Frittella (2017, 2019a), this approach results in a classification of errors that takes into account the impact of the misinterpreted numeral on the transmission of meaning—for instance, it highlights whether numerals in the delivery are mutually contradictory or implausible. The same study also implies that the interpreted numerical information would be classified as a "functional error" if the delivery contained a correctly interpreted numeral but substantially distorted the communicative intention conveyed by the speaker—for instance, if a numeral used to support the speaker's argument was interpreted as an argument against it. Evaluation from a communicative perspective also implies a differentiation between nonstrategic omissions, which compromise the transmission of the message, and strategic omissions, which do not affect the rendition of the message—for instance, when a numeral appearing in the speech for the third time is omitted but hinted at by anaphoric reference (cf. Cheung, 2009; Frittella 2017, 2019a).

The communicative approach, albeit still being developed, may be the most adequate to holistically explore the delivery and its accuracy. It may also be the most productive when it comes to yielding pedagogical implications. In fact, thus far, studies aligned with a communicative approach have succeeded in shedding light on recurring problems and on how they may be addressed through interpreting strategies (Cheung, 2008; Frittella, 2017, 2019a). This knowledge informs the development of interventions on the interpretation of numbers (Cheung, 2009; Frittella, 2019b). On the contrary, studies that try to respond to the pedagogical challenge through a cognitive analysis (Mazza, 2001; Pinochi, 2009) conclude that "there does not seem to be any real solution to this problem" (Pinochi, 2009, p. 55).

The major limitation of this approach is that it is still being developed, and, hence, its methods are still only loosely defined. Frittella (2017, 2019b), for instance, proposes a communicative classification of error and strategy but suggests in the conclusion that a revision may be needed. Cheung (2009) proposes a classification of strategies based, however, on limited observations and only one specific language combination (Chinese-English). Therefore, in its current state, the communicative approach presents a shortage of validated methods and research instruments.

Final Methodological Considerations

The literature review above identifies three main research approaches to the interpretation of numbers (cognitive, syntactic, and communicative) and discusses their corresponding methodological aspects. These are summarized in the table at the end of this section. A few final remarks concerning the analysis are needed.

First, in the discussion, all methodological elements perfectly match the specified approach. Although the alignment of methodological components to the general research approach is usually the sign of a robust research design (cf. Creswell, 2018), this is not the case for every study. For instance, one study may use a unit of analysis corresponding to the syntactic approach to answer a research question about cognitive processes. This is not necessarily a weakness and may sometimes even be necessary to achieve the study's aim. However, the rationale for certain methodological choices should be clearly motivated in scientific work.

Second, the cognitive (microanalytical) approach and the communicative (macroanalytical) approach may be best regarded as two extremes of a scale rather than a strict dichotomy. Although studies may be categorized as belonging to one or the other approach based on their general orientation, they may actually be located at various points on this continuum. The syntactic approach itself may be regarded as an intermediate point on the scale.

Third, it is argued that the design of the test speech should be purposeful and match the unit of analysis and research question of the study. It should be stressed that if the test speech should make the observation of the impact of a specific variable possible, the design should also exclude the impact of undesired confounding variables. Furthermore, if the researcher aims to perform a quantitative analysis, a sufficient number of data points for each variable should be collected.

Fourth, transparency in the discussion of how data are collected and analyzed, which implies explaining the exact characteristics of the speech units in which numerals occur (i.e., providing a detailed description of the input variable considered), is of paramount importance for replication purposes and to ensure the comparability of findings. For instance, Kajzer-Wietrzny et al. (2021) report that in their study, number-dense passages do not correspond to an increase in error rates, contrary to what has been observed in other studies (e.g., Frittella, 2017, 2019a; Mazza, 2001). However, if the specific characteristics of the speech passage are unknown, it is not possible to ascertain whether the conditions are comparable across these studies (i.e., whether their findings all refer to passages of equal density and complexity).

Finally, each and every study should carefully frame its findings within the limitations inherent to its approach and research design to avoid compromising its reliability. Although this is true of all research, it may be particularly important in the context of the interpretation of numbers because, as discussed before, the concept itself varies across studies.

Approach	Conceptualization of the research issue.	Sample RQ	Unit of analysis	Speech designs	Evaluation	Limitations
Cognitive	Interpreting numbers means transcoding numerals from the SL to the TL.	How are numerals mentally processed during SI? What factors influence the transcoding process?	Numeral	Numerals of different sizes of types.	*Error*: The numeral is incorrectly rendered. *Omission*: The SL numeral is missing in the delivery.	Not reflective of overall delivery accuracy.
Syntactic	Interpreting numbers means conveying the numerals and their linguistic context.	How are numerals and the elements they refer to processed? How are different referents influential?	Numeral and referent combination. The whole NIU	Different types of referents. NIUs of different lengths and syntactic structures.	*Error*: Some NIU components are incorrectly rendered. *Omission*: Some NIU components are missing in the delivery.	Not reflective of overall delivery accuracy.
Communicative	Interpreting numbers means communicating a message.	How is the whole numerical information rendered? What strategies may be used?	Numerical information (text, context, and function)	Speech units of varying completely (e.g., defined by the concept of objective redundancy)	*Error*: The numerical information is incorrectly rendered. *Omission*: Nonredundant numerical information is missing in the delivery.	Methods still ill-defined and possible subjective interpretation of evaluation criteria.

*SL = source language, TL = target language
Table 1. Summary of approaches and methodological issues

Study Design

• *Aim*

The aim of the present paper is to make a methodological contribution to research on the CAI tool–supported SI of numbers as well as to the interpretation of numbers in general. The literature review identifies different approaches to the conceptualization of the research issue, which lead to different data analysis and speech design methods. The present study aims to highlight the impact of key methodological choices on the results. It is hoped that the discussion will help guide the study design and interpretation of findings of future research on the topic.

• *Research Question*

Empirical studies on the interpretation of numbers have been characterized by different approaches, and some criticisms appear in scholarly work concerning the adequacy of specific methods to describe the accuracy of the delivery of numbers, as presented in the literature review. However, due to the lack of data, the impact of one or the other approach on the study results may only be conjectured. Two main conjectures are explored empirically in this study.

First, it may be conjectured that different analysis methods may influence the results, causing a possible measurement bias. The first research question of the present study is as follows:

> RQ1: Do results vary quantitatively and/or qualitatively if the data set is evaluated by the *cognitive approach* (method 1) and the *communicative approach* (method 2)?

The choice of contrasting these two approaches is motivated by the fact that they may be regarded as two extremes on a scale. Furthermore, the syntactic approach is included within the communicative approach, which represents its further expansion, as argued earlier in the paper. Therefore, the qualitative analysis is expected to help define some limitations of the syntactic approach, too, although this is not analyzed separately.

Second, it may be conjectured that the complexity of the speech unit in which the numeral occurs (i.e., the variable of *task complexity*) may influence the effectiveness of the interpreter's use of CAI tools and, hence, the delivery. The second research question is as follows:

> RQ2: Does task complexity affect study participants' rendition of numbers in the CAI tool–supported SI task, and how?

• *Context*

The results of the analysis presented in this paper are related to data gathered during the pilot study of our research on the usability of the CAI tool *SmarTerp*,[4] developed through the EIT Digital[5] grant (cf. Frittella, in press). The overall aim of the research project is to expand the field's understanding of the ways that CAI tools support interpreters in the cognitively taxing task of interpreting numbers simultaneously, to derive implications for CAI tools' user interface (UI) design as well as for the training of conference interpreters. The pilot study, which was conducted remotely between May 31 and June 19, 2021, aimed to validate the study design, methods, and materials developed by the lead researcher and author of this paper. For the purpose of the present paper, the data set has been analyzed to answer the above-presented research questions.

• *Participants*

The participants in the pilot study were five Italian conference interpreters with English as their B/C language, two males and three females, ages 26–35, holders of a postgraduate degree in conference interpreting, and members of a professional association (AIIC,[6] AITI,[7] and Assointerpreti[8]). Except for one participant, all declared to have a yearly assignment volume as English-Italian simultaneous conference interpreters above 30 and to have completed more than 30 RSI (remote simultaneous interpreting) assignments over the previous year. The participants were recruited through the researcher's professional network. In the recruitment process, they signed an informed consent and filled out an enrollment questionnaire for profiling purposes (i.e., collecting such information as their qualifications and professional experience). Each chose a pseudonym for themselves to grant anonymity. Participation in the pilot study was voluntary, but the participants were rewarded with a €50 gift card.

Participation in the study involved the following tasks:

(1) → Completion of an *asynchronous e-learning module* (approx. 1.5 hours) with a theoretical introduction to ASR-supported CAI tools, the interface of SmarTerp, and a practical exercise similar to the test task. The e-learning module was developed by the author of the present paper with the aim to prevent first-time CAI tool use from affecting the test results.

(2) → Participation in the *remote test* (approx. 1.5 hours), consisting of an SI interpreting task with the support of SmarTerp (10 minutes), an evaluation questionnaire (10 minutes), and an interview (approx. 60 minutes).

• *The CAI Tool SmarTerp: UI Design for Numbers*

The study used a mock-up of the SmarTerp prototype. SmarTerp is a "third-generation" (*CAI – EABM*, 2021) CAI tool that makes use of ASR and AI technology to recognize problem triggers in the source speech and display them on the interpreter's laptop screen in real time, currently with a 2-second latency. A screenshot of the interface may be seen in Figure 2. Refer to Frittella (in press) for a comprehensive discussion of the UI design and the underlying rationale.

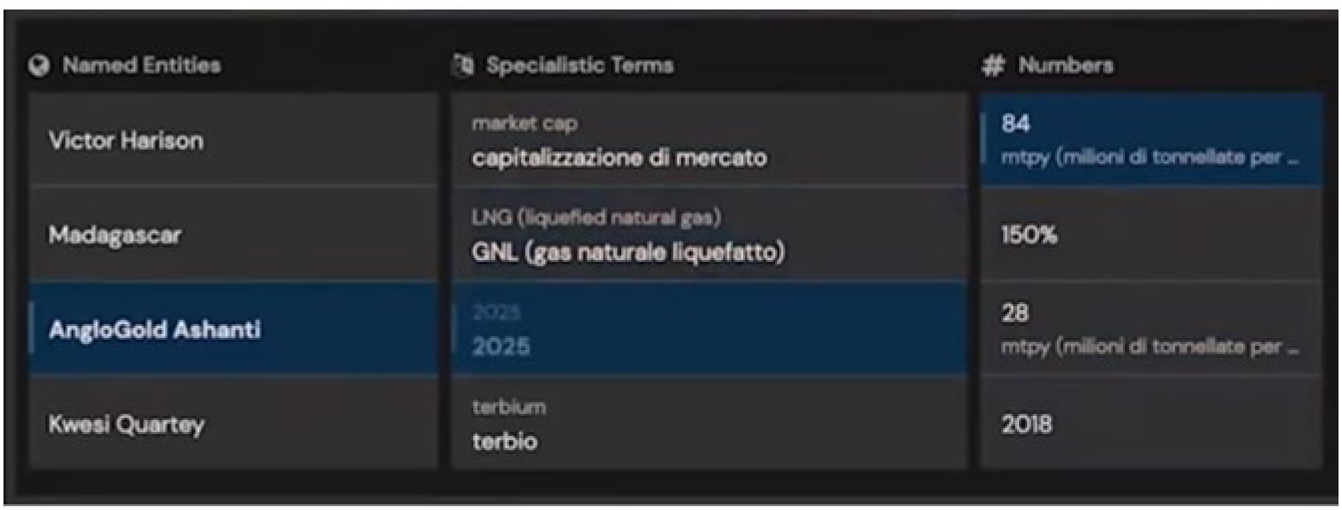

Figure 2. SmarTerp CAI tool interface

• *Materials*

• *Design of the Test Speech*

The design of the test speech implied a choice about the trade-off between control of variables and ecological validity. As pointed out by Prandi (2017), examining the impact of precise input variables on interpreters' deliveries requires a carefully designed speech that may present structural and prosodic characteristics different from unplanned speech typical of most real-life interpreting assignments. Cognizant of these limitations, the author of this paper chose this option because a real-life speech would have been unsuitable to answer RQ2, which is concerned with the impact of task complexity.

The first step in the speech design was the conceptual definition of tasks of increasing complexity. In the present paper, a "task"—more precisely, a *numerical (interpreting)* task— is defined as a representative problem situation in the SI of numbers. The idea of evaluating interpreting performance based on representative tasks is derived from usability testing (Barnum, 2020) as well as expertise studies, which are based on the observation of how experts deal with purposefully crafted challenges within the test (Ericsson et al., 2018). The notion of "representativeness" implies that the numerical tasks constituting this test should reflect the problems encountered by interpreters in real life, especially those challenging situations in which they are most likely to seek support from the CAI tool.

The second step was establishing the criteria for specific tasks of varying complexity that would constitute the speech. The degree of task complexity was defined based on the concept of *objective redundancy* presented in the theoretical section of this paper. Furthermore, every task needed to be challenging enough to prompt CAI tool use. Table 2 shows a list of the test tasks and their descriptions.

Task Code	Task name	Numerical task description
NU	Numeral	Interpreting a complex numeral (i.e., three-digits, order of magnitude = "trillion") in a simple sentence.
NR	Numeral and referent	Interpreting a complex numeral (i.e., an acronym / named entity / specialized term / numerical value) associated with it.
NIU	Numerical information unit	Interpreting a complex NIU consisting of (a) a complex referent, (b) a complex unit of measurement (i.e., an acronym / named entity / specialized term / numerical value), and (c) several numerals as in the following structure: amount increased/decreased by (X%) from Y (time 1) to Z (time 2).
NCR	Redundant number cluster	Interpreting a number cluster with redundant elements, which presents the following characteristics: (a) the passage contains three subsequent NIUs, (b) the time and place references remain unvaried and are repeated in each NIU, (c) the unit of measurement and the referent remain unvaried, but the referent is expressed with a different synonym in each NIU, and (d) the numeral changes in each NIU.
NCN	Nonredundant number cluster	Interpreting a number cluster without redundancy, which presents the following characteristics: (a) the passage contains three subsequent NIUs, (b) time, place, referent, unit of measurement, and numeral change in each NIU, and (c) either the referent or the unit of measurement are complex.

Table 2. Description of numerical tasks in the test speech

The next steps in drafting the speech were choosing a communicative context and then designing a speech unit matching the characteristics detailed in the description of the numerical task, as shown in Table 3. The full description is provided in the appendix of the paper.

Task code	Task name	Numerical task description	Numerical task
NU	Numeral	Interpreting a complex numeral (i.e., three digits, order of magnitude = "trillion") in a simple sentence.	The continent currently has a gross domestic product of USD 3.42 trillion.

Table 3. Example of numerical task

As a final step, the passages corresponding to each numerical task were arranged into a logical sequence to create a blueprint for the test speech. This step moved from the method used by Prandi (2017), drawing on Seeber and Kerzel (2012), of designing a speech with a "fixed internal structure that allows us to focus on the sentence level without sacrificing ecological validity completely" (Prandi, 2017, p. 85). The aim was to prevent excessive cognitive load that would arise from an uninterrupted series of problem triggers, on the one hand, and the *spill-over effect* (Gile, 2009) that may confound the analysis of relationships between input variables and delivery, on the other hand. The speech design method consisted of alternating *target sentences* that contained the input variable under study and *control sentences* that were free of any problem trigger and used to provide context.

In this study, each numerical task was enclosed in control sentences before and after. The introductory and the closing control sentence(s) were of 20–30 words in length and presented neither a problem trigger nor syntactic or conceptual complexity, as in the example below. When the speech was recorded, each closing sentence was followed by a 0.3-second pause:

> [Our objective is to] accelerate the political and social-economic integration of the continent. There are several signs that we are on the right track ((*introductory control sentences*)). The continent currently has a gross domestic product of USD 3.42 trillion ((*target sentence*)). This represents a remarkable achievement if we consider the fast pace of our economic growth over the past decades ((*closing control sentences*)). [3-second pause]

• ***Video***

The speech used for the pilot study was video recorded by the lead researcher and author of this paper in a nativelike pronunciation (New Zealand English). The average reading speed was 110 words per minute. Both the audio and the video were high resolution. The speech was well articulated and read in a natural and emphatic tone. The recorded video was entered into the SmarTerp prototype, and the running prototype was video recorded. A recorded video rather than the live tool was used as test materials for three main reasons: (1) to avoid an additional complexity in the remote testing procedures; (2) to create equal conditions for all participants; and (3) to prevent technical problems (e.g., tool failure, issues with participants' connection speed) from creating an additional variable to our analysis.

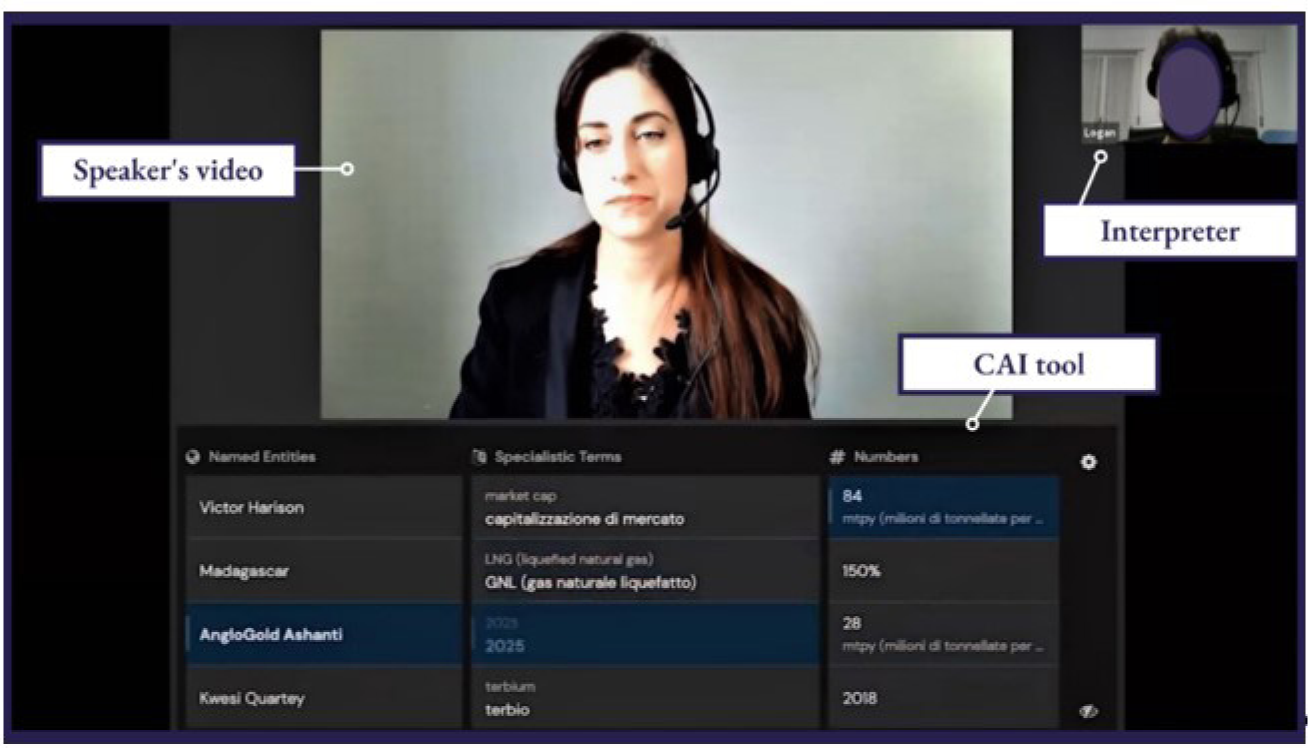

Figure 3. Remote test

• *Procedure*

Our pilot study was conducted between May 31 and June 7, 2021, as a remote testing procedure, given the ongoing COVID-19 pandemic. The lead researcher and author of this paper invited the participants to a remote testing session using the web conferencing platform Zoom. After an introduction, participants were sent a link to access the test speech video and were asked to share their screen. They were given 1 minute to read some generic information about the communicative context of the speech but could not search for additional information and terminology. They were then asked to interpret the speech simultaneously from their B/C language, English, into their mother tongue, Italian. Their webcam, delivery, and screencast were recorded as a single integrated visualization. The interpreter's webcam and audio were also saved as a separate file to zoom in on their faces, thus making it possible, in case of doubt, to ascertain whether they were looking at the tool when numerals were presented.

• *Data analysis*

• *Transcriptions*

Relevant parts of study participants' delivery were transcribed for data-analysis purposes with the following procedure. First, the test speech was segmented according to the units considered in our analysis:

- The numerical tasks represent the fundamental conceptual and semantic units as well as the input variables.
- The NIU is the smallest unit of meaning within the numerical task.
- The numeral is the arithmetic value itself.

Then, the source speech was organized into an Excel spreadsheet accordingly; see Table 4 for an example related to the numerical task *redundant number cluster* (NCR):

Numerical information unit	Numeral	Source speech
NCR-1	NCR-1-a	By 2030, the African continent would add about 295 million new people aged 15 to 64.
	NCR-1-b	
	NCR-1-c	
NCR-2	NCR-2-a	The growth would push the number of 15- to 64-year-old Africans up by 40% by 2030.
	NCR-2-b	
	NCR-2-c	
NCR-3	NCR-3-a	By 2030, Africa would hence be home to nearly 1 billion people of 15 to 64 years of age.
	NCR-3-b	
	NCR-3-c	

Table 4. Segmentation of the source speech

Finally, relevant parts of participants' deliveries were transcribed in a column to the right of the "source speech" column.

• *Comparison of Methods*

To answer RQ1, the data set was coded twice, using two distinct methods:

(1) *Method 1—cognitive approach*: The delivery was evaluated as accurate solely based on the rendition of the bare numeral; all omitted numbers were evaluated as an error, and more specifically as an omission, irrespective of whether the omitted number was redundant.

(2) *Method 2—communicative approach*: The delivery was evaluated as accurate only if the accuracy requirements of a communicative approach to the analysis (detailed below) were satisfied; the omission of a redundant number was not evaluated as an error if the interpreter adopted a strategy not to change the meaning of the utterance.

By method 1, the numerals in the delivery were evaluated as either accurate or inaccurate. Inaccuracies were broken down into errors (coded as e) and omissions (coded as o).

By method 2, each delivery unit was evaluated as either accurate or inaccurate based on the following criteria derived from an adaptation of the *redundancy ladder model* (Frittella, 2019a, cf. literature review):

(1) *Numeral*: Is the rendition of the bare numeral accurate?
(2) *IU*: Is the rendition of numeral, referent, unit of measurement, relative value, time, and geographic location accurate?
(3) *Text*: Is the interpreted numerical task internally consistent in logic and externally congruous in meaning with the source speech?
(4) *Context*: Is the interpreted numerical information plausible?
(5) *Function*: Is the interpreted numerical information functionally equivalent with the source speech, and does the delivery convey the communicative intention expressed by the speaker?
(6) *Strategy*: Did the interpreter use a strategy to tackle a number-related difficulty without altering the meaning of the utterance or causing a loss of information?

Error categories were identified at each level of analysis, first based on the literature review and then through a preliminary analysis of the data set, in a process similar to the identification of themes in thematic analysis (Braun & Clarke, 2006). The resulting evaluation criteria are presented in the results section.

• *Impact of Complexity*

To answer RQ2, participants' performance on each task was aggregated to calculate the *success rate*. Like the idea of testing performance on representative tasks, the success-rate concept was borrowed from usability testing (Barnum, 2020). The criteria used in this stage of analysis are reported in Table 5. The success rate of numerical tasks comprising several NIUs was calculated as the mean of the success rates of the constituting NIUs.

Delivery	Success rate	Criterion	Example (source→target)
Correct rendition	100%	All elements of the NIU were rendered accurately, and all other accuracy criteria in method 2 (see above) were satisfied.	In 2021 → in 2021
Omission of redundant component lexical substitution	100%	A component of NIU was omitted, but the key statement is accurate.	In 2021 → this year
Partial rendition	Proportional to the content of the NIU	Some elements were omitted, but the key statement is accurate.	Analysts forecast that African production of LNG will increase by 150% from 28 mtpy in 2018 to reach 84 mtpy by 2025. → Analysts forecast that African production of LNG will increase b 150% to reach 84 mtpy by 2025. =80%
Generalization, summarization	30%	The numerals were omitted and the information was summarized by the interpreter.	Analysts forecast that African production of LNG will increase by 150% from 28 mtpy in 2018 to reach 84 mtpy by 2025. → Analysts forecast that African production of LNG will increase substantially in the next years.
Semantic error	0%	Regardless of whether the numeral and other NIU components were rendered accurately, the delivery substantially contradicted the original meaning.	Population will increase *by* 1 billion. → Population will to increase *to* 1 billion.
Complete omission	0%	The whole numerical task was omitted.	In 2021 → ø

Table 5. Evaluation criteria for task accuracy

• *Results*

• *Categories of Error and Strategy*

Table 6 summarizes the categories of error and strategy identified on each level of analysis. It also reports how many instances of each category (column "Tot cases") were identified and in how many participants' delivery they occurred (column "Tot p."). A more detailed table with delivery examples is provided in the appendix of the paper.

Level	Category	Explanation	Total cases	Total p.
Numeral	*Error*	The interpreted numeral was incorrect.	3	3
	Omission	The numeral was omitted.	23	5
NIU	*Wrong referent*	The interpreted referent differed from the referent in the SL numerical task.	2	1
	Wrong unit of measurement (UoM)	The interpreted UoM differed from the SL one.	7	5
	Wrong relative value	The interpreted relative value differed from the SL one.	3	2
	Misattribution of components	The semantic links between the components of the interpreted NIU did not correspond to the SL ones.	4	3
	Sentence fragment	A NIU did not express a complete thought, as one or more of its essential components were missing.	7	4
	Omission of the NIU	The whole NIU was omitted.	1	1
Text	*Inconsistent numerals*	The numerals within the interpreted numerical task were contradictory.	1	1
	Distortion of information	The meaning of the interpreted numerical task, albeit internally consistent and plausible, differed substantially from the SL.	5	3
Context	*Plausibility error*	The interpreted numerical information seemed unreasonable and improbable against the world knowledge of an informed listener.	9	4
Function	*Functional error*	Although the numeral and all components of the NIU corresponded to the SL NIU, the interpreted message differed from the original one in its function.	0	0
Strategy	*Omission of redundant item*	The interpreter omitted an item (the numeral or another component of the NIU) that was repeated within the numerical task.	4	3
	Abbreviation of acronym	The interpreter simplified the referent or the UoM by using an acronym.	3	2
	Lexical substitution	The interpreter replaced a component of the NIU with its non-numeric equivalent or through anaphoric reference.	4	2
	Generalization of the numeral	The interpreter replaced the numeral with a general expression to form a sentence of finite meaning.	4	3
	Summarization	The interpreter summarized the meaning of the numerical information.	3	1

Table 6. Categories of error and strategy

• *Method Comparison*

• *Quantitative Difference*

Table 7 reports the results of the analysis conducted on the data set, using method 1 (the cognitive approach—that is, evaluating only the rendition of the numeral) and method 2 (the communicative approach—that is, considering all other levels of analysis). Participants interpreted a total of 110 numerals (22 numerals × five deliveries).

		Total number of errors (/110)	Error rate
Method 1: cognitive	*Numeral*	27	24.5%
	Error	3	2.7%
	Omission	24	21.8%
Method 2: communicative	*Other level*	34	30.9%

Table 7. Results by method

Seven more errors were detected by the second method, representing 25.9% of all errors detected by method 1 and 20% of errors detected by method 2. Twelve of the 27 errors (44%) at the numeral level were not evaluated as errors by method 2, while 19 of the 34 errors (55%) at all other levels were evaluated as accurate by method 1. This means that a total of 31 inconsistencies was identified in the evaluations performed on the data set by the two methods. It is noticeable that no delivery was exempt from incongruent evaluations.

• *Qualitative Difference*

The impact of method on results may be best understood through a qualitative analysis. Starting from the instances that were evaluated as numeral errors by method 1 but as accurate by method 2, all incongruencies correspond to omitted numbers: 12 in a total of 27 errors identified by method 1. These correspond to strategies adopted by interpreters for redundant numbers without changing the meaning of the utterance and making the NIU discernible from the context of the delivery, as in the example below, representing a case of anaphoric reference through lexical substitution:

Eg. *(1)* → Source (NCR-2-a): The growth would push the number of 15- to 64-year-old Africans ((repeated item)).
Target (Carlo): People in this age range ((anaphoric reference to numeral expressed in the previous sentence)).

More specifically, using method 2, these 12 omissions were classified as generalization (four cases), lexical substitution (four cases), omission of redundant item (two cases), and summarization (two cases). See the table in the appendix for examples of each error category.

For the instances that were evaluated as accurate by method 1 and inaccurate by method 2, 19 incongruencies were registered out of 34 errors identified by the latter method. These incongruencies relate to errors at the level of the NIU, which break down into the following categories: *wrong unit of measurement* (three cases across three participants), *wrong relative value* (three cases across two participants), *misattribution of components* (two cases across two participants), and *sentence fragment* (two cases across two participants). The incongruency in evaluation outcomes by the two methods is clarified by the example below: Although the interpreted numeral was correct, the participant omitted the referent (oil) and misinterpreted part of the unit of measurement (million *barrels* per day), transforming it into the referent of the NIU, which substantially changes the meaning of the utterance:

Eg. *(2)* → Source (NCN-1-b): Africa produced nearly 8.41 mbd [million barrels per day] of oil.
Target (Carlo): We produced approximately 8.41 million barrels per day.

Considering the text level, segments evaluated as correct by method 1 were found to correspond to a *distortion* of the original meaning (five cases across three participants) and *inconsistent numerals* (one case) when evaluated by method 2. The example below of inconsistent numerals clearly demonstrates the limitations of the syntactic approach (i.e., confining the analysis of the SI of numbers to the numeral or NIU level). Although the numerals all correspond to the original, and the NIUs are sentences of finite sense, there is an internal contradiction in the delivery:

Eg. *(3)* → Source (NCR-1, -3): By 2030, the African continent would add about 295 million new people aged 15 to 64. . . . By 2030, Africa would hence be home to nearly 1 billion people of 15 to 64 years of age.
Target (Minerva): Again, by 2030, the African continent will have about 295 million people aged 15 to 64. . . . Again, by 2030, Africa will have 1 billion inhabitants aged 15 to 64.

Considering the context level, a numeral assessed as correct by method 1 corresponded to a *plausibility error* in five cases across three participants. In the example below, 1 billion was evaluated as an accurate rendition, but the omission of part of the referent substantially changes the meaning of the utterance, making the whole message implausible (the population of Africa stands at approximately 1.3 billion as of 2021, so it is implausible to say that the population will soar to 1 billion by 2030):

Eg. *(4)* → Source (NCR-3): By 2030, Africa would hence be home to nearly 1 billion people aged 15 to 64.
Target (Diana): By 2030, Africa will be home to over 1 billion people.

No error was found at the level of the function of the message. This may depend on the architecture of our speech, which made the function of numerals clear through the introductory and closing control sentences.

• *Impact of Task Complexity*

Results concerning the impact of task complexity on delivery accuracy are reported in Table 8.

Code	Task	P1	P2	P3	P4	P5	Mean	Median	Mode
NU	Isolated numeral	100%	100%	100%	0%	100%	80%	100%	100%
NR	Numeral and referent	100%	100%	100%	100%	100%	100%	100%	100%
NIU	Numerical information unit	80%	40%	80%	20%	20%	48%	40%	80%-20%
NCR	Redundant number cluster	100%	67%	0%	0%	100%	53%	67%	100%-0%
NCN	Non-redundant number cluster	33%	33%	38%	80%	25%	42%	33%	33%

Table 8. Impact of task complexity on accuracy levels

Based on the results of our analysis of the impact of task complexity on accuracy rates, participants' accuracy seems to tendentially decrease with the increase of task complexity, although this trend is not clear for NIU and NCR. Accuracy rates are very consistent in tasks of lowest complexity—the only outlier who made a mistake in the NU task declared in the post-task interview that she got distracted at that stage. However, within-subject variability is considerable in tasks of higher complexity, as in the instance of P3 (Minerva) and P4 (Diana), who scored higher in NCR than NCN.

Is it possible that a redundant number cluster may be easier to process for these interpreters than a complex NIU? May these phenomena be attributed to idiosyncratic factors and skill gaps, as hypothesized by previous studies (Alessandrini, 1990; Frittella, 2019a; Kajzer-Wietrzny et al., 2021)? Or is the CAI tool's UI at least partly responsible for participants' errors? Given the small sample, it is not possible to advance a possible answer to these questions. However, they are worth noticing and deserve further exploration in future studies.

• Discussion

The results of our analysis highlight the impact of methodological choices on research results concerning interpreters' delivery accuracy for numbers with CAI tool support during SI.

• *Method Comparison: Implications for the Choice of Research Approach*

In response to the research question RQ1—Do results vary quantitatively and/or qualitatively if the data set is evaluated by the cognitive and the communicative approach?—our analysis suggests that evaluating delivery accuracy solely by the accuracy of the bare numeral (*method 1: cognitive approach*) or considering the rendition of the NIU as well as textual, contextual, and functional dimensions of the message (*method 2: communicative approach*) does produce quantitative and qualitative differences in the results. Although the *syntactic approach* is not the specific focus of the analysis, the qualitative results suggest that the latter approach may not be sufficient to detect all errors in the data set.

As for the quantitative difference, seven more errors were identified by method 2 than by method 1—a total of 34 versus 27 errors out of a total 110 interpreted numerals. Although the difference is rather small in absolute terms, in relative terms, it corresponds to the detection of 25.9% more errors relative to the total of 27 and of 20% more errors relative to the total of 34. The total number of inconsistencies is even higher (31). A larger or smaller impact of method on the results may be identified in a larger data set—a possible question to be answered in future investigations.

The discrepancy between methods may be best understood by considering the qualitative characteristics of this difference. Already, Cheung (2009, p. 66) has raised the question of the validity of research findings based solely on the accuracy of the interpreted numeral. We report cases in which correctly interpreted numerals actually corresponded to a plausibility error (e.g., Africa will be home to 1 billion people by 2030) or a complete distortion of the sense of the original message (e.g., Africa produces 8 million *barrels per day* instead of barrels *of oil* per day). These cases further reinforce the argument that the accuracy of interpreted numerals is not an adequate measurement of delivery accuracy. While the choice of the numerals as the unit of analysis (i.e., the *cognitive approach*) may be adequate for studies focused on cognitive processes, it seems unsuitable for exploring the overall impact of numbers and the use of CAI tools on the delivery. Researchers should take care to contextualize their findings within this limitation.

Our analysis also suggests that the *syntactic approach*, which takes the numeral and referent combination as the measurement of delivery accuracy, is also not sufficient to accurately and reliably measure the accuracy of numbers' delivery. This is exemplified by the example reported of inconsistent numerals—the two adjacent NIUs in the interpreter's delivery make perfect sense when evaluated individually but should still be evaluated as inaccurate because they are also mutually contradictory. Such inconsistency errors may only be detected at a communicative analysis at the text level.

• *Impact of Task Complexity: Speech Design Methods*

Responding to the research question RQ2—Did task complexity affect study participants' rendition of numbers in the CAI tool–supported SI task, and how?—our analysis points to a tendency in study participants' delivery accuracy to decrease at the increase of task complexity. Given the small data set, the results may not be generalized in a statistical sense and should be corroborated with further data. However, these observations provide arguments for the need to consider the variable of task complexity in the design of test speeches and in the interpretation of results on the CAI tool–supported SI of numbers as well as the interpretation of numbers in general.

• *Limitations*

The limitations of the present work concern the generalizability of its findings, the shortcomings of our speech design, the ecological validity of findings, and insufficient validation of proposed methods.

Starting with the generalizability of findings, it should be stressed that our data set was rather small, which is why no statistical generalizability was claimed for the findings. The significance that we claim for the work is in the methodological principles it highlighted rather than in the quantitative results obtained.

Further limitations may be identified in the speech design. In particular, we saw that the fixed internal structure of the speech, with opening and closing control sentences making the point of view of the speaker clear, may have prevented the observation of possible functional errors that may arise in all situations in which the function of the number must be inferred by the interpreter (Frittella, 2019a). This limitation further reinforces the argument that the speech used to evaluate interpreters' rendition of numbers should be in line with the research questions.

A further limitation arises from the choice of a high-constraint research design, which involved a moderate degree of experimental control (although we would define our approach as mixed-method and tendentially exploratory/qualitative rather than experimental/quantitative). As rightly pointed out by Prandi (2017), the disadvantage of a carefully designed speech and experimental control variables is that it limits the ecological validity of findings—that is, it yields findings that may not be reflective of accuracy rates achieved in real-life assignments. This choice was deliberate, as constraining the number of variables was necessary to answer our research questions. However, readers interpreting this study's findings should bear in mind the limitations implied by its underlying design choices.

Finally, as stressed several times in this contribution, the methods proposed in this paper are novel and may require further empirical validation and refinement. Although they may provide a valuable starting point for future research, they might need to be refined and adapted to the specific research question addressed in future studies.

• Conclusion

Following the recent integration of ASR and AI technology into CAI tools, the interpretation of numbers assumes new relevance in the research landscape. The present study aims to address two major methodological issues related to the choice of evaluation methods and the design of test speeches. The fact that these aspects vary sensibly across studies reduces the comparability and reproducibility of findings. Unless appropriately addressed, these limitations may yield a distorted interpretation of findings and risk threatening the reliability of studies. For instance, claims that a CAI tool can successfully support the SI of numbers are unwarranted unless a range of crucial influential variables is included in the test speech design (such as the complexity of the speech passage in which numerals occurs) and the delivery is evaluated holistically rather than focusing only on the bare numeral.

Carefully accounting for these methodological issues (and, possibly, other issues that the present paper leaves unaddressed) is fundamental to generating a reliable knowledge base. In this new "technological turn" (Fantinuoli, 2018), the dissemination of findings based only on a partial understanding of the CAI tool–supported SI of numbers may generate false expectations in professionals who, when disillusioned, may lead to a counterproductive closure toward technology. Moreover, a deeper understanding of the challenges inherent to this novel and complex task is needed to inform the development of a training solution.

The present paper argued that the exploration of the (CAI tool–supported) interpretation of numbers has been thus far guided by three main approaches: *cognitive*, *syntactic*, and *communicative*. Through an empirical analysis of the same data set by distinct methods, the impact of these approaches on the results was unveiled and contrasted. The results seem to support the choice of a communicative approach to explore the broad impact of numbers and the use of CAI tools on delivery quality. They also point to the need to purposefully design test speeches by manipulating the variable of task complexity. By providing the study materials as annex, this paper aims to encourage peer scrutiny and offers a concrete example of how speech design principles discussed in the paper were applied.

It is the author's hope that other researchers interested in studying the interpretation of numbers (with and without CAI tool support) may find in the present paper some guidance on the development of a methodological framework for the exploration of this complex and fascinating topic. Below, the key methodological recommendations emerging from the paper are summarized. They are consistent with the recommendations for high-quality research design proposed in leading manuals (e.g., Creswell, 2018). Concrete examples of how to apply these recommendations may be found in the literature review:

1. Conceptualize the *research issue* and choose a corresponding *research approach*: Respond to the question "What does 'interpreting numbers' entail?" Is it a transcoding process (*cognitive approach*)? Is it the rendition of a numeral and the other elements constituting the information unit (*syntactic approach*)? Or is it the delivery of a message with a semantic and pragmatic dimension (*communicative approach*)?
2. Formulate the *research question* accordingly: What research question is relevant for the exploration of the research issue, as defined by your approach?
3. Identify and clearly define your *unit of analysis* based on the chosen approach.
4. *Design your test speech* to include the variables that may affect interpreters' rendition of "numbers" (based on your conceptualization) and that are relevant to your research question.
5. Define *evaluation methods* pertinent to your unit of analysis and adequate to respond to your research question; think of how errors and omissions will be evaluated within your paradigm.
6. Discuss your findings within the *limitations* inherent to your methodology; in particular, specify to what extent they may be regarded as reflective of broadly conceived "delivery accuracy."

Finally, it is recommended that these steps be described with clarity and detail to encourage peer scrutiny and allow research consumers to identify the scope of applicability of reported findings. It may be advisable to refer to studies aligned with one's approach to help readers contextualize the study. It is also recommendable to provide the test speech, or the segments containing numerals, for transparency and to allow replication.

• References

Alessandrini, M. S. (1990). Translating numbers in consecutive interpretation: An experimental study. *Interpreters' Newsletter*, *3*, 77–80.

Barnum, C. M. (2020). *Usability testing essentials: Ready, set . . . test!* (2nd ed.). Morgan Kaufmann.

Braun, S., & Clarici, A. (1996b). Inaccuracy for numerals in simultaneous interpretation: Neurolinguistic and neuropsychological perspectives. *The Interpreters' Newsletter*, *7*, 85–102.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

CAI – Ergonomics for the Artificial Booth Mate (EABM). (2021). https://www.eabm.ugent.be/cai/

Canali, S. (2018). *Utilizzo del riconoscimento vocale come supporto durante l'interpretazione simultanea dei numeri [ASR as a support in the SI of numbers]* [master's thesis]. Unint.

Chernov, G. (2004). *Inference and anticipation in simultaneous interpreting*. John Benjamins Publishing.

Cheung, A. K. (2008). Simultaneous interpreting of numbers: An experimental study. *Forum-Revue Internationale d'Interpretation Et De Traduction*, *6*(2), 23–38. https://doi.org/10.1075/forum.6.2.02kfc

Cheung, A. K. (2009). Numbers in simultaneous interpreting: An experimental study. *Forum-Revue Internationale d'Interpretation Et De Traduction*, *7*(2), 61–88. https://doi.org/10.1075/forum.7.2.03che

Creswell, J. W. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.

Defrancq, B., & Fantinuoli, C. (2020). Automatic speech recognition in the booth. *Target*, *1*(33), 1–30. https://doi.org/10.1075/target.19166.def

Desmet, B., Vandierendonck, M., & Defrancq, B. (2018). Simultaneous interpretation of numbers and the impact of technological support. In C. Fantinuoli (Ed.), *Interpreting and technology* (pp. 13–27). Language Science Press.

Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams, M. A. (Eds.). (2018). *The Cambridge handbook of expertise and expert performance*. Cambridge University Press. https://doi.org/10.1017/9781316480748

Fantinuoli, C. (2018). Interpreting and technology: The upcoming technological turn. In C. Fantinuoli (Ed.), Interpreting and technology (pp. 1–12). Language Science Press. https://langsci-press.org/catalog/book/209

Frittella, F. M. (2017). Numeri in interpretazione simultanea: Difficoltà oggettive e soggettive [The simultaneous interpretation of numbers: Objective and subjective difficulties.]. Europa Edizioni.

Frittella, F. M. (2019a). "70.6 billion world citizens": Investigating the difficulty of interpreting numbers. *Translation and Interpreting*, *11*(1), 79–99. https://doi.org/10.12807/t&i.v11i1.686

Frittella, F. M. (2019b). Numbers: From stumbling block to training tool. *Interpreters' Newsletter*, *24*, 35–56. https://doi.org/10.13137/2421-714X/29524

Frittella, F. M. (in press). *Designing interpreting technology for usability: A case study of SmarTerp.* Language Science Press (accepted in May 2021).

Gile, D. (2009). Basic concepts and models for interpreter and translator training. John Benjamins Publishing.

Gotri, S. (2003). *Figures in simultaneous interpretation from French into Italian* [unpublished master's thesis]. Università degli studi di Bologna / Forlì.

Jones, Roderick. (1998). Conference Interpreting Explained. Manchester: St. Jerome.

Kajzer-Wietrzny, M., Ivaska, I., & Ferraresi, A. (2021). "Lost" in interpreting and "found" in translation: Using an intermodal, multidirectional parallel corpus to investigate the rendition of numbers. *Perspectives*, *29*(4), 469–488. https://doi.org/10.1080/0907676X.2020.1860097

Korpal, P., & Stachowiak-Szymczak, K. (2018). The whole picture: Processing of numbers and their context in simultaneous interpreting. *Poznan Studies in Contemporary Linguistics*, *54*(3), 335–354. https://doi.org/10.1515/psicl-2018-0013

Korpal, P., & Stachowiak-Szymczak, K. (2019). Interpreting accuracy and visual processing of numbers in professional and student interpreters: An eye-tracking study. *Across Languages and Cultures*, *20*(2), 235–251. https://doi.org/10.1556/084.2019.20.2.5

Korpal, P., & Stachowiak-Szymczak, K. (2020). Combined problem triggers in simultaneous interpreting: Exploring the effect of delivery rate on processing and rendering numbers. *Perspectives*, *28*(1), 126–143. https://doi.org/10.1080/0907676X.2019.1628285

Mazza, C. (2001). Numbers in simultaneous interpretation. *Interpreters' Newsletter*, *11*, 87–104.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*(1–2), 107–157. https://doi.org/10.1016/0010-0277(92)90052-j

Mead, P. (2015). Numbers. In F. Pöchhacker (Ed.), *Routledge encyclopedia of interpreting studies* (1st ed., pp. 286–288). Routledge. https://www.book2look.com/book/Vayscboatn

Moratto, R. (2011). Number processing in Chinese to Italian simultaneous interpreting. *FORUM. Revue Internationale d'interprétation et de Traduction / International Journal of Interpretation and Translation*, *9*(2), 209–231. https://doi.org/10.1075/forum.9.2.10mor

Pellatt, V. (2006). The trouble with numbers: How linguistic, arithmetical and contextual complexity affect the interpretation of numbers. 口译的专业化道路：国际经验和中国实践, *18*, 350–365.

Pinochi, D. (2009). Simultaneous interpretation of numbers: Comparing German and English to Italian. An experimental study. *Interpreters' Newsletter*, *14*, 33–57.

Prandi, B. (2017). Designing a multimethod study on the use of CAI tools during simultaneous interpreting. In J. Esteves-Ferreira, J. Makan, R. Mitkov, & O.-M. Stefanov (Eds.), *Translating and the computer 39, conference proceedings*. Editions Tradulex.

Seeber, K. G., & Kerzel, D. (2012). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism*, *16*(2), 228–242. https://doi.org/10.1177/1367006911402982

• Appendix

• *Appendix 1. Numerical Tasks in the Test Speech*

Task code	Task name	Numerical task description	Numerical task
NU	Numeral	Interpreting a complex numeral (i.e., three digits, order of magnitude = "trillion") in a simple sentence.	The continent currently has a gross domestic product of USD 3.42 trillion.
NR	Numeral and referent	Interpreting a complex numeral and the complex referent (i.e., an acronym / named entity / specialized term / numerical value) associated with it.	This year, the market cap of AngloGold Ashanti—the largest mining company headquartered in Africa—was USD 12.13 billion.
NIU	Numerical information unit	Interpreting a complex NIU consisting of (1) a complex referent; (2) a complex unit of measurement (i.e., an acronym / named entity / specialized term / numerical value); (3) several numerals, as in the following structure: amount increased/decreased by (X%) from Y (time 1) to Z (time 2).	Analysts forecast that African production of LNG [liquefied natural gas] will increase by 150% from 28 mtpy [million tonnes per year] in 2018 to reach 84 mtpy by 2025.
NCR	Redundant number cluster	Interpreting a number cluster with redundant elements, which presents the following characteristics: (1) the passage contains three subsequent NIUs; (2) the time and place references remain unvaried and are repeated in each NIU; (3) the unit of measurement and the referent remain unvaried, but the referent is expressed with a different synonym in each NIU; and (4) the numeral changes in each NIU.	Africa's working-age population is growing rapidly and is projected to surpass that of any other continent by 2030: • By 2030, the African continent would add about 295 million new people aged 15 to 64. • The growth would push the number of 15- to 64-year-old Africans up by 40% by 2030. • By 2030, Africa would hence be home to nearly 1 billion people of 15 to 64 years of age.
NCN	Non-redendant number cluster	Interpreting a number cluster without redundancy, which presents the following characteristics: (1) the passage contains three subsequent NIUs; (2) time, place, referent, unit of measurement and numeral change in each NIU; and (3) either the referent or the unit of measurement is complex.	Let us not forget that Africa has a wealth of natural resources: • In 2019, Africa produced nearly 8.41 mbd [million barrels per day] of oil. • Madagascar alone produced approximately 58,000 metric tons of nickel in 2021. • Namibia's diamond production amounted to 2.52 million carats in 2018."

• *Appendix 2. Error Categories with Examples*

Level	Category	Explanation	Example	Total cases	Total p.
Numeral	*Error*	The interpreted numeral was incorrect.	*Source (NU)*: The continent currently has a gross domestic product of USD 3.42 trillion. *Target (Diana, It)*: Il continente ha attualmente il prodotto interno lordo di 3,42 miliardi di dollari. *Target (Diana, En)*: The continent's gross domestic product currently stands at 3.42 billion dollars.	3	3
	Omission	The numeral was omitted.	*Source (NCN-2-b)*: in 2018 *Target (Sally)*: Ø	23	5
NIU	*Wrong referent*	The interpreted referent differed from the referent in the SL numerical task.	*Source (NCN-1-b)*: The growth would push *the number* of 15- to 64-year-old Africans . . . *Target (Diana, It)*: La *percentuale* di questa fascia di età . . . *Target (Diana, En)*: The proportion of this age group . . .	2	1
	Wrong unit of measurement (UoM)	The interpreted UoM differed from the SL one.	*Source (NIU)*: from 28 *mtpy* ((million tonnes *per year*)) . . . *Target (Carlo, It)*: da 28 *milioni di tonnellate* . . . *Target (Carlo, En)*: from *28 million tonnes* . . .	7	5
	Wrong relative value	The interpreted relative value differed from the SL one.	*Source (NCR-2)*: The growth would push the number of 15- to 64-year-old Africans up by 40% by 2030. *Target (Minerva, It)*: E questo sposterà la fascia demografica tra i 15 e i 64 anni *al* 40% entro il 2030. Target *(Minerva, En)*: This will move the 15- to 64-year-old population *to* 40% by 2030.	3	2
	Misattribution of components	The semantic links between the components of the interpreted NIU did not correspond to the SL ones.	Source (NCN-1,-2): In 2019, Africa produced nearly 8.41 mbd of oil. Madagascar alone produced approximately 58,000 metric tons of nickel in 2021. Target (Sally, It): Nel 2019, l'Africa ha prodotto 8,41 milioni di barili di petrolio al giorno, così come 58 (.) mila tonnellate di nickel. Target (Sally, En): In 2019, Africa produced 8.42 million barrels per day of oil, as well as 58,000 tonnes of nickel ((misattribution to the referent and time location of the previous NIU)).	4	3
	Sentence fragment	A NIU did not express a complete thought, as one or more of its essential components were missing.	*Source (NCN-3-a)*: Namibia's diamond production amounted to 2.52 million carats. *Target (Sally, It)*: La Namibia ha prodotto 2,52 milioni di carati. *Target (Sally, En)*: Namibia produced 2.52 million carats ((referent missing)).	7	4
	Omission of the NIU	The whole NIU was omitted.	*Source (NCR)*: By 2030, Africa would hence be home to nearly 1 billion people of 15 to 64 years of age. *Target (Sally)*: Ø	1	1

Texts	*Inconsistent numerals*	The numerals within the interpreted numerical task were contradictory.	*Source (NCR-1, -3)*: By 2030, the African continent *would add about 295 million new people aged 15 to 64. . . .* By 2030, Africa would hence *be home to nearly 1 billion people of 15 to 64 years of age.* *Target (Minerva, It)*: Sempre entro il 2030, il continente Africano *avrà circa 295 milioni di abitanti in età dai 15 ai 64 anni. . . .* Sempre entro il 2030, l'Africa *avrà 1 miliardo di abitanti in età dai 15 ai 64 anni*. *Target (Minerva, En)*: Again, by 2030, the African continent *will have about 295 million people aged 15 to 64. . . .* Again, by 2030, Africa will have *1 billion inhabitants aged 15 to 64*.	5	3
	Distortion of information	The meaning of the interpreted numerical task, albeit internally consistent and plausible, differed substantially from the SL.	*Source (NCN-1-b)*: Africa produced nearly 8.41 *mbd of oil*. *Target (Carlo, It)*: Abbiamo prodotto circa 8,41 *milioni di barili al giorno.* *Target (Carlo, En)*: We produced approximately 8.41 *million barrels* per day.	9	4
Context	*Plausibility error*	The interpreted numerical information seemed unreasonable and improbable against the world knowledge of an informed listener.	*Source→ (NCR-3-b,-c)*: Africa would hence be home to nearly *1 billion people aged 15 to 64*. *Target (Diana, It)*: l'Africa darà domicilio a *oltre 1 miliardo di persone.* *Target (Diana, En)*: Africa will be *home to over 1 billion people.*	0	0
Function	*Functional error*	Although the numeral and all components of the NIU corresponded to the SL NIU, the interpreted message differed from the original one in its function.	*No example found in this study; the example below was reported in Author (2019a, p. 93).* *Source*: First, let me thank our more than 66,500 employees for making our success in 2013 possible. *Target*: We have over 66,000 employees.	4	3

Strategy	*Omission of redundant item*	The interpreter omitted an item (the numeral or another component of the NIU) that was repeated within the numerical task.	*Source (NCR-1,-2,-3)*: By 2030, . . . by 2030 . . . by 2030. *Target (Carlo, It)*: Entro il 2030, . . . sempre entro lo stesso anno . . . Ø *Target (Carlo, En)*: By 2030, . . . by that same year . . . Ø		
	Abbreviation of acronym	The interpreter simplified the referent or the UoM through the use of an acronym.	*Source (NU)*: The continent currently has a *gross domestic product* of USD 3.42 trillion. ((CAI shows: prodotto interno lordo)) *Target (Logan, It)*: Il continente ha attualmente un *PIL* che è di 3,42 bilioni di dollari. *Target (Logan, En)*: The continent currently has a *GDP* of 3.42 trillion dollars.	3	2
	Lexical substitution	The interpreter replaced a component of the NIU with its non-numeric equivalent or through anaphoric reference.	Source (NCR-2-a): The growth would push the number of 15- to 64-year-old Africans ((repeated item)) . . . Target (Carlo, It): E le persone in questa fascia di età . . . Target (Carlo, En): People in this age range ((anaphoric reference to numeral expressed in the previous sentence)) . . .	4	2
	Generalization of the numeral	The interpreter replaced the numeral with a general expression to form a sentence of finite meaning.	*Source (NCN-3-b)*: in 2018 *Target (Minerva, It)*: negli ultimi anni *Target (Minerva, En)*: over the past years	4	3
	Summarization	The interpreter summarized the meaning of the numerical information.	*Source (NCN-2-a)*: Madagascar alone produced approximately 58,000 metric tons of nickel. *Target (Carlo, It)*: Il Madagascar è una grande risorsa per il nickel, che continuerà a crescere. *Target (Carlo, En)*: Madagascar represents a great resource for its nickel ((production: in the previous sentence)), which will continue to increase.	3	1

Endnotes

1 This is testified by the birth of and strong interest around research projects, such as EABM–*Ergonomics for the Artificial Booth Mate* (eabm.ugent.be), led by the Johannes-Gutenberg University of Mainz/Germersheim and the University of Ghent, and the EU-funded Innovation Activity *SmarTerp* (smarter-interpreting.eu).
2 www.smarter-interpreting.eu
3 The discussion of methodological issues is informed by the principles contained in such manuals as Creswell (2018).
4 https://smarter-interpreting.eu/
5 eit.europa.eu
6 International Association of Conference Interpreters: aiic.org
7 Italian Association of Translators and Interpreters: aiti.org
8 Italian Association of Conference Interpreters: assointerpreti.it

All examples were translated from Italian into English by the researcher. Original samples are provided in the appendix.