Effective Editing in Respeaking: Unveiling Uniquely Human Skills in Live Speech-to-Text Transformation

Tomasz Korybski 1[0000-0003-2353-0816] and Elena Davitti 2[0000-0002-7156-9275]

1 University of Surrey, Guildford, UK
2 University of Surrey, Guildford, UK
t.korybski@surrey.ac.uk
e.davitti@surrey.ac.uk

Abstract. Set in the context of a rapid technological development in translation, interpreting and speech recognition technology, this paper takes a close-up view of an area where human input still seems to be an indispensable and very hard-to-replace: the area of live speech transformation. To contribute to answering inevitable questions about the future role(s) of humans in real-time language-related practices, we focus on effective edition (EE, Romero-Fresco and Pöchhacker 2018) as the uniquely human skill to deviate from the source message in a way that does not cause any loss of information and may even improve the final output. Firstly, we discuss this concept in the context of earlier research on paraphrasing in translation, reduction in respeaking, and reformulation in interpreting. Secondly, we explore this form of language transformation in practices characterised by immediacy in the way content is conveyed, focusing on the spoken-to-written form. To show the predominant types of effective editions (EEs), we use authentic data from human intra- and interlingual respeakers collected during two related research projects conducted at the Centre for Translation Studies (University of Surrey) in the years 2019-2023. The data in the paper show a variety of EEs live human interventions that can be mostly grouped into forms of condensation, re-expression, and compensation. Our discussion will centre on the diverse form that EEs can take in various practices, and the importance to recognise that they are context-dependent and can entail dynamic content shifts. We will conclude by emphasising the importance of understanding and categorising the range of these human-performed edits for evaluation, training, and the future advancement of automated solutions powered by large language models.

Keywords: Effective Editions, Interpreting, Respeaking, NER Model, NTR Model, Live Speech-to-Text
Rationale and Objectives of this Paper

The object of analysis of this paper is the phenomenon of effective editing in the live speech-to-text practice of intra- and interlingual respeaking. Respeaking can be defined as a “technique whereby a respeaker listens to the original sound of a (live) programme or event and respeaks it in [in the same language, or translates it in another language and in real time], including punctuation marks and some specific features for the deaf and hard-of-hearing audience, to a speech recognition software, which turns the recognized utterances into subtitles displayed on the screen with the shortest possible delay” (adapted from Romero-Fresco 2011: 1). This is a challenging task, which requires rapid decision-making to be able to let the original meaning travel through a language (from source to target language) and a diamesic transfer (from spoken input to written output).

To this end, respeakers may reformulate in real time, either in the same language (in intralingual respeaking) or in another language (in interlingual respeaking), and this phenomenon is the focus of this paper. Romero-Fresco and Pöchhacker (2017: 159), define Effective Editions (EEs) in respeaking as deliberate “deviations from the source text that do not involve a loss of information or that even enhance the communicative effectiveness of the subtitles”. Such ‘deviations’ are representations of respeakers’ flexibility in transforming the source for the audience without a loss of meaning. To contextualise the phenomenon, we notice its affinity with ‘paraphrasing’ and ‘quasi-paraphrasing’ – interventions aimed at tackling the lexical, grammatical, delivery-related, or stylistic challenges posed by source content. EEs are highly context-specific, which results in a set of distinctive features of effective editing that we present in the examples (section 2.3), seeking to highlight the differences between the live speech transformation workflow within one language (intralingual respeaking) and across two languages (interlingual respeaking).

Rather than striving to produce an exhaustive catalogue of editing phenomena in respeaking, our datasets have led us to suggest a broader, tri-partite classification of EEs that may prove useful in evaluation and training. The extensive dataset gathered in one of the projects that have contributed data for this paper, SMART (see section 2.1), has enabled us to identify potential patterns bottom-up and develop a taxonomy from them. Importantly, in SMART we discovered a statistically significant relationship between the number of EEs and performance, indicating that a higher use of EEs leads to higher content accuracy. As a result, we believe there is a need for more comprehensive research into EEs, which could offer valuable insights for training and quality assessment, as well as for informing human-like transformation practices in speech-to-text and speech-to-speech communication.

Furthermore, after reviewing the existing research on live transformation of speech by human professionals working in live assignments (intra- and interlingual respeakers, interpreters), we have identified a gap in the current body of knowledge on the topic that can be (at least partially) filled with the data we have accumulated from two research projects presented in section 2.1 below. Firstly, although researchers have recognised similarities in paraphrasing/transformation activities implemented by respeakers and interpreters, no taxonomy or classification has been created so far to capture what is generically expressed by Romero Fresco’s and Pöchhacker’s (2017) definition.
Following the near-equivalence paradigm and the concept of quasi-paraphrasing, we therefore take a more specific look at effective editing phenomena to classify them in a pragmatic, evidence-based and bottom-up manner. We claim that the notion of effective editing is useful to capture and describe a range of transformation practices applied by respeakers and, by extension, other real-time speech-to-text practices, because it goes beyond the definition of paraphrasing, reformulation and explicitation previously applied in research. Secondly, to add to the existing research on similarities and differences between intralingual and interlingual respeaking, we show how respeakers use effective editing strategies in concrete contexts. Thirdly, we present live editing strategies as a prime example of a (still) uniquely human skill that may pose a challenge for automated solutions, especially in real-time scenarios. We situate the presentation of EEs from our research within the functional paradigm in translation and interpreting studies, specifically Gile’s (1990), Seleskovitch’s and Lederer’s (1975) théorie du sens and Vermeer’s (1978) skopos theory. According to this approach, the language professional’s decision-making process in determining what is a functional and, therefore, non-omissible component is contextual, and may vary on a moment-by-moment basis, thus impacting how target messages are formulated also in respeaking practice.

1.1 Relevant Theories in Language Transformation

All speakers of a language possess the ability to transform speech and text using a range of methods, which may vary depending on specific communication and situational needs. From basic paraphrasing to advanced summarization, explicitation, or reduction skills, humans transform language in a variety of ways, within and across languages (as in translation, interpreting) as well as modalities (as in intra/interlingual respeaking). Moreover, different modes of content delivery can be used in this process, such as the use of visual representations like pictograms or emojis in place of letters of the alphabet.

The transformation skills discussed here became a focal point for research already in the 1950s, when the term paraphrasing was first widely used and researched by Chomsky (1956). Chomsky’s Transformational-Generative model of language explained the transformation of language through syntactic operations, such as reordering and substituting words. The notion of a single underlying structure of language that can be subjected to different rules to modify it gained traction and impacted further investigation of how language can be transformed (including early research on Machine Translation, Bar-Hillel 1953). Since then, research into linguistic paraphrasing has expanded, with new theories and methods being developed. An impactful theory of linguistic paraphrasing is the Meaning-Text Theory proposed by Zholkovsky and Mel’čuk (1965). This theory suggests that meaning is derived from the context of language, with a lot of emphasis on the semantic (lexical) layer which is the foundation for syntactic and morphological processing based on nodes and arguments linking lexemes.

More recently, with the development of machine translation and natural language processing in general, a linguistic analysis of paraphrasing has become an area of interest for computational linguists. In their presentation of paraphrasing types, Bhagat and Hovy (2013) contend that a more elaborate definition of paraphrases is necessary, departing from the generic understanding of paraphrases as utterances that carry the same
meaning as the original utterance but expressed with a different wording. To this end, Bhagat and Hovy build on linguists such as De Beaugrande and Dressler (1981), Mel’chuk (2012) and Clark (1992) to use the notion of *approximate conceptual equivalence* of paraphrases or ‘quasi-paraphrases’, which moves away from the unrealistic expectation of strict synonymy that may have been inferred from the generic definition. Bhagat and Hovy introduce a set of 25 operations to specify how paraphrasing can be performed – all of which fall under the supra-categories of re-expression and condensation that we propose later in this paper. They argue that although paraphrasing is a complex and multifaceted transformation activity, it can be structured by listing different paraphrasing phenomena in a systematic way. We have adopted a similar approach to categorising EEs but decided to introduce supra-categories including the category of compensation which is intricately linked to live speech transformation and as such is not captured by Bhagut and Hovy’s analysis. The value of having a simple taxonomy is that it facilitates bottom-up analyses and may lead to streamlining both qualitative and quantitative analysis of live speech-to-text and speech-to-speech data.

1.2 Effective Editions, Paraphrasing and Strategic Reformulation in Real-Time Language-Related Practices

The classification quoted above is important for our discussion of transformation of live speech as it offers a broad catalogue of phenomena, many of which we have encountered in our data, too. However, it does not capture the specificity of live speech transformation in full – hence this contribution. Before discussing the categories of effective editing from our data, it is important to define the relevant professions and the settings where live transformation of speech takes place, and how different yet related concepts have been defined to refer to relevant phenomena of language transformation.

**Simultaneous Interpreting** is when a human interpreter listens to and simultaneously interprets the speech into a target language, usually with special equipment (AIIC 2013). Interpreting is referred to in this paper because researchers have underlined its affinity with respeaking (Robert and Remael 2017, Eugeni 2008, Romero-Fresco and Pöchhacker 2018, Davitti & Sandrelli 2020). Drawing up a competence model for respeakers, Pöchhacker and Remael (2019) emphasize the elements that interlingual respeaking shares with interpreting, i.e., the interlingual and real-time nature of the effort, with the added and challenging component of editing. Due to the required immediacy of simultaneous interpreting and respeaking, the decisions of the language professional performing the task must be made with minimum latency, and the transfer between two languages means that complex equivalence challenges often need to be resolved within seconds or even fractions of a second. This promotes paraphrasing as an efficient strategy to satisfy the audience’s communication needs while meeting the time pressure involved in this technique. Researchers interested in the process of interpretation, and especially its Production Effort component (Gile 1995) have looked at paraphrasing as an effective and indispensable tool for interpreters. Paraphrasing and strategic reformulation exercises have long been present in the core of most interpreter training curricula (Moser Mercer 1998). Furthermore, Seleskovitch (1978) and other researchers proposed a model of interpreting theory that includes a *deverbalisation* phase between
understanding the source message and re-expressing its content in the target message. This means that content is processed semantically, focusing on its sense rather than on its lexical form. In a related vein, Gumul’s (2017) extensive analysis of explicitation techniques in interpreting builds on the work of more than 10 interpreting researchers and offers a detailed breakdown of explicitation-type reformulations into categories which are also highly relevant for the bottom-up analyses we performed in both our projects.

**Respeaking.** Language transformation in respeaking has been a topic of research for over a decade now (e.g., Romero-Fresco 2011). An important part of the research (and the pragmatic perspective on respeaking as a service with concrete clients and end users) has been the evaluation of respeakers’ output. The approach most widely used in the assessment of accuracy of intra- and interlingual respeaking is based on errors (the NER and the NTR models, respectively – see 2.2 below for more details) identified in a close-up linguistic analysis of (fragments) of transcribed source inputs and target outputs. However, the two recognised models for product accuracy assessment include a component of edition evaluation (called correct editions in the NER, and effective editions in the NTR) intended as positive shifts in the target text with respect to their corresponding source. As previously explained, these represents a positive deviation from the source, which are not penalised in the model since they do not alter the meaning or result in any loss of information. In fact, such changes may even improve the effectiveness of the message being conveyed.

Even though paraphrasing has been mentioned as a common denominator for practices where content is transferred across two languages, it is important to emphasize that in the context of written translation the source text is often ‘clean’ and in its final version. Conversely, source speeches performed in a live settings are frequently abundant in ad-hoc formulations, redundancies, disfluencies, overlapping speech, or simply sub-optimal quality of audio due to speaker’s pronunciation and background noise, all of which place an additional burden on the human professional’s speech transformation. Importantly, this overarching challenge is shared by both interpreters and respeakers, leading to a unique set of transformation phenomena discussed in this paper. Researchers emphasize the similarities between the two professions as regards preparation, processing, and delivery: the ‘simultaneity’ of reception of audio and own production, balancing the capacity of one’s own working memory, processing and output monitoring are indispensable for the service to meet its communication objectives. These similarities have led Chmiel et al. (2018) to conduct a study on paraphrasing skills in interpreters, respeakers and bilinguals where interpreters tended to be better at semantic redundancy elimination and the ensuing production of concise output, although the advantage was not very clearly pronounced (Chmiel et al. 2018). In a study focusing on reduction practices in (intralingual) respeaking, Luyckx et al. (2010) found that condensations and omissions seem to be a planned process actively shaped by respeakers depending on external factors such as source text speed and availability of ‘omissible’ respeaking units. The same study also presents a list of source reduction strategies that enable the respeaker to retain (much of) the original information despite a significantly condensed output. These include some of the strategies that are also included in our
classification in later sections: omissions, shifting questions into affirmatives, and simplifying indicators of modality *(ibid*, pp. 31-33).

2 Effective Editing in Respeaking Datasets – Classification and Comparison

2.1 Sources of Data

For the purpose of this paper, we looked at EEs across the empirical datasets from two distinct projects carried out at the Centre for translation Studies (University of Surrey). Below we provide a brief overview of both projects.

The SMART project *(Shaping Multilingual Access through Respeaking Technology*, Economic and Social Research Council UK, ES/T002530/1, 2020-2023) aimed to investigate the emerging technique of *interlingual respeaking* for delivering real-time speech-to-text services across languages. Specifically, the project analysed the interlingual respeaking performances of 51 language professionals from various backgrounds, including consecutive and simultaneous interpreting, written translation, pre-recorded subtitling, and live subtitling. The study aimed to explore the competences underlying such complex process, the level of accuracy achievable by language professionals with at least 2,000 hours of practice in one or more of the disciplines listed previously, and how to optimise upskilling based on empirical research insights. Participants received 25 hours of bespoke training-for-testing course, which exposed them to both intralingual and interlingual respeaking. They were required to speak English paired with Italian, Spanish, and/or French, with at least one of these languages as their mother tongue. They also underwent six tests, including three intralingual and three interlingual ones, designed to expose participants two different scenarios: speed (fast speakers), planned/unplanned delivery (partially improvised, partially prepared speech), and multiple speakers (interview scenario). This paper refers to the interlingual dataset generated by SMART, which comprises 153 performances of 15 to 20 minutes each, totalling over 2,300 minutes of recorded and analysed performance. This dataset provides a rich source of interlingual respeaking data with aligned sources and targets which were evaluated using the NTR model for accuracy evaluation.

The MATRIC project *(Machine Translation and Respeaking in Interlingual Communication*, Expanding Excellence in England, 2019-2022) investigated an alternative speech-to-text interlingual communication workflow involving a human intralingual respeaker providing live subtitles in English via speech recognition software (Dragon) and machine translation (the EU’s eTranslation) to translate the English subtitles into multiple languages: French, Spanish, Italian, Polish, German, and Romanian. First, an evaluation of intralingual respeaking data from four professional respeakers performing a total of 12 speeches (each professional respoke intralingually three authentic English language speeches from the European Parliament’s speech database) enabled us to single out the best performances, which were later machine-translated to produce a total of 18 performances. 12 of these (for Italian, Spanish, French and Polish) were analysed and compared with the corresponding 12 transcripts of actual performances of four interpreter tandems from the European Parliament’s booths working at the same events (also available through the EPTV database). Source speech duration
varied between 3 and 14 minutes, representing the typical range of length of Europarl debate interventions. Additionally, care was taken to select speeches with varied speech types (improvised and partially improvised, varied topics, speakers with different speeds and accents including one native English accent). Overall, approximately 200 minutes of speech were analysed using both the NER and NTR methods, as explained in the next section.

2.2 Capturing Effective Editions – the NER and NTR Models

Both projects focused on accuracy, measured via the NER and NTR models for intralingual and interlingual evaluation respectively. The NER model (Romero-Fresco 2011, Romero-Fresco and Martínez 2015) is a recognised method for evaluating the accuracy of live subtitles produced through intralingual respeaking in media or live event broadcasts. The letters in the model’s acronym stand for the total number of words in the live subtitles (N), edition errors (E), and recognition errors (R). The percentage of accurate content is calculated by subtracting the E and R values from the N value, and then dividing it by N. The NTR model (Romero-Fresco and Pöchhacker 2017) is based on the NER model and used to evaluate interlingual respeaking data. In the NTR model, translation errors (T) replace the edition errors to account for the accuracy of interlingual transfer. Errors captured with the NTR include omission, addition, and substitution errors (content errors), as well as correctness and style (form errors). Despite a different naming convention in each of the models, the levels of severity in both the NER and the NTR are as follows: minor (0.25 penalty point), standard/major (0.5 penalty point), and serious/critical (1 penalty point). Importantly and crucially for the models’ differentiation from WER-like models, despite the focus on errors, they leave scope for capturing EEs (also referred to as Correct Editions in the NER model). The analysis grid we used was adapted by Davitti and Sandrelli (2020) from the NER score spreadsheet for the evaluation of intralingual respeaking data. The spreadsheet used for NER and NTR evaluation allowed for segmentation and alignment of source and target’s idea units, to enable investigation at a micro level. The grid also features a special column for capturing and commenting on EEs, which makes it possible to count them as well as extract and examine more closely examples of interventions by language professionals. This repository was analysed qualitatively by six researchers with extensive experience in respeaking and interpreting research. Importantly, as the SMART project offered very extensive data, initial characterisation and taxonomy was done on the basis of SMART’s dataset and then applied to MATRIC. As a result of the analysis, we understood that, firstly, it is possible to distinguish predominant categories of EEs, and secondly, that the current definition of EEs could benefit from expanding and specification based on our authentic data. Below we present these categories with brief definitions and examples sourced from our data.

2.3 Classification and Examples of Effective Editions

Upon separate analysis of the datasets for the purpose of their respective projects, we discovered a clear pattern in the EE types identified, which allowed us to categorise them into three distinct supra-categories. The first category consists of non-penalised
omissions, which typically resulted in effective condensation, as not all omissions lead to information loss. The second category is re-expression, which can be broken down into two major dimensions: lexico-semantic and structural. The third category identified in the data is compensation, which occurs when missing information from a previous fragment is compensated in a later fragment. The following examples and extended definitions will provide a more comprehensive explanation of our tripartite classification. Please note that for the purpose of succinctness and consistency in this paper, we focus on English into English intralingual respeaking and Spanish into English / English into Spanish interlingual respeaking. However, we discovered the same patterns across all other language combinations explored in the two projects. Rather than producing an exhaustive catalogue of all possible instances of condensations or re-expressions, in this paper we focus on presenting examples that showcase the general mechanics of effective editing. The sequence of examples in each supra-category is as follows: examples from the intralingual dataset followed by examples from the interlingual dataset. The effectively edited fragments are in bold print. ‘BT’ stands for back-translation (in the interlingual examples).

**Condensations**

Condensations are the predominant category in the datasets. At micro level, they occur when a source idea unit is compressed and expressed in a shorter form in the target idea unit without (any considerable) loss of information. In our data, condensation was implemented primarily through omissions of redundant information, deictic expressions, or grammatical interventions. Condensations at macro level occur when a target idea unit can be expressed in a much more concise way by referring to the preceding idea unit or utilizing elements of content from it, and when a target idea unit captures more than one source idea unit. This type of condensation is typically achieved using pronouns to replace names, or deictics such as this, that, these, those, now, then, here.

The examples below show two types of condensation sourced from the datasets. All the source data transcriptions contain all the words that were in fact uttered by the original speaker. As a result, repetitions and redundancies are present in the source transcripts to demonstrate the full scope of the human intervention in the target.

**Example 1 (intralingual respeaking)**

_source:_ Despite the the dramatic and I would even say tragic events **we are just going to discuss in a minute** the first time we see each other after the after the Christmas break and **therefore I really** would like to wish all of you **all the citizens you represent and all the European Union** all the best in the New Year and happiest 2020

target:_ Despite the dramatic and I would even say tragic events, it's the first time we have seen each other after Christmas and I would like to wish all of you all the best in the New Year and the happiest 2020.

Example 1 was significantly condensed by the respeaker: the original 330 characters with spaces were turned into just 205 characters with spaces in the respoken version. This resulted in shorter, more readable captions, with the message largely intact.
Thanks to the respeaker’s intelligent interpretation of the context, they were able to leave out a whole string of words that did not contribute important information, thus streamlining the entire process for themselves and for the recipients. This depth of intervention could not be expected from an ASR solution – although some redundant elements such as the obvious repetitions resulting from hesitation or uncertainty would have been cut out by currently available speech recognition tools.

**Example 2 (interlingual respeaking)**

**Source:** Entiendo que puede ser algo confuso así que quizá debería explicarlo mejor  
**BT:** I understand that it can be somewhat confusing, so perhaps I should explain it better.  
**Target:** It might seem confusing, so I should clarify that.

Example 2 features condensation on two levels: firstly, the rapport-building ‘I understand’ which has an important function in conversations, but not necessarily in live captions, is dropped. Secondly, ‘quizá / perhaps’ is dropped, leaving the transfer of the sentence’s modality to the modal verb ‘should’. Arguably, these changes are justified, although the lack of the ‘softening’ of the sentence’s style by means of ‘perhaps’ results in a more direct affirmative clause. Overall, however, and in the wider context of this example, these modifications bring no detriment to the original’s message and result in improved readability. Although there may be contexts where even a slight shift in modality will interfere with the message conveyed (e.g., a legal context), the respeaker’s judgment of the situation (often supported by an assignment brief and preparation) should limit editing to the fragments of source text that lend themselves well to such transformation.

The condensations shown above show the uniquely human skill of editing based on split-second decisions and contextual judgment that result in well-readable output either in the same or a different language. Although today’s large language models and text summarization solutions cope well with text reduction (leading to condensation), there remain at least two major areas of human superiority. Firstly, apart from redundant content, all content words in automatically summarized fragments of texts tend to stay in the output. Humans, in turn, are able to “filter out” non-crucial words, including content words, based on the context. This results in succinct and easy-to-read captions. Secondly, human respeakers can switch their condensation practice on and off very dynamically within one assignment, while an automated solution would need to be prompted separately for selected fragments of the source speech to be able to mimic a human respeaker’s behaviour. These two challenge areas, of course, come on top of all the existing challenges related to ASR (such as overlapping speech, background noise) providing input for any human-like editing operations in a cascaded system.

**Re-expressions**

Re-expression is about effectively using the lexico-semantic and structural possibilities afforded by the source content to produce successful renditions suitable for live captions. Re-expression can take the form of semantic, syntactic and stylistic interventions
that do not interfere with the message and provide a more readable output thanks to, for example, voice change, sentence splitting, sentence merging or additions or substitutions that are deemed positive and are, therefore, not penalised in NTR evaluation.

Example 3 (intralingual respeaking)
Source: On Sunday yesterday my colleague Commissioner Lenarcic reached out and spoke to the Australian minister Littleproud and be reiterated the Union’s readiness to assist Australia in this moment of crisis
Target: On Sunday, yesterday, my colleague Commissioner Lenarcic reached out and spoke to the Australian Minister Littleproud and stated the union’s readiness to support Australia in the crisis.

Example 3 shows a situation where successful re-expression at a syntactic level (merging two utterances) and lexico-semantic level (verb substitution) leads to slight condensation of the target, too. In fact, these two categories of effective editing often co-occur in our data, although re-expression may also lead to the opposite of condensation, i.e., extension of the target if content explicitation is necessary following the respeaker’s assessment of the caption reader’s needs. In Example 3, both merging the sentences and verb substitution produce a more succinct target: the subject does not need to be repeated, and the verb ‘state’ reads much shorter than ‘reiterate’. The omission of ‘this moment’ is also justified by the use of the definite article and the noun ‘crisis’, which provide enough clarity given the context.

Example 4 (interlingual respeaking)
Source: Entonces, ¿significa esto que las personas que se dedican al rehablado intralingüístico pues que no tienen experiencia como intérpretes pueden dedicarse al rehablado interlingüístico?
BT: So, does this mean that people who perform intralingual respeaking, since they do not have experience as interpreters, can still perform interlingual respeaking? I would say yes, absolutely
Target: So this means that people who are intralingual respeakers who are not interpreters can work as interlingual respeakers if they work hard.

In Example 4 we can see re-expression through two interventions: at a structural level, i.e., the respeaker changed the source’s rhetorical interrogative into an affirmative – as the answer was provided by the speaker in the next sentence, the shift was possible; and at a lexico-semantic level, through an addition which is perfectly plausible on the basis of the context preceding this segment and actually clarifies it further. Such processing is evidence of the respeaker’s understanding of the speaker’s intention, and their ability to operate syntax and form to express meaning to the best of their abilities.

Compensations
Working with live input provides respeakers with unique opportunities to change the sequence of information presentation if such a change is possible, i.e., in situations when chronology of the presentation is not crucial. Compensation consists in providing
missing information from a previous idea unit (which can be either its part or even an entire idea unit) in later target idea units. Although typically compensations span adjacent idea units, it is also possible for language professionals to compensate for content provided a few idea units earlier.

**Example 5 (intralingual respeaking)**

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>And, of course, we also have the pension pay gap which is very very serious.</td>
<td>[no rendition]</td>
</tr>
<tr>
<td>The current gender pension gap in Europe stands as over double the gender pay gap at 35.7%</td>
<td>The current gender pension gap in Europe is serious and stands at over double the gender pay gap, at 35.7%.</td>
</tr>
</tbody>
</table>

**Example 6 (interlingual respeaking)**

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entonces, ¿significa esto que las personas que se dedican al rehablado intralingüístico pues que no tienen experiencia como intérpretes pueden dedicarse al rehablado interlingüístico?</td>
<td>[no rendition]</td>
</tr>
<tr>
<td>Yo diría que sí absolutamente pero necesitas superar algunos obstáculos mentales del síndrome del impostor</td>
<td>I would say yes, absolutely, but you need to overcome some of the imposter syndrome’s mental hurdles</td>
</tr>
</tbody>
</table>

Examples 5 and 6 show that source information can be compensated if the respeaker’s working memory allows it. Please note that a compensation can also include a different editing category such as re-expression – as is the case in Example 6. Importantly, due to its non-chronological nature, compensation is also uniquely human.

### 2.4 Discussion

The examples above serve to identify the editing strategies typically employed by intralingual and interlingual respeakers. Although such strategies can be found ‘in isolation’, it is natural for them to co-occur within the same idea unit, as in Example 6. Furthermore, the presence of EEs in interlingual respeaking data may be more pronounced than in intralingual respeaking data. This can be mostly ascribed to the language transfer component, which makes a fully verbatim approach difficult to implement.
Based on the quantitative analysis conducted in SMART, we found EEs were a positive predictor of accuracy, thus confirming their impact on improving interlingual respeaking output, including the readability aspect. This therefore means that learning more about EEs is important for more precise evaluation, goal-oriented training / up-skilling of language professionals, and possible future (partial) automation of the task.

3 Conclusions, Impact and Further Research

Impact on Data Evaluation and Training. The examples from respeaking data we have cited in this paper show that live spoken text can be subject to a varied range of transformations that we have classified into three macro-categories as condensations, re-expressions and compensations. The extent to which these transformations are applied can depend on many factors, including the briefing of the respeaker, the features of the original speech including speed, structural complexity, genre, and, last but not least, the professional’s individual inclinations that may result from one’s idiosyncratic disposition, experience and/or training. Our awareness of the range of effective transformations that are applied by respeakers can impact how these professionals are trained. The three broad categories we have proposed may prove useful for providers in identifying any existing strengths and possible training needs for language professionals, and the range and modulation of shifts that can be suitable in a specific scenario. Furthermore, a deeper understanding of EEs can contribute to a more comprehensive and objective application of the NER and NTR models: for example, in some live settings the positive impact of successfully applied EEs may offset some shortcomings of the target (such as minor content omissions). In the near future we plan an experimental re-calculation of the NTR scores from the SMART project including a positive value for EEs to gauge their impact on overall scoring. In the context of education, although strategic reformulation exercises have been part and parcel of the training of various language professionals for decades now, there appears to be the need to study them further in relation to hybrid practices. This is particularly relevant as traditional disciplinary boundaries are blurring, with multimodal practices (like respeaking) emerging and requiring further investigation. Moreover, there is a need to develop a deeper understanding of the frequency of occurrence of each of EE categories to decide what to prioritise in training. Existing and future data sets can be analysed quantitatively and additional variables such as the impact of source genre on effective editing will need to be taken into consideration when shaping future training courses.

Impact on Resource Development and Automation. Apart from training and evaluation, the classification we have proposed may be used to showcase any shortcomings of existing prototypes in machine interpreting and ASR-based live captioning services. The recent rapid progress in large language models such as GPT 4 also offers promising potential for future development of automatic human-like editing of live speech for the purpose of live subtitling or even interpreting. However, to effectively teach the model and expect consistently good outcomes, we first must understand the phenomenon
thoroughly and collect sufficient data. What we need as a next step is to create larger databases of EEs across different live language-related practices to gain a better quantitative and qualitative understanding of the phenomenon. This will then enable us to identify best practices from human data to train automated large language models.

References


