CANADIAN NER EVALUATION GUIDELINES

Version: 2018-11-02

PURPOSE AND USE OF THESE GUIDELINES

These Guidelines establish a standard set of practices for NER evaluators in Canada to enable consistent scoring. Details differ from NER as used in the UK, to adapt to Canadian broadcasting and captioning practices, but the developer of NER, Dr. Pablo Romero-Fresco, has reviewed them and attests that scores will be comparable.

NER SCORING INTERPRETATION GUIDE

NER "ERRORS"

NER measures the accuracy of captions by comparing the caption viewer's experience with that of the hearing viewer. NER evaluation begins with the preparation of a verbatim transcript and a caption transcript for ten minutes of live-captioned content in a television program. Details of this preparation are described below. The NER evaluator then scores captions by comparing the two transcripts and assigning an "error" type – with its score deduction – to any differences. One or several errors may apply, chosen from six possible error types, when:

- there is a loss or change of meaning, or;
- there are words/phrases that do not alter meaning, but cause the caption reader to interrupt the reading experience to attempt to understand them.

The six error types and their deductions are:

- **Correct Edition (CE)** (-0.0). CE is scored when captions are different from the verbatim audio but retain its full meaning, without interrupting words/phrases.
- **Omission of Main Meaning (OMM)** (-0.5)
  The captions have lost the main idea presented in the audio's Independent Idea Unit (see page 2).
- **Omission of Detail (OD)** (-0.25). Here, the captions have lost one or more modifying meaning, affecting a Dependent Idea Unit (see page 2) but not the main idea.
- **Benign Error (BE)** (-0.25) A captioned word or phrase is incorrect, causing an interruption in the reading. However, the viewer can readily figure out the original meaning, from context (in video) or similarity to the real word.
- **Nonsense Error (NE)** (-0.5) A captioned word or phrase is incorrect and the viewer can't figure out the original meaning. If the impact of the word/phrase is to alter or omit the meaning of the idea unit, OD, OMM or FIE will be scored instead.
- **False Information Error** (**FIE**) (-1.0) The captions make sense, but the information they present is different from the verbatim. The caption viewer cannot tell that the meaning is false.

IDEA UNITS AND ERRORS

NER divides the transcripts into "Idea Units" [1] since it deals primarily with comparisons of meaning, and the grammatical divisions of sentences and clauses often do not correspond exactly to meaning. In oral communication a single idea may be spread over several sentences. Also, one sentence may contain an idea that functions as a modifier of another sentence.

Idea Units in NER are of two types – independent and dependent.

An **independent** idea unit, such as "The blaze started this morning at the front of the house", is the oral equivalent of a sentence, makes sense as a full, independent message and may be composed of several dependent idea units, such as "this morning" and "at the front of the house".

> "Omission of Main Meaning" is an error that often consists in the omission of a full independent idea unit (which may not be noticed by the viewer) or of a dependent idea unit that renders the independent unit meaningless or nonsensical.

A **dependent** idea unit is often a complement or modifier and it provides information about the "when", the "where", the "how", etc. of an independent idea unit.

> "Omission of Detail" is less destructive of meaning than OMM because its impact is limited to dependent idea units – words or phrases that modify or add detail to the main idea. They do not render the independent unit meaningless or nonsensical.

"Nonsense Errors" and "Benign Errors" do not apply to idea units because they are about interruptions, not meaning. These errors are scored when a word or phrase is misspelled, garbled, or in some way interrupts the reading experience while the viewer considers it.

- A "Benign Error" is scored when viewers can understand and sometimes reconstruct the original words. Typical cases are missing apostrophes, missing punctuation (such as "?"), misspellings that are close to the original, etc.
- A "Nonsense Error" applies when the word or phrase cannot be understood, causing greater reading disruption.

Note that NER scoring does not consider what it is reasonable to expect of a captioner. If sounds are part of the hearing viewer's experience (i.e. are intelligible), and their meaning is not in the captions, an error will be scored. The difficulty of captioning a show – due to rapid speech or other factors – should be noted in the evaluator's comments but does not affect the scoring.

---

[1] The definition of "idea unit" is drawn from the base document in NER: *Accuracy Rate in Live Subtitling – the NER Model by* Pablo Romero-Fresco and Juan Martínez, 2014.

THE DECISION TREE

Deciding which error applies can be clarified by using a "decision tree" approach, in which the evaluator looks at an idea unit with caption-verbatim discrepancies and asks a sequence of questions that lead eventually to the right "error". The evaluator then moves on to the next idea unit with a discrepancy and applies the same process.

"Does the caption convey the full meaning of the idea unit, without words/phrases that interrupt the reading experience?"					
yes			Score **CE, move to next Idea Unit**		
no	**"Does the caption for this idea unit present false and understandable information?"**				
	yes		Score **FIE, move to next Idea Unit**		
	no	First, **"Does the caption viewer lose meaning that was available to the hearing viewer?"**			
		yes	**"Is the lost meaning the main idea or a modifying detail?"**		
			main idea	Score **OMM**	
			modifying detail	Score **OD**	
		no	*No score, next question*		
		Second, **"Does any word/phrase in the idea unit interrupt the reading experience (because of misspellings, missing punctuation, etc.)?"**			
		no	*no score*		
		yes	**"For each such word/phrase, has it already caused a scored OD or OMM?"**		
			yes	*no additional score*	
			no	**"Can the interrupting word/phrase be understood with a moment's thought?"**	
				yes	Score **BE**
				no	Score **NE**

COMBINING ERRORS

Since NER considers both distortions of meaning and interruptions to the reading experience, it is possible to score both types of error within a single idea unit.

For example, if the meaning of a whole idea unit is lost, 1xOMM is scored, but a BE or NE interruption error may also be scored within the same idea unit. After considering OD or OMM in the "First" section of the decision tree, the evaluator then moves to the "Second" section to consider BE and NE. However, if the possible BE or NE _caused_ the loss of meaning, OMM/OD alone is the accurate measure of the error, and the interruption error is not scored. In such cases, scoring the interruption is considered a "double-count" of the same error, so it is not used.

However, both an OD/OMM and an NE may be scored in the same idea unit when an unrelated garbled word is added to captions that have a separately-caused issue of meaning.

Also, one may score multiple BEs and/or NEs if there are several interrupting words/phrases in the idea unit.

NE on its own is a relatively rare score, since an incomprehensible word/phrase will typically affect the meaning of the idea unit, and OD/OMM will be scored instead. However, it is possible for an idea unit to be fully

communicated while still containing a superfluous garbled word that causes a long interruption in reading, in which case NE alone is the correct score.

OD and OMM do not combine. If an independent idea unit contains one or more dependent idea units, and the meaning of the main idea is lost, we do not score additional ODs for loss of details, but 1xOMM.

ADDING UP .25 ERRORS TO EQUAL MORE THAN A .5 ERROR.

Multiple ODs may be scored, even when their total errors count more than the error scored for an OMM. Following UK practice, 3xOD or 4xOD is a possible score within a single independent idea unit.

LISTS AND IDEA UNITS

Each element in a list is usually seen as a separate dependent idea unit. Their omission would be 1xOD or more depending on how many units' meaning was not transmitted. The fact that they are "like" elements does not justify a 1xOD score if more than one is left out.

Note that the loss of what appears to be a detail can affect the meaning of the whole independent idea unit. For example, leaving a crucial ingredient out of a recipe could ruin it, and therefore its impact could be 1xOMM, or even FIE.

IDEA UNITS IN PLAY-BY-PLAY

In sports play-by-play, a sequence of actions by a single player, or a single play sequence, should be treated as one Independent Idea unit. Elements within it, if omitted, are Dependent Idea Units and their omission is scored OD or multiple ODs if more than one is omitted. If the whole sequence is omitted or made meaningless, the score is 1xOMM.

FIES

FIE is scored when the captions replace the meaning of the verbatim with another clear but false meaning. The resulting caption must make sense and be clear to be FIE. It does not communicate false information if the caption viewer doesn't readily understand it, or if it suggests a wrong meaning but does not communicate it clearly. If the result of the caption is confusing, it will not be an FIE but may be another kind of error.

The assessment of errors in NER is never dependent on the importance of the topic. Since we cannot know the viewer's priorities, all subjects are equal. By this token, changes in meaning, even if not "important", are still FIEs.

NER does not consider whether the viewer might know the statement is false from out-of-program background knowledge. If the captions make sense, and provide wrong information according to the verbatim audio, then there is an FIE.

However, in-program context matters: if the misinformation is corrected with an on-screen graphic, or is corrected in following captions, then it is not an FIE, but may be another kind of error. Sometimes correct meaning or correct words can be supplied from visuals on the screen – player numbers, a score bug, a visual of the scene being

described, surrounding text. The video may provide information that can be edited from the captions without loss, e.g. temperatures in a weather report, if they are on-screen.

GRAMMAR AND PUNCTUATION ERRORS

PUNCTUATION

Punctuation should be added to be verbatim transcript as described below. Punctuation discrepancies should be scored.

Missing question marks are usually scored BE: the viewer can tell from the word order that a question has been asked but has to stop for a moment because the "?" is omitted. This causes an interruption and therefore BE should be scored. A period that replaces a comma, breaking up a sentence, can cause a similar pause and be BE.

However, in cases like play-by-play audio, where complete sentences are rare and the punctuation might as easily be "…" or "-" between phrases, meaning is usually not affected by punctuation, and the caption reader does not have to pause to understand.

A missing comma will often make no difference in meaning, so CE may be appropriate, where the comma has been included in the verbatim transcript.

Misplaced or missing apostrophes often change the meaning of words. However, sometimes a missing apostrophe can be understood and become simply an interruption – BE.

GRAMMAR

Spoken audio is often ungrammatical – incorrect grammar in the captions is not scored as an error if it reflects the usage in the verbatim. Captions are not expected to correct the grammar of the spoken word. However, an edit or paraphrase that corrects the grammar of the speech while retaining meaning is scored CE.

Grammatical errors created by the captions may change meaning, or may cause a pause in the viewer's reading flow, so they may cause BEs or other errors. But if the grammatical error is a common usage that does not cause an interruption in reading or a change in meaning, it may be CE.

MEANING AND SOUNDS

Some sounds are not meaningful words – e.g. UH, UM – and their omission does not affect meaning, so CE is appropriate if they are in the verbatim transcript.

However, some non-words are meaningful sounds, and their omission from captions is meaningful. LAUGHTER, APPLAUSE, etc. are part of the hearing experience, and should be in the verbatim transcript. Their omission may be OMM or OD depending on context. If other captions supply the meaning of the sounds, however, or the commentator notes their presence, the evaluator may choose to score the omission CE.

When sounds obscure speech, the missing words are not part of the hearing experience, so they will not be in the verbatim transcript, and their absence from the captions is not an error.

THROW-AWAY WORDS AND QUIPS

"You know", "Thanks for your report" or "THANKS, ERIKA" when the weather announcer takes over from the host, and other similar phrases may be included in the verbatim but add no meaning. Omitting them is CE.

Quips and banter may be essential to the show experience, however. Comments that are very light in nature may be meaningful.

CORRECTIONS

CORRECTION OF MISTAKES

Correction of a captioning mistake, if it is accurate, will not be an error in itself, but a Correct Edition.

WRONG NOUNS CORRECTED IN AN ON-SCREEN GRAPHIC

This is a Benign Error (BE), worth -0.25. The caption is wrong and causes a reading interruption, but the viewer can see that and, with a moment's thought, still understand it.

VERBATIM INFO IS IN A GRAPHIC BUT NOT CAPTIONS

This is a CE. The Captions don't include it (a sports score, for example, or a news reader reading a quotation) but the caption viewer experiences it. No reading interruption is caused and the meaning is conveyed.

IDENTIFYING PEOPLE

WHEN ARE SPEAKER IDS REQUIRED IN ADDITION TO CHEVRONS?

- Whenever it's hard to identify a speaker with their captions, speaker IDS are needed.
- Whenever a speaker changes, chevrons (>>) are needed. If they are missing, some kind of error is indicated.

Typically, when there are only two speakers, only one ID is needed for each speaker, when they first speak, after which chevrons will suffice to identify them. Or, if the speaker is identified in the captions, an ID is not needed when they speak, though chevrons are.

But multiple speakers, off-screen speakers, rapid interchange with lagging captions can all make it difficult to know who uttered what caption, so speaker IDs with the chevrons may be required.

The nature of an ID omission can vary like any other error. And, as elsewhere, context and meaning are important. Since there are a number of possibilities here, a decision tree approach can be helpful. Assuming that chevrons and/or speaker ID are missing from the captions, causing identification issues, the evaluator asks:

Does the video, with lag time considered, make the speaker ID clear?				
	Yes	**BE**		
	No	Is the result to misidentify the speaker?		
		Yes	Does the misidentified speaker appear to say something that, in program context, they don't agree with?	
			Yes	**FIE**
			No	**OD**
		No	The speaker ID is simply unclear, but otherwise the meaning is communicated.	**OD**
			The loss of speaker ID means the captions lose the main meaning.	**OMM**

MISIDENTIFYING A PLAYER IN PLAY-BY-PLAY

This is a Benign Error when the viewer can figure it out easily from the visuals. But misidentifying a player or their actions can be an FIE, if the visuals do not correct it clearly.

MISSPELLED NAMES

A misspelled name can lead to different scores.

- If the misspelling makes the idea unit meaningless, OD or OMM are scored.
- If the correct name can be understood from the caption – e.g. "Deniss" for "Dennis" – it's a BE.
- If the meaning of the idea unit is fully communicated, but a name can't be understood – "DDNnn" for "Dennis" – it's NE.
- If the name is known to be someone else's name, or a comprehensible word, it might be an FIE, if the result is a clear misstatement.
- American spellings are normally BEs, (e.g. "honor" for "honour"), but if it's a US name or proper noun, the US spelling is acceptable (e.g. The Marine Honor Guard").

THE NAME OF A FAMOUS INDIVIDUAL WAS EDITED OUT, BUT THEIR TITLE REMAINS

If part of a name or title is omitted, but it is in nearby captions or an onscreen graphic, it is a Correct Edition. Words have been omitted, but the meaning is intact. Without such supplementary sources, the omission would be OD. If the omitted name/title is essential to the main point, it would be OMM."

In some shows the captioner may omit the last name of reporters or regular guests, even though the last names are used in the audio. These omissions are OD, unless the last name is in nearby captions or visuals, since NER does not consider viewers' possible background knowledge.

PREPARATION OF TRANSCRIPTS

The preparation of the transcripts is the key to scoring errors and must be done carefully.

- NER requires that ten minutes of live content be transcribed and scored. It need not be continuous if it bridges over commercials and other pre-captioned material; such material is not scored or included in the transcripts.
- All of the audible words should be correct in the verbatim transcript. Where a word is inaudible, i.e. it would not be part of the hearing experience, so [inaudible] or a similar phrase can be used to signal the evaluator that these omissions are not scored.
- Chevrons, Speaker IDs and bracketed notes (e.g. [applause] and [music] should be present in the verbatim transcript, as needed.
- Punctuation must be added, since spoken audio carries no punctuation. In general, the captioner's punctuation can be kept in the verbatim unless there is a reason to add or alter it because the captioner's punctuation choice has caused a difference in meaning that might cause OD, OMM, or FIE. Missing or incorrect punctuation that might cause a reading interruption should also be inserted.
- The verbatim transcript should not attempt to revise ungrammatical audio to make it correct written English via punctuation.
- It is often difficult to know what correct grammar would require, so it may be best not to add commas to the verbatim transcript unless meaning requires them. For example, there are the rare classic examples where missing commas can cause an FIE: "He eats shoots and leaves." or "Let's eat, Grandma."
- Quotation marks should be in the verbatim if they are needed for the reader to understand the correct meaning without interruption.
- Musical lyrics should be in the verbatim transcript if they are clearly audible. However, if they are masked by dialogue, too fast to be understood, or just unclear, they may be omitted in favour of musical symbols or the note [music].
  - Note that a paraphrase or description of the music lyrics that captures their meaning is usually CE, unless, in context (e.g. national anthems) the exact words are necessary.
- Off-camera speech should be included in the transcript with some form of source ID if it can be clearly heard.
- The verbatim transcript need not include meaningless utterances such as, "Uh" or "Um". These are not words, and needn't be included. On the other hand, false starts to a sentence like, "I – I said" or meaningless interjections like "you know" and pro forma words, like "Thanks" should be included though their omission from captions will probably be scored CE.
- Usage of hyphens varies; if needed for meaning they are an error and should be in the verbatim transcript.

Care is also needed with the caption transcript. Even if text is supplied from a broadcast logger, proof-reading is necessary. NER is about the viewer experience and evaluations must be based on what was seen by the viewer.

THE NER SCORE CALCULATION

The NER score is calculated as follows.

NER SCORE = (Words – NER Deductions) / Words

Where:

1. "Words" equals the total number of captioned words in the segment evaluated.
   a. For these purposes, "words" includes chevron groups and punctuation marks. Words in brackets like [light applause and laughter] count as a single word[2].
2. "NER Deductions" equals the total of all error deductions assessed by the Evaluator in the segment.

---

[2] This definition was arrived at after discussion with Dr. Pablo Romero-Fresco and is identical to the definition used in all NER practice.