Transcription And Annotation Handbook

Excerpt
(Transcription Rules)

Deliverable ID:	D3.1
Dissemination Level:	CO
Project Acronym:	HAAWAII
Grant:	884287
Call:	H2020-SESAR-2020-2
Topic:	SESAR-ER4-18-2019
Consortium Coordinator:	DLR
Edition date:	07 January 2022
Edition:	01-00-02
Template Edition:	02.00.02

1 Rules for Word Transcription

In this section, we define the rules for transcribing air traffic control utterances (from controllers or pilots) to a sequence of words.

The rules defined here for transcription also define the output of the Speech-To-Text (STT) block, which is the input into the ATC concept extraction block.

1.1 English speech transcription

- Try to **transcribe the utterance sentence by sentence and in each sentence word by word**. This prevents omitting transcription of a part of the utterance
- Use **US English spellings**
- **Do not use special symbols like** ´ ` " " ". Usage of **'** is permitted. Example: **I'll** not **I´ll**.
- **Write every word in lowercase,** with the exception of the ATC phrases provided in section 1.3.
- **Do not use punctuation, . ? ! . Start sentences** of **utterances** with **lower-case letter** (not *Hello world*. but like this `hello world`)
- Try to **write it as it sounds** e.g. `wanna` (not *want to*), `I'll` (not *I will*). But these still **have to be existing** or **otherwise meaningful** words, e.g. `descend` (not *decend*), `thousand` (not *tousand*), `three` (not *tree*) And as a special case `nine` (not *niner*). Use the NATO alphabet for the spoken letters as defined in section 1.7.3.

1.2 Numbers

- Write out **numbers in words exactly as was spoken** (123.4 as `one hundred and twenty three point four` **or** `one two three decimal four`, depending on the exact pronunciation).

1.3 Procedures, acronyms, codes

- **Procedure words** like **wilco, roger, clear, copy, affirmative, ok** should be written in lowercase format `wilco, roger, clear, copy, affirmative, ok`
- **Use these compound-words 'without space in between':** **`takeoff`** not **`take off`** , **`callsign`** not **`call sign`** Look into appendix "Preferred writing of some expressions" for more examples and exceptions.
- **Common acronyms/abbreviations spelled by English alphabet are written as ALL CAPITAL:** pronounced <u>ee tea ey</u> -> transcribed **`ETA`**. pronounced <u>ay el es</u> -> transcribed **`ILS`**. pronounced <u>kju en aitsh</u> -> transcribed **`QNH`**. pronounced <u>ar nav</u> -> transcribed **`Rnav`**

See the table of common acronyms in sec. 1.8, Figure **1**.

 If it is pronounced as a "word" instead of spelling it, transcribe it in lowercase: **pronounced ey tea ay es -> transcribed `ATIS`** , **pronounced atis (/ah tt ee ss/) -> transcribed as `atis` not `ATIS`**), acronyms pronounced as vadin, inlux, transcribe as `vadin`, `inlux` `(see` Figure **1**`)` **.**

- For **very short English words**, the above rules need to also be applied:
  - **pronounced ai confirm -> transcribed `I confirm,`**
  - **pronounced six oh four -> transcribed `six O four`**.
- **Local codes** (runway 04R) should be transcribed as `runway zero four right`. However, if just "runway zero four" is said, we transcribe it as "runway zero four". We transcribe such codes using entire words and not using short forms such as `04R`.

1.4 Other Unclear speech, pauses and noise

- **Use star * to indicate swallowed or unfinished words pronunciation**.
  `*inavian` for swallowed "Scandinavian", `goodb*` for interrupted "goodbye".
  Do not use "*" as stand-alone character. You should probably refer to "[unk]" then.
- **Annotate restarts / repetitions properly**. The transcription should not look like "`hi, how are you`" but "`hi how how are you`" in case of repetition or "`hi h* h* how are you`" in case of restart.
- **Use [unk] in case you cannot understand word(s)**. Do not use "[unk] [unk]" twice after each other as [unk] can contain multiple words. This means we map the word to a "garbage model" in the Speech-to-text model training (a popular way to "exclude" it from the training). Or, we remove the whole utterance, if there are no other words.
- Use `[hes]` in case there is clear **hesitation** (umm, uhh, uhm, hmmm, aaah, aeh, err, em). Use `[noise]` for **non-speaker noise** (alarm, keying, etc), `[spk]` for **speaker noise** (laugh, cough, etc).
- **Silence:** In case there is a pause in between an utterance, do not add any tags and continue transcribing the next word uttered.

1.5 Non-English speech

- Encapsulate non-English speech into **[NE] [/NE] tokens**, or **[NE langID] [/NE] tokens** (`here is prague airport [NE Czech] dobry den [/NE] provide your call sign please`).
  1. If you do not identify the language nor can transcribe it, use `[NE] [unk] [/NE]` sequence.
  2. If you identify the language, but cannot transcribe use `[NE French] [unk] [/NE]`

3.   If you identify the language and can transcribe use `[NE Czech] dobry den [/NE]`

- While transcribing non-English speech, use <mark>only ASCII code letters</mark> (remove diacritics, accents and other additional marks)
  Spanish *buenos días* is transcribed `as buenos dias`,
  Czech *dobrý den* is transcribed as `dobry den`
- **Prefer the American English version of local names**, if possible.
  mark as `cleared to budapest` (English) and not *cleared to budapescht* (Hungarian pronunciation), even if the pronunciation was not properly "English". However, transcribe as `here is [NE German] wien [/NE] airport` if "vienna airport" has clearly not been said.

1.6 Speech quality

- **Delete segments, where there is no speech** (automatic recognizer can generate some random text over music/noise).
- If the whole audio file is really bad or can only be transcribed as "[unk]", just refuse[1] it. A crosstalk utterance "XT" is an exception from this rule – do not delete crosstalk.

1.7 Material to be used for Transcription

1.7.1 ICAO codes:

-   ICAO Doc 8585 (with costs; regular updates)
-   https://en.wikipedia.org/wiki/List_of_airline_codes

1.7.2 Airline/callsign search:

http://www.avcodes.co.uk/airlcodesearch.asp

1.7.3 Alphabet:

https://www.icao.int/Pages/AlphabetRadiotelephony.aspx

alfa bravo charlie delta echo foxtrot golf hotel india juliett kilo lima mike november oscar papa quebec romeo sierra tango uniform victor whiskey x-ray yankee zulu

1.7.4 ICAO standard phraseology (less formal document):

https://www.skybrary.aero/bookshelf/books/115.pdf

---

[1] CoCoLoToCoCo moves an audio file to the "removed" folder and SpokenData excludes the part of the audio chunk and does not generate a transcription file, respectively.

1.7.5 ICAO Aeronautical telecommunications (formal document, detailed):

https://www.icao.int/Meetings/anconf12/Document%20Archive/AN10_V2_cons%5B1%5D.pdf

1.8 Transcription of ATC Acronyms

Figure **1** lists acronyms often used in ATC communication. Most of them are **spelled** (column "Main way of pronunciation") and thus should be written with **upper case letters**. However, some of them might be **spoken as a word** which then needs to be transcribed with **lower case letters**.

Acronyme	Main way of pronounciation	Signification
ACAS	spoken	Airborne Collision Avoidance System
AFIS	spelled	Aerodrome Flight Information Service
AFIS	spelled	Aerodrome Flight Information Service Officer
AIP	spelled	AeronauticalInformation Publication
AMSL	spelled	Above Mean See Level
ATC	spelled	Air Traffic Control
ATIS	spoken	Automatic Terminal Information Service
ATS	spelled	Air traffic Service
ATZ	spelled	Aerodrome Traffic Zone
FIS	spelled	Flight Information Service
POB	spelled	Person On Board
PTT	spelled	Press To Transmit
QDM	spelled	Magnetic heading (assuming no wind) to a VDF station
QDR	spelled	Magnetic bearing of aircraft from a VDF station
QFE	spelled	Atmospheric pressure at aerodrome level (or runway thresh-old)
QNH	spelled	Altimeter subscalesetting to indicate elevation (AMSL) when on the ground and altitude when in the air
QTE	spelled	True bearing
RTF	spelled	Radiotelephony
RVR	spelled	Runway Visual Range
SSR	spelled	Secondary Surveillance Radar
VDF	spelled	Very high frequency Direction Finding
VHF	spelled	Very high frequency (30 to 300 MHz)
VFR	spelled	Visual Flight Rules
VOLMET	spoken	Meteorological information for aircraft in flght

Figure 1 Acronyms for Transcriptions.

1.9 Tags Cheat-Sheet

Tag for Transcription	Explanation
[unk]	Word(s) is/are not intelligible / understandable

[hes]	Clear hesitation (umm, uhh, uhm, hmmm, aaah, aeh, err, em)
[noise]	Non-speaker noise (alarm, etc.)
[spk]	Speaker noise (laugh, cough, etc.)
[key]	Double-pressed PTT
[NE] … [/NE]	Non-English, language not identified
[NE langID] … [/NE]	Non-English, language identified, e.g., "[NE German] … [/NE]"

Table 1: Tags for Transcription with Explanation

Annotating speaker identity in SpokenData	
UNK-1, UNK-2 etc.	Unknown "speaker" identity (callsign); if identity is known, annotate as e.g., "DLH123"
ATCo, pilot	ATCo or pilot is speaking in the utterance
XT	Crosstalk (blocking), two or more speakers at the same time

Table 2: Tags for Transcription with Explanation

1.10 Preferred Writing of some Expressions

When using the TransAnnoOnlineChecker integrated in CoCoLoToCoCo, these substitutions are automatically performed (see Table 3).

Correct	Incorrect
takeoff	take off, take-off
callsign	call sign, call-sign
descend (if verb is meant)	decent, descent (if verb is meant), but it is "rate of descent" (as noun is meant)
line up	lineup, line-up
read back	readback
flight level	flightlevel
standby	stand by (to not be mixed up with "taxi to your stand bye/by")
good bye	goodbye

Table 3: Preferred Writing of some Expressions

1.11 Callsign Designators

How should the airline designator in "nor shuttle three one kilo now turn right ..." be transcribed? Different possibilities exist, e.g. "nor shuttle three ... "or "norshuttle three ..." (or even "nor_shuttle three ..." from manual transcription is theoretically possible). If the word "nor" is not spoken, then it is of course "shuttle three ...".

Following these rules try to make it unique, so that the resulting WER (word error rate) calculation is also unique:

1. Multiple word sequences for the same three letter airline designator are allowed, e.g., "nor shuttle" or "shuttle" or "nor". However, **transcriptions must contain only words which were spoken.** Transcribing "nor shuttle three ..." when only "nor three" was spoken is not acceptable.

2. If we have multiple word sequences for an airline designator, the transcriptions must be consistent, i.e., they must be unique, if we remove underscores and blanks, e.g., "hansa" and "lufthansa" are allowed at the same time, but "lufthansa" and in addition "luft hansa" or "luft_hansa" are not allowed at the same time.

3. Do not use dashes, hyphens, digits, special characters or underscores in the transcription of airline designators.

4. Separating meaningful words in the English language is preferred. Transcribe `easy jet` not `easyjet`, `nor shuttle` not `norshuttle`

5. There should not be any exceptions in spelling airline names or writing them in one or multiple words (knowing that there will be rare yet unknown cases that have problems or do not follow the no-exception-rule). However, the formerly announced exceptions (D3.1 version 1.0) are now deprecated, i.e., it is now used in two words "finn air", "bee line", "bel air", "top jet", "sun turk", "air baltic", "iran air", elit avia", "korean air", "sky fox", "ryan air", "travel air", "sky travel" as well as "speed bird". Hence, the words "air", "wings", "avia", "sky", "jet", "fly" should be written as a single word if they represent a meaningful syllable in an airline designator. However, existing longer words comprising one of the aforementioned should still be written in one word such as "airline" and "airlines" as well as "stingray". Airline names in non-English languages are normally written in one word independent of their syllables/meanings, i.e., "lufthansa" is correct. Furthermore, spelled letters need to be written upper case (this can normally only be decided after some utterances contained such airlines due to a lack of knowledge of pronunciation), i.e., "KLM", "CSA lines", "UPS", "GW air" are correct[2]. Special signs such as underscore "_", hyphen "-", and characters "ä,ö,ü,ß,é" shall be replaced by the most reasonable version.

6. As "nor shuttle" is not an exception due to the above rules, "norshuttle" and "nor_shuttle" should be automatically replaced, i.e., only "nor shuttle" is accepted.

---

[2] We build our transcription airline list on a designator file in json format that was initially not optimized according to any transcription rule. Hence, there are first all hyphens ("-") to be replaced by a blank " ", second some deletions to subtract from designators and third some additions to be added in order to stick to the defined rules.

An automatic checking and replacement was implemented in CoCoLoToCoCo if the TransAnnoOnlineChecker is executed, it will replace a non-preferred callsign to the preferred form (i.e., replace one-word version of airline designator by two words or vice versa). However, it cannot (and it is also not desirable) replace an incomplete callsign by complete one – see rule 1.

The complete list of airline codes together with their callsigns (spoken version) are given by files designators.json, minus deletions.json, plus additions.json defined in D1.3 section 3.13.

Another information, which can be useful for correct transcription of waypoints and holding points and other local names, transcriber might not be familiar with, can be found in files defined in D2.1 section 4.