Transatlantic Approaches for Automatic Speech Understanding in Air Traffic Management

Hartmut Helmke, Oliver Ohneiser, Matthias Kleinert
Institute of Flight Guidance
German Aerospace Center (DLR)
Braunschweig, Germany
Hartmut.Helmke@dlr.de, Oliver.Ohneiser@dlr.de,
Matthias.Kleinert@dlr.de

Shuo Chen, Hunter Kopald, Robert M. Tarakan
Center for Advanced Aviation System Development
The MITRE Corporation
McLean, United States
chen@mitre.org, hkopald@mitre.org, rtarakan@mitre.org

Abstract—**MITRE in the US and European stakeholders under the leadership of DLR have independently developed ontologies for representing the meaning of controller-pilot ATC (air traffic control) radio communications. With the intent of benefiting from each other's work, and possibly harmonizing the two ontologies in the future, DLR and MITRE performed a structured comparison of the two ontologies. To explore how local differences in vocabulary and ATC procedures influence ontology design, both parties exchanged the transcripts of 100 ATC radio transmissions from their terminal airspaces. This paper summarizes similarities and differences of the ontologies from MITRE and Europe. Overall, despite a 12% difference in word level representation, 80% of the instructions from both data sets have common semantic representation in both ontologies.**

Keywords—automatic speech recognition; natural language understanding; semantic interpretation; air traffic control; radio communications; intent representation; semantic ontology

I. Introduction

Over the last few years, there have been striking advances in the use of automatic speech recognition (ASR) within the Air Traffic Management (ATM) domain. In Europe, as part of the Single European Sky ATM Research (SESAR) joint undertaking, the German Aerospace Center (DLR) brought together scientists, engineers, and air traffic control operators from civil, academic, and commercial institutions. In the United States, in support of the Federal Aviation Administration (FAA), the MITRE Corporation Center for Advanced Aviation System Development (MITRE CAASD) has repeatedly demonstrated the value of large-scale post-operations analysis of radio communications to improve aviation safety and efficiency. The developed prototypes from DLR and MITRE demonstrated the efficacy of using automatic speech recognition and understanding to improve controller efficiency, enhance controller automation interactions, and reduce controller workload during real-time operations. One prototypic example application that MITRE [1] and DLR [2] developed independently was readback error detection, demonstrated on controller-pilot radio transmissions recorded in the control room.

Both DLR and MITRE recognize that ASR is just the first step in utilizing the information in ATC speech. ASR is the process of translating the audio signal containing speech to the closest phonetic sequence of written text. There is still a sizable jump between automatically transcribing speech and understanding its meaning. For example, a perfect ASR system could accurately transcribe all the words in an Air Traffic Controller (ATCo) radio transmission, "*speed bird twenty two sixty one one eight zero knots until five to tower eighteen nine bye*", but an assistant system might still not be able to understand the meaning, or semantics, of this transmission. Therefore, we shift to using the term Automatic Speech Recognition and Understanding (ASRU) henceforth to indicate both the automatic speech-to-text conversion as well as the text-to-meaning interpretation. Robust text-to-meaning interpretation is especially important if some of the spoken words are wrongly recognized, or if the speech contains disfluencies, local colloquialisms, or is dependent on external context.

European ATM stakeholders led by DLR [3] and MITRE [4] have independently developed ATC language ontologies to support specific applications of ASRU. DLR and MITRE jointly analyzed similarities and differences in their ontological representations of ATC radio communications to identify areas for improvement and harmonization towards a common ontology.

A common ontology could enable better sharing and reuse of data, models, algorithms, and software between the US and Europe. Its implementation may be captured through a variety of techniques, e.g., rule-based parsing, machine learning models, human interpretation. A common ontology could enable more effective collaboration on the research and development of these techniques, foster discovery and sharing of best practices, and enable ATC application researchers to more effectively take advantage of Natural Language Processing (NLP) capabilities developed by other ATC researchers or those in other fields. Furthermore, a common ontology could pave the way to a more global standard that would be needed when operational ASRU systems achieve widespread adoption at controller positions and in aircraft cockpits in the future.

Manually generating semantic meaning labels on ATC transcripts requires an understanding of ATC operations, airspace, and phraseology, and can be even more costly and complicated than just transcribing ATC audio recordings alone because of the required subject matter expertise. Although advances in NLP have reduced the amount of data required to train effective models, a sizable corpus of labeled data is still required to capture the breadth and complexity of ATC semantics. A common ontology could enable researchers to share and consolidate their respective labeled corpora to build more robust, advanced NLP models. Furthermore, ontologies play an instrumental role in disambiguating language for semantic analysis, i.e., language understanding [5]. As such they will be critical to the adoption of artificial intelligence technologies that require advanced language understanding and linguistic capabilities.

This paper summarizes related work in section II, then defines our understanding of an ontology for ATC transmissions in section III. Section IV describes the effect of different rules for word level transcript. Section V presents our analysis of the similarities and differences between US and European ATC communications and language understanding representations, focusing on terminal airspace. We use representative transcripts of radio communications from the US and Europe to exemplify the concepts and structure that can be captured in a common ontology. In section VI, we discuss rationale for differences we discovered and how they affect future harmonization of a common ontology. Section VII describes opportunities for harmonizing the European and MITRE ontologies. Section VIII presents our conclusions.

II. Related Work

Voice communications are an essential part of ATC, because they are the primary means of communicating intention, situation awareness, and environmental context. Over the last decade, researchers have invested tremendous effort into advancing the accuracy and sophistication of in-domain ASR and Natural Language Understanding (NLU) capabilities to enable human-machine teaming that improves aviation safety and efficiency.

Early applications of ASR and NLU focused on simulation pilots for high-fidelity controller training simulators, because these applications presented controlled environments with well-defined phraseology and a limited set of speakers [6], [7]. Later applications in lab settings expanded to simulation pilots for human-in-the-loop simulations in ATM research measuring workload [8]. With the adoption of electronic flight strips in ATC facilities, reference [9] applied ASRU to demonstrate the effectiveness of speech assistants in reducing controller workload and improving efficiency. Prototypes demonstrating the use of ASRU to enhance safety in live operations also emerged. ASRU can support the detection of anomalous trajectories [10]. It can also support the detection of closed runway operations and wrong surface operations in the tower domain [11]. The efficacy of using ASRU to automatically detect readback discrepancies was analyzed in the US [1], and in Europe [2], [12]. A safety monitoring framework that applied ASR and deep learning to flight conformance monitoring and conflict detection has been proposed by [13]. The growing prevalence of unmanned aerial vehicles has also led to use cases in autonomous piloting. Text-to-speech and NLP can enable communications between human controllers and autonomous artificial intelligence pilots as advocated by [14]. Finally, the accuracy and robustness achieved by mature in-domain ASR has enabled mining of large-scale ATC communication recordings for post-operational analyses. Reference [4] measured approach procedure usage across the U.S. National Airspace System using automatically transcribed radio communications in post analyses. Similarly, reference [15] assessed the quantity of pilot weather reports delivered over the radio against the quantity of pilot reports manually filed during the same time frame.

A common theme across all these applications is the use of a language understanding layer that distills and disambiguates semantic meaning from the text transcripts generated by ASR. Although there is variability in the semantic structures and concepts relevant to each use case, almost all extracted semantics relate to the representation of controller and pilot intent or situation awareness. Currently, research groups in US and Europe create and maintain their own semantic taxonomies or ontologies to define the elements and relationships that represent intentions or situational context relevant to their specific use cases. These elements usually cover ATC concepts like aircraft callsigns, command types, command values in structured human-readable and machine-readable formats.

The European ontology was defined by fourteen European partners from the ATM industry as well as by air navigation service providers (ANSPs) funded by SESAR 2020 [16]. The ontology was refined through use by different projects, such as STARFiSH [17], "HMI Interaction Modes for Airport Tower" [18] [19] in the tower environment, "HMI Interaction modes for approach control" [20], and HAAWAII [21], which expanded the ontology to support pilot utterances [12].

The MITRE ontology was developed and matured over several years, with many contributing projects. Our earliest ontology was created for the simulation pilot component of an enroute ATCo trainer [6]. It was later expanded to incorporate tower domain phraseology for projects like the Closed Runway Operations Prevention Device [11]. More recently, to support the varied use cases required of our large-scale, post-processing capability [4], the ontology was expanded to cover most of the phraseology for standard operations documented in [22]. With each iteration we made it more robust and flexible to cover regional phraseology variations across the operational domains, i.e., tower, terminal, and enroute airspace.

III. What We Mean By Ontology

Communications can be considered in terms of four levels of a computer interaction model consisting of *lexical*, *syntactical*, *semantic*, and *conceptual* levels [23].

The **lexical level** (or word level in this case) deals with words and distinguishes between synonyms – words with the same meaning that are spoken differently, such as *nine* vs. *niner*, and *speed bird* vs. *speedbird*. These words are the building blocks for ATC radio transmissions. This level of an ontology specifies the universe of words (i.e., the vocabulary) that may appear in a transmission and can have a large impact on ASR accuracy as well as on complexity of extracting meaning from text transcripts. The vocabulary consists of general-purpose words such as *climb*, *descend*, *cleared* and *contact*, as well as names, such as those for airline callsigns, location identifiers, navigational aids, and procedure names. Ideally, these terms are static and can be defined as part of an official vocabulary list or dictionary in the ontology. However, there are situations with ad hoc introduced words such as for special callsigns or waypoint names, e.g. "gndlf", "yebuy", "isace".

The **syntactical level** deals with grammar and distinguishes between similar meaning phrases that are worded differently. For example, the phrases *runway two seven left cleared to land*, and *cleared to land two seven left* are syntactically different because they have different word ordering, however they have the same meaning, which is to convey clearance to land on runway two seven left.

The **semantic** representation deals with meaning despite differences in vocabulary or grammar that do not affect the meaning of the communication. For the ontology, we deal with the meaning associated with individual transmissions and allow the understanding of a transmission to include information not explicitly spoken but implied in the transmission. Both phrases from the syntactical level example may be mapped to an agreed form such as *CTL RWY 27L* or *RW27L CLEARED_TO_LAND*. Later in this paper, we discuss how these semantics are represented in the European and MITRE ontologies.

The **conceptual** level deals with a higher level of understanding that goes beyond the semantic level. It captures the bigger picture, the gestalt, which can be bigger than the sum of the individual transmissions. An example of an event at the conceptual level is the concept of an aircraft being in the arrival phase of flight. For some applications, this is more important than knowing the particular set of altitude and speed reductions an ATCo issued. Another example is the speech associated with a go-around, which might involve a back-and-forth discussion between an ATCo and pilot followed by a series of ATCo instructions. The scope of this paper does not include this level of knowledge in the ontology.

In this paper, we define ontology as the set of meaning entities, attributes, and relations required to represent the semantic interpretation of ATC communications. This means that the ontologies we discuss primarily address the lexical and semantic level described above.

Sometimes a detail that seems to be relevant only at the lexical or syntactic level is important for understanding the meaning of the transmission. For example, ATCos delivering a transfer of control (i.e., contact instruction) often terminate the transmission with a farewell expression in the language of the airline that is transferring away, such as *bon voyage* to an Air France pilot or *servus* to an Austrian Airlines pilot. While these expressions are not critical to the meaning of the transmission, they can help associate the transmission with the correct aircraft or help identify it as a transfer of control instruction when that information cannot otherwise be extracted from the transcribed transmission. For this reason, ontologies should cover any information that could be relevant for the application, even as each step through the levels of representation from lexical to syntactical to semantic to conceptual introduces abstractions that remove irrelevant information.

An ideal ontology is independent of its software implementation and will support a wide range of downstream and end-user applications.

IV. Comparison of Word Level Representation

Both MITRE and DLR have designed and implemented ATC ontologies on the lexical and semantic levels to distill and convey meaning for speech-dependent applications and data analyses. In this section, we briefly discuss representation at the lexical level, i.e., at word, level.

Consider this simple, artificial ATCo transmission: "*eurowings 1 3 9 alpha cleared I L S approach oh 8 right auf wiedersehen*". Suppose also that there is a sound suggesting speaker hesitation after the word "eurowings" and a cough after the word "approach". Note that the words "auf wiedersehen" (meaning "good bye") are in German, unlike the rest of the transmission, which is in English. The design of the lexical ontology can result in very different textual representations of this transmission's content.

The European ontology [24] for lexical representation transcribes this transmission as: "*euro wings [hes] one three nine alfa cleared ILS approach [spk] O eight right [NE German] auf wiedersehen [/NE]*". While MITRE's word level ontology represents this transmission as: "*eurowings uh one three niner alfa cleared i l s approach oh eight right auf wiedersehen*". There are obvious differences between the two representations in terms of both spoken words and non-speech sounds.

Two obvious and important differences are how the two ontologies represent hyphenated words (e.g., *eurowings* versus *euro wings*) and initialisms (e.g., *ILS versus i l s*). For most transmissions a transformation from one interpretation to the other is possible without loss of information. However, sometimes, information can be lost (e.g., the European approach maps both "*nine*" and "*niner*" to "*nine*") and it depends on the application whether this loss is harmful to understanding.

Some representations make semantic interpretation easier. For example, suppose "*speed bird one two alfa*" was spoken in a transmission and a speech-to-text engine transcribed it as "*speed one two **zero** alfa*", incorrectly deleting the word "bird" and injecting the word "zero". Furthermore, suppose we know that aircraft with callsigns "BAW12A" and "AFR10A" are in the airspace at the time of the transmission. Then how these symbolic callsigns are expanded into spoken form, i.e.,

- "*speed bird one two alfa*", "*air france one zero alfa*" or
- "*speedbird one two alfa*", "*airfrance one zero alfa*",

could affect the Levenshtein [25] word distance between the automatically recognized text and the expected correct transcript. In the latter expansion form, both contextual callsigns would have a word distance of two to the automatically transcribed callsign. In contrast, the first expansion form distinguishes the callsign BAW12A as the more plausible spoken callsign, because it has a word distance of two from the automatically transcribed callsign, which is smaller than the word distance of three between the callsign AFR10A and the automatically transcribed callsign.

Note that it is impossible to entirely separate the word-level ontology from the ASR software implementation. While more traditional ASR techniques depend on pronunciation dictionaries that map pronunciations onto words, emerging end-to-end ASR technologies do not have explicit pronunciation dictionaries. The primary motivation for using this new technology would be for better ASR accuracy, but it would also have the impact, good or bad, of transcribing words not in the ontology.

V. Comparison of Semantic Interpretation

This section presents a comparison of the two ontologies at the semantic level (SL), using examples derived from typical ATCo and pilot transmissions. The following tables show radio communication transcripts in European word level format in the first gray table row. The rows SL_{US} and SL_{EU}. present the semantic interpretations in the MITRE ontology and European ontology, respectively.

A. *Transmissions with Callsign and Command Type*

The example in Table I shows a simple transmission containing a callsign and an approach procedure clearance. The aircraft callsign in the US semantic interpretation, SL_{US}, consists of up to four distinct elements: (i) airline code, (ii) tail or flight number, (iii) wake vortex category, such as heavy, and (iv) operation type such as commercial, general aviation (GA), military or air taxi aircraft. The European version SL_{EU} condenses the callsign to a single string, as it would appear in a radar label of, e,g., a controller working position's situation data display.

TABLE I. Semantic Interpretation of Simple Transmission

november three mike victor cleared ILS runway two one approach	
SL_{US}	Callsign: {N, 3MV, GA}, Cleared: {21, ILS}
{"GACallsign": {"TailNumber": "3MV", "GAAircraftType": "N"}, "Cleared": {"Runway": "21", "ApproachType": "ILS"}}	
SL_{EU}	N123MV (CLEARED ILS) 21
{"csgn":"N123MV","type":"CLEARED","sndT":"ILS","valu":"21"}	

Beyond the difference in representation components, another difference between the SL_{EU} and SL_{US} representations for callsign is the extent to which external (speech or nonspeech) context information is used to help interpret the callsign words explicitly spoken in the transmission. SL_{US} distinguishes text-only interpretation and with-context interpretation to preserve lexical differences that could indicate misspoken callsigns or readback error and incorporates context information separately to infer the intended callsign at the semantic meaning level. SL_{EU} applies context at the time of interpretation and expands the spoken truncated callsign to the full callsign where applicable. It does not preserve lexical differences for callsigns.

Both interpretations distinguish command type and parameter values. In the European version, the type can be extended with a second optional type modifier (e.g., ILS for type CLEARED or ALTITUDE for type MAINTAIN). The US version codes the additional type modifier information in parameter values.

Table I shows the semantic interpretations in both a simplified human-readable format and a machine-readable JSON format with blue highlighted keywords for the succeeding values. For the remainder of the paper, we will use only the human-readable format for brevity.

Table II shows an example with a wrongly spoken or misrecognized callsign. As described earlier, SL_{US} first interprets the callsign as it appears in the transcript without context; then it generates a second callsign interpretation that corrects the lexical interpretation to an inferred interpretation with context information. SL_{EU} ignores the wake vortex type and corrects the digit "5" in the callsign immediately to "4" based on context awareness that the closest matching callsign in the airspace is FDX482.

TABLE II. Callsign Correction and Unit Inference

fedex five eighty two heavy maintain four thousand three hundred	
SL_{US}	Callsign: (FDX, 582, H, Commercial}, Maintain: {Feet, 4300}
SL_{EU}	FDX482 (MAINTAIN ALTITUDE) 4300 none

In the altitude command representation, SL_{US} adds "*Feet*" as the inferred qualifying unit on the altitude based on the range of the number, whereas SL_{EU} extracts unit as "none", because the altitude unit is not lexically present.

In both examples above, SL_{US} explicitly differentiates between derived or inferred content and lexically present content. Derived or inferred elements are highlighted in gray in the example tables. *GA* in Table I is highlighted in gray to indicate that it is derived information because although the aircraft does not explicitly state that it is a general aviation aircraft, this information can be inferred from the format of the callsign. Similarly, *Feet* in Table II is highlighted in gray because the altitude unit "*Feet*" is inferred.

B. *Complex ATCo Transmissions*

Table III shows a transmission with up to three different commands. At the semantic level. SL_{EU} associates the callsign of the aircraft as an attribute to each instruction concept extracted from the transmission, whereas SL_{US} treats the callsign as a standalone concept. SL_{EU} treats "HEADING" as the command type with an optional qualifier "RIGHT", whereas SL_{US} encodes the command and its modifier as the command type, "TurnRight", and the "Heading" as a qualifier on the parameter value.

TABLE III. Transmission with Multiple Commands

good day american seven twenty six descend three thousand feet turn right heading three four zero	
SL_{US}	Courtesy, Callsign: {AAL, 726, Commercial}, Descend: {3000, Feet}, TurnRight: {340, Heading}
SL_{EU}	**AAL726** GREETING, **AAL726** DESCEND 3000 ft, **AAL726** HEADING 340 RIGHT

Table IV shows a transmission without a callsign. SL_{EU} explicitly states "none" for certain command types when expected interpretation slots cannot be populated, because they are not lexically present in the transcript – neither "*left*" nor "*right*" was specified on the heading vector. Both SL_{US} and SL_{EU} derive the command type, shaded in gray in SL_{US}, from the spoken digits. Only heading values are prefixed with a leading zero allowing interpretation of "zero four zero" as a heading in this example. If the heading value had been "three four zero", then the number could have been a speed, heading, or altitude and too ambiguous to interpret. The different handling of ambiguity in both ontologies will be discussed further in Figure. 4. SL_{EU} also contains a parameter value of "*none*" for command type *CLEARED ILS* to indicate that the runway is not lexically present. SL_{US} represents only what is lexically present and does not call out parameters that might have been intentionally omitted by the speaker.

TABLE IV. TRANSMISSION WITHOUT CALLSIGN

zero four zero cleared ILS approach	
SL	

zero four zero cleared ILS approach	
SL_{US}	Fly: {040, Heading}, Cleared: {ILS}
SL_{EU}	NO_CALLSIGN HEADING 040 none, NO_CALLSIGN (CLEARED ILS) none

Table V shows a transmission which does not contain a command. SL_{EU} explicitly states "NO_CONCEPT" as the command type. In SL_{EU}, a "NO_CONCEPT" entry is only possible if the transmission contains no other modeled command. SL_{US} has no required interpretation slots and thus represents the callsign present in the transcript as the only concept.

TABLE V. TRANSMISSION WITHOUT COMMMAND

lufthansa one two charlie go ahead	
SL_{US}	Callsign: {DLH, 12C, Commercial}
SL_{EU}	DLH12C NO_CONCEPT

C. ATCo Instructions with Conditions and Advisories

Table VI shows how each ontology represents conditions in instructions. The limiting condition, "until established", on the approach clearance is currently not covered in the SL_{US} ontology but could be added by qualifier-parameter pairs as described in subsection V.E. The position-based condition is currently ignored by SL_{US} ontology. SL_{EU} extracts the parameter qualifying unit "ft" because "feet" is lexically present in the transcript.

TABLE VI. ALTITUDE INSTRUCTION WITH LIMITING CONDITION

maintain four thousand feet until established	
SL_{US}	Maintain: {Feet, 4000}
SL_{EU}	(MAINTAIN ALTITUDE) 4000 ft (UNTIL ESTABLISHED)

SL_{EU} has a predefined Condition component in its instruction ontology for the purpose of capturing conditions associated with any command type. SL_{US} has selective coverage for position-based conditions by command type.

Table VII shows two instructions with position-based conditions. SL_{US} does not support the position-based conditions on approach procedure clearances because they are rarely used in the US, but does support them on altitude instructions, where they are used more in operations.

TABLE VII. INSTRUCTIONS WITH POSITION-BASED CONDITIONS

at dart two you are cleared ILS runway two one left	
SL_{US}	Cleared: {21L, ILS}
SL_{EU}	(CLEARED ILS) 21L (WHEN PASSING DART2)
leaving baggins descend and maintain one four thousand feet	
SL_{US}	Descend: {14000, Feet, leaving, BGGNS}
SL_{EU}	DESCEND 14000 ft (WHEN PASSING BGGNS)

Table VIII shows examples of instructions with traffic advisories and standalone situation awareness advisories (for brevity, we deliberately omit the callsign in these examples). SL_{US} represents traffic and wake turbulence advisories as ontology concepts with various parameter values. SL_{EU} condenses these into either "*INFORMATION TRAFFIC*" if they are about traffic or "*INFORMATION MISCELLANEOUS*" if they are not.

TABLE VIII. INSTRUCTIONS WITH ADVISORIES

maintain two fifty knots for traffic	
SL_{US}	Maintain: {Knots, 250, for traffic}
SL_{EU}	(MAINTAIN SPEED) 250 kt, (INFORMATION TRAFFIC) none
traffic twelve o'clock two miles same direction and let's see the helicopter	
SL_{US}	Traffic: {Distance: 2, OClock: 12, TrafficType: helicopter}
SL_{EU}	(INFORMATION TRAFFIC) HELICOPTER
caution wake turbulence one zero miles in trail of a heavy boeing seven eighty seven we'll be going into this [unk]	
SL_{US}	Wake: ()
SL_{EU}	(INFORMATION MISCELLANEOUS)

D. Pilot Readbacks, Reports, and Requests

Table IX shows an example of a classic pilot readback to an ATCo instruction. Unlike previous examples, the transmission example originates from a pilot speaker. SL_{US} represents the descend readback as an instruction concept but does not represent the callsign because it is too vague to interpret as a callsign without a tail or flight number. SL_{EU} represents the transmission with the labels NO_CALLSIGN and PILOT, to convey speaker information.

TABLE IX. PILOT TRANSMISSION WITH REASON

descend flight level one seven zero silver speed	
SL_{US}	Descend: {FL, 170}
SL_{EU}	NO_CALLSIGN PILOT DESCEND 170 FL

E. Formal Definition of Semantic Ontologies

The above examples should illustrate some of the differences and commonalities between the US and European semantic interpretation ontologies. In this subsection, we formally define the ontology structures and detail their differences and similarities.

1) MITRE Ontology

Figure. 1 illustrates the ontology of SL_{US} in graph format. At the highest level, SL_{US} starts with a concept called *Command Interpretation* that represents an instruction, and it has a mandatory attribute called *Command Type*. The *Command Type* attribute declares the type of the instruction, such as an aircraft maneuver like "*climb*" or a clearance to fly a procedure like "*cleared ILS two one approach*".

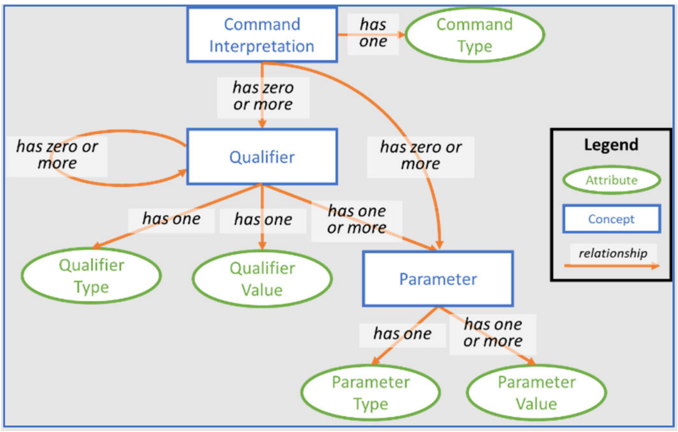

Figure. 1. Graphical representation of SL_{US} ontology.

Each *Command Interpretation* can have zero or more child concepts called *Qualifiers* and *Parameters*. Both characterize, modify, and/or add values to the instruction. *Qualifiers* disambiguate or characterize *Parameters* by representing value units that are lexically present in the transcript, e.g., "*flight level*", "*heading*", "*knots*", etc. *Qualifiers* can be nested to represent deeper, hierarchical relationships. For example, to represent the condition "*until the dulles VOR*", the highest-level *Qualifier* would represent the preposition "*until*", its child *Qualifier* would represent the waypoint type "*VOR*", and its child Parameter would represent the name of the waypoint "*dulles*".

Parameters represent the value payloads for instructions that require a value, such as a heading (in degree) for a turn instruction or an altitude (in feet or flight level) for a climb instruction. A *Parameter* may exist without a *Qualifier* parent if the format of the *Parameter* value or the instruction's command type makes the *Parameter* inherently unambiguous. For example, in the instruction "*climb three four zero*", the command type "*climb*" allows us to infer that an altitude must be represented in the *Parameter* and the value format in three digits allows us to infer that the altitude is in flight level even though a unit is not explicitly stated.

Figure. 2 illustrates the SL_{US} ontology as a block diagram for comparison to the SL_{EU} ontology in Figure. 3.

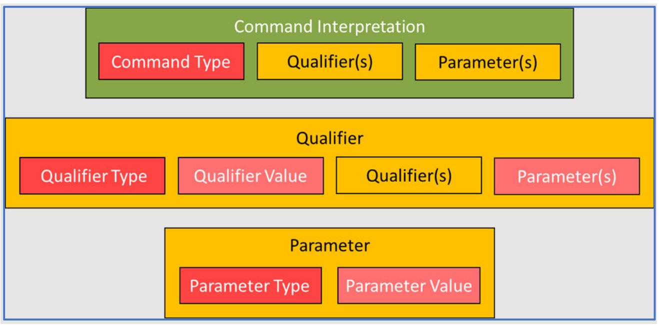

Figure. 2. Block diagram of SL_{US} ontology, with the highest level concept in green, mandatory elements in red and pink, and optional elements in orange

2) European Ontology

Figure. 3 illustrates the ontology of SL_{EU}. At its highest level, SL_{EU} starts with a concept called *Instruction*, i.e., a command and optional conditions. A *Command* concept always has a *Type* attribute that declares the type of instruction represented. When no instruction is found in a transcript, a *Command* concept with *Type* "*NO_CONCEPT*" is created. Depending on the *Type*, one or more *Values* can follow. If a *Value* is available, the optional attributes *Unit* and *Qualifier* can follow. The optional *Condition* component can be present for any Type and more than one may be associated with one *Command*.

Figure. 3. Block diagram of SL_{EU} ontology, optional elements in orange. The Type is mandatory and the Value(s) may also be mandatory, depending on the Type.

Type can consist of a subtype as illustrated by the command *CLEARED ILS*. The *Speaker* attribute can have the values "*ATCO*" or "*PILOT*". If not specified, it is *ATCo* or can be derived from additional available context information. The *Reason* attribute is only relevant for pilot transmissions. Then the values "REQ=REQUEST", "REP=REPORTING" or an empty value are possible. The empty value, i.e., the default value, in most cases contains a pilot's readback. The Reason attribute is motivated by the examples in Table IX.

VI. Analysis and Discussion

A. General Similarities and Differences

The European and MITRE ontologies largely represent the same information, but in different formats. The laws of aerodynamics and fundamental principles of air traffic control are universal, thus the similarities. However, because they were developed separately, for different applications and different speaker populations, there are differences in the design details. These differences make exchanging training data or comparing analysis results and findings between Europe and MITRE more difficult than they would be with a single, common ontology.

1) Differences at the Word Level

The differences that we observed at the word level can be summed up as fitting into the following categories:

- Identical words with different spelling (e.g., *Juliett* versus *Juliet*)
- How initialisms are handled (e.g., *ILS* versus *i l s*)
- Words with similar meaning and different pronunciations and spelling (e.g., *nine* versus *niner*)
- Words absent from one ontology or the other (e.g., the word "altimeter" is not included in the European ontology and the corresponding ICAO term "QNH" is absent from the MITRE ontology)
- Whether speech disfluencies and coarticulation are captured at the word level (e.g., *cleartalan* versus *cleared to land*)
- Words not represented in the US English language (e.g., the German word *wiederhören* for a farewell)

These differences can have an impact on ASR speed and accuracy performance, and on the end user or downstream software application.

2) *Similarities and Differences at the Semantic Level*

Both SL_{US} and SL_{EU} define highest order semantic concepts at the instruction level, but with small variations in the designs. Both have concepts that equate to commands like clearance to fly an approach procedure, climb or descend to an altitude, and turn to a heading. One noticeable difference at this semantic level is that SL_{EU} attaches the callsign as an attribute to each instruction while SL_{US} represents the callsign as another highest order concept on par with an instruction.

Another aspect with respect to callsign representation was already highlighted in the previous section in Table II. SL_{EU} represents one final interpretation generated from context-aware parse whereas SL_{US} distinguishes the interpretation of the text alone and interpretation of the text with context as distinct and separate representations. Thus, SL_{US} models the layers of interpretation, preserving whether the callsign is lexically present in its entirety in the transcript, partially present in the transcript, or corrected as part of context-aware comparison.

In contrast, SL_{EU} seeks to represent the ground truth intention of the speaker and associates a plausibility value to each interpreted concept and attribute, i.e., callsign, type, value, etc., that ranges between 0 and 1.0 to indicate the representation's proximity to the truth. Thus, the plausibility value of a corrected or derived callsign is always less than 1.0. This is because the target output of the European ontology is the ground truth intention expected from a perfect speech-to-text system and a perfect semantic interpreter, respectively. If the final combined output from speech to text and semantic interpretation does not represent the speaker's intention, then that is considered an error. For example, if the ATCo said "*speed bird one two alfa*", then the ground truth intention representation should be BAW12A. If, however, the ATCo correctly said, "*speed bird one two alfa*", but "*speed bird one three alfa*" is recognized, the target ground truth representation is BAW12A. Anything else is deemed an error, regardless of whether the error comes from the speech-to-text or the semantic interpretation. If "*speed bird one three alfa*" is said and recognized, but only BAW12A is in the air, the output should still be BAW12A under the rules of SL_{EU}. We do not know, what was said, but only what was recognized.

The ontologies also differ in how they represent absence of an instruction or callsign. SL_{US} represents what is present, but not what is absent, e.g., when a transcript does not contain a callsign, then a callsign concept does not appear in the semantic interpretation. In contrast, SL_{EU} states explicitly when the callsign is absent with a unique attribute value "NO_CALLSIGN" attached to each instruction. Furthermore, SL_{EU} also notes, when no instructions were detected in the whole transcript with the unique command type "NO_CONCEPT".

As shown in the examples in Tables VI, VII and in Figure. 3 SL_{EU} has a dedicated ontology element for the conditional component of any instruction, whereas SL_{US} currently allows custom conditions that are curated to fit specific command types. Conditions in SL_{EU} are initiated by a conditional keyword from the finite set of conjunctions "*WHEN*", "*UNTIL*", "*AFTER*", and "*IF*", and followed by the requirement.

The ontologies also differ in how they represent ambiguity. SL_{US} imposes requirements on what *Parameters* must be present for an instruction to be unambiguous and represented as a *Command Interpretation*. The *Parameters* required differ by *Command Type*. For example, the standalone transcript "*maintain three four zero*" would be considered ambiguous and not evaluate to a *Command Interpretation* under SL_{US}, because the *Parameter Type* of the instruction cannot be disambiguated to a speed in knots, an altitude flight level, or a vector heading.

SL_{EU} generates all possible semantic interpretations on the ambiguous text and assigns a plausibility value below 1.0 to each. In the example, "*maintain three four zero*", SL_{EU} would generate three possible semantic interpretations, (1) "*MAINTAIN HEADING 340 none*", (2) "*MAINTAIN ALTITUDE 340 none*", and (3) "*MAINTAIN SPEED 340 none*", and then use aircraft and dialogue context to associate a plausibility value. Figure. 4 shows the ambiguous output of this example in JSON format.


```
{
  "filename": "2022-09-15__09-36-08-04.wav",
  "hypotheses": [{
    "abstraction_layer_word_sequence": "maintain three four zero",
    "abstraction_layer_ontology_command": [{
      "commands": [{
        "csgn": "NO_CALLSIGN","valu": "340", "unit": "none",
        "type": "MAINTAIN", "sndT": {"v": "SPEED","p": "0.3"}
      }]
    }]
  }, {
    "abstraction_layer_word_sequence": "maintain three four zero",
    "abstraction_layer_ontology_command": [{
      "commands": [{
        "csgn": "NO_CALLSIGN","valu": "340", "unit": "none",
        "type": "MAINTAIN", "sndT": {"v": "HEADING","p": "0.8"}
      }]
    }]
  }, {
    "abstraction_layer_word_sequence": "maintain three four zero",
    "abstraction_layer_ontology_command": [{
      "commands": [{
        "csgn": "NO_CALLSIGN","valu": "340", "unit": "none",
        "type": "MAINTAIN", "sndT": {"v": "ALTITUDE","p": "0.8"}
      }]
    }]
  }]
}
```


Figure. 4. SL_{EU} JSON structure for ambiguous instructions.

The differences that we observed at the semantic level can be summed up as fitting into the following categories:

- How callsigns are represented.
- The extent of and representation of inferred and implied information in the semantic representations.
- The level of detail represented for advisory-type transmissions (e.g., traffic advisories, pilot call-in status information).
- Which less-common ATCo instructions have defined representations.
- How ambiguous ATCo instructions are represented.

B. Quantifying the Differences

MITRE and DLR exchanged 100 transmissions, with transcripts and semantic annotations, from the terminal area of a major US airport and a European hub airport. The US transcripts and annotations were manually transformed into the European format and vice versa. We assessed the word level differences on the transcript level in terms of Levenshtein distance [25].

Out of 1554 total words in the transmissions; 187 of them required modification to adhere to the other party's ontology, i.e., 12.0% of words were modified through substitution (89), deletion (35), and insertion (63). We omit uppercase to lowercase transformation from this measure. Figure 5 shows a sample transcript and its transformation.

```
all····right·cleared·for·the·ils····two·five·
alright······cleared·for·the·i···l·s·two·five·
```

Figure. 5. Word Level Difference between European (first row) and US (second row) transcripts resulting in a Levenshtein distance of 5.

In the following bullets we list and explain some of the most often occurring cases from the 200 transcripts that are represented differently at the word level in the MITRE and European ontologies as sketched in subsection VI.A:

- Separation and combination of words/letters
  - ”ILS” vs. “i l s” (23 times)
  - “southwest” etc. vs. “south west” etc. (19 times)
- Different spellings
  - “nine” vs. “niner” (9 times)
  - “juliett” vs. “juliet” (6 times)
  - “OK” vs. “okay” (4 times)
- Special sounds and their notation
  - “[unk]” vs. no transcription (7 times)
  - “[hes]” vs. “uh” (7 times)

Table X shows the overlap of commands represented by the MITRE and European ontologies at the semantic level, after analyzing 121 ATCo instructions from Europe and 120 from the US. *DESCEND* in SL_{EU} corresponds to *Descend in* SL_{US}. *MAINTAIN ALTITUDE* with Value and Unit in SL_{EU} corresponds to *Maintain* with the US Qualifier *feet* or *FL*. The *Cleared ILS Z* in SL_{US} now corresponds to *CLEARED ILSZ in* SL_{EU}. *GREETING* and *FAREWELL* in SL_{EU} correspond to *Courtesy* in SL_{US}. SL_{US}'s *Radar Service Terminated* is currently not modelled in SL_{EU}. In contrast, SL_{US} does not model SL_{EU}'s *CALL_ YOU_BACK* command type.

TABLE X. PERCENTAGE OVERLAP BASED ON ANALYSIS OF 241 ATCO INSTRUCTIONS

Type of Semantic Comparison	Overlap of Concepts
Concept present in both ontologies, before adaptation	82%
Corresponding concept after small adjustments	95%
Achievable match with existing model structures	100%

C. Operations Related Differences

1) Different Units for Air Pressure

In the US, ATC Handbook [22] specifies the rules and guidelines for air traffic radio communications, including vocabulary, phraseology, and meaning. In Europe, a set of ICAO documents including [26] serve the same purpose.

One difference between FAA and ICAO communications is how altimeter information is conveyed between an ATCo and pilot. Europe uses hectopascals (hPa) for the QNH, whereas the US uses inches of mercury. For example, 29.92 inches of mercury correspond to 1013.21 hPa. The European ontology was not able to semantically model “*burbank altimeter is two nine nine two*”, whereas the US ontology does not have a representation for “swanwick QNH zero nine nine five”. Note that both ontologies could be expanded to handle both the FAA and ICAO versions. The authors were pleasantly surprised that they did not observe more differences of this type already in the 200 transcripts. Table XI shows a possible ontology enhancement for both versions e.g., SL_{US} introduces two new Qualifiers that explicitly state the air pressure units. Both ontologies do not model the leading zero.

TABLE XI. EXTENSION OF BOTH ONTOLOGIES TO MODEL AIR PRESSURE

burbank altimeter is two nine nine two	
SL_{US}	Altimeter:{2992, inHg}
SL_{EU}	(INFORMATION ALTIMETER) 2992
swanwick QNH zero nine nine five	
SL_{US}	Altimeter:{995, hPa}
SL_{EU}	(INFORMATION QNH) 995

2) Special ILS approach procedures

US hub airports can have different ILS approach procedures for the same runway. The word sequences “*cleared ILS* ***zulu*** *runway one nine approach*” and “*cleared ILS* ***yankee*** *runway one nine approach*” would be mapped in European ontology to (CLEARED ILS) RW19. This is a loss of information, because *zulu* or *yankee* is an additional identifier suffix added to approach plates to distinguish them from other approaches of the same type to the same runway. Table XII shows a suggestion of how to extend SL_{EU}. Version 1 creates a new second type for the command type CLEARED. Version 2a creates a new runway name, version 2b allows two values for the CLEARED ILS command type, and version 2c (looks the same as version 2b) interprets the *Z* as a qualifier instead of a further value. Which one to use in the future is not yet decided, but it shows that SL_{EU} is flexible enough to adapt to special cases in US airspace.

TABLE XII. EXTENSION OF ONTOLOGY TO MODEL ILS ZULU ETC.

cleared ILS zulu runway one nine approach	
SL_{US}	Cleared:{19, ILS, Z},
SL_{EU}	(CLEARED ILSZ) RW19 or (CLEARED ILS) RW19Z or (CLEARED ILS) (RW19 Z)

VII. Opportunities for Harmonization

The authors recognize that significant resources would be needed to migrate to a new common ontology that harmonizes the two different ontologies presented in the previous sections. Thus, instead, we explore smaller steps that could bring the European and MITRE ontologies closer without major disruption to either's downstream applications.

One easy win is to merge concepts where each ontology handles non-overlapping variations. For example, MITRE can add the barometric altimeter setting (QNH) in hectopascals to its existing Altimeter concept while Europe can add the equivalent in inches of mercury to its QNH concept. Another relatively easy way to bring the two ontologies closer together is to coordinate future enhancements before we add them to our ontologies and implement them in software.

We expect that new downstream applications will dictate new needs for additional semantic information to be added to the ontology. In particular, we expect to add specification of additional types and values to accommodate rarer instructions and different airspaces. As new use cases arose for MITRE's large-scale post-processing capability, new data types were added to support analysis of missed approaches, drone activity, airspace excursion advisories, and other less common events. We anticipate that there will be a need to expand the ATC ontologies to support end-users' future needs. Here are some specific areas where we think close coordination of ontology enhancements could be beneficial:

- New ATM procedures, such as Interval Management
- New rules for the incorporation of Unmanned Aerial Systems (i.e., drones) into the airspace
- Accommodation of commercial space launches

The European ontology supports two types of information that are not currently supported in the MITRE ontology:

- *n*-best interpretations of transmissions
- Confidence scores

Ideally, MITRE can incorporate these in a way that is compatible with the European implementation.

The scopes of both ontologies are targeted to the information in single transmissions. MITRE has found that many analyses depend on bringing together information from the entire ATCo-pilot dialogue and linking this dialogue with non-speech information, such as that found in flight plans and surveillance data. There are many design decisions involved in representing entire dialogues. This is an opportunity for European ATM stakeholders, MITRE, and others to collaborate.

Both Europe and MITRE have attempted to design ontologies that are independent of ASR and NLP algorithms and implementations. Ideally, the ontology and semantic data representation model are independent of ASRU implementation. However, emerging deep-learning neural network (DLNN) ARSU implementations may introduce new challenges to our current approach to defining ontology. For example, a DLNN end-to-end ASR outputs letters, not words, allowing for sequences of letters (i.e., words) that are not pre-defined in an ontology lexicon. The ontology would need to be adapted for the challenges emerging that NLP technologies introduce. By working together, Europe and MITRE can avoid having their ontologies diverge further while benefiting from our combined expertise.

VIII. Conclusions

Our comparative analysis of the two ontologies showed that both serve similar purposes and capture largely the same information.

Analyzing differences at the word level, we identified a 12% difference in word representations, i.e., out of 1,554 words in the US data, 187 were represented differently in the SL_{EU} word level ontology. These differences primarily result from alternate conventions for hyphenated words, initialisms, and spelling. Although transformation is easily possible in most cases without loss of information, this is a first candidate for future harmonization between US and European ontologies.

At the semantic level, we identified information that is currently represented in one ontology but not in the other. These omissions fit into two categories: those that were absent, and those that were included at a higher-level of detail in one than the other. We also discovered that Europe and MITRE have different conventions for when external information is incorporated into the semantic representation. This is primarily noticeable with aircraft callsigns.

With this in mind, we concluded that it is possible to convert between the European and MITRE semantic representations of transmissions, but sometimes with a loss of information and sometimes with differences in the underlying meaning. We also concluded that an easy opportunity for harmonization is to coordinate on definitions as new ATM procedures and end user applications dictate the need for additions to the ontologies.

Future efforts supporting global harmonization for safety could create a global standard for controller-pilot voice communication ontology. Our ultimate goal is still to facilitate seamless sharing and reuse of data, models, algorithms, and software between the US and Europe so that our pooled resources will lead to ever more advanced applications of artificial intelligence and sophisticated human-machine teaming in the ATC domain.

Acknowledgment

DLR acknowledges the support of the SESAR Joint Undertaking for developing the first version of the ontology in PJ.16-04-W2 SESAR2020 (grant agreement No 734141) and for funding the extended version in the HAAWAII project (grant agreement No 884287).

MITRE acknowledges the support from the FAA for the research funding and for allowing access to recorded ATCo-Pilot live-operations radio voice communications.

REFERENCES

[1] S. Chen, H. Kopald, R. S. Chong, Y.-J. Wei and Z. Levonian, "Readback error detection using automatic speech recognition," in *12th USA/Europe Air Traffic Management Research and Development Seminar (ATM2017)*, Seattle, WA, USA, 2017.

[2] H. Helmke et al., "Readback error detection by automatic speech recognition and understanding: results of HAAWAII project for Isavia's enroute airspace," in *12th SESAR Innovation Days*, Budapest, Hungary, 2022.

[3] H. Helmke et al., "Ontology for Transcription of ATC Speech Commands of SESAR 2020 Solution PJ.16-04," in *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)*, San Antonio, TX, USA, 2021.

[4] S. Chen, H. Kopald, R. Tarakan, G. Anand and K. Meyer, "Characterizing national airspace system operations using automated voice data processing," in *13th USA/Europe Air Traffic Management Research and Development Seminar (ATM2019)*, Vienna, Austria, 2019.

[5] Y. Lin, "Spoken Instruction Understanding in Air Traffic Control: Challenge, Technique, and Application," *Aerospace,* vol. 8, no. 3, p. 65, 2021.

[6] R. Tarakan, K. Baldwin and N. Rozen, "An automated simulation pilot capability to support advanced air traffic controller training," in *The 26th Congress of ICAS and 8th AIAA ATIO*, Anchorage, AK, USA, 2008.

[7] S. Schultheis, "Integrating advanced technology into air traffic controller training," The MITRE Corporation, McLean, VA, USA, 2014.

[8] J. M. Cordero, N. Rodriguez, J. M. de Pablo and M. Dorado, "Automated speech recognition in controller communications applied to workload measurement," in *3rd SESAR Innovation Days*, Stockholm, Sweden, 2013.

[9] H. Helmke, O. Ohneiser, J. Buxbaum and C. Kern, "Increasing ATM efficiency with assistant-based speech recognition," in *12th USA/Europe Air Traffic Management Research and Development Seminar (ATM2017)*, Seattle, WA, USA, 2017.

[10] S. V. Subramanian, P. F. Kostiuk and G. Katz, "Custom IBM Watson speech-to-text model for anomaly detection using ATC-pilot voice communication," in *2018 Aviation Technology, Integration, and Operations Conference*, Atlanta, GA, USA, 2018.

[11] H. Kopald and S. Chen, "Design and evaluation of the closed runway operation prevention device," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Chicago, IL, USA, 2014.

[12] H. Helmke et al., "Readback error detection by automatic speech recognition to increase ATM safety," in *14th USA/Europe Air Traffic Management Research and Development Seminar (ATM2021)*, Virtual Conference, 2021.

[13] Y. Lin, L. Deng, Z. Chen, X. Wu, J. Zhang and B. Yang, "A real-time ATC safety monitoring framework using a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems,* vol. 21, no. 11, pp. 4572-4581, November 2020.

[14] M. Lowry, T. Pressburger, D. A. Dahl and M. Dalal, "Towards autonomous piloting: communicating with air traffic control," in *AIAA Scitech 2019 Forum*, San Diego, CA, USA, 2019.

[15] S. Chen, H. Kopald, B. Avjian and M. Fronzak, "Automatic pilot report extraction from radio communications," in *2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC)*, Portsmouth, VA, USA, 2022.

[16] H. Helmke et al., "Ontology for transcription of ATC speech commands of SESAR 2020 solution PJ.16-04," in *IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, London, United Kingdom, 2018.

[17] Bundesministerium fur Bildung und Forschung, "KI-in der Praxis," Bundesministerium fur Bildung und Forschung, [Online]. Available: https://www.softwaresysteme.pt-dlr.de/de/ki-in-der-praxis.php.

[18] Deutsches Zentrum fur Luft und Raumfahrt (DLR), "Virtual/augmented reality applications for tower (SESAR Solution PJ.05-W2-97.1)," [Online]. Available: https://www.remote-tower.eu/wp/project-pj05-w2/solution-97-1/.

[19] Sesar Joint Undertaking, "Industrial research project: digital technologies for tower," SESAR 3 Joint Undertaking, [Online]. Available: https://www.sesarju.eu/projects/DTT.

[20] Europaische Kommission, "PJ.10 W2 Separation Management and Controller Tools," [Online]. Available: https://cordis.europa.eu/programme/id/H2020_SESAR-IR-VLD-WAVE2-10-2019/de.

[21] Deutsches Zentrum für Luft- und Raumfahrt (DLR), "HAAWAII: highly automated air traffic controller workstations with artificial intelligence integration," [Online]. Available: https://www.haawaii.de/wp/.

[22] Federal Aviation Administration, "JO 7110.65Z - Air Traffic Control," Federal Aviation Administration, Washington, DC, USA, 2021.

[23] J. D. Foley and A. Van Dam, Fundamentals of interactive computer graphics, 1st edition, Reading, MA, USA: Addison-Wesley Publishing Company, 1982.

[24] H. Helmke et al., D3.1 Transcription and Annotation Handbook, version 1, HAAWAII Project, November 5, 2020.

[25] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet physics. Doklady,* vol. 10, no. 8, pp. 707-710, 1966.

[26] International Civil Aviation Organization, "Procedures for Air Navigation Services (PANS) - Air Traffic Management (Doc 4444)," International Civil Aviation Organization, Montreal, Canada, 2016.