Grammar for
Natural Language Recognition

Adaptation to Air Traffic Phraseology

IB-Nummer 112-2011/03
Zugänglichkeitsstufe A/I

Institute of Flight Guidance
Director: Prof. Dr. Dirk Kügler

Document Control Information

Responsible Project Manager/ Department Head:	Jürgen Rataj
Responsible Author:	Dagi Troegner
Additional Author(s):	
Project / Research Task:	Natural Language Recognition
Classification:	A/I (only for use in the Institute of Flight Guidance)
Filename:	IB_NatürlicheSprache.pdf
Release:	1.0
Save date:	2011-01-17
Total Pages:	19

Release Note:
Document is released in accordance to Document Control Sheet

© 2010, DLR, Institute of Flight Guidance, Germany
This document with all its parts is protected by copyright. Any use within or without the domain of the copyright act is illegal without a written consent of the DLR, Institute of Flight Guidance and will be prosecuted. This applies in particular to copying, translations, microfilm reproductions or converting, processing and storing this information on digital systems.

Change Log

Release	Date	Changed Pages or Chapters	Comments
1.0	04.01.2011	all	initial release

Zusammenfassung (Executive Summary)

Natürliche Sprache ist ein wesentliches Element im Verständnis menschlicher Handlungen. Im Bereich des Flugverkehrs und insbesondere in der Flugführung ist gesprochene Sprache weiterhin das vorwiegende Kommunikationsmittel. Die von der International Civil Aviation Organization (ICAO) vorgegebene Phraseologie bestimmt hierbei die Terminologie und den Wortlaut im Sprechfunk. In diesem Bericht wird ein Ansatz zur Formulierung einer Grammatik vorgestellt, die die verwendete natürliche Sprache in der Kommunikation im Luftverkehr definiert. Die dargestellte Grammatik basiert auf einer abstrakten Syntax zur Beschreibung von natürlicher Sprache für eine Verwendung in der Computerlinguistik. Die abstrakte Syntax dient als Basis zur Ableitung einer Grammatik für die Englische Sprache, basierend auf einer Formalisierung der Universalgrammatik(UG) von Noam Chomsky. Die abgeleitete Grammatik wurde modifiziert und an die individuellen Erfordernisse einer Fluglotsen- und Piloten-Phraseologie angepasst. Unvollständigkeiten, Mehrdeutigkeiten und Parenthesen (Zwischenbemerkungen), die mit natürlicher Sprache verbunden sind, wurden in die resultierende Grammatik *ATCCommands* integriert. Der Bericht enthält eine Beschreibung der Grundlagen zur Ableitung der vorgestellten Grammatik sowie eine detaillierte Beschreibung der definierten Elemente der Grammatik *ATCCommands*.

Scope of Document

Natural Language is a fundamental element in the comprehension of human actions. In the area of air traffic control speech is still a basic means of communication. A specified phraseology by the International Civil Aviation Organization (ICAO) predetermines the terminology and wording in radiotelephony. In this paper an approach is presented to formulate a grammar to specify the natural language that is used in air traffic communications. The grammar presented is based on an abstract syntax describing natural language for an application in computer linguistics. It is used as a foundation to deduce a grammar for the English Language, based on a formalization of the Universal Grammar (UG) by Noam Chomsky. The deduced grammar is adapted and modified to differences given by the air traffic controllers' and pilots' phraseology. Incompleteness, ambiguity and parentheses (incidental remarks) associated with natural language are integrated into the resulting grammar *ATCCommands*. The definition of the grammar and its fundamentals will be described in this paper.

Contents

1 Grammar for Natural Languages

Natural languages are concepts that enable a phrasing of thoughts and considerations for self-expression and communication in verbal and written form. The functionality of natural languages is defined in [1] as a medium for planning, problem solving, conclusion, prediction and memory. A special characteristic of natural languages compared to formal languages used in computer sciences is besides ambiguity and incompleteness, the ability to reference objects out of an actual sentence sequence. From a linguistic point of view, the information sciences can be regarded as the application of computer technologies as medium for communication between different stakeholders [2]. From this it follows that in the area of application of Arrival Management Systems (AMAN) a software system can be regarded as an implementation of a communication interface between air traffic controllers and pilots. In this chapter an abstract syntax for the description of natural languages for the use in computer linguistics is described. Based on the abstract language model a grammar for the English language can be deduced. The grammar for the English language is based on a formalization of the Universal Grammar (UG) by Noam Chomsky. The deduced grammar is adapted and modified for the purpose of defining a language for the phraseology used in air traffic communications.

1.1 Syntax and Semantics of Natural Languages

The set of allowable expressions is defined by formal principles and rules (syntax) of a language. A syntax describes possible sequences of characters to form grammatically correct words and sentences. Natural languages underlie incessantly changes, which leads despite syntactical correctness to a development of ambiguity in a language.

The fundamental elements of a language can be assigned to syntactical categories. A syntactical category is a class of elements (constituent) with similar morphosyntactic characteristics. A constituent describes a linguistic element as part of a bigger entity. In linguistics lexical and phrasal categories can be differentiated. Whereas simple constituents (words) are part of the lexical category, the phrasal category contains combinations of complex constituents (phrases). The elements of the lexical category can therefore be defined for a natural language by the word classes Noun (N), Verb (V), Adjective (A), Adverb (Λ) and Preposition (P).

The phrasal category is composed of sequences of words and can therefore be regarded as an extension of the lexical category. Sequences of words or phrases are always affected by a certain word class. A phrase can be formulated as a Verb Phrase (VP), a Noun Phrase (NP), a Prepositional Phrase (PP), an Adverb Phrase (ΛP) or an Adjective Phrase (AP). A Verb Phrase is a sequence of associated words, that have a verb as a core. It is formulated by an extension of the lexical category Verb (V). Similarly, the meanings and formulations of the further phrasal categories can be derived; they as well are sequences of words that are build around a respective core (Noun, Verb, Adjective, Adverb, Preposition) and can be considered as an extension of their lexical category. Complement Phrases (CP) are another element of the phrasal category. A Complement Phrase can be used to embed parentheses within a main clause by preceding a complement word (complementizer). A complementizer is a subordinate conjunction, like *that, although* or *after*. The words described so far have a descriptive meaning. In contrast, the

Table 1.1: Lexical and Syntactic Categories of Natural Languages [2]

Category	Subcategory	Symbol	Description
Lexical ($\mathcal{L}$)			
	Noun	N	Entities and abstract objects
	Verb	V	Actions, states and possessions
	Adjective	A	Properties of a noun
	Adverb	Λ	Properties of a verb
	Preposition	P	Designated relations in space or time
Functional ($\mathcal{F}$)			
	Determiner	τ	the, a, this, these etc.
	Degree word	δ	too, so, very, more etc.
	Qualifier	κ	almost, always, often etc.
	Auxiliary	α	will, can, should etc.
	Conjunction word	γ	and, or, that etc.
	Negative	$\neg$	not
Phrasal ($\mathcal{P}$)			
	Noun Phrase	NP	τN[PP]
	Verb Phrase	VP	V[NP]
	Adjective Phrase	AP	[δ]A[PP]
	Adverb Phrase	ΛP	[Λ]V — V [Λ]
	Prepositional Phrase	PP	[δ]P[NP]
	Complement Phrase	CP	Supplemental part of N/NP, V/VP, A/AP or P/PP
Relational ($\mathcal{R}$)			
	Sequential	$\rightarrow$	and, then
	Branch	|	or
	Parallel	||	and, simultaneously(actions by the same subjects)
	Embedded	$\rightarrowtail$	that, which, whether etc.
	Concurrent	$\oiint$	and, simultaneously(actions by different subjects)
	Interleave	|||	alternatively
	interrupt	$\lightning$	when, while, during etc.

elements of the functional category solely represent grammatically relevant characteristics of a complex phrase [4]. Another category is represented by the subgroup of relational connectors. Connectors are linguistic expressions that correlate sentences in a shared semantical relationship. In table 1.1 the specified lexical and syntactic categories are listed. Optional elements are labeled with squared brackets.

The syntactic elements that have been specified in 1.1 can be classified according to [2] into a lexical ($\mathcal{L}$), functional ($\mathcal{F}$), phrasal($\mathcal{P}$) and relational ($\mathcal{R}$) category. This leads to the following definition of the syntactic and lexical elements of a natural language.

Definition 1.

$$\begin{array}{lll} \mathcal{S} & \widehat{=} & (\mathcal{L}, \mathcal{F}, \mathcal{P}, \mathcal{R}) \\ & = & \{N, V, A, \Lambda, P\} \\ & \| & \{\tau, \delta, \kappa, \alpha, \gamma, \neg\} \\ & \| & \{NP, VP, AP, \Lambda P, PP, CP\} \\ & \| & \{\rightarrow, |, \|, \mapsto, \oiint, |||, \lightning\} \end{array}$$

The syntactical analysis of a language is based on a predefined set of rules, the grammar of a language. Languages are normally infinite objects [5]. Therefore a finite method of description for languages is essential to an algorithmic ability for processing. A grammar is defined according to U.Schöning [5] as follows:

Definition 2. *A grammar G is defined by the 4-tuple G = (V, Σ, P, S), where*

- *V is a finite set of non-terminal characters or variables with* $V \cap \Sigma = \varnothing$.
- Σ *is a finite set of terminals, disjoint from V, which make up the actual content of the sentence. The set of terminals is the alphabet of the language defined by the grammar.*
- *P is a finite set of rules or productions. Formally P is a finite subset of* $(V \cup \Sigma)^+ \times (V \cup \Sigma)^*$.
- *S is the start variable (or start symbol) with* $S \in V$.

The language L(G) of a grammar G is defined by the set $L(G) = \{w \in \Sigma^* : S \overset{*}{\Rightarrow} w\}$. Therefore, all words w that can be derived from the start symbol S are considered to be element of the language L(G) defined by G.

1.2 Universal Grammar by Noam Chomsky

The Universal Grammar (UG) is a theory in linguistics that suggests that there are properties that all known natural human languages have. The theory suggests that some rules of grammar are hard-wired into the brain, and manifest without being taught. The Universal grammar consists of a set of unconscious constraints that let us decide whether a sentence is correctly formed. This mental grammar is not necessarily the same for all languages. But according to theories by Noam Chomsky, the process by which, in any given language, certain sentences are perceived as correct while others are not, is universal and independent of meaning. According to O'Grady and Archibald [6] five basic principles shared by all Natural Languages can be identified. These basic principles are known as generality, parity, universality, mutability, and inaccessibility.

- Generality - All natural languages can be defined by a grammar
- Parity - All grammars are equivalent in terms of their expressive capacity
- Universality - All languages share basic principles and properties
- Mutability - Grammars are constantly changing over time
- Inaccessibility - a grammatical knowledge of the mother tongue is built at the subconscious layer of the brain

These five principles build the foundation for a theory describing a Universal Grammar. The Universal Grammar (UG) is a system of categories, mechanisms, and constraints shared by all human languages. The UG is perceived to be innate in all humans providing a generic type of syntactic mechanisms. This generic mechanism provides the ability to combine words and phrases in an appropriate structure. The principle of Parity states that all grammars are equivalent, which leads to the assumption that all grammars have to follow a common high-level rule. According to Yingxu Wang [2], the top-level structure of a sentence in any natural language can be formalized as in (1.1).

$$S \rightarrow [Subject]Predicate|S \tag{1.1}$$

Using the top-level sentence structure defined in (1.1) with S being the start symbol and the lexical and syntactical elements described in table 1.1, a grammar for the English Language can be derived. This grammar and its derivation are described in the following section.

1.3 Grammar for the English Language

The grammar of the English Language is an instantiation of the Universal Grammar. In [2] a deductive grammar to define the syntactical rules of a language is formally described. Based on this deductive grammar valid sentences of a language can be derived. The Deductive Grammar of the English Language (DGE) uses the fundamental sentence structure in equation 1.1 for the definition of a start symbol and the listed elements in table to constitute variables, a terminal alphabet and a set of productions.

An excerpt of the grammar G_{NL} for the English Language based on [2] is defined by the equations (1.2)-(1.6). The productions defined by the grammar are represented by rules in the form

$$\textit{left-hand side} \rightarrow \textit{right-hand side}.$$

Furthermore, optional elements are labeled with a question mark (?) and arbitrary elements with an asterisk (*). With the usage of a vertical hyphen alternative right-hand sides of a production can be combined. Terminal words and symbols are comprised in lexical rules that are described by right-hand sides of productions enclosed by angle brackets (<>).

$$G_{NL} = (V, \Sigma, P, S), with \tag{1.2}$$

$$\begin{aligned} V = \{&alpha, delta, gamma, kappa, not, predicate, rule\_A, rule\_AP,\\ &rule\_Lambda, rule\_N, rule\_NP, rule\_Object, rule\_P,\\ &rule\_PP, rule\_S, rule\_V, rule\_VP, rule\_start, subject, tau\} \end{aligned} \tag{1.3}$$

$$\Sigma = \{terminal alphabet\ of\ including\ numbers,\ special characters\} \tag{1.4}$$

$$S = \{sentence\} \tag{1.5}$$

$$P = \{sentence \rightarrow ((subject)?predicate)?(gamma\ sentence)^*\} \tag{1.6a}$$
$$\cup\ \{subject \rightarrow\ rule\_NP\} \tag{1.6b}$$
$$\cup\ \{predicate \rightarrow\ rule\_VP\} \tag{1.6c}$$
$$\cup\ \{rule\_NP \rightarrow\ ((tau)?(rule\_AP)?rule\_N(rule\_PP)?\}$$
$$\cup\ \{rule\_AP)? \mid tau\ rule\_N)^*)(gamma\ rule\_NP)^*\} \tag{1.6d}$$
$$\cup\ \{rule\_PP \rightarrow\ (lambda)?rule\_P(rule\_NP)?\} \tag{1.6e}$$
$$\cup\ \{rule\_AP \rightarrow\ (delta \mid kappa)^*((lambda)?rule\_A|lambda)?\} \tag{1.6f}$$
$$\cup\ \{rule\_VP \rightarrow\ (alpha?eta?rule\_V(rule\_AP)?(rule\_Object)?$$
$$\mid (lambda\ rule\_P)?rule\_V(eta)?(rule\_Object)*$$
$$\mid eta$$
$$\mid (rule\_Object)^*(lambda\ rule\_P)?)(gamma$$
$$rule\_VP(alpha)?(eta)?rule\_V(rule\_Object)^*)^*\} \tag{1.6g}$$
$$\cup\ \{rule\_Object \rightarrow\ rule\_NP\} \tag{1.6h}$$
$$\cup\ \{rule\_N \rightarrow < NOUNS >\} \tag{1.6i}$$
$$\cup\ \{rule\_V \rightarrow\ (< TO >)?(< to\_be > \mid < to\_have >$$
$$\mid < to\_do >)\ eta?\} \tag{1.6j}$$
$$\cup\ \{rule\_P \rightarrow\ propositions\}$$
$$\cup\ \{delta \rightarrow\ degree\_words \mid (comparison)?$$
$$(< INT > \mid < FLOAT >)$$
$$< UNIT\_SPECIFICATION > \mid comparison\} \tag{1.6k}$$

The grammar G_{NL} is based on the grammar for the DGE presented in [2]. Modifications and adaptations were made to enhance the expressiveness of the resulting language definition. The grammar G_{DGE} is simplified to reduce complexity and so particularly does not address person rules of nouns, time rules of verbs or the matching of nouns and verbs in sentences. These simplifications are basically addressed in G_{NL}. The following adjustments have been made:

- Enhancement of possible sentence structures
- Matching of Nouns and Verbs for a sentence
- Rules to derive tenses of regular verb forms
- Semantic predicates for the identification of a sentence's meaning
- Numerical comparisons of objects in sentences
- Plural determination

A detailed description of the enhancements in grammar G_{NL} can be found in [11]. G_{NL} was formulated at first to enable the recognition of constraints formulated in a natural language. These constraints are used to restrict solution spaces in model checking and model completion algorithms for domain-specific modeling languages in the area of Arrival Management. The application area is extended to include these constraints in the calculations of an Arrival Management System and therefore enables operators to formulate distinct preferences and procedural methods for their individual system. This approach is presented in [11].
Based on the grammar G_{NL} a definition for the phraseology used in air traffic can be derived with little basic modifications. The phraseology and the derivation of an according grammar for the language are described in the following chapter.

2 Adaptation to Air Traffic Phraseology

2.1 Phraseology according to ICAO

The need for clear and unambiguous communication between pilots and Air Traffic Control (ATC) is vital in assisting the safe and expeditious operation of aircraft. It is important, therefore, that due regard is given to the use of standard words and phrases and that all involved ensure that they maintain the highest professional standards when using RTF (Radiotelephony). Phraseology has evolved over time and has been carefully developed to provide maximum clarity and brevity in communications while ensuring that phrases are unambiguous. However, while standard phraseology is available to cover most routine situations, not every conceivable scenario will be catered for and RTF users should be prepared to use plain language when necessary following the principle of keeping phrases clear and concise. The standard words and phrases (phraseology) used in air traffic is subject to a set of international and national regulations. At international level the regulations according to aeronautical radio services are accountable to two organizations [15]:

- The International Civil Aviation Organization (ICAO) with the Annex 10 (Aeronautical Telecommunications) [14] and further ICAO documents (e.g. ICAO Doc9432 [13], ICAO Doc4444 [12]).
- The International Telecommunication Union (ITU) or L'Union Internationale de Télécommunications (UIT). The ITU is responsible for assigning frequency ranges and the registration of frequencies. The organization is involved in conducting studies and enunciating recommendations concerning the telecommunication sector.

National regulations in Germany are specified in the Air Traffic Act (Luftverkehrsgesetz (LuftVG)), the legal basis to enact statutory orders. Based on the LuftVG further procedures are specified, e.g. procedures according to radiotelephone traffic are specified in the air traffic statutory orders (Luftverkehrs-Ordnung(LuftVO)).

In Annex 10 the basic specifications for the air traffic communication are comprised. This includes the language, time system and alphabet that has to be used in any communication. Phrasing and words are specified in the ICAO Documents and can be classified into six basic categories.

1. Clearance and Taxi (ground-based procedures)
2. Take-off and Departure (procedures for phase of flight transition)
3. Read-back (confirming and repeating procedures)
4. Climb, Cruise and Descent (procedures altering flight attitude)
5. Approach and Landing (procedures for phase of flight transition)
6. Emergency Communications (procedures preventing accidents)

Phrasing examples for each of the six categories can be found in ICAO Doc4444. In this paper a focus on defining a grammar for the categories Read-Back (3), Climb, Cruise and Descent (4) and Approach and Landing (5) is set. The grammar is used to recognize commands given by air traffic controllers to arrival aircraft in Terminal Maneuvering Areas (TMA). The commands are translated and can be used to appraise an air traffic controller's intention. This can be usefully implemented in a variety of systems used in air traffic control. In arrival management systems these information enables the system to adapt more rapidly to procedure variations initiated by the air traffic controller or any pilot in his responsibility. The phraseology used in air traffic communication differs from spoken language in many aspects. The characteristics of this phraseology and the resulting grammar are presented in the following section.

2.2 Grammar *ATCCommands*

The phraseology specified by the ICAO differs from spoken language in basically four aspects. These are *Incompleteness*, *Sentence structure*, *Parentheses* and *Restricted Vocabulary*. In comparison to Natural Languages the ICAO phraseology uses a more clear and concise wording. This leads to a more compact sentence structure, mostly without using prepositions, qualifier words or articles similar to imperatives. The basic sentence structure can be reduced to

$$< Noun >< Verb >< Degree\ Word >.$$

An example of this a compact phrasing could be an air traffic controller command calling to an aircraft to change its altitude in the approach for a landing.

"DLH123 DESCEND 5000 FEET".

This compact sentence structures can be described as *incomplete* compared to sentences formulated in a Natural Language. To adapt the grammar defined by G_{NL} in section 1.1 several variables in the productions are labeled as optional, allowing the phraseology to use only elementary sentence elements and therefore be more compact and precise.
Another difference between the ICAO-phraseology and a Natural Language is the altered sentence structure. In Natural Languages the top-level sentence structure is formed by an optional *Subject* followed by a *Predicate*. The phraseology requires however the ability of a concatenation of more than one sequenced *Predicate*. This leads to an adjusted *sentence structure* compared to G_{NL}. The top-level production is therefore altered to equation 2.1.

$$sentence \rightarrow (((subject)?predicate)?predicate?(gamma\ sentence)^{*}). \tag{2.1}$$

A pilot confirming his change of altitude with information of actual flight altitude can be an example of the described predicate concatenation.

"DLH123 LEAVING FLIGHT LEVEL 90 DESCENDING ALTITUDE 5000 FEET".

The ICAO phraseology does not include polite addressing forms that are often included in initial calls or last calls while sector handovers. This could be greetings, leave-takings or just comments integrated into an official phrasing. These optional insertions can be included at any position in the sentence. In linguistics they are called *Parentheses*. The grammar for the phraseology is adapted to define parentheses by adding a further production to the lexer *CONTACT-PHRASE*. The production contains some often used expressions. As they are not containing any relevant information for the translation, they are marked by an instruction as hidden tokens to

the lexer.
Compared to Natural Language with a wide vocabulary, the ICAO-phraseology is restricted to specified phrases and a terminology. The noun of a sentence is exclusively a callsign of an aircraft or a ground station. Both have a defined sequence of predetermined words or numbers. A callsign is composed of an airline name (or its equivalent 3-letter code) and a sequence of two to four numbers, whereas the ground station is composed by an radar or tower control center and its function designator (Apron, Control, Arrival, Tower etc.). The airline names are predefined in ICAO Doc8585, where all actual airline names and their according 3-letter codes are listed. The restricted vocabulary in the phraseology constitutes a more compact terminal alphabet and therefore more compact lexer rules.

In Natural Languages as well as in ICAO-phraseology, simple sayings are included in the grammar. The sayings defined by the ICAO, see Annex 10, are included into the grammar by enhancing the set of productions by the production *simplePhrase*. It contains a list of terminal words and phrases allowed in the phraseology. One of the sayings defined in ICAO Doc8585 is used to ask for the transmission quality:

"HOW DO YOU READ".

The complete list of sayings is included in the corresponding lexer rule. The usage of simple phrases, like sayings, is integrated into the grammar in the top-level sentence structure as an alternative production. From this it follows that equation 2.1 is enhanced by an alternative production. The result is equation 2.2.

$$sentence \rightarrow simplePhrase|(((subject)?predicate)?predicate?(gamma\ sentence)^{*}). \quad (2.2)$$

The resulting grammar after the described modifications is a grammar G_{ATC} defining a large part of the ICAO-phraseology according to ICAO Annex 10. The complete definition of grammar G_{ATC} is listed in Annex B.

Annex A

Abbreviations

Abbr.	Abbreviation
AIP	Aeronautical Information Publication
AMAN	Arrival Management System
ATC	Air Traffic Control
DLR	Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Centre)
ICAO	International Civil Aviation Organization
LuftVG	Luftverkehrsgesetz
RTF	Radiotelephony
TMA	Terminal Maneuvering Area
UG	Universal Grammar

Annex B

Grammar ATCCommands

$$G_{ATC} = (V, \Sigma, P, S), with \tag{2.3}$$

$$V = \{alpha, delta, gamma, kappa, not, predicate, rule\_A, rule\_AP, rule\_Lambda, rule\_N, rule\_NP, rule\_Object, rule\_P, rule\_PP, rule\_S, rule\_V, rule\_VP, rule\_start, subject, tau\} \tag{2.4}$$

$$\Sigma = \{Terminology\ according\ to\ ICAO\} \tag{2.5}$$

$$S = \{sentence\} \tag{2.6}$$

$$P = \{sentence \rightarrow < SIMPLEPHRASE > |(((subject)?predicate)?predicate?(gamma\ sentence)^*)\} \tag{2.7a}$$

$$\cup\ \{subject \rightarrow\ rule\_NP\} \tag{2.7b}$$

$$\cup\ \{predicate \rightarrow\ rule\_VP\} \tag{2.7c}$$

$$\cup\ \{rule\_NP \rightarrow\ ((tau)?(rule\_AP)?rule\_N(rule\_PP)?\} \cup\ \{rule\_AP)? \mid tau\ rule\_N)^*)(gamma\ rule\_NP)^*\} \tag{2.7d}$$

$$\cup\ \{rule\_PP \rightarrow\ (lambda)?rule\_P(rule\_NP)?\} \tag{2.7e}$$

$$\cup\ \{rule\_AP \rightarrow\ (delta \mid kappa)^*)(lambda)?(rule\_A)?\} \tag{2.7f}$$

$$\cup\ \{rule\_VP \rightarrow\ (alpha?eta?rule\_Vrule\_P?(rule\_Object)^*(rule\_AP)? \mid (lambda\ rule\_P)?rule\_V(eta)?(rule\_Object)* \mid eta \mid (rule\_Object)^*(lambda\ rule\_P)?)(gamma\ rule\_VP(alpha)?(eta)?rule\_V(rule\_Object)^*)^*\} \tag{2.7g}$$

$$\cup\ \{rule\_Object \rightarrow\ rule\_NP\} \tag{2.7h}$$

$$\cup\ \{rule\_N \rightarrow\ nouns|prep\_Noun\} \tag{2.7i}$$

$$\cup\ \{rule\_A \rightarrow\ adjective(< INT >< INT >< ALPHABET >?adjective)?\} \tag{2.7j}$$

$$\cup\ \{nouns \rightarrow\ ground\_station|callsign|alphabet|general\} \tag{2.7k}$$

$$\cup\ \{prep\_Noun \rightarrow < PREPNOUN > (< INT >< INT >< ALPHABET >?)?\} \tag{2.7l}$$

$$\cup\ \{rule\_V \rightarrow < VERBLIST >^+ < DIRECTION >?\} \tag{2.7m}$$

$$\cup\ \{callsign \rightarrow\ airline\} \tag{2.7n}$$

$$\cup\ \{rule\_P \rightarrow\ propositions\} \tag{2.7o}$$

$$\cup\ \{delta \rightarrow\ DEGREE\_WORDS \mid (COMPARISON)? (< INT > \mid < FLOAT >) < UNIT\_SPECIFICATION >? \mid < UNIT\_SPECIFICATION > (< INT > \mid < FLOAT >) \mid comparison\} \tag{2.7p}$$

$$\cup\ \{alphabet \rightarrow < ALPHABET >\} \tag{2.7q}$$

$$\cup\ \{airline \rightarrow\ (< AIRLINES > \mid < ALPHABET >^+) < INT >< INT >< INT >? < INT >?\} \tag{2.7r}$$

$$\cup\ \{ground\_station \rightarrow < AIRPORT >? < STATIONS >\} \tag{2.7s}$$

$$\cup \{adjective \rightarrow < ADJECTIVES >\} \quad (2.7t)$$

$$\cup \{gamma \rightarrow < COMMA >? < CONJUNCTION\_WORDS > (< CONJUNCTION\_WORDS >)?\} \quad (2.7u)$$

$$\cup \{tau \rightarrow < DETERMINER\_WORDS >\} \quad (2.7v)$$

$$\cup \{lambda \rightarrow < ADVERBS >\} \quad (2.7w)$$

$$\cup \{kappa \rightarrow < QUALIFIER\_WORDS >\} \quad (2.7x)$$

$$\cup \{alpha \rightarrow < AUXILIARY\_WORDS >\} \quad (2.7y)$$

$$\cup \{eta \rightarrow < NEGATION >\} \quad (2.7z)$$

Lexer-Rules L according to Σ:

$$L = \{< DEGREE\_WORDS > \rightarrow 'TOO'|'SO'|'VERY'|'MORE'|'QUITE' |'AROUND'|'ABOUT'|'APPROXIMATELY'|'CIRCA'|'NEARLY'\} \quad (2.8a)$$

$$\cup \{< COMPARISON > \rightarrow ('GREATER THAN'|'MORE THAN' |'LESS THAN')('OR EQUAL')?|'EQUAL TO'\} \quad (2.8b)$$

$$\cup \{< SIMPLEPHRASE > \rightarrow 'SAY AGAIN YOUR CALLSIGN' |'HOW DO YOU READ'|'SAY AGAIN'\} \quad (2.8c)$$

$$\cup \{< AIRPORT > \rightarrow 'LANGEN'|'BREMEN'|'BERLIN'|'KARLSRUHE' |'MUENCHEN'|'BERLIN TEGEL'|'DRESDEN'|'DUESSELDORF' |'ERFURT'|'FRANKFURT'|'HAMBURG'|'HANNOVER' |'KOELN BONN'|'LEIPZIG'|'MUENSTER'|'NUERNBERG' |'SAARBRUECKEN'|'STUTTGART'\} \quad (2.8d)$$

$$/* \textit{Complete List of Airlines according to ICAO Doc } 8585 */$$

$$/* \textit{Excerpt from Airline Lexer Rule} : */$$

$$\cup \{< AIRLINES > \rightarrow 'AMERICAN AIRLINES'|'BRITISH AIRWAYS' |'LUFTHANSA'|'SWISS AIR'|'IBERIA'|'UNITED AIRLINES'\} \quad (2.8e)$$

$$\cup \{< ADVERBS > \rightarrow < WVCLASS >\} \quad (2.8f)$$

$$\cup \{< ADJECTIVES > \rightarrow ('GEDERN'|'SPESSART'|'ROLIS' |'TRANSITION'|'FINAL')\} \quad (2.8g)$$

$$\cup \{< ALPHABET > \rightarrow ('ALPHA'|'BRAVO'|'CHARLIE'|'DELTA' |'ECHO'|'FOXTROT'|'GOLF'|'HOTEL'|'INDIA'|'JULIETT'|'KILO' |'LIMA'|'MIKE'|'NOVEMBER'|'OSCAR'|'PAPA'|'QUEBEC'|'ROMEO' |'SIERRA'|'TANGO'|'UNIFORM'|'VICTOR'|'WHISKEY'|'XRAY' |'YANKEE'|'ZULU')\} \quad (2.8h)$$

$$\cup \{< VERBLIST > \rightarrow ('ACKNOWLEDGE'|'AFFIRM'|'AFFIRMING' |'APPROVED'|'CANCEL'|'CHECK'|'IS CLEARED'|'CLEARED' |'CONFIRM'|'CONTACT'|'IS CORRECT'|'CORRECT'|'CORRECTION' |'DISREGARD'|'GO AHEAD'|'I SAY AGAIN'|'MONITOR'|'NEGATIVE' |'READ BACK'|'RECLEARED'|'REPORT'|'REQUEST'|'ROGER' |'SPEAK SLOWER'|'SQUAWK'|'STANDBY'|'WILCO'|'REDUCE' |'REDUCING'|'TURN'|'TURNING'|'DESCEND'|'DESCENDING' |'INDENTIFIED'|'PROCEED'|'PROCEEDING'|'LEAVE'|'LEAVING'$$

$$|'REQUESTING'|'RECEIVE'|'RECEIVING'|'RECEIVED'|'DIRECT'$$
$$|'FLY'|'FLYING'|'HOLD'|'ENTER'|'ENTERING'|'PASS'|'PASSING'$$
$$|'CLIMB'|'CLIMBING'|'CONTACTING'|'INTERCEPT'|'IDENTIFY'$$
$$|'IDENTIFIED'|'OVER'|'OUT')\} \quad (2.8i)$$

$$\cup \{< DETERMINER\_WORDS >\rightarrow 'THE'|'A'|'THESE'|'THIS'|'AN'\} \quad (2.8j)$$

$$\cup \{< DIRECTION >\rightarrow 'LEFT'|'RIGHT'\} \quad (2.8k)$$

$$\cup \{< CONJUNCTION\_WORDS >\rightarrow 'AND'|'OR'|'IF'|'IN\ CASE'$$
$$|'ELSE'|'OTHERWISE'|'THEN'|'THAT'|'WHICH'|'WETHER'\} \quad (2.8l)$$

$$\cup \{< EOS >\rightarrow ';'|('BREAK\ BREAK')\} \quad (2.8m)$$

$$\cup \{< STATIONS >\rightarrow 'CONTROL'|'APRON'|'RADAR'|'ARRIVAL'$$
$$|'DIRECTOR'|'PRECISION'|'TOWER'|'GROUND'\} \quad (2.8n)$$

$$\cup \{< QUALIFIER\_WORDS >\rightarrow 'ALMOST'|'ALWAYS'|'OFTEN'$$
$$|'PERHAPS'|'NEVER'\} \quad (2.8o)$$

$$\cup \{< AUXILIARY\_WORDS >\rightarrow 'WILL'|'MUST'|'CAN'|'MAY'$$
$$|'SHOULD'|'COULD'\} \quad (2.8p)$$

$$\cup \{< NEGATION >\rightarrow 'NOT'\} \quad (2.8q)$$

$$\cup \{< PREPNOUN >\rightarrow 'SPEED'|'POSITION'|'HEADING'$$
$$|'HOLDING\ PATTERN'|'ALTITUDE'|'INDICATED\ AIRSPEED'$$
$$|'TRUE\ AIRSPEED'|'FREQUENCY'|'QNH'|'RUNWAY'|'ILS'$$
$$|'INBOUND'|'APPROACH'\} \quad (2.8r)$$

$$\cup \{< UNIT\_SPECIFICATION >\rightarrow 'FLIGHT\ LEVEL'|'FEET'$$
$$|'DEGREES'|('NAUTICAL\ MILES')|'KNOTS'$$
$$|'HOURES'|'MINUTES'|'SECONDS'\} \quad (2.8s)$$

$$\cup \{< NUMBER >\rightarrow 'ZERO'|'ONE'|'TWO'|'THREE'|'FOUR'|'FIVE'$$
$$|'SIX'|'SEVEN'|'EIGHT'|'NINER'|'HUNDRED'|'THOUSAND'\} \quad (2.8t)$$

$$\cup \{< INT >\rightarrow '0'..'9'+\} \quad (2.8u)$$

$$\cup \{< FLOAT >\rightarrow '0'..'9'+ < DECIMAL >' 0'..'9'+\} \quad (2.8v)$$

$$\cup \{< DECIMAL >\rightarrow '.'\} \quad (2.8w)$$

$$\cup \{< WVCLASS >\rightarrow 'HEAVY'|'MEDIUM'$$
$$|'LIGHT'|'SUPERHEAVY'\} \quad (2.8x)$$

$$\cup \{< DECIMAL >\rightarrow '.'\} \quad (2.8y)$$

$$\cup \{< CONTACTPHRASE >\rightarrow 'GOOD\ MORNING'$$
$$|'GOOD\ AFTERNOON'|'GRUETZI'|'HAVE\ A\ GOOD\ DAY'$$
$$|'HALLO'|'HELLO'|'GOOD\ BYE')\} \quad (2.8z)$$

Bibliography

[1] J. L. Casti, *Complexity, Language, and Life - Mathematical Approaches*, Springer-Verlag GmbH, 1986.

[2] Y. X. Wang, *A formal syntax of natural languages and the deductive grammar*, Fundamenta Informaticae, vol. 90, no. 4, pp. 353–368, 2009.

[3] S. Barnert, M. Boeckh, D. M. Delbrück, and W. Greulich, *Der Brockhaus - Computer und Informationstechnologie*, F.A. Brockhaus GmbH, 2003.

[4] T. Gallmann, Peter Lindauer, *Funktionale Kategorien in Nominalphrasen*, Beiträge zur Geschichte der Deutschen Sprache und Literatur (PBB), vol. 116, p. 1–27, 1994.

[5] U. Schöning, *Theoretische Informatik - kurzgefasst*, Spektrum Akademischer Verlag, 2001.

[6] W. O'Grady and J. Archibald, *Contemporary Linguistic Analysis: An Introduction*, Addison Wesley Publishing Company, 2000.

[7] N. Chomsky, *Some Concepts and Consequences of the Theory of Government and Binding*, MIT Press, 1982.

[8] M. Lesk, *Lex - a lexical analyzer generator, computing science technical report no.39*, tech. rep., AT&T Bell Laboratories, 1975.

[9] S. Johnson, *Yacc - yet another compiler compiler, computing science technical report no.32*, tech. rep., AT&T Bell Laboratories, 1975.

[10] T. Parr, *ANTLR Version 2.7.3 Reference Manual*, 2004.

[11] D. Troegner, *Generic Constraint Generation for Domain-Specific Modelling Languages*, PhD Thesis, Leibniz University Hannover,(to appear).

[12] International Civil Aviation Organization (ICAO), *ICAO Doc4444, Procedures for Air Navigation Services - Air Traffic Management*, Fourteenth Edition, 2001.

[13] International Civil Aviation Organization (ICAO), *ICAO Doc9432, Manual of Radiotelephony*, Fourth Edition, 2007.

[14] International Civil Aviation Organization (ICAO), *Annex 10 - Aeronautical Telecommunications*, Volume II, Sixth Edition, 2001.

[15] H. Mensen, *Moderne Flugsicherung - Organisation, Verfahren, Technik*, Sppringer Verlag, 2004.

List of Tables