International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

F.746.10

(08/2020)

SERIES F: NON-TELEPHONE TELECOMMUNICATION SERVICES

Multimedia services

Architecture for spontaneous dialogue processing system for language learning

Recommendation ITU-T F.746.10

ITU-T F-SERIES RECOMMENDATIONS

NON-TELEPHONE TELECOMMUNICATION SERVICES

TELEGRAPH SERVICE	
Operating methods for the international public telegram service	F.1–F.19
The gentex network	F.20–F.29
Message switching	F.30–F.39
The international telemessage service	F.40–F.58
The international telex service	F.59–F.89
Statistics and publications on international telegraph services	F.90–F.99
Scheduled and leased communication services	F.100–F.104
Phototelegraph service	F.105–F.109
MOBILE SERVICE	
Mobile services and multidestination satellite services	F.110–F.159
TELEMATIC SERVICES	
Public facsimile service	F.160–F.199
Teletex service	F.200–F.299
Videotex service	F.300–F.349
General provisions for telematic services	F.350–F.399
MESSAGE HANDLING SERVICES	F.400–F.499
DIRECTORY SERVICES	F.500–F.549
DOCUMENT COMMUNICATION	
Document communication	F.550–F.579
Programming communication interfaces	F.580–F.599
DATA TRANSMISSION SERVICES	F.600–F.699
MULTIMEDIA SERVICES	**F.700–F.799**
ISDN SERVICES	F.800–F.849
UNIVERSAL PERSONAL TELECOMMUNICATION	F.850–F.899
ACCESSIBILITY AND HUMAN FACTORS	F.900–F.999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T F.746.10

Architecture for spontaneous dialogue processing system for language learning

Summary

Recommendation ITU-T F.746.10 describes the architecture, functional entities, and interfaces for a spontaneous dialogue processing system for language learning. The scope of this Recommendation is focused on describing the architecture with different functional components in a spontaneous dialogue processing system, which are: input/output management, dialogue understanding, dual dialogue management and generation, dialogue knowledge management, incremental dialogue knowledge learning, unstructured spontaneous speech recognition management and language learning function.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T F.746.10	2020-08-13	16	11.1002/1000/14327

Keywords

Dialogue processing, dialogue system, language learning framework, spontaneous speech.

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

Recommendation ITU-T F.746.10

Architecture for spontaneous dialogue processing system for language learning

1 Scope

This Recommendation describes the architecture, functional entities, and interfaces for a spontaneous dialogue processing system for language learning.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T F.746.5] Recommendation ITU-T F.746.5 (2017), *Framework for language learning system based on speech and natural language processing (NLP) technology*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 dialogue act [ITU-T F.746.5]: The user's intention or purpose of the utterances in a dialogue. Example: request for information, command for action, agreement.

3.1.2 natural language processing [b-ITU-T F.746.3]: A method that analyses text in natural languages through several processes such as part-of-speech recognition, syntactic analysis and semantic analysis.

3.1.3 speech recognition [b-ITU-T H.703]: A kind of user interface for translation of spoken words into text.

3.2 Terms defined in this Recommendation

This Recommendation defines the following term:

3.2.1 natural language understanding: A method that analyses text in natural languages to extract information through several processes such as part-of-speech recognition, syntactic analysis and semantic analysis.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ASR Automatic Speech Recognition

DB Database

IR Information retrieval

LM Language Model

PLU Phone-Like Unit

POS Part of Speech

QA Question Answering

TTS Text to Speech

5 Conventions

In this Recommendation:

– The keywords "is required to" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this Recommendation is to be claimed.

– The keywords "is prohibited from" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this Recommendation is to be claimed.

– The keywords "is recommended" indicate a requirement which is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance.

– The keywords "is not recommended" indicate a requirement which is not recommended but which is not specifically prohibited. Thus, conformance with this specification can still be claimed even if this requirement is present.

– The keywords "can optionally" indicate an optional requirement which is permissible, without implying any sense of being recommended. This term is not intended to imply that the vendor's implementation must provide the option and the feature can be optionally enabled by the network operator/service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with the specification.

6 Functional architectures for spontaneous dialogue processing system for language learning

6.1 Architectural framework

Figure 1 describes a general architecture of a spoken dialogue system which comprises the following modules:

– Speech recognition module to generate the user utterance sentence by transcribing the user utterance (speech to text),

– Natural language understanding module to generate several candidates for user utterance intention, which analyses the sentences spoken by a user to codify a representation of its meaning,

– System dialogue managing module to search for a system utterance intention and a pattern by referring to the dialogue knowledge database (DB) for meaning representation of several user utterance intention candidates,

– User dialogue managing module to search the dialogue knowledge DB for the subsequent several candidates for user utterance intention after the system dialogue managing module generates the system utterance intention and the pattern after a current user utterance,

– Dialogue generation module to search the dialogue knowledge DB for the system utterance intention and the pattern selected by the system dialogue managing module or the user dialogue managing module and a dialogue pattern with respect to the several candidates for user utterance intention, and to generate a system utterance text using the found dialogue pattern;

– Speech synthesis (or TTS (text to speech)) module to output the generated dialogue text using speech.

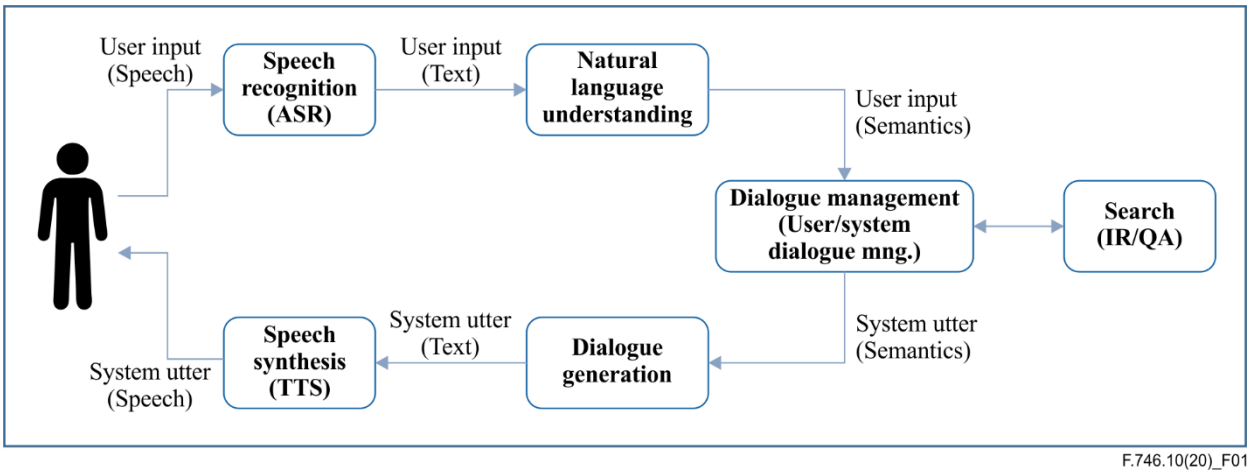

Figure 1 – General architecture of spoken dialogue system

6.2 Functional entities of spontaneous dialogue processing system for language learning

The functional entities that comprise the spontaneous dialogue processing system for language learning are described in Figure 2 as follows:

– Input/output management module,
– Dialogue understanding module,
– Dual dialogue management and generation module,
– Dialogue knowledge management module,
– Incremental dialogue knowledge learning module,
– Unstructured spontaneous speech recognition management module,
– Language learning module.

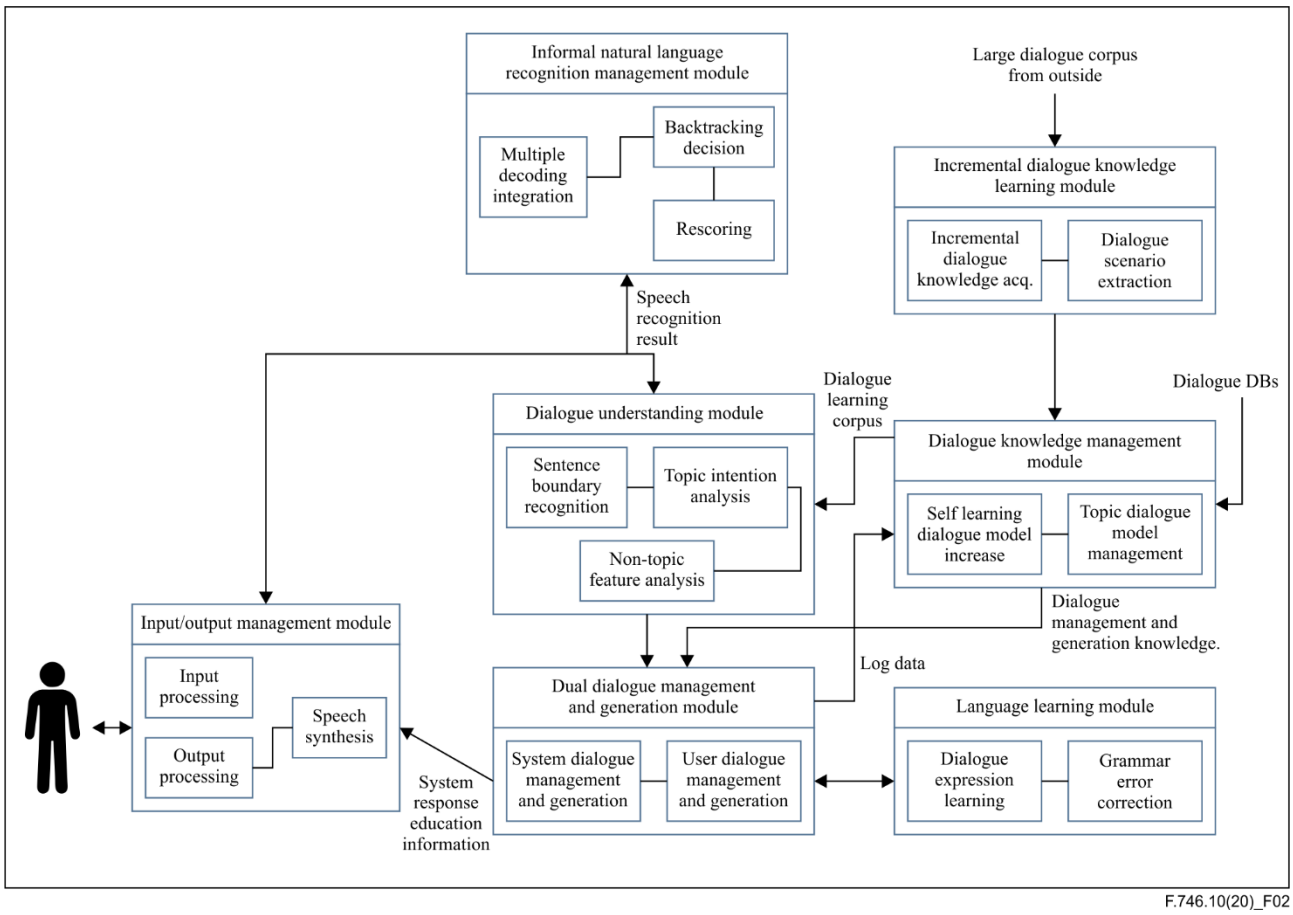

Figure 2 – Functional entities of spontaneous dialogue system for language learning

6.3 Input/output management module

The Input/output management function provides the audio input interface function which processes the speech data input from the microphone and the output interface function which recognizes speech data and sends back the recognition result to the user. The input/output management module receives the information from an unstructured spontaneous speech recognition management module, a dialogue understanding module as well as a dual dialogue management and generation module.

6.4 Dialogue understanding module

The dialogue understanding module performs dialogue act understanding function to express the dialogue intention/act of the utterances by analysing user's utterance strings and understanding their meaning. This module also performs dialogue act training function to automatically train dialogue patterns and to provide statistical classification of training information based on the domain dialogue corpus annotated with the dialogue acts. The goal of these module activities is to understand the participating user's intention.

There are three sub-modules in the dialogue understanding module as follows:

– Sentence boundary recognition sub-module: in this sub-module the sentences are processed so that the boundary of each sentence is recognized and sent to the next processing sub-module,

– Topic intention analysis sub-module: from the preceding sentence boundary recognition sub-module, each sentence arrives at this module and the topic intention is analysed using deep neural network.

– Non-topic feature analysis: in this sub-module, features other than the topic related ones such as dialogue acts are analysed, and the resulting information is produced for more analysis.

6.5 Dual dialogue management and generation module

Dual dialogue management function is to analyse both system dialogues and user dialogues for n-best dialogue acts and meaning expressions to decide the best dialogue act to generate the relevant system dialogue and manage the dialogue situations. Dialogue generation function selects the correct system response templet for the user's intention and decides meaningful information values to fill in the slots of the selected templet. Finally, it generates the appropriate system dialogue sentences.

Dual dialogue management and generation function provide the following sub functions:

– User intention analysis function to understand the user's intention correctly for the current topic,

– Auxiliary topic handling function to provide responses if the topic is outside the main topic and to guide the dialogue towards the appropriate topic flow,

– Topic management knowledge search function to find relevant response patterns, slots, or task knowledge to achieve the objective of the current topic,

– Dialogue history management function to store changing tasks, slots and intentions and monitor the intention flows of the dialogue,

– Next utterance recommendation function to suggest the next intention and utterances of the user according to the ongoing dialogue flow,

– System reply generation function to provide responses according to the user's intention,

– Dialogue management log record function to record the dialogue flow, the history changes or search conditions. This log information is used to check the errors in the system and dialogue knowledge.

6.6 Dialogue knowledge management module

Dialogue knowledge management module performs the function of storing the dialogue knowledge required to achieve the dialogue goal of a domain task in the dialogue knowledge DB. The dialogue knowledge management module includes the function of updating the dialogue knowledge DB, based on hierarchical task, with the subtask dialogue knowledge that is constructed or newly generated by a dialogue service designer. Dual dialogue management function in a spoken dialogue system uses a hierarchical dialogue task library for a system response suitable for a user utterance input. A spoken dialogue processing system performs a dialogue with a user by repeatedly performing the process of recognizing a user utterance as a user utterance sentence generating a system utterance text referring to the dialogue knowledge DB in order to achieve the dialogue goal of the domain task, and thereby producing as speech output the generated system utterance text.

6.7 Incremental dialogue knowledge learning module

The incremental dialogue knowledge learning module performs the function of adding new dialogue knowledge in the dialogue management database by pseudo situation simulation and incremental learning of situation states. The sub-functions are automatic extraction sub-function of conversation scenarios and incremental dialogue learning sub-functions.

Automatic extraction of conversation scenarios sub-function is dedicated to collect big data dialogues among people from the scenarios of movies and dramas and to perform the speaker boundary recognition, relevance assessment and topic classification.

Incremental dialogue learning sub-function is dedicated to recognize the topic of the dialogue scenarios, augment with the slot information, intention and subjects according to the topic, and to produce automatically or by human validation of the dialogue knowledge that can be used for the dialogue system.

6.8 Unstructured spontaneous speech recognition management module

Unstructured spontaneous speech includes speech errors, stuttering, murmuring hesitation, such as "hmm" and it is more difficult to recognize it. Unstructured spontaneous speech recognition management module integrates multiple decoding results, deciding on the partial backtracking section and rescoring of the selected section to get the best speech recognition results. When the multiple decoding results are integrated, the following decoding and detecting results are aligned and integrated according to the time frames (see Figure 3):

– Spontaneous speech recognition results from multi-model-based decoding module,

– Multiple detector results from multi-model-based decoding module,

– Prosodic feature-based speech detection results from multi-model-based decoding module,

– Frame unit utterance verification results from phone-like unit-based (PLU-based) decoding module,

– PLU detector results from PLU-based decoding module,

– Symbol recognition results per time from symbol unit recognition module.

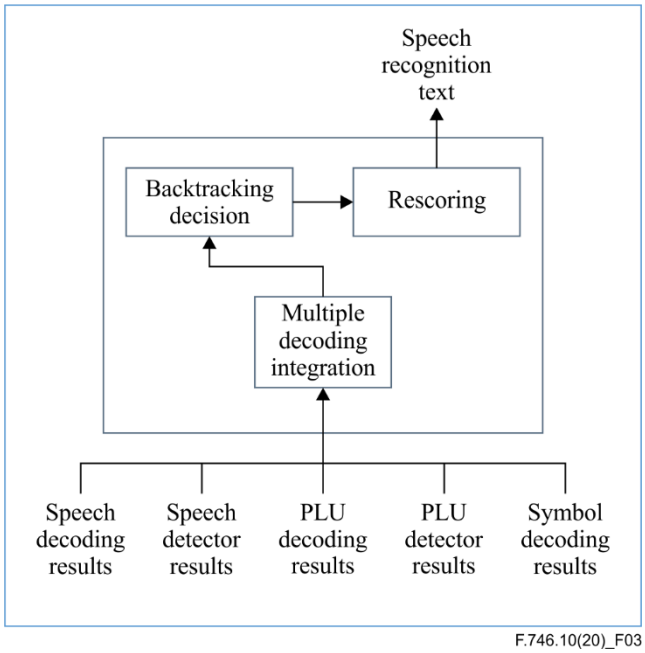

Figure 3 – Sub-functions for unstructured spontaneous speech recognition management

In the multiple decoding integrating function, the decoding results are selectively chosen for integration if the decoding results are to contribute to the speech recognition performance. In addition, the main speech recognition results from the spontaneous model-based decoding are managed separately for integration from those results of the sub-decoding functions.

The function of deciding on the partial backtracking section is to predict the error section of spontaneous speech recognition using the integrated multiple decoding results based on the time frame and then to decide reconsidering of the decoding results which does not overlap with the error section. The sub-recognition results are aligned to the time frames first and then the main recognition results are aligned to be used for the backtracking decision.

Rescoring function serves to reconsider speech recognition decoding results. Then the new recognition candidates are searched for the word boundary that is not in the error section. The target of the rescoring function constitutes the main recognition results. For this purpose, the main recognition results are stored by time frame and are subject to rescoring the words in the section designated by the backtracking function.

6.9 Language learning module

As described in [ITU-T F.746.5], the language learning module (see Figure 4) takes the user's speech as input and automatically detects grammatical errors, corrects them and then sends the corrected sentence to the user as a feedback to improve his competence of foreign languages. The language learning module consists of a grammar error correction function and a dialogue expression learning function. The grammar error correction function detects errors included in the user's spoken sentences and corrects them, while the dialogue expression learning function provides feedback of the grammar errors and suggests better expressions to the user.

– Grammar error detection sub-function is dedicated to detecting grammatical errors in the spoken sentences of the users. The grammatical errors are limited to those defined in the system and multiple errors can be detected in one sentence. Various error detection models are used for the grammar error detection.

– Grammar error correction sub-function is dedicated to providing the correction for the detected grammar errors. The error correction function produces the grammar error correction information by using the part of speech (POS) of the error words and context information around the words.

– Error selection and correction sub-function is dedicated to selecting the best error candidate among those from different error detection models and to generate error correction information on the selected candidate error.

– Error feedback providing sub-function informs the user of the grammar errors and suggests better expressions. The explicit/implicit error feedbacks are used accordingly depending on the context in which errors occur.

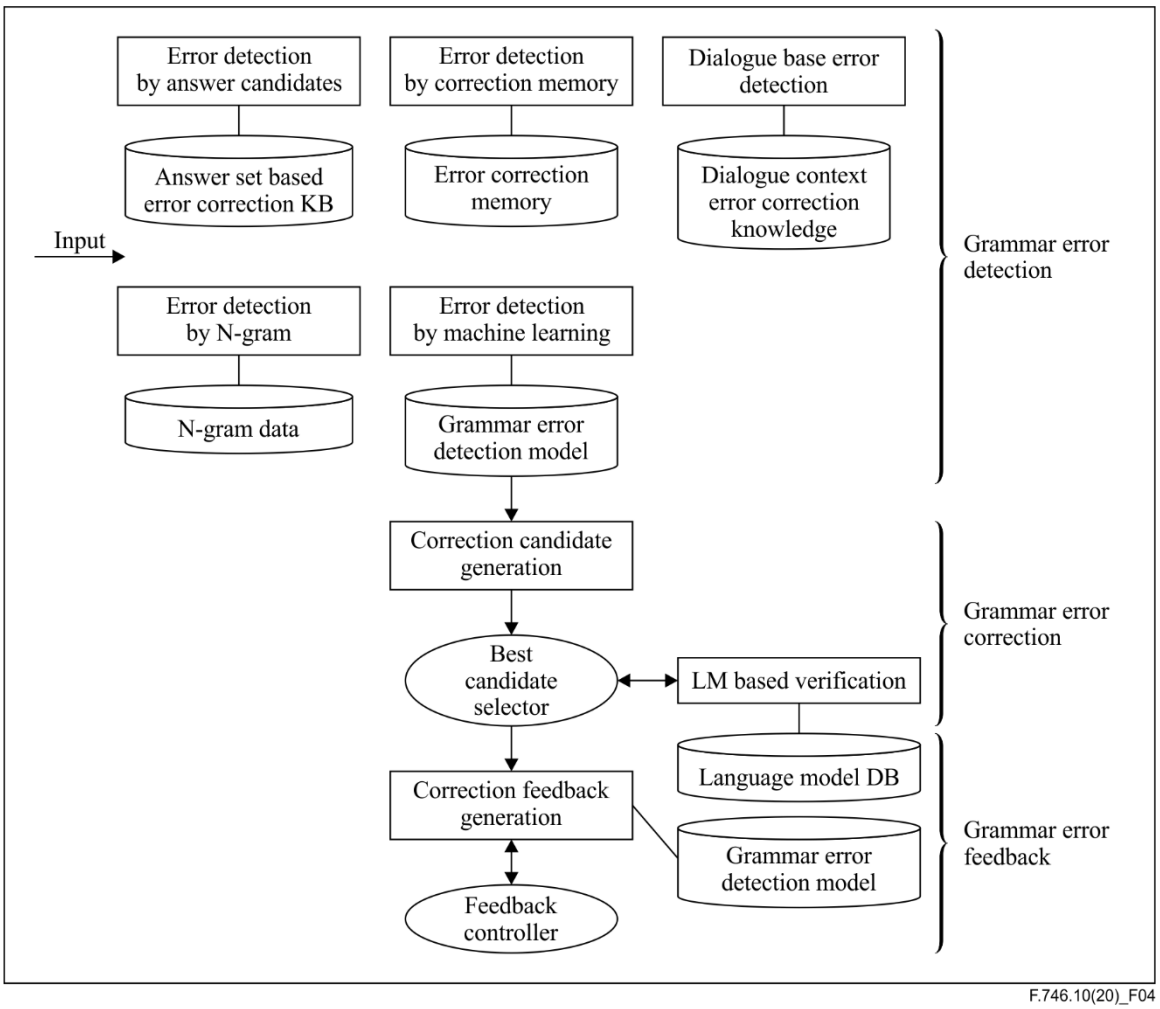

Figure 4 – Language learning module [ITU-T F.746.5]

Dialogue expression learning function as shown in Figure 5 consists of the following functional blocks: an input recognizer, a speech recognizer, a native language intent analyser, a learning target language intent analyser, a translator and a dialogue processor that comprise a two-language free dialogue system for language learning. When the input form recognized by the input recognizer is a speech, the speech recognizer converts the speech into a text by speech recognition function in the input/output management module. The native language intent analyser analyses a dialogue intent from the recognized native language when the recognized type of the input language is a native language.

Meanwhile, the learning target language intent analyser analyses a dialogue intent from the recognized learning target language when the recognized type of the input language is a learning target language. This intent analysis function is performed through the topic intention analysis function in the dialogue understanding module by interaction among involved modules. When a dialogue intent of the user is recognized as a translation request through the native language intent analyser, the translator translates a translation target into the learning target language. The dialogue processor provides a system response according to the results of an analysis of the dialogue intent performed by the learning target language intent analyser. The system response is also made based on the result of processing the dialogue of a user native utterance translated into the learning target language through the translator.

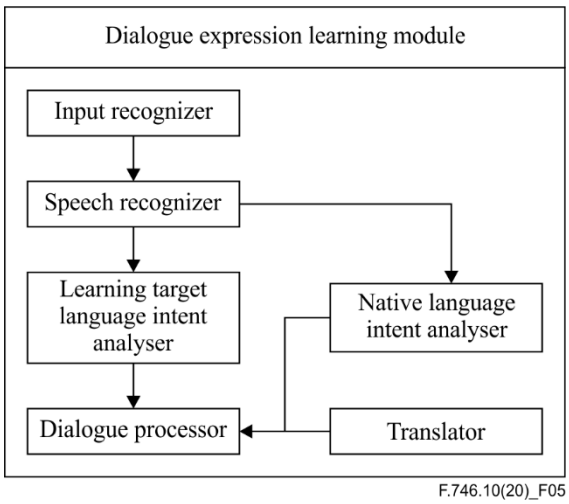

Figure 5 – Dialogue expression learning module

Bibliography

[b-ITU-T F.746.3] Recommendation ITU-T F.746.3 (2015), *Intelligent question answering service framework*.

[b-ITU-T H.703] Recommendation ITU-T H.703 (2016), *Enhanced user interface framework for IPTV terminal devices*.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	Tariff and accounting principles and international telecommunication/ICT economic and policy issues
Series E	Overall network operation, telephone service, service operation and human factors
Series F	**Non-telephone telecommunication services**
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling, and associated measurements and tests
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2020