One Word, One Command, One Translation: Combining GenAI and Traditional Approaches in Simulating In-Situ Spoken Human Interaction for Medical and Engineering Applications

Christina Alexandris
National and Kapodistrian University of Athens
calexandris@gs.uoa.gr

Abstract

The proposed deployment of knowledge graphs bridges neural networks and other forms of data-driven processing with traditional strategies such as template-based, slot-filling frameworks and controlled-language like restrictions. The present approach is based on knowledge graphs enabling the role of context-specific dimensions of singular words uttered in HCI applications. The cases concern medical and engineering applications and speaker-related / environmental factors disrupting communication. Paralinguistic features such as deictic gestures and other types of implied information not uttered / not correctly processed by Speech Recognition constitute an additional parameter in the proposed approach.

Accessible Information for Everyone for Individual and Social Well-Being

Heavily data-based state-of-the-art Natural Language Processing (NLP) approaches are proposed to include the targets of making more types of information accessible to more types of recipients and user groups and making more types of services accessible and user-friendly to more types of user groups (Alexandris 2024), contributing both to individual and social well-being. In particular, the integration of customized written and spoken text output in Human-Computer Interaction (HCI) applications for special purposes and/or targeted user groups continues to face various challenges and is even often discouraged, especially if features such as precision, clarity, user friendliness and/or politeness are an essential requirement. This is often the case when NLP application design is heavily based on big data and, in contrast, domain-specific heuristic and rule-based approaches such as Controlled Languages (a traditional strategy for text types such as technical texts) are less commonly integrated. This may impact less experienced user groups and/or users who may face difficulties in interacting with a system due to circumstantial reasons (i.e. fatigue) or due to their physical or mental condition. When the use of HCI applications is not supported and not encouraged for Special Text Types and/or User Groups, it may result in contributing to the worsening of social inequalities in respect to Artificial Intelligence (AI) and AI applications. In the case of NLP applications where user requirements and data/application customization are of particular interest and importance, data may be chosen (or data input may be controlled or processed) according to application type and user groups, if necessary. Typical cases include less-resourced languages, less experienced users, and less agile users (Alexandris 2024). Less-resourced languages may range from natural languages that have slightly less resources than languages with the largest amounts of resources (such as English) (this is the case of most official –standard languages in the European Union) to natural languages that have very little or no resources. The latter case includes widely-spoken languages spoken by large and dynamic populations in the developing markets across the World. As cited by Ranta et al. 2020: "Data austerity is particularly important for low-resource languages, which may be unlikely ever to have enough data for statistical methods" (Ranta et al. 2020). Less experienced users may require guidance by the System for an efficient interaction and/or correct choice of input. They may also belong to particular age groups or social groups (Alexandris 2024). Less agile users may be users who face difficulties in interacting with a system or application due to factors such fatigue or mental state or due to characteristics such as age or some mild form of physical impairment (Alexandris 2024). This category excludes users with severe physical or mental disabilities requiring specialized approaches, equipment and applications.

Furthermore, three general categories of recipient / user types are distinguished, reflecting recipient knowledge and/or user expectations from the System (Alexandris 2020, Wiegers and Beatty 2013): Experienced Users (existing knowledge, part of profession, culture and/or life style, Inexperienced Users (no/limited knowledge) and Distantiated

Users (existing knowledge yet not part of profession, culture and/or life style) (Alexandris 2020).

Speech Interfaces: Combining GenAI and Rule-Based Approaches

Interactive spoken input applications directed towards a broad varied user group target to achieve the so-called "Human-Compatible AI" by ensuring "understandability", correctness, appropriateness and efficiency of the terms used and often involve reformulation of terms, integration of additional, explanatory information in generated processed (i.e. translated) spoken texts and/or additional modules in the user's interaction with the Chatbot / Dialogue System. A characteristic example targeting to user-friendliness for a broad and varied user group are applications such as banking products, as presented by researchers at IBM more than a decade ago (Lewis 2009), where special emphasis is placed on the choice of the appropriate terminology and vocabulary, guidelines for using the appropriate expressions and sentence structure. These features target to achieve user friendliness and the system's appropriate adaptation to different user behaviors and error recovery, discretely encouraging user interaction and by giving the impression that the interaction is "moving forward", not "stalling" due to user / system error (Lewis 2009).

As with most types of communication, the achievement of clarity and precision in compliance to the Maxims of the Gricean Cooperative Principle (Grice 1975, Grice 1989) are important features both in Machine Translation and in spoken dialogue systems. This is especially important in applications involving technical texts (i.e. in the airline industry) where the traditional employment of Controlled Languages ensures the achievement of precision and clarity in system-output, with the appropriate choice of linguistic features (i.e. grammar categories, sentence types, vocabulary-terminology) (Lehrndorfer 1996, Wojcik and Holmback 1996, Kuhn 2014, Fuchs 2021, Marzouk 2021). However, the nature of spoken data – especially in situations of urgency and/or mental stress, calls for special adaptations to the types of Controlled Languages that rely heavily on syntactic components and syntax logic (Fuchs 2021).

Regarding less resourced languages, it is considered that a rule-based Grammatical Framework (GF) (Ranta et al. 2020), Explainable Machine Translation (XMT) and Explainable Natural Language Processing (NLP) (Ranta et al. 2020) can solve a variety of existing problems in NLP tasks. However, the time and resources needed to construct complete grammars discourages their direct deployment in spoken interfaces of applications that need to be immediately developed and used, due to special needs and circumstances.

Here, we present two types of HCI applications with spoken interaction in personal decision making processes where precision, clarity and user friendliness are an essential requirement and may target either Experienced or Inexperienced/ Distantiated Users, along with the factor of less resourced languages and the factor of limited agility due to fatigue, weakness, mental state etc. The first case concerns Experienced Users in the domain of commercial and military aircraft maintenance. The second case concerns Inexperienced/ Distantiated Users in the domain of medical information / chatbots. In the first case, concerning expert knowledge, information is more easily recognizable and retrievable in instances of errors, missing/implied information or disruptions in spoken communication. In the second case, no expert knowledge is presumed, in contrary, spoken user-input is not predicted to include (correct/precise) medical terms (if any, at all) and, additionally, system-generated medical information and terminology may even be misinterpreted. These two cases serve as a comparison between user types with varying degrees of expert/world knowledge. Other potential domains, use-cases and applications may give rise to additional parameters and issues.

Both cases concern spoken interaction that may be preferable over written interaction due to practical issues and/or special circumstances. On the other hand, the directness of spoken interaction is also connected to situations like urgency that can also be accompanied by complications in communication related to factors such as a noisy environment, (Factor 1: Noise-Signal) or a person's inability to communicate due to physical pain, mental state (Factor 2: Physical/Mental Stress) or language barriers (Factor 3: Language-Comprehension). In addition, in both cases, parameters regarding speech data are taken into account, namely implied information not spoken / not correctly processed by Speech Recognition (ASR) (Siegert and Krueger 2021, Cohn and Zellou 2021, Du et al. 2023) – especially in multilingual applications (Zellou and Lahrouchi 2024, Nakamura 2009) - and paralinguistic features concerning information not uttered. Since the present approach has not been implemented and tested in real-life situations in both types of applications and user groups, it is considered that any method of performance evaluation (by default) includes sets of parameters typically used in Speech Recognition (ASR) and Spoken Dialogue Systems. These include correct recognition of spoken input (i.e. noisy environment factors), speaker-related factors, success-completion of interaction and user satisfaction.

The domain-specific nature of both applications allows a domain specific controlled-language like approach within a traditional directed-dialog, system-initiative framework (Lewis 2009, Jurafsky and Martin 2025), placing special emphasis on the role of singular words within the spoken utterance. The pivotal role of singular words connected to arguments / frame-slots is characteristic in a variety of multilingual NLP approaches (Jurafsky and Martin 2025), including– past research, such as the Universal Words (UWs)

of the Universal Networking Language (UNL) of the United Nations Research Center (Uchida, Zhu, and Della Senta 2005). In light of the multiple challenges faced by approaches targeting to the correct processing and/or translation both in regard to spoken data and to Controlled Languages, the role of the singular word in spoken utterances allows the by-passing of typical issues / problems encountered (Alexandris 2023, Jurafsky and Martin 2025) and accounts for unspoken linguistic / paralinguistic information.

Knowledge Graph – Neural Network Approach

As presented in previous research (Alexandris 2023, Alexandris, Du, and Floros 2022), the deployment of knowledge graphs capturing all dimensions of singular spoken words – including unspoken linguistic information and information of paralinguistic features - is proposed. A finite set of knowledge graphs constructed for the specialized domain of the application concerned is proposed to serve as initial training data. In particular, these features contained in the knowledge graphs can be integrated into state-of-the art data-driven approaches involving Machine Learning and neural networks (Wang, Qiu, and Wang 2021, Mountantonakis and Tzitzikas 2019, Tran and Takashu 2019, Mittal, Joshi, and Finin 2017). They can function as training data and seed data (i.e. in Graph Neural Networks, Ye et al. 2022) in NLP applications, where user requirements and customization are of particular interest and importance (Alexandris, Du, and Floros 2022). This approach is compatible with recent applications where knowledge graphs and neural networks are combined to resolve a wide variety of practical problems across disciplines and professional domains (i.e. Physics, Finance) (Antaris, Rafailidis, and Girdzijauskas 2021, Yerramsetti and Yerramsetti 2023), even across multilingual data and applications (Tam et al. 2022). The target of by-passing the need of large amounts of labelled data with the use of the appropriately modelled knowledge graphs as training data, as presented in current approaches, demonstrates the versatility and potential of the combination of knowledge graphs with neural networks (Tam et al. 2022). Furthermore, the knowledge graphs deployed, capturing all dimensions of singular spoken words (Alexandris 2023) are characterized by a small and shallow set of nodes, avoiding deep, convoluted structures, resulting to their easier training by the chosen models (Antaris, Rafailidis, and Girdzijauskas 2021, Yerramsetti and Yerramsetti 2023, Tam et al. 2022).

Contribution to Human-Compatible AI and Socially-Responsible AI for Well Being

The present “one word, one command, on translation” approach focuses on taking full advantage of the recognition and processing of the minimal possible input of utterances, in particular, singular words. One of its targets is requiring the minimal effort of users-speakers facing challenges or disadvantages in communication due to various factors (1). These include underlying physical / mental health issues, old age, non-native / non-standard language speech pronunciation, current mental state due to stress, fatigue (i.e. from driving), physical pain or environmental factors such as background noise or problematic speech / network signal. In addition, this approach allows the information recognized from singular words to be combined with paralinguistic features (i.e. prosodic emphasis, gestures) that are informative in an in-situ Human-Human communication context, therefore, contributing to Human-Compatible AI application goals (2). Another target of the present approach is requiring the minimal resources for adaptation in other languages, especially less resourced languages, with the construction of small datasets (i.e. translations) in contrast to big data. However, these datasets can also serve as seed data for possible future development and integration in neural network processing (3). The proposed approach is not limited to particular types of software tools / processing techniques: it can be implemented with the resources available to the developers, independently from their location, professional network or budget. An additional target of the “one word, one command, on translation” approach is to account for communication challenges in speech interfaces of applications where safety and security and/or physical and mental well-being are of crucial importance (4). Achieving and guaranteeing safety and security constitute the foundation of communication in the domain of Aircraft Maintenance described here. Safety and physical/mental well-being are an essential element in Medical Chatbots for the broad public, as presented here. These features (1-4) are compatible to "Socially Responsible / Human-Compatible AI for Well-being”.

Simulating Human Behavior in Aircraft Maintenance Training

Previous research (Chatzipanayiotidis and Alexandris 2023) focused on the multiple challenges encountered in spoken technical texts – and the related decision making processes, in particular, the achievement of correctness, clarity and precision in the field of commercial and military aircraft maintenance and aircraft maintenance training- education, especially for their processing and possible translation. These applications, concerning technical texts and engineering, also focus in the avoidance of Cognitive Bias (i.e. Confidence Bias, Confirmation Bias - Azzopardi 2021, Hilbert 2012) by engineers and pilots. In this case, correctness, clarity, and precision are of critical importance for securing safety, precision, efficiency and direct deployment and understandability for non-native speakers and/or translation in different languages. In the aircraft industry, the strict, domain-specific framework of Controlled Languages is one of the typical and traditional approaches for achieving these

goals (Lehrndorfer 1996, Wojcik and Holmback 1996, Kuhn 2014). Previous research (Chatzipanayiotidis and Alexandris 2023) is based on typical practices for processing spoken input in monolingual / multilingual applications with template-based, slot-filling strategies (Jurafsky and Martin 2025). Template-based slot-filling frameworks are based on the recognition and extraction of word-level speaker intent keywords along with their relevant entities/slots in utterances. Various types of strategies have been implemented in template-based slot-filling frameworks over the last decade, with the most recent approaches combining rule-based methods and neural networks. For example, Okur et al. 2019 have implemented term-frequency and rule-based mapping mechanisms from word-level intent keywords extraction to utterance-level intent recognition and have improved system performance by investigating various neural network (RNN) architectures and developed a hierarchical (2-level) model to recognize speaker (passenger) intents along with relevant entities/slots in utterances.

In previous research (Chatzipanayiotidis and Alexandris 2023), the task-oriented spoken interaction in aircraft maintenance and aircraft maintenance training is based on Speech act recognition, in combination with the recognition of the appropriate word groups. The processing strategy also targets to maintain the minimum possible size of the sets in keyword recognition, since varying degrees of quality of ASR systems across widely implemented and less widely implemented natural languages – as well as possible noise from the environment, constitute additional, unpredictable factors. The strategy employed constitutes a template-based, slot-filling framework which is based on keyword recognition with keywords in the form of tuples (x,y) (Chatzipanayiotidis and Alexandris 2023), namely the mandatory recognition of two keywords in complimentary relation (x,y). According to evaluations performed, this relation is observed to ensure the correct identification of speech act type and the correct identification of content type in the spoken utterances. The recognition of two keywords in complimentary relation also accounts for cases where speech act type and content type coincide or overlap, a possibility in some types of commands / utterances for safety regulations and alerts (Chatzipanayiotidis and Alexandris 2023).

The keywords extracted from the content of each spoken utterance are integrated to a finite set of slots, linked to the respective Speech Act. For example, the "generic-intent" slot of the template-based, slot-filling framework contains information types (classes and subclasses) corresponding to the intention of the Speaker and the respective Illocutionary Speech Act. For example: (flights: (flight schedule), (flight tracking)), (maintenance-task: (aircraft inspection), (fueling), (cleaning), (aircraft relocation), (sampling), (testing), (repair), (replacement), (material inspection)), (safety measures: (safety measures-request)), (resources: (aircraft availability), (aircraft location), (personnel for task), (spare part)) , (manual-references: (instructions)), (maintenance-report: (report-request)).

Typical examples - related utterances are the following:

- fueling (aircraft, aircraft_part, number) (example: "Fill fuel tank in aircraft 2255"),
- repair (aircraft, aircraft_part, material, number, time) (example: "repair damage in aircraft 2255"/ "repair damage in fuel tank"),
- aircraft_inspection (aircraft, place, number, time) (example: "inspect aircraft SN 520"/ "inspect fuel tank in aircraft SN 520") (Chatzipanayiotidis and Alexandris 2023).


"Fill fuel tank in aircraft 2255" "Repair damage in aircraft 2255" "Repair damage in fuel tank" / " Inspect aircraft SN 520" / " Inspect fuel tank in aircraft SN 520»

Figure 1: Example of spoken commands for frame slots (aircraft, aircraft_part, number), repair (aircraft, aircraft_part, material ,number, time) and aircraft_inspection (aircraft, place, number, time).

This template-based, slot-filling framework places special emphasis on the role of individual words in a task-specific domain. In other words, the non-language-specific but strictly context-specific dimension of a word can also be domain-specific. For example, a particular word may imply a specific role or action. It may be noted that this allows possible implementations within a "frame-slot" framework in domain-specific HCI / HRI for processing spoken utterances. In this case, the mere utterance of a single word may imply a domain-specific type of information consisting a complete phrase or sentence – or one or more possible domain-specific alternative types of implied information. This is of importance in speech applications where the factors Noise-Signal (Factor 1), Physical/Mental Stress (Factor 2) and/or Language-Comprehension (Factor 3) are involved.

Since template-based slot-filling frameworks are based on the recognition and extraction of word-level speaker intent keywords along with their relevant entities/slots in utterances, the function of these entities/slots in utterances can be connected or substituted by nodes of a knowledge graph, as described below. This general concept is not dependent on the deployment of a particular type of template-based slot-filling framework: Choice of implementation may vary depending on available tools and system requirements.

As proposed in previous research (Alexandris 2023, Alexandris, Du, and Floros 2022), context-specific additional dimensions of individual spoken words may be described as a context-specific information (atmo) "sphere" surrounding the spoken word. The concrete meaning – actual semantic content of the word (retrievable and processable in Natural Language Processing-NLP) is surrounded by two context-

specific layers, with its context-specific and language-specific dimensions in the inner layer of the sphere (A) and its context-specific and non-language-specific / domain-specific dimensions in the outer layer of the "sphere" (B) (Alexandris 2023). These dimensions are integrated in knowledge graphs, with its subsequent use in vectors and other forms of training data as dataset for training a neural network for NLP tasks. In the knowledge graphs, unspoken information is represented by the "Context" relation (Alexandris 2023) and by the distinctive nodes it connects.

The "sphere" model serves as a means of describing the types of context-/-domain specific linguistic and paralinguistic information and positioning their distinct relations in regard to the concrete meaning – actual semantic content (denotation) of a spoken word (constituting the "nucleus" of the sphere). In other words, the distinct types of ("Context") relations connecting the nodes in the proposed knowledge graphs can be described by the "sphere" model.

Although the present approach does not specify one particular type of implementation for the interaction between knowledge graphs and neural networks, it is compatible to results involving the construction of knowledge graph models for neural network training stating that "models with additional information, such as node attributes, node types, relationship types, prior knowledge and so on, have better performance."(Wang, Qiu, and Wang 2021). The present approach requires extensive testing and evaluation across large and varied datasets in order to produce accurate and safe results. For example, in their research linking knowledge graphs and neural networks, Wang, Qiu and Wang, 2021 have tested the performance of several models (i.e. from "older" model ConvE to RotatE and HypER, among others).

Here, the "Context" relation (Alexandris 2023) connects nodes with information types implied by unspoken linguistic or paralinguistic features, co-occurring with the spoken word in the utterance. In other words, the "Context" relation connected to an individual (spoken) word in a knowledge graph can shed light into the possible dimensions of the spoken word (Alexandris 2023), namely unspoken, implied information of linguistic / paralinguistic nature – especially if recent approaches integrating knowledge graphs into spoken dialogue systems are taken into account (Deng et al. 2023).

In the case of (unspoken) domain-specific information (namely the outer layer ("B") of the "sphere"), the "Context: W" relation (Alexandris 2023) connects the spoken word with the information it implies. In the case of the frame-slot framework of the task-oriented spoken interaction in aircraft maintenance and aircraft maintenance training, the "Context: W" relation concerns the relations of the information in the frame slots. For example, the word "fuel tank" (aircraft_part /place) is connected to the commands "fueling" (i.e. "fill"), "repair" and "aircraft_inspection" (i.e. "inspect") (Fig.2) (Alexandris 2023). In other words, the utterance of the single word "fuel tank" (Fig. 2) is connected to a restricted set of information types, within the domain-specific template-based, slot-filling framework (Fig. 2).

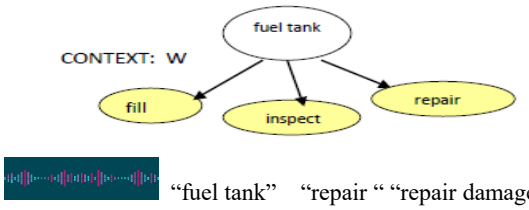

Figure 2: Knowledge graph fragment for spoken utterance "fuel tank" and commands "fill[fuel tank]", "inspect[fuel tank]","repair[fuel tank]" (Alexandris 2023) and example of one / two word fragments of spoken commands for frame slots (Chatzipanayiotidis and Alexandris 2023): (aircraft, aircraft_part, number), repair (aircraft, aircraft_part, material ,number, time) and aircraft_inspection (aircraft, place, number, time).

The domain-specific frame-slot approach transformed into knowledge graphs allows a single word (i.e. "fuel tank") to be connected to a small set of possible words – arguments, allowing the System to proceed to verification or explanatory generated utterances, for example "Repair fuel tank – Correct?", " Repair fuel tank in which aircraft?"

As in the case of the "Context: W" relation, the "Context: P" relation concerns the non-language-specific / domain-specific dimensions in the outer layer of the "sphere" (B) and involves the paralinguistic features, co-occurring with the spoken word in the utterance. Examples of non-language-specific / domain-specific paralinguistic features are prosodic emphasis and deictic gestures.

A characteristic example of non-language-specific features comprising additional dimensions of information content of words is the case of specific words receiving prosodic emphasis within the discourse and/or domain of the spoken interaction. Prosodic emphasis may stress (i.e. in urgency) and/or clarify the semantic content of the spoken utterance in a broad range of interaction types (Alexandris 2020). For example, prosodic emphasis on the word "fuel tank" may also imply damage or even fire.

Integration of Deictic Gestures

Deictic gestures constitute another characteristic example of non-language-specific features comprising additional dimensions of information content of words. As in the case of prosodic emphasis, deictic gestures may stress and/or clarify the semantic content of the spoken utterance in a broad range of interaction types (i.e. "fuel tank", Fig. 3). For example, if the single word "Gas" / "Gas!" is uttered (with or without prosodic emphasis) (Alexandris 2023), a deictic gesture pointing towards the control panel may either imply

the “check gas” or, especially if there is prosodic emphasis, “gas is low” (urgency) , whereas a deictic gesture pointing towards the tank is more likely to imply “fill gas” (Fig. 4).

Additionally, the domain-specific frame-slot approach transformed into knowledge graphs, enabling the one-word one command possibility facilitates both speech recognition (ASR) and, even more so, translation – especially in less-resourced languages, in machine or human translation.

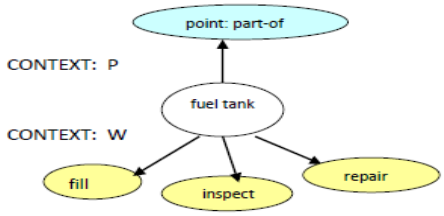

Figure 3: “Fuel tank”: Fragment of knowledge graph for a singular spoken word “fuel tank” and context-specific and non-language-specific “CONTEXT**:** W” relation for domain-specific information in HCI applications. Word “fuel tank” is marked with implied possible information for the commands “fill[fuel tank]”, “inspect[fuel tank]”,“repair[fuel tank]”and related frame-slots, along with the paralinguistic deictic gesture “point:part-of” (pointing to fuel tank) corresponding to the “CONTEXT**:** P” relation.

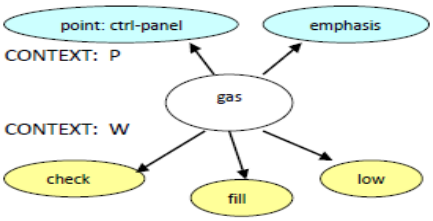

Figure 4: “Gas!”: Fragment of knowledge graph for a singular spoken word “gas” [with prosodic emphasis (Alexandris 2024)] and context-specific and non-language-specific “CONTEXT**:** W” relation for domain-specific information in HCI applications. Word “gas” is marked with implied possible information “(the [gas] is) low”, “fill [gas]” and “check [gas]”, with the paralinguistic deictic gesture “point:ctrl-panel” (pointing to control panel) corresponding to “CONTEXT**:** P” relation.

Medical Chatbots

Other applications, beyond the specialized nature of the task-oriented spoken interaction in aircraft maintenance and aircraft maintenance training, can benefit from the knowledge graph based approach where an individual (spoken) word can be connected to unspoken, implied information of a linguistic or paralinguistic nature. Applications providing medical information towards a broad public – and facilitate personal decision-making- can benefit from the one-word one command possibility, facilitating both speech recognition (ASR) and translation – especially in less-resourced languages, in machine or human translation.

A characteristic example of an implemented interactive system designed to provide medical information to a general public and broader user group is the "ATHENA" system – application (Medical Dialog System / Chatbot) - intended as a preparatory step for medical appointments and (medical) care establishments. (Malonas 2024), where the first stage of interaction is verbal. serving as a user friendly welcoming stage before the activation of the System’s interactive symptom checker (presenting the user with a set of probabilities in relation to the selected symptoms and the corresponding condition) and chatbot. A user friendly initial verbal interaction targets to facilitate as many categories of users as possible, including the most inexperienced, elderly users and those who experience temporary interaction difficulties due to manual work (i.e. driving) or due to time pressure, stress or fatigue (Malonas 2024). The design of the “ATHENA" System is intended to facilitate adaptations to languages other than English (i.e. less resourced languages), at least in the spoken dialog and interactive symptom checker components (i.e. Greek) (Malonas 2024). For the spoken interaction, this is achieved with the restrictive framework of a traditional directed-dialog approach with a single word answer (i.e. “Yes”/”No”) and keyword recognition (Malonas 2024), and also with the possibility of further adaptation to a Controlled-Language-like simplified version of spoken interaction for upgrading usability (Fig. 5), avoiding Lexical Bias (Trofimova 2014) with choice of appropriate expressions.


SYS: Hello. This is ATHENA. I am a virtual Medical assistant. I will ask you some questions. [Please, tell me,] do you feel well? (Yes/No) >>> USR: My belly hurts
SYS:[KEYWORD: Abdominal pain]. You have belly pain– Correct? (Yes/No) >>> USR: Yes:
SYS: [Confirm KEYWORD: Abdominal pain]. Say “Yes” for more information {ATHENA ChatBot}. Say “Stop” if you want to stop this conversation.

Figure 5: Spoken interaction in the “ATHENA” System and Medical ChatBot”: Adapted simplified version.

The above-described features take into account the previously mentioned complications in communication related to factors such as a noisy environment, or a person’s inability to communicate due to physical pain, mental state or language barriers (Hatim 1997). Here, deictic gestures may stress and/or clarify the semantic content of the spoken utterance in a broad range of interaction types. For example, if the single word “pain” (Fig. 6 and 7) is uttered (with or

without prosodic emphasis), a deictic gesture pointing towards the abdomen has a high probability to imply "part-of-body", whereas no deictic gesture may also imply "general" (an unspecified physical pain or a general pain all over the body) or even "grief" (psychological pain). As another example (Fig. 8), for the recognition / utterance of the single word "pain", a deictic gesture pointing towards the throat coupled with the registered paralinguistic feature of a cough ("cough"), has a high probability to imply "part-of-body" and gives information related to possible ailments.

As in the case of aircraft maintenance (training), the possibility of processing one single word - with or without the recognition of paralinguistic features and related information- takes into account the factors Noise-Signal (Factor 1), Physical/Mental Stress (Factor 2) and/or Language-Comprehension (Factor 3) and facilitates speech recognition and machine / human translation – especially in less-resourced languages. Unlike the traditional frame-slot framework for spoken interaction, the context-specific nodes of knowledge graphs also account for information in prosodic emphasis and deictic gestures detected within the context-specific dimensions ("atmosphere") of spoken words presented in previous research. This enables realistic simulations of in situ human interaction and behavior, accounting for unspoken information - which can also be related to a single spoken word.

SYS: Hello. This is ATHENA. I am a virtual Medical assistant. I will ask you some questions. [Please, tell me,] do you feel well? (Yes/No) >>> USR: Pain
[VISUAL INPUT: point: abdomen]
SYS: [KEYWORD: Abdominal pain]. You have belly pain– Correct ? (Yes/No) >>> USR: Yes:
SYS: [Confirm KEYWORD: Abdominal pain]. [Please] Say "Yes" for more information {ATHENA ChatBot}. Say "Stop" if you want to stop this conversation.

Figure 6: Spoken interaction in the "ATHENA" System and Medical ChatBot"- Adapted version with visual input.

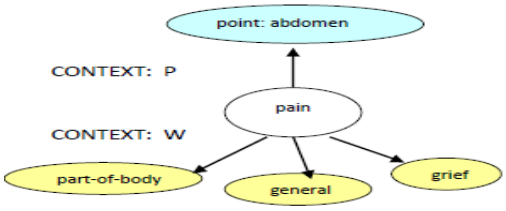

Figure 7: Fragment of knowledge graph for a singular spoken word "pain" (with deictic gesture) and context-specific and non-language-specific "CONTEXT: W" relation for domain-specific information. Word "pain" is marked with implied possible information "part-of-body [pain]", "general [pain]" and "grief [pain]". The latter two information possibilities are excluded, with the paralinguistic deictic gesture "point:abdomen" (pointing to abdomen) corresponding to "CONTEXT: P" relation.

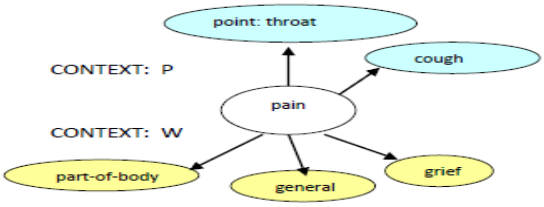

Figure 8: Fragment of knowledge graph for a singular spoken word "pain" (with deictic gesture) and context-specific and non-language-specific "CONTEXT: W" relation for domain-specific information. Word "pain" is marked with implied possible information "part-of-body [pain]", "general [pain]" and "grief [pain]". The latter two information possibilities are excluded with the paralinguistic deictic gesture "point:throat" (pointing to throat) and paralinguistic prosodic feature "cough".

Conclusion and Further Research

Since knowledge graphs can be adapted into neural networks, according to current research work, the presented information they contain contributes to the enrichment of GenAI approaches. The proposed deployment of knowledge graphs bridges traditional strategies with GenAI - neural networks and other forms of data-driven processing and is intended to be tested in a larger and more varied group of data for further research and evaluation.

References

Alexandris C. 2024. GenAI and Socially Responsible AI in Natural Language Processing Applications: A Linguistic Perspective. In Proceedings of the AAAI Symposium Series 3(1): 330-337. https://doi.org/10.1609/aaaiss.v3i1.31230.

Alexandris, C. 2023. Processing Information Unspoken: New Insights from Crowd-Sourced Data for Sentiment Analysis and Spoken Interaction Applications. In Proceedings of AAAI-SRAI, Socially Responsible AI for Well-being (SS-23-09) collocated with the 2023 AAAI Spring Symposium. San Francisco CA. https://ceur-ws.org/Vol-3527/Paper_456.pdf.

Alexandris, C., Du, J., Floros, V. 2022. Visualizing and Processing Information Not Uttered in Spoken Political and Journalistic Data: From Graphical Representations to Knowledge Graphs in an Interactive Application. In *Lecture Notes in Computer Science* 13303, edited by M. Kurosu, 211–226. Cham: Springer. doi:10.1007/978-3-031-05409-9_16.

Alexandris, C. 2020. *Issues in Multilingual Information Processing of Spoken Political and Journalistic Texts in the Media and Broadcast News*. Newcastle upon Tyne: Cambridge Scholars.

Antaris, S.; Rafailidis, D.; Girdzijauskas, S. 2021. Knowledge distillation on neural networks for evolving graphs. *Social Network Analysis and Mining* 11(1). doi.org/10.1007/s13278-021-00816-1.

Azzopardi, L. 2021. Cognitive Biases in Search: A review and reflection of cognitive biases in Information Retrieval. In Proceedings of the 2021 ACM SIGIR Conference on Human Information Interaction and Retrieval. New York: Association for Computing Machinery. https://dl.acm.org/doi/10.1145/3406522.3446023.

Chatzipanayiotidis, S., Alexandris C. 2023. Processing Speech Acts: Spoken Communication for Aircraft Maintenance. In Proceedings of the 14th International Conference in Experimental Linguistics ExLing 2023. Athens, Greece. https://exlingsociety.com/wp-content/uploads/2024/07/14_0007_000601.pdf.

Cohn, M., Zellou, G. 2021. Prosodic differences in human-and Alexa-directed speech, but similar local intelligibility adjustments. *Frontiers in Communication* 6 (675704).

Deng, C.; Tong, B.; Fu, L.; Ding, J.; Cao, D.; Wang, X.; and Zhou, C. 2023. PK-Chat: Pointer Network Guided Knowledge Driven Generative Dialogue Model. arXiv preprint. arXiv:2304.00592v1 [cs.CL]. Ithaca, NY: Cornell University Library.

Du, Y. Q.; Zhang, J.; Fang, X.; Wu, M. H.; Yang, Z. W. 2023. A semi-supervised complementary joint training approach for low-resource speech recognition. *IEEE/ACM Transactions on Audio Speech and Language Processing* 31: 3908-3921. doi: 10.1109/TASLP.2023.3313434.

Fuchs, N. E. 2021. Reasoning in Attempto Controlled English: Mathematical and Functional Extensions. In Proceedings of the Seventh International Workshop on Controlled Natural Language (CNL 2020/21). https://aclanthology.org/2021.cnl-1.8.pdf.

Grice, H.P. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Grice, H.P. 1975. Logic and conversation. In *Syntax and Semantics* 3, edited by P. Cole, and J. Morgan. New York: Academic Press.

Hatim, B. 1997. *Communication Across Cultures: Translation Theory and Contrastive Text Linguistics*. Exeter: University of Exeter Press.

Hilbert, M. 2012. Toward a Synthesis of Cognitive Biases: How Noisy Information Processing Can Bias Human Decision Making. *Psychological Bulletin* 138(2): 211-237.

Jurafsky, D., Martin, J. H. 2025. Speech and Language Processing, an Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, 3rd edition Draft: https://web.stanford.edu/~jurafsky/slp3/ed3book_Jan25.pdf. Accessed: 2025-01-21.

Kuhn, T. 2014. A Survey and Classification of Controlled Natural Languages. *Computational Linguistics* 40(1): 121-170.

Lehrndorfer, A. 1996. *Kontrolliertes Deutsch: Linguistische und Sprachpsychologische Leitlinien für eine (maschniell) kontrollierte Sprache in der technischen Dokumentation*.Tübingen: Narr.

Lewis, J.R. 2009. Introduction to Practical Speech User Interface Design for Interactive Voice Response Applications, IBM Software Group, USA. Tutorial T09 presented at HCII 2009. San Diego, CA, July 19-24.

Malonas, C. 2024. A Medical ChatBot. Master's Thesis, Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens, Greece.

Marzouk. S. 2021. An in-depth analysis of the individual impact of controlled language rules on machine translation output: a mixed-methods approach. *Machine Translation* 35:167–203. doi.org/10.1007/s10590-021-09266-0.

Mittal, S.; Joshi, A.; and Finin, T. 2017. Thinking, Fast and Slow: Combining Vector Spaces and Knowledge Graphs. arXiv preprint. arXiv:1708.03310v2 [cs.AI]. Ithaca, NY: Cornell University Library.

Mountantonakis, N., Tzitzikas, Y. 2019. Knowledge Graph Embeddings over Hundreds of Linked Datasets. In *Communications in Computer and Information Science* 1057, edited by E. Garoufallou, F. Fallucchi and E. William De Luca, 150-162. Cham: Springer. doi: 10.1007/978-3-030-36599-8_13.

Nakamura, S. 2009. *Overcoming the language barrier with speech translation technology*. Science and Technology Foresight Center, National Institute of Science and Technology Policy, Japan.

Okur, E.; Kumar, S.H.; Sahay, S.; Esme, A.A.; and Nachman, L. 2019. Natural Language Interactions in Autonomous Vehicles: Intent Detection and Slot Filling from Passenger Utterances. arXiv preprint. arXiv:1904.10500v1 [cs.CL]. Ithaca, NY: Cornell University Library.

Ranta, A.; Angelov, K.; Gruzitis, N.; Kolachina, P. 2020. Abstract Syntax as Interlingua: Scaling Up the Grammatical Framework from Controlled Languages to Robust Pipelines. *Computational Linguistics* 46(2):425–486. doi.org/10.1162/coli_a_00378.

Siegert, I., Krüger, J. 2021. Speech Melody and Speech Content Didn't Fit Together - Differences in Speech Behavior for Device Directed and Human Directed Interactions. In *Advances in Data Science: Methodologies and Applications, Intelligent Systems Reference Library* 189, edited by G. Phillips-Wren, A. Esposito and L.C. Jain, 65–95. Cham: Springer. doi.org/10.1007/978-3-030-51870-7_4.

Tam, N.T.; Trung, H.T.; Yin, H.; Vinh, T.V.; Sakong, D.; Zheng B.; Hung, N.Q.V. 2022. Entity alignment for knowledge graphs with multi-order convolutional networks. *IEEE Transactions on Knowledge and Data Engineering* 34(9): 4201-4214. doi: 10.1109/TKDE.2020.3038654.

Tran, H. N., Takashu, A. 2019. Analyzing Knowledge Graph Embedding Methods from a Multi-Embedding Interaction Perspective. arXiv preprint. arXiv:1903.11406v4 [cs.LG]. Ithaca, NY: Cornell University Library.

Trofimova, I. 2014. Observer Bias: An Interaction of Temperament Traits with Biases in the Semantic Perception of Lexical Material. *PLoSONE* 9(1): e85677.

Uchida, H.; Zhu, M.; Della Senta, T. 2025. *Universal Networking Language.* The UNDL Foundation, Tokyo, Japan.

Wang, M.; Qiu, L.; Wang, X. 2021. A Survey on Knowledge Graph Embeddings for Link Prediction. *Symmetry* 13(3): 485. doi:10.3390/sym13030485.

Wiegers, K., Beatty, J. 2013. *Software Requirements*. London: Pearson - Microsoft Press.

Wojcik R. H., Holmback, H. 1996. Getting a Controlled Language Off the Ground at Boeing. In Proceedings of CLAW– 1996: 22–31. Leuven, Belgium.

Ye, Z.; Kumar, Y. J.; Sing, G. O.; Song, F.; Wang, J. 2022. A Comprehensive Survey of Graph Neural Networks for Knowledge Graphs. *IEEE Access* 10: 75729-75741. doi:10.1109/ACCESS.2022.3191784.

Yerramsetti, M., Yerramsetti, A.D. 2023. Graph neural networks for link prediction in dynamic knowledge graphs. *International Journal of Scientific Research in Engineering and Management IJSREM* 7(7). doi: 10.55041/ijsrem24523.

Zellou, G., Lahrouchi, M. 2024. Linguistic disparities in cross-language automatic speech recognition transfer from Arabic to Tashlhiyt. *Scientific Reports* 14: 313. doi.org/10.1038/s41598-023-50516-3.