International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 15, No.1 (Jan-24)

http://dx.doi.org/10.12785/ijcds/150102

Ontological Concepts and Logical Listeners for Dysarthric Speech Understanding

Benard Alaka[1] **and Bernard Shibwabo Kasamani**[1]

[1]*School of Computing and Engineering Sciences,Strathmore University, Nairobi, Kenya*

Received 30 Mar. 2023, Revised 23 Nov. 2023, Accepted 23 Dec. 2023, Published 1 Jan. 2024

Abstract: Depending on the social features of the speaker and the social setting in which they are speaking, the relationship between meaning and context might alter. In order to interpret the meaning from dysarthric speech, this paper proposes a theoretical framework for employing speech-event representations, also known as situational projections. The multi-layered approach has been broken down into four main components: a few-shot learner that builds up speaker familiarity; a situational projection component that marshals natural sentences and the built-up familiarity markers into a vector triple; a contextualizer that builds up ontological concepts of the input triple; and finally, a transducer that assumes the function of a logical listener.

Keywords: Dysarthria, Speech meaning, Speech context, Familiarity, Contextualizer, Situational Projection, Listener, Dialogue Acts

1. INTRODUCTION

Speech comprehension during discourse demands common information between the speaker and the listener. Even while context can occur both linguistically and non-linguistically, it is frequently challenging to comprehend some speech in its entirety independent of its specific linguistic context [1]. The variations get considerably more complex when a single speaker can use a single language in a range of contexts [2], [3].

The models that are now available for language context treat it as a purely linguistic problem, placing an excessive focus on understandability, especially in context-aware Neural Machine Translators (NMT) that use inter-sentence referencing as a method of context inference [4]. The inclination to view context as a merely linguistic issue [5] is in direct opposition to the likelihood that a phrase's meaning at one point in time would have a completely different meaning at a later time or in a different location.

As such, listeners have mostly been left to infer the speaker's intended meaning by using prior utterances, context about the speaker, objects, and concepts [5], [6]. Whilst visual short-term memory bottlenecks and issues with complicated reference in conceptual models are caused by the human ability for processing unknown information, these issues are not unique to humans [7], [8]. The listener's familiarity with the speaker is primarily determined by the circumstance, event, and topic expertise. Ontology extracts have lately been utilized to explicitly describe events in structural models [9]–[13].

This article is structured as follows. The related works reviews concepts of speech comprehension and its alignment with the concept of familiarity, the approaches used for ontology formulation and logical listeners are discussed. The results of the model proposed are reported followed by the discussion and conclusion of the study.

2. RELATED WORKS

This study examines the speech comprehension issues associated with the speech disorder dysarthria; a weakness of the cheek, tongue, or throat muscles as a result of a neurological system problem [14]–[16]. The proposed theoretical model is based on Malinowki's "Context of Situation" theory, which asserts that, attempts to translate context-dependent languages word-for-word using dictionary equivalences are doomed to failure and also reveal false assumptions about what words mean and how they have meaning [17], [18]. When it comes to speech, the communication partners' shared knowledge about the occasion, setting, topic, intent, or any other feature of the context in which the utterance occurs is often referred to as the context or scenario [19].

It has been demonstrated that different contexts in speech event help in understanding or deriving meaning

E-mail address: balaka@strathmore.edu, bshibwabo@strathmore.edu

from speech, contrary to the notion that context is an abstract construct with a theoretical framework [20], [21]. This is where the idea of situation comes into play as an abstract representation of environment where the relevance of spoken words or set of words exist [2].

Therefore, it is implied that a sentence may be correctly comprehensible but have a misleading meaning or context. It has been demonstrated that context-free Neural Machine Translators (NMT) frequently translate solitary texts in a logical manner. These translations end up being incompatible with one another when combined in a document [4]. Sentence-level NMT, which ignores discourse phenomena and encodes the individual source sentences without using contexts, is another example that stands out [22]. According to the distributional theory, words that are used in comparable settings often have similar meanings. In contrast to the contexts of other words, this theory has been used to determine that the context of an unintelligible word would most likely be identical to circumstances of its own different occurrences [23].

The listener's familiarity with the dysarthric speaker's language choices across a set of situational markers would typically help or hinder communication between a dysarthric speaker and listener [24], [25]. Otherwise, the listener would have to search a dimensionally larger prediction space in order to understand some dysarthric speech. Theoretically, it is conceivable that the optimum way to determine familiarity would be to deduce a commonality in vocabulary frequency when referencing to a certain meaning or collection of meanings. Yet, the plain usage of word-event is recommended as a simple way for inferring familiarity given restricted access to sufficient vocabulary.

Therefore, to define familiarity, we represent the word feature vectors that considers various situational markers, such as topic-events and emotional-events. We assume that the listener has no prior knowledge of the dysarthric speaker, but they have general discourse knowledge about how some of the phrases would be generically contextualized given the nature of dialogue between the listener and the dysarthric speaker.

We then discuss emotional Dialogue Acts (DA), which are a combination of emotional speech indicators and Dialog Act (DA). Via its interaction, the dialog act, the smallest unit of language communication, represents the speaker's process of language communication. To create a faultless dialogue system, dialogue act recognition (DA) is essential. DA serve as powerful representations for the illocutionary force of the utterance, which in some ways represents the speaker's intention. This study consequently focuses on the listener's remediation of two frequent speech markers—the speaker's emotion during discourse and the dialogue act—which when combined determine the familiarity class.

3. METHODOLOGY

We take a typical domain free space into consideration in order to develop the speaker's ontology. Due to the fact that any topic may come up during the talk, the speech of a dysarthric speaker is not constrained to any one domain. Furthermore, the ontology developed is conventional and does not account for the absence of objects, subjects, attributes, or relationships among them, which may be the pattern used by the majority of dysarthric speakers. Such abnormalities are accommodated for by the familiarity function's inference, which is covered in the following section. In order to create a concept lattice and concept pathway for the supplied partially ordered sets, formal concept analysis (FCA) is utilized.

We examine an example of comments from a dysarthric speaker [26] as presented in Table I to demonstrate how FCA is applied. The sentences chosen came from a set of unrelated, Dysarthric contextual suggestions, as is typically the case in everyday conversation.

Definition 1: (*Situation*) We define a situation as a (formal) context which is as a triple *(O,A,R)* where O and A are collections of objects o and attributes a respectively with R being a binary relation between O and A, that is; $R \subseteq O \times A$. Additionally, $(o, a) \in R$ is read as: the object o has the attribute a.

The concept pathway is initially set up to show a list of qualities within the speech and equivalent generated natural language expression that are depicted therein. This list of attributes is then used to create the conceptual hierarchy. Figure 1 serves as an illustration of this, illustrating the similarities and contrasts between the starting concept and its child concepts through the use of generated natural language (NL) statements at the beginning of a thought route. According to [27], these NL fragments represent a user journey along a notion pathway. The concept pathway as tree structure that will build as the speaker speaks and the listener tries to navigate between related formal concepts as well as their semi-concepts is meant to be used by the ad hoc approach for formulation of ontologies for the suggested solution

Table II is a formal concept table that lists the concepts that have been chosen and are not ambiguous to the ontology development process. For describing the semantic relationship between an object and an attribute that provides basic NL processing rules and templates, the predicate technique may be used to define vagueness. A Parser, which provides procedural guidance on how a class of attributes should be derived from free text, also determines the vagueness of a statement.

As seen in Table II, the specified concepts use a universal strategy unless the universal class is absent, as in the case of the concept pipe. The parser decides to leave the term "pipe" exactly as it is since any references to plumbing or

TABLE I. Dysarthria Speech Prompt

Prompts
You wished to know all about my grandfather.[break] Don't ask me to carry an oily rag like that. [break] Well, he is nearly ninety-three years old;[break] We have often urged him to walk more and smoke less pipe;[break] Except in the winter when the ooze or snow or ice prevents,[break] dresses himself in an ancient black frock coat, [break] usually minus several buttons; [break]. A long, flowing beard clings to his chin ;[break]Twice each day he plays skilfully and with zest upon our small organ.

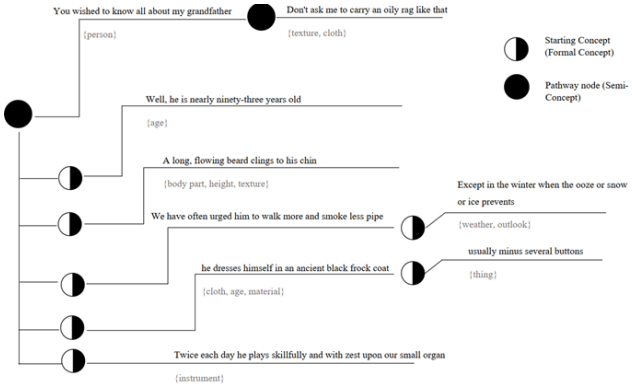

Figure 1. Formal Concept Pathway

TABLE II. Formal Concept Table

Attributes	Concepts					
	Person	Cloth	Body Part	Weather	Instrument	Pipe
Texture		X	X			
Age	X					
Height			X			
Smoke	X					X
Move	X			X		
Prevent	X					
Colour		X				
Plays	X				X	
Size			X		X	
Material		X				

classes that are linked to tubes could violate the semantic association constraints. On the other hand, attributes are kept in their original primitive form since there are an endless number of ways that one notion could connect to another, making generalization dangerous.

Definition 2: (*Context*) A triple K = (O, A, R) is considered a formal context, if O and A are non-empty sets of objects and attributes, respectively, and $R \subseteq O \times A$ is the incidence (binary) relation between O and A.

The hierarchy of formal concepts that adhere to a partial ordering principle are determined by the concept lattice, which is constructed from the incidence matrix (formal context table). The idea lattice then provides generalization and specialization between the concepts.

To generate concept lattices, we adapt the Viterbi algorithm [28] to construct the lattice without necessarily considering the transducer output as the concept path parser. This algorithm is illustrated in Table III and an example of a lattice for the Dysarthric speech prompt is illustrated in Figure 2.

TABLE III. Lattice Construction Algorithm

CreateLattice

Require: SituationalContext, DysarthricText
G←generateDirectedGraph()
for index = 0 to size(**Require:** Situational$_{C}$*ontext*, *DysarthricText*)
do
generateNode(G,index)
end for
for index = 0 to len(text)-1 **do**
if index=0 or nodeInDegree(G,index)>0
updateContext(context,G)
updateLattice(G,index)
end if
end for
cleanup(G,length(text), true)
return G

4. RESULTS

The use of meaning extraction models typically built using ordinary speech is substantially hampered by the speaker variance inherent in Dysarthric speech.Transfer learning is proposed as a strategy for transferring knowledge between one (source) topic to the next (target) topic, in this case the listener, to avoid starting from scratch while developing a new meaning extraction system [29].

A. Listener and Familiarization

A familiarization module along a listener make up this model. This model's objective is to extract and learn events that help the listener become more familiar with the speaker, and then to create embeddings that can be used to discover the speaker's ontologies. Figure 2 provides a detailed illustration of this. As was previously said, familiarization involves learning events from scratch using only the speaker's collection of words—both known and new.

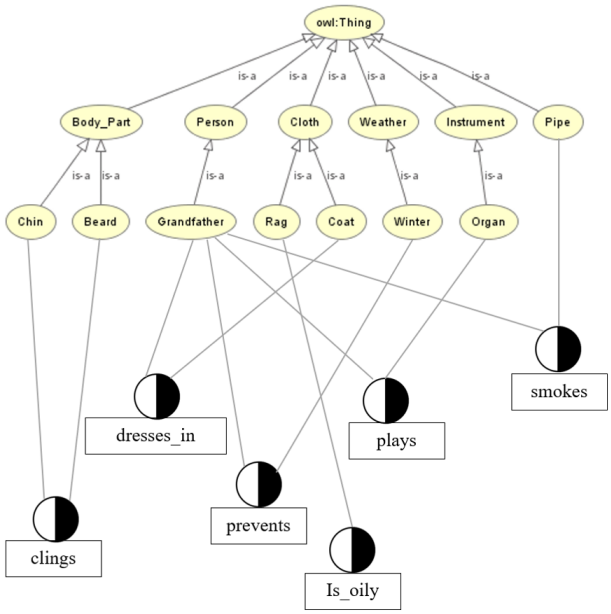

Figure 2. Concept Lattice for Sample Dysarthric Speech

Only two events—topic knowledge and voice emotion—are included in this study.

On the basis of the frequency of appearances in the full sentence as a structurally fundamental (atomic) unit of analysis, the probabilities of a phrase appearing in the text are used to measure topic knowledge. As a result, the probability that a sentence will appear in text is calculated as the sum of the probabilities that a sentence will appear in a particular topic and that a topic will appear in a particular text:

$$P(s \mid d) = \sum_{t \in T} P(s \mid t) x P(t \mid d) \quad (1)$$

Each individual text is utilized to create a subject with a specific probability distribution $P(t \mid d)$, and the topic samples are then used to create a phrase with a given probability distribution $P(s \mid t)$. Each sentence thus has a single topical reference. In order for this to work, it is assumed that sentences with similar themes will also have similar embeddings [30], [31].

A voice conversion (VC) approach is utilized to extract values of voice emotion in order to address the issue of low data availability for Dysarthric speakers. This strategy was first put forth by [33], [34] and is based on partial least squares. It makes use of a phoneme-discriminative characteristic to transform a dysarthric voice into non-dysarthric speech. Following this, the fundamental frequency (F0), spectrogram (SP), and aperiodic spectrum (AP) are extracted from the converted speech samples and transformed into multi-dimensional Mel-cepstral coefficients (MCEPs), which are then put through a 3-layer Backpropagation neural network to identify four emotions. The consequence is that the NN's input and output layers have 12 and 4 nodes, respectively. 10 nodes have been chosen for the hidden layer.

B. Situational Projection and Contextualization (Embedding)

Situational projection in this context refers to the creation of embeddings that include the hypothetical events suggested by the familiarization module. The model evaluates both the events and the set of triples at once, in various iterations, in an arbitrary sequence to produce embeddings. As a result, each time an ontological fact is read, the model fetches the familiarity embeddings of the random instance that appears in the triple and feeds them into a contextualization layer to learn the fetched ontologies.

A familiarity incident is very likely to lead to a new ontological path. If this occurs, it is advantageous to update the ontology and then the embedding's inferences, resulting to a richer knowledge base despite the wider predictive space provided. To encode both facts and conclusions about the instances they represent, the embeddings are gradually updated.

The embedding tagging goal is therefore defined as follows: given a natural language phrase that corresponds to a formal representation, we wish to apply a tag to each sentence and, in turn, an ontological path. In the formal representation, the tag is used to display the familiarity markers within the phrases. A one-hot vector is created to start training: (the i-th word in a vocabulary of j words) is formulated [32]. Each vocabulary index is converted into an embedding vector, a low-dimension vector, via the single hot-vector. The vector embedding is then fed into the recurrent neural network (RNN) shown in Figure 3 for training. The embedding vector is a tripple $e = (s,e,t)$ that is fed into the network that consists of:

- a natural language *sentence s*, namely a sequence of words *(s)*.
- event markers learnt from the familiarity zero-shot learner *(e)*
- topic knowledge markers learned from the familiarity zero-shot learner *(t)*

A few shot learner is used to learn dialogue acts and determine how familiar a speaker is to the audience. Both the known $X_k$and unknown X_u. word classes are used by this learner. The embedding learner would therefore automatically include these classes because, theoretically, some markers proposed by the speakers may be unknown to the listener insofar as the raw input would be the aforementioned tripple. The RNN Network has hidden states across a specified dimension $h_d^{<k>}$ that are used to learn the tripple to infer and produce vector embeddings that

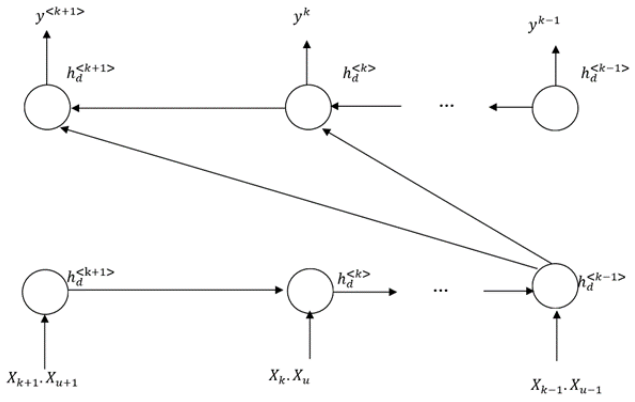

Figure 3. RNN Learner for Tagging Embeddings

display topic knowledge (event) and emotional events for a particular sentence. This network's output is consequently known vectors y^k , which are typically updated by:

$$h_d^{<k>} = f\left(h_d^{<k-1>}, x_d\right) \tag{2}$$

Where x_d is a tripple at a given instance and f is a non-linear activation function [33].

As shown in Figure 3, we modify the Recurrent Network Transducer method [34], which employs a transducer as a link between two recurrent networks [35]. The listener serves as a decoder network in this situation, learning the speaker's familiarity and embedding it onto ontologies that are output and used in a joint network. The familiarity embeddings are normalized by applying a SoftMax over such, thereby modeling a probability distribution across all the embeddings.

The transducer makes use of the conveyed data from the source domain that is absent from the target domain. This adheres to the idea that the speaker is oblivious of the listener's attempts to familiarize themselves and, as a result, inherently alludes to a global (infinite) ontology while speaking to the listener. The*i*-th hidden state in the output domain h_i^T is therefore computed by:

$$h_i^T = RNN\left(h_{l-1}^{\widetilde{T}}, x_i^T; \theta_T\right) \tag{3}$$

Where $h_{l-1}^{\widetilde{T}}$ comprises both the information that was transmitted from the source domain *(i)* and the information that was passed from the destination domain h_{l-1}^T at the previous time step. We figure it out as:

$$h_{l-1}^{\widetilde{T}} = f\left(h_{i-1}^T, \psi_i|\theta_f\right) \tag{4}$$

Where θ_f is the parameter for f and ψ_i being the alignment computed from the look up inference logic as highly probably embedding from the source domain.

5. DISCUSSION

A. *Summary of Findings*

In order to assist the logical listener in condensing the predictive space, we have developed a novel theoretical framework in this study that converts natural phrases into formal context ontologies distinguished by familiarity indicators. The suggested solution employs a multi-layered strategy made up of four main parts. The few-shot learner, an artificial neural network, is the first and teaches speakers' familiarity in terms of both subject matter expertise and vocal emotions.

The second component is the situational projection, which gathers natural phrases and acquired familiarity markers into a vector triple that is given into the speech contextualizer which is the third component. With the aim of locating formal contexts that are relatively close to the listener's approximation, the speech contextualizer learns the ontological concepts of the input tripple.

A recurrent neural network transducer serves as the fourth component, linking between the situational projection (speaker side) and the contextualizer. This transducer serves as a logical listener or actuator that feeds the familiarity learner and the contextualizer with triples and embeddings from the external environment, respectively. The secondary function of the transducer is computation, which involves producing inference based on the contextualized speech in the form of the original natural sentence, any potential familiarity markers, and the rebuilt formal context, which is the meaning of the natural phrase.

B. *Implication of Findings*

The use of familiarity markers suggests that a different strategy would result in a wider predictive space for the contextualizer and a higher incidence of incorrect meaning extraction. As a result, the markers serve as a kind of filter for the expected triple vectors. Despite this advantage, it is also known that adding familiarity markers to overfiltered data may provide data that encourages overfitting in contextualization models. It is therefore recommended that the number of familiarity markers per discourse instance be restricted to a number less than the dimension of the triple, in this case two markers per training. This, however, ultimately depends on the model's performance.

While familiarity has the connotation of prior knowledge, this study suggests a new technique that uses a zero-shot learner and assumes absolutely no prior knowledge of the familiarity markers. This is advantageous since it increases the likelihood that the model will generalize more effectively, especially in the case where one listener is attempting to deduce meaning from various Dysarthric speakers.

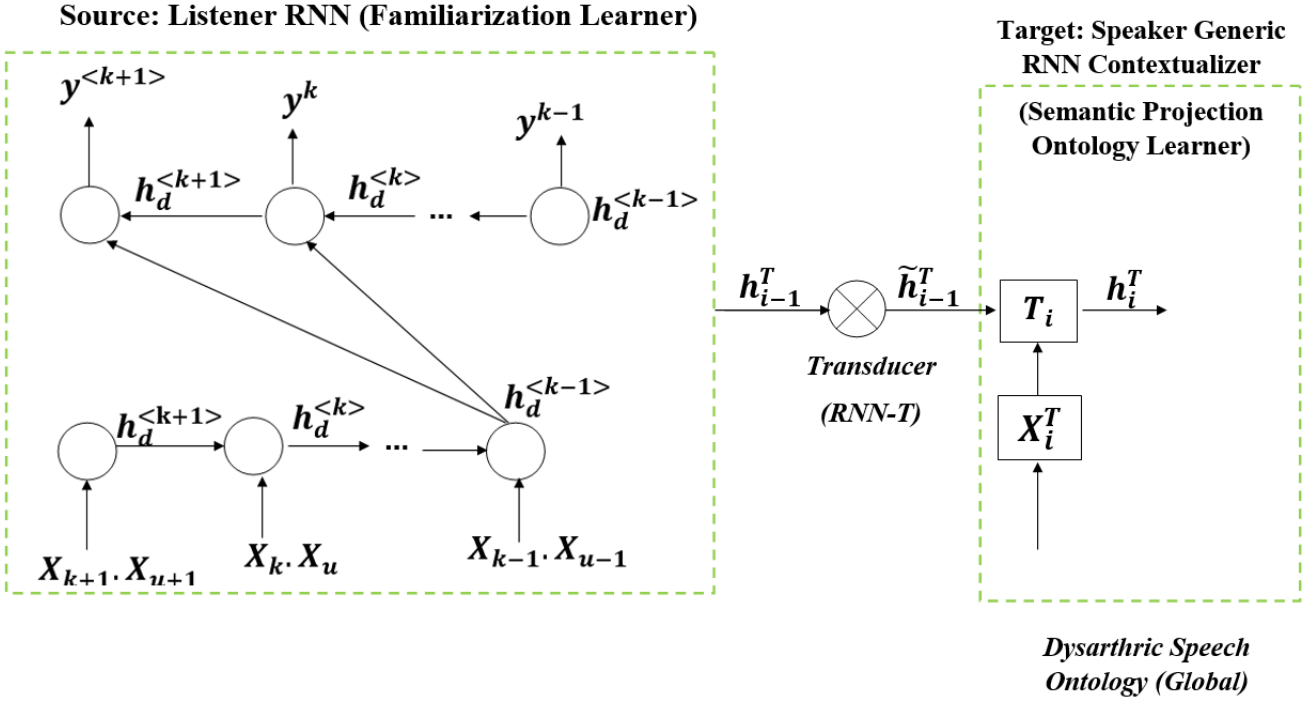

Figure 4. Listener Speaker Joint-Trainer

The situational projection makes use of the natural sentence and familiarity indicators to create deeper embeddings for the contextualize to deal with. This clarifies the possibility that the same sentence may appear more than once due to various situational factors (familiarity markers in this case). This is frequently the situation in the everyday speech of dysarthric speakers, who must restrict their vocabulary, reuse it in many circumstances, and rely on the listener to deduce the seemingly new alluded meaning.

The ontology learner that is suggested in the contextualization component is a joint learner that learns both global and ontologies that are embedded in domains of familiarity and is shown in Figure 4. Hence, an inference engine is required, in this case, a recurrent neural network transducer (RNN-T). Due to the RNN-dual T's open nature, it was decided to use it. It has already been utilized as an encoder-decoder tool in earlier studies, but in this work, it is modified to function as both a data conveyor between the speaker and the listener, ultimately producing the meaning of the original natural language.

C. Limitation of Study

The Dysarthric sentence prompts feature distinct but connected statements from the study by [26], as shown in Table I. This is in line with the little discourse information of dysarthric speakers. The natural workaround is selecting carefully linked statements, which can be tedious. The model has access to more data thanks to cooperative learning, which calls for the incorporation of global (non-dysarthric speech data) discourse data. The second restriction, the use of a logical listener rather than a real listener, is connected to this. While the proposed module serves as both an inference engine and a data parser from and forth between the various components of the proposed system, the logical listener is used to minimize the absence of listener data.

Another drawback of the study is that familiarity is consistent, since this is not always the case. For example, topic knowledge can alter dramatically throughout routine conversation. However, only two familiarity indicators were taken into consideration for this investigation, and a static (non-updating) strategy was used in order to reduce the possibility of overfiltering and overfitting the model.

6. CONCLUSION

The study suggests an integrated method for understanding dysarthric speech. The method discussed includes a contextualizer, a recurrent neural network transducer, a zero-shot learner for familiarity learning, a situational projection component, and a learner for familiarity learning from dysarthric speech. Further empirical research will be required to support the results, as being an effective and accurate method of obtaining speech understanding, particularly for the unintelligible Dysarthric speech, given the theoretical character of this study. The familiarity learner will be treated and made more dynamic in the future with regard to updating the familiarity markers on an as-needed basis, which will further extend this work. The recurrent neural network transducer should also be expanded in order to accommodate live or real listeners as required by

comparable data availability.

References

[1] S. Maruf, A. F. T. Martins, and G. Haffari, "Selective Attention for Context-aware Neural Machine Translation." [Online]. Available: http://arxiv.org/abs/1903.08788

[2] Z. Ling and T. Chuanmao, "Study on the Role of Context in Discourse Analysis from the Viewpoint of "Make" in Different Sentence Meanings," vol. 5, no. 6, pp. 1818–1825.

[3] J. Tiedemann and Y. Scherrer, "Neural Machine Translation with Extended Context." [Online]. Available: http://arxiv.org/abs/1708.05943

[4] E. Voita, R. Sennrich, and I. Titov, "When a Good Translation is Wrong in Context: Context-Aware Machine Translation Improves on Deixis, Ellipsis, and Lexical Cohesion," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 1198–1212.

[5] J. Garten, B. Kennedy, K. Sagae, and M. Dehghani, "Measuring the importance of context when modeling language comprehension," vol. 51, no. 2, pp. 480–492. [Online]. Available: http://link.springer.com/10.3758/s13428-019-01200-w

[6] M. T. Pilehvar and J. Camacho-Collados, "WiC: the Word-in-Context Dataset for Evaluating Context-Sensitive Meaning Representations." [Online]. Available: http://arxiv.org/abs/1808.09121

[7] M. Basaldella, L. Furrer, C. Tasso, and F. Rinaldi, "Entity recognition in the biomedical domain using a hybrid approach," vol. 8, no. 1, p. 51. [Online]. Available: https://jbiomedsem.biomedcentral.com/articles/10.1186/s13326-017-0157-6

[8] G. Figueiredo, A. Duchardt, M. M. Hedblom, and G. Guizzardi, "Breaking into pieces: An ontological approach to conceptual model complexity management," in *2018 12th International Conference on Research Challenges in Information Science (RCIS)*. Ieee, pp. 1–10.

[9] F. Ali, D. Kwak, P. Khan, S. H. A. Ei-Sappagh, S. M. R. Islam, D. Park, and K.-S. Kwak, "Merged Ontology and SVM-Based Information Extraction and Recommendation System for Social Robots," vol. 5, pp. 12 364–12 379. [Online]. Available: http://ieeexplore.ieee.org/document/7962152/

[10] J. a. P. A. Almeida, R. A. Falbo, and G. Guizzardi, "Events as Entities in Ontology-Driven Conceptual Modeling," in *Conceptual Modeling*, A. H. F. Laender, B. Pernici, E.-P. Lim, and J. P. M. de Oliveira, Eds. Springer International Publishing, vol. 11788, pp. 469–483, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/978-3-030-33223-5%5F39

[11] M. N. Asim, M. Wasim, M. U. G. Khan, W. Mahmood, and H. M. Abbasi, "A survey of ontology learning techniques and applications," vol. 2018.

[12] J. Chen, G. Alghamdi, R. A. Schmidt, D. Walther, and Y. Gao, "Ontology Extraction for Large Ontologies via Modularity and Forgetting," in *Proceedings of the 10th International Conference on Knowledge Capture*. Acm, pp. 45–52.

[13] J. Chen, P. Hu, E. Jimenez-Ruiz, O. M. Holter, D. Antonyrajah, and I. Horrocks, "OWL2Vec∗: embedding of OWL ontologies," vol. 110, no. 7, pp. 1813–1845. [Online]. Available: https://link.springer.com/10.1007/s10994-021-05997-6

[14] S. Knuijt, J. G. Kalf, B. G. van Engelen, B. J. de Swart, and A. C. Geurts, "The Radboud Dysarthria Assessment: Development and Clinimetric Evaluation," vol. 69, no. 4, pp. 143–153. [Online]. Available: https://www.karger.com/Article/FullText/484556

[15] S. Pinto, R. Cardoso, J. Sadat, I. Guimarães, C. Mercier, H. Santos, C. Atkinson-Clement, J. Carvalho, P. Welby, P. Oliveira, M. D'Imperio, S. Frota, A. Letanneux, M. Vigario, M. Cruz, I. P. a. Martins, F. Viallet, and J. J. Ferreira, "Dysarthria in individuals with Parkinson's disease: a protocol for a binational, cross-sectional, case-controlled study in French and European Portuguese (FraLusoPark)," vol. 6, no. 11, p. e012885. [Online]. Available: https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2016-012885

[16] M. Summaka, H. Harati, S. Hannoun, H. Zein, N. Koubaisy, Y. Fares, and Z. Nasser, "Assessment of non-progressive dysarthria: practice and attitude of speech and language therapists in Lebanon," vol. 21, no. 1, p. 450. [Online]. Available: https://bmcneurol.biomedcentral.com/articles/10.1186/s12883-021-02484-2

[17] M. Dudek, "Not So Long Time Ago Before Malinowski: The Puzzle of Lotar Dargun's Influence on Bronislaw Malinowski," in *Bronislaw Malinowski's Concept of Law*, M. Stépień, Ed. Springer International Publishing, pp. 3–20. [Online]. Available: http://link.springer.com/10.1007/978-3-319-42025-7%5F1

[18] A. Lukin, "Language and Society, Context and Text: the Contributions of Ruqaiya Hasan," in *Society in Language, Language in Society*, W. L. Bowcher and J. Y. Liang, Eds. Palgrave Macmillan UK, pp. 143–165. [Online]. Available: http://link.springer.com/10.1057/9781137402868%5F6

[19] C. K. Porcaro, P. M. Evitts, N. King, C. Hood, E. Campbell, L. White, and J. Veraguas, "Effect of Dysphonia and Cognitive-Perceptual Listener Strategies on Speech Intelligibility," vol. 34, no. 5, pp. 806.e7–806.e18. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0892199719300165

[20] S. Akinkurolere, "Exploring the Significance of Context in Meaning: Speech Act Features of Performative Political-Speeches of President Umaru Musa Yar'Adua," vol. 7, no. 1, pp. 65–84. [Online]. Available: https://www.athensjournals.gr/humanities/2020-7-1-3-Akinkurolere.pdf

[21] K. C. Hustad, "The Relationship Between Listener Comprehension and Intelligibility Scores for Speakers With Dysarthria," vol. 51, no. 3, pp. 562–573. [Online]. Available: http://pubs.asha.org/doi/10.1044/1092-4388%282008/040%29

[22] M. Li, "Mixed-channel inventory policy between retailer and e-retailer stores in two-echelon supply chain," in *2009 16th International Conference on Industrial Engineering and Engineering Management*. Ieee, pp. 1525–1529.

[23] B. Pintér, G. Vörös, Z. Palotai, Z. Szabó, and A. Lőrincz, "Determining Unintelligible Words from their Textual Contexts," vol. 73, pp. 101–108. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1877042813003212

[24] S. Bloch and C. Saldert, "Person Reference as a Trouble Source in Dysarthric Talk-in-Interaction," in *Atypical Interaction*, R. Wilkinson, J. P. Rae, and G. Rasmussen, Eds. Springer International Publishing, pp. 347–372. [Online]. Available: http://link.springer.com/10.1007/978-3-030-28799-3%5F12

[25] K. L. Lansford, S. Luhrsen, E. M. Ingvalson, and S. A. Borrie, "Effects of Familiarization on Intelligibility of Dysarthric Speech in Older Adults With and Without Hearing Loss," vol. 27, no. 1, pp. 91–98. [Online]. Available: http://pubs.asha.org/doi/10.1044/2017%5FAJSLP-17-0090

[26] F. Rudzicz, A. K. Namasivayam, and T. Wolff, "The TORGO database of acoustic and articulatory speech from speakers with dysarthria," vol. 46, no. 4, pp. 523–541. [Online]. Available: http://link.springer.com/10.1007/s10579-011-9145-0

[27] T. Wray and P. Eklund, "Context and Natural Language in Formal Concept Analysis," in *Modeling and Using Context*, P. Brézillon, R. Turner, and C. Penco, Eds. Springer International Publishing, vol. 10257, pp. 343–355, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/978-3-319-57837-8%5F28

[28] N. Barrett and J. Weber-Jahnke, "Building a biomedical tokenizer using the token lattice design pattern and the adapted Viterbi algorithm," p. 11.

[29] F. Xiong, J. Barker, Z. Yue, and H. Christensen, "Source Domain Data Selection for Improved Transfer Learning Targeting Dysarthric Speech Recognition," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Ieee, pp. 7424–7428.

[30] L. Liu, L. Tang, W. Dong, S. Yao, and W. Zhou, "An overview of topic modeling and its current applications in bioinformatics," vol. 5, no. 1, p. 1608. [Online]. Available: http://springerplus.springeropen.com/articles/10.1186/s40064-016-3252-8

[31] O. Kozbagarov, R. Mussabayev, and N. Mladenovic, "A New Sentence-Based Interpretative Topic Modeling and Automatic Topic Labeling," vol. 13, no. 5, p. 837. [Online]. Available: https://www.mdpi.com/2073-8994/13/5/837

[32] G. Petrucci, C. Ghidini, and M. Rospocher, "Ontology Learning in the Deep," in *Knowledge Engineering and Knowledge Management*, E. Blomqvist, P. Ciancarini, F. Poggi, and F. Vitali, Eds. Springer International Publishing, vol. 10024, pp. 480–495, series Title: Lecture Notes in Computer Science. [Online]. Available: http://link.springer.com/10.1007/978-3-319-49004-5%5F31

[33] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation." [Online]. Available: http://arxiv.org/abs/1406.1078

[34] A. Graves, "Sequence Transduction with Recurrent Neural Networks." [Online]. Available: http://arxiv.org/abs/1211.3711

[35] T. Makino, H. Liao, Y. Assael, B. Shillingford, B. Garcia, O. Braga, and O. Siohan, "Recurrent Neural Network Transducer for Audio-Visual Speech Recognition." [Online]. Available: http://arxiv.org/abs/1911.04890

Benard Alaka is a Doctoral Candidate at Strathmore University's School of Computing and Engineering Sciences. His research interest include Computer Vision, Speech Processing, Speech Comprehension and applications of Machine Learning.

Dr. Bernard Shibwabo Kasamani is a Senior Lecturer at Strathmore University's School of Computing and Engineering Sciences. He also serves as the Director of Graduate Studies and is an active researcher in the areas of Machine Learning, Embedded Systems, Full Stack Applications, Dataspaces and Data Engineering.