

Conversational artificial intelligence – demystifying statistical vs linguistic NLP solutions

Kulvinder Panesar

York St University, York, United Kingdom

Corresponding author: k.panesar@yorksj.ac.uk

Received: 24 February 2020 / Accepted: 5 April 2020 / Published: 22 May 2020

Abstract

This paper aims to demystify the hype and attention on chatbots and its association with conversational artificial intelligence. Both are slowly emerging as a real presence in our lives from the impressive technological developments in machine learning, deep learning and natural language understanding solutions. However, what is under the hood, and how far and to what extent can chatbots/conversational artificial intelligence solutions work – is our question. Natural language is the most easily understood knowledge representation for people, but certainly not the best for computers because of its inherent ambiguous, complex and dynamic nature. We will critique the knowledge representation of heavy statistical chatbot solutions against linguistics alternatives. In order to react intelligently to the user, natural language solutions must critically consider other factors such as context, memory, intelligent understanding, previous experience, and personalized knowledge of the user. We will delve into the spectrum of conversational interfaces and focus on a strong artificial intelligence concept. This is explored via a text based conversational software agents with a deep strategic role to hold a conversation and enable the mechanisms need to plan, and to decide what to do next, and manage the dialogue to achieve a goal. To demonstrate this, a deep linguistically aware and knowledge aware text based conversational agent (LING-CSA) presents a proof-of-concept of a non-statistical conversational AI solution.

Keywords: conversational artificial intelligence, knowledge representation, machine/deep learning, natural language understanding/processing, role and reference grammar

1. INTRODUCTION

We have seen a resurgence of artificial intelligence (AI) (Woolf 2019), from 2010 onwards due to

extensive computing power and speed (Moore's law) and huge dataset available from the rise of the Internet and complex system architectures, handling petabytes of data, with access to abundant amounts of data. AI is when a machine mimics the cognitive functions that humans associate with other human minds such as learning and problem solving, reasoning, knowledge representation, social intelligence and general intelligence. Popular conversational systems have seen the disruption of the smart machine industry, but we are faced with the long standing issue of interpretation of the accuracy meaning in human-computer interaction. However, it is forecasted that by 2021, more than 50% of enterprises will spend more per annum on bots and chatbot creation than traditional mobile app development.¹ This future investment is driven by the hype that conversational AI technologies will give customers a seamless and low-effort experience. The focus of this paper is to pursue this claim, and look into its accuracy of meaning, unpacked in stages, via the objectives of this paper: (1) to address the fundamentals to conversation which is natural language - its processing, understanding and generation of conversation; (2) a review of the underpinning NLP technologies and advancements, and the underlying architectural models that drive them and the impact and limitations of the solutions; (3) to analyse through examples the mathematical impact of increasing knowledge and knowledge representation in statistical conversational AI solutions; (4) to discuss statistical vs linguistic conversational solutions; (5) for comparison, to explore the requirements and implementation of a conversational software agent based on a linguistic solution; (6) to present a range of conclusions and a future vision of conversational AI.

2. NLP AND NLP ADVANCEMENTS

2.1. Introduction to NLP

The analysis "what happens in an Internet Minute"² highlights that more than 80% of the data in this world is unstructured in nature, which includes text, which requires text mining and Natural Language processing (NLP) to make sense out of this data. NLP (our focus) is one of the main problems of AI as well as reasoning, knowledge planning, and learning. NLP is defined as: 'as a theoretically motivated range of computational techniques for analysing and representing natural occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or application' (Liddy 2001, 2). Figure 1 demonstrates the NLP component stack identifying a pipeline for phonology, morphology, syntax, semantic and reasoning. The challenge of NLP and NLU is due to the range of ambiguities from lexical to pragmatic. Various NLP tools are used: sentiment analyser, Parts of Speech (POS) Taggers, Chunking (shadow parsing, Named Entity Recognition (NER), Emotion Detection, Semantic Role Labelling (SRL) using parse trees and assigns roles to the arguments in the sentence. Nuances of meaning make natural language understanding difficult as the text's meaning can be influenced by context and reader's "world view" (Sharda, Delen, and Turban 2019).

¹ <http://www.artificial-solutions.com>

² <https://www.visualcapitalist.com/what-happens-in-an-internet-minute-in-2019/>

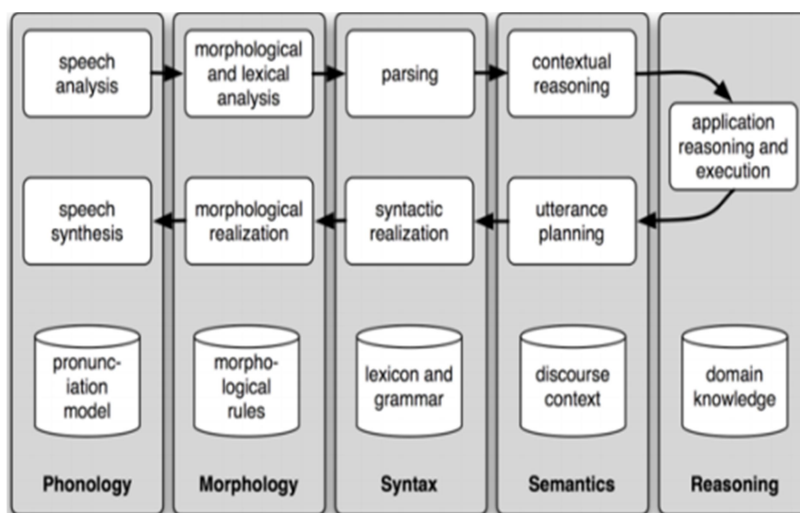


FIGURE 1. TRADITIONAL NLP COMPONENT STACK (BIRD, KLEIN, AND LOPER 2009, 24).

2.2. NLP in AI and current timeline

Singh (2018) confirms that AI is a broader umbrella under which Machine Learning (ML) is subset of AI and Deep Learning (DL) is subset of ML. For clarity, AI is a program that can sense, reason, act and adapt. ML is extracting knowledge from the data with supervised learning (SL) algorithms – human teaching machines with labels to produce the correct outputs (for example: is there a traffic sign in the image); DL is a prominent topic, known for its breakthroughs in fields like computer vision (facial key point detection) and game playing (Alpha GO), surpassing human ability. Unsupervised learning (UL) – learning without a teaching, also considered DL whereby networks capable of learning unsupervised from data that is unstructured or unlabelled, and linked to as deep neural learning or deep neural network (NN).

Saifee (2019a) confirms that 2018 was widely touted as the ‘Imagenet moment’ in the field of NLP, due to the various big players such as Google, OpenAI, Facebook releasing various pre-trained model and open source algorithms – causing advancements in NLP, which have happened over the past 2-5 years. Sharda, Delen, and Turban (2019) detail a series of examples of NL applications: (1) answering NL questions via IBM Watson; (2) analysing documents and producing short summaries (abstracts) to assist search results and guide the reader; (3) speech synthesis (speech-to-text), speech recognition (text-to-speech), and inter-language text-to-text translation; (4) collaborative filtering—used to implement recommender systems (“if you liked this movie, you might also like...”); (5) text classification - classifying news articles by categories, such as world news, sports, or business, etc.; (6) topic modelling in documents; (7) sarcasm/hate and crime/cyberbullying detection using sentiment analysis; (8) text simplification; (9) speech to sign language and vice versa; (10) lip-reader technology, for people who cannot speak, converts lip movement to text or speech to enable conversation. Some of these NLP examples were underpinned by the popular sequence-to-sequence model (seq2seq) (Sutskever, Vinyals, and Le 2014). It is a DL model useful for text summarisation in that it takes sequence of words and outputs another sequence of items. This trained model contains an encoder and decoder. The encoder takes each item in the entire input sequence and processes it into a vector (context). The encoder sends the context over to the decoder and generates the

output sequence one word at a time (Alammar 2018). For this paper, we focus on the seq2seq model and its developments for which it powers applications like Google Translate, voice-enabled devices and online chatbots. In the next section, we discuss a range of NLP advancements.

2.3. NLP historical technology context

Prior to 2012, Saifee (2019a) highlights the main NLP processing of: (1) bags of words representation without any ordering; (2) popular techniques of latent semantic indexing (LSI), SVM (Support Vector Machine) and Hidden Markov Models (HMM); (3) NLP - semantic similarity – focusing on ‘topics’ rather than words achieved by techniques such as LSI, LDA (Latent Dirichlet allocation) and PCA (principal component analysis); (4) continued standard NLP tricks such as POS tagging, NER (named entity recognition). The next era from 2013 to 2016 saw the introduction of word embeddings, both frequency and prediction based (neural network), in particular Word2Vec and alternatives of Glove and FastText. They also help in analysing syntactic similarity among words. *Example 1.* Word2Vec: a Python dictionary: ['Word', 'Embeddings', 'are', 'Converted', 'into', 'numbers']. A vector representation of a word may be a one-hot encoded vector where 1 stands for the position where the word exists and 0 everywhere else. The vector representation of “numbers” in this format according to the above dictionary is [0,0,0,0,0,1]. *Example 2.* GloVe: to test [women+king_man=queen] has model code [model.most_similar(positive=['woman','king'],negative=['man'],topn=1)] with an output score 'queen: 0.508'; *Example 3.* GloVe: to test [degree of similarity between two words] has model code [model.similarity('woman','man')] with output score = 0.73723527; *Example 4.* GloVe: test [probability of a text under the model] with model code: model.score(['The fox jumped over the lazy dog'.split()]) with score output score 0.21. *Example 5:* visual word embedding in Figure 2.

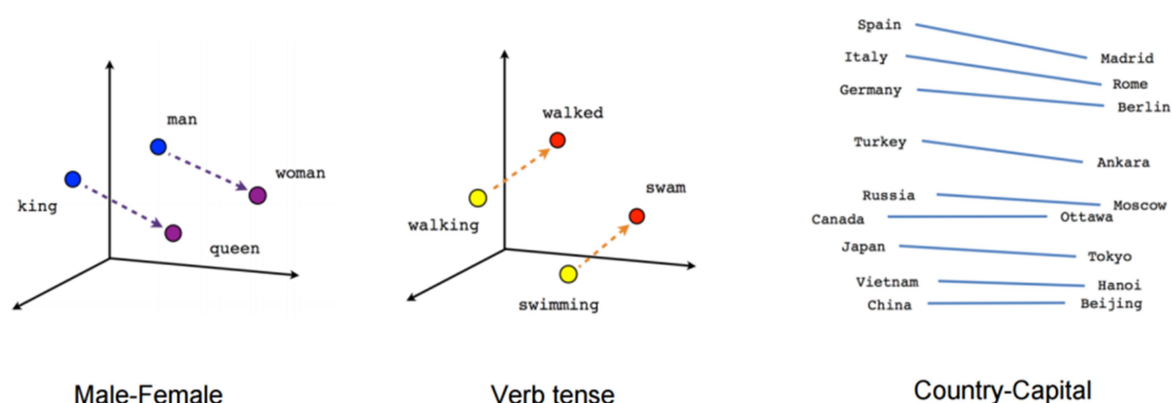


FIGURE 2. WORD EMBEDDING.³

The 2014-2016 period saw the introduction of DL and the comeback of recurrent neural networks (RNN). This is an old technique, which specialises in dealing with problems whose input/output have a sequential manner, having the potential for processing a sentence with

³ <https://www.analyticsvidhya.com/blog/2019/09/demystifying-bert-groundbreaking-nlp-framework/>

order determined by a grammar –a challenge due to memory issues. The RNN sentence concept involves reading a word at a time, retaining the prior word/context word in memory, and which update your understanding on these new words. Whilst iterating through the elements of the input sequence, it encodes and maintains an internal ‘state’; which is subsequently reset when processing two different and independent sequences, but subject to the vanishing gradient problem (dilution of information) due to being slow.

To solve this problem, there are two variants, namely, LSTM (long short term memory) and GRU (Gated Recurrent Unit). Both word embedding and RNN provided the way for rapid progress in NLP and significantly improved the performance on various benchmark datasets. After 2017, there has been a range of exciting development of transformers – attention models, in response to the key challenges of LSTM (challenge 2) and GRU (challenge 1 and 3) as discussed in Table 1.

TABLE 1. REVIEW OF NLP ADVANCEMENTS.

<i>Type</i>	<i>Concept</i>	<i>Benefit/Drawback</i>
WordVec (Mikolov et al. 2013) Google	Uses Neural Network (NN) Variants of Continuous bag of words (CBOW) - based a neighbouring word predict a given word Skip gram – reverse of CBOW. Given a target work – predict the neighbouring	<i>Benefits</i> Training (fast), Vocabulary (millions of words), Re-use of pre-calculated word embedding (learned representations impact the performance on other NLP tasks. Unsupervised approach – ease of annotated data; small cosine distance for similar words
GloVe Global Vector for Word Representation (Pennington, Socher, and Manning 2014) Stanford	Combines benefit with global matrix factorisation method (LSA) – statistical results and WordVec for analogy tasks	<i>Benefits</i> Instead of predicting a word, GloVe vectors have inherent meaning – derived from word co-occurrences See Table 1. This evaluation scheme favours models that produce dimensions of meaning, thereby capturing the multi-clustering idea of distributed representations.
FastText (Joulin et al. 2016) Facebook	Similar to Word2Vec – but embedding created from n-gram of characters: for example: ‘fast’ is <f, fa, as, st,t>	<i>Benefits</i> Allows you to generalise to unknown words, not in the training data; requires less training
LSTM (Sundermeyer, Schlüter, and Ney 2012)	Heavy processing of information over many timestamp using three gates	<i>Drawbacks</i> (1) Lack of parallelisation hurt performance; (2) Learning long term dependency (loss of context) remains a challenge compared to shorter sentences; (3) A linear increase in the number of operations required for the distance between the positions.
GRU (Cho et al. 2014)	Using two gates and less parameters than LSTM.	
Attention (Bahdanau et al. 2016)	Architecture to address challenge 2 above – loss of context	<i>Benefits</i> Training times are significantly reduced – as the distance between words is not computationally relevant

Transformer (Vaswani et al. 2017)	Architecture to address (1) and (3) as well as tackling generic tasks of sequence to sequence tasks in deep learning.	<i>Benefits</i> Based on the attention mechanism – no CNN or RNN units. Super quality results on language translation from previous benchmarks. Very useful to other tasks due to its generalisation More parallelisation <i>Drawbacks</i> Only fixed-length sentences Larger sentence fragmented to feed into the model which causes “context fragmentation”
BERT <i>Bidirectional Encoder Representations from Transformers.</i> (Devlin et al. 2018)	Language representation model: Novel bi-directional training and popular for the concept of pre-training/transfer learning in NLP. Designed to pre-train deep bidirectional representations from unlabelled text by jointly conditioning on both left and right contexts in all layers.	<i>Benefits</i> The pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task specific architecture modifications. Powerful results for eleven NLP tasks.

The architecture of post-2017 NLP models is discussed in the next section.

2.4. Attention, transformers, transfer learning and BERT

2.4.1. Attention concept used in transformers

As mentioned in Section 2.2, RNNs are central to language understanding tasks such as language modelling, question answering and in machine translation. However, for machine translation, sequential processing of words left-to-right and right-to-left requires RNNs to perform multiple computations depending how far the words are apart. For instance, in the example utterance ‘*I arrived at the bank after crossing the...*’ requires knowing the context that the sentence will end in ‘*.. road.*’ or ‘*... river.*’ (Uszkoreit 2017) . Here heavy computational processing is not necessary with a Transformer which requires a small number of calculated steps. Using the example above, each step in the transformer will apply a self-attention mechanism originally designed for machine translation will model the relationship between the words regardless of the relative position in the utterance (Vaswani et al. 2017). An example of its use is translating from English to French; each word in the output French sentence is informed by a context based on the original input English sentence. It is composed of varying degrees of attention/importance, as demonstrated in Figure 3 with two different sentences pairs and where the demonstrative lexical ‘it’ refers to a different noun in each, similar to our human brains in thinking the first one refers to an animal and the second one is to the street.

The animal didn't cross the street because **it** was too tired.
L'animal n'a pas traversé la rue parce qu'**il** était trop fatigué.

The animal didn't cross the street because **it** was too wide.
L'animal n'a pas traversé la rue parce qu'**elle** était trop large.

FIGURE 3. TWO UTTERANCES (USZKOREIT 2017, 5).

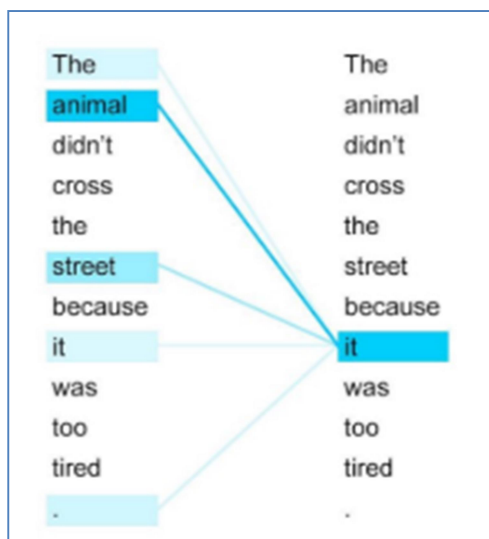


FIGURE 4. 1ST UTTERANCE MAPPING CONTEXT (USZKOREIT 2017, 5).

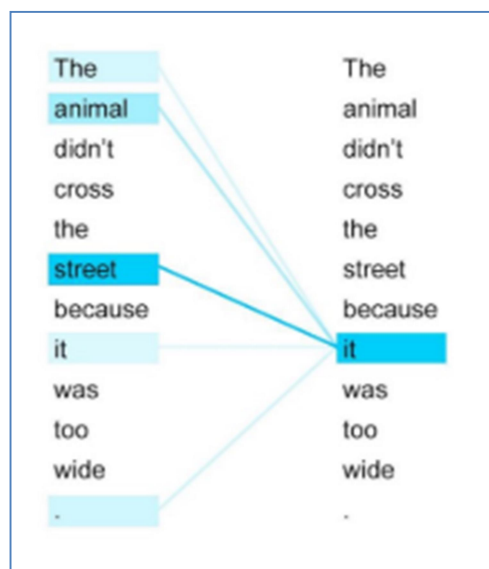


FIGURE 5. 2ND UTTERANCE MAPPING CONTEXT (USZKOREIT 2017, 5).

Figure 4 and Figure 5 illustrate that the darker colour coding is underpinned with varying attention weights (scores) learned from the NN, but calculated as a function of input/hidden states. Here, the problem of long-term dependency is solved by using the most appropriate context at each step.

As mentioned in Table 1, the transformer architecture by Vaswani et al. (2017) was created to

address challenges of LSTM and GRU as well as tackling generic tasks of sequence to sequence tasks such as syntactic constituency parsing in deep learning.

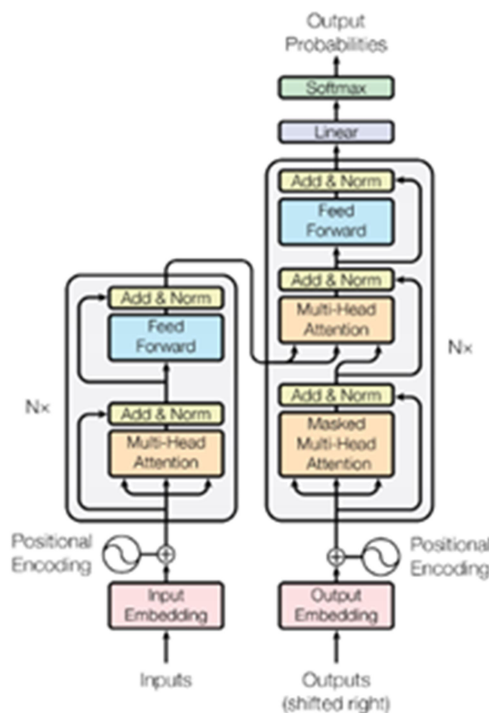


FIGURE 6. TRANSFORMER MODEL ARCHITECTURE (VASWANI ET AL. 2017, 3).

Figure 6 illustrates the transformer architecture with high level details including: (1) encoding for reading the input sentence and decode paradigm for the output sentence; (2) auto regressive – inputs are derived from previous outputs; (3) no RNN – just attention layers; (4) embedding – each word to numbers; (4) positional encoding for the order of the original sequence; (5) multi-head attention layers; (6) 3 types of attention - Encoder self-attention, Encoder-Decoder Attention and Decoder self-attention; (7) self-attention – for example the Encode self –attention (left bottom attention block) will work out the attention weight for the output of the previous layer (of the Encoder) to learn the relationship/context for each word in the input and with other word in the input; (8) Encoder-decoder attention - right top – works out how relevant is each word in the input for each word in the output.

2.4.2. Transfer learning concept

At one point, machine learning and data mining generally worked on training and future data in the same feature space and same distribution, however, there was an opportunity and a need at the same time for knowledge transfer learning between task domains. The impact of this was the elimination of expensive data labelling and the improvements in the performance of learning based on the source domain (Pan and Yang 2009). The transfer learning (TL) concept is where a model trained on one source task/domain uses the knowledge gained from it, and is shared to a re-purposed model on a second related target task/domain. Bhashkar (2019) identifies that TL is used with the concept of policy learning based on Deep Q-Networks (DQN)

to support the user’s goal and ensure dialogue management in conversational AI bots.

2.4.3. BERT and advancements

In reference to Table 1, BERT (Bidirectional Encoder Representations from Transformers) is a framework inspired by various other architectures, training approaches, and language models and based on the Transformer (Devlin et al. 2018). It has changed the NLP landscape. Features of BERT include: (1) bi-directional training looks at the words from left to right and from right to left for better context – compared to other language models in one direction; (2) pre-trained on large huge text corpuses (about 3 billion) to fine-tune one additional output layer using your own parameter and own training data – sentence classification, question answering system, named entity recognition etc. without having to do too many task specific changes in model architecture; (3) training objective is based on un-supervision with BERT jointly training on two tasks: (i) “predict a word given the words on its left and right (with some details on masking to avoid leakage) and (ii) predicting the next sentence given a sentence” (Saifee 2019b); (4) larger model size as Base (110M parameters) and Large (340M parameters) performed better due to the pre-training on a huge corpus. Figure 7 illustrates the use of BERT for extracting the embedding for each tweet and using the embedding to train a text classification model for hate speech (Bhashkar 2019).

Various BERT notable modifications include: RoBERTa and mBERT by Facebook AI, XLNet by researchers at CMU and ERNIE (Baidu). These heavy lifting methods by the giants provide an opportunity gateway for easier fine-tuning for NLP tasks. Bourke (2018) reviews the current state of DL and interviewed Andrew NG who states “AI is the new electricity” and he predicted a lot of success from transfer learning than unsupervised and reinforcement learning (Ruder 2016) – but the question remains to what degree in NLP and conversational AI?

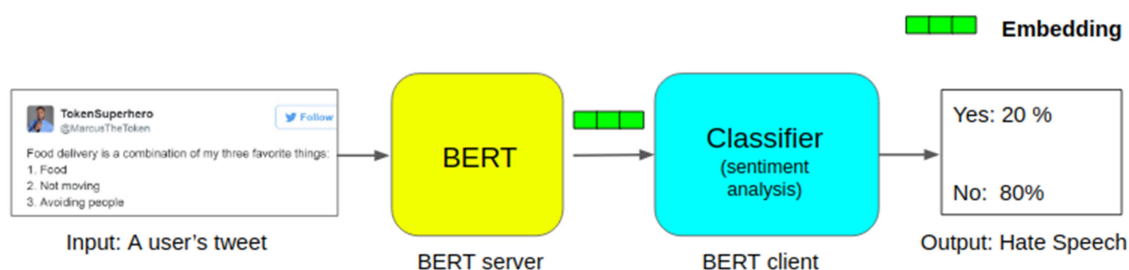


FIGURE 7. BERT FOR TWEET CLASSIFICATION (BHASKAR 2019, 14).

3. CONVERSATIONAL AGENTS: A DETAILED LOOK REVIEW OF THE SPECTRUM OF CONVERSATIONAL INTERFACES

3.1. Current Conversational AI Interest

Conversational AI is becoming popular and exists on the hype cycle for Emerging Technologies in 2018, at the tip of innovation trigger and the peak of inflated expectation (Gartner 2018).

Hence, both academic research efforts and the commercial space are very interested in the conversational AI and the accessibility of large volumes of data. Subsequently this has led to a vast range of opportunities for advancing conversational systems. Examples include delivering to the consumer with smartphone technology familiarisation, and consumer support agents, with a better user experience and hence greater value (Yan 2018). Exciting conversational AI initiatives to drive this interest include: (1) NLP Hackathon – to create a chatbot to help with everyday tasks/voice bots using open source framework Raza (Ekzarian 2018); (2) talks – Chatbot Summit, Chatbot and AI, AI & ML, Chatbot conference NL (Chatbot ecosystem), AI and Big Data (Chatbots) – similar clarifications; (3) chatbots ‘available-as-a-service’ – Wit.ai, Dialogflow.api, Google NL, LUIS, Watson Conversation Service; (4) collaborative research by key industry partners, Microsoft, and Google – discussed later (Gao, Galley, and Li 2019) .

A long-standing issue within conversational AI systems is refining the accuracy of the interpretation of meaning to provide a realistic dialogue to support the human-to-computer communication, conversational agents (CA), a class of dialog systems (DS) (Radziwill and Benton 2017), have been a subject of research in communications for decades. Interactive Voice Response (IVR) systems (e.g. “Press or Say 1 for English”) are also DS, but are not usually considered CA since they implement decision trees (McTear, Callejas, and Griol 2016) . There are two streams of CA development identified by O’Shea, Bandar, and Crockett (2010) - ‘embodied’ CSAs and ‘linguistic’ CSAs. Embodied CA’s may have an animated humanoid body and facial expressions; use of speech and gestures with scope for deeper relationships and human collaboration. From current AI and NLP resources it is evident that researchers use terms CSAs interchangeable based on the approaches used, which are categorised into chatbot-based systems, NLP-based dialogue management systems and goal-oriented (O’Shea 2010), and hence, generalised into **two** types of conversational AI: task-oriented chatbots, alternatively non-task-oriented CA. The former refers to completing tasks for the user, from train scheduling to restaurant reservation – hence software systems that mimic interactions with real people, and the latter, engaging the user in the human-computer-interaction in an open domain conversation (Yan 2018).

3.2. Conversational AI environment, building and examples

Chatbot and conversational AI environments generally have quite similar technologies and architectures. Since the introduction of smartphones and mobile applications, apps for short, the term chatbot is mostly used for messenger apps rather than for pure computer programs (AbuShawar and Atwell 2015). With this innovation, Feldman (2018) confirms there are three models of chatbot and conversational AI interaction and integration emerge: (1) chat as a layer – ubiquitous resource – customer support; (2) chat as a pillar – messaging core part of UI – conversations and then conversation with a human agent; (3) chat as a backbone – uses today’s messaging platform (via dialogs, templates and web review).

For clarity, conversational interfaces are very popular with virtual/personal assistants (PA) to automate tasks such as scheduling meetings; reporting weather and requesting songs. PAs include: Amazon Alexa, Apple Siri, Google Home, and Microsoft Cortana, are task-oriented - ‘unbound’, ‘stateful’; bots, used generally for home use and outside our scope and also use ML

and DL. There is a variety of chatbots. (1) Simple chatbot – E-commerce, geared to help users shop on an e-commerce website, greeted with ‘Hello’ and subsequent stages with heavy lifting in the background, but domain limited. Other examples to book a flight: Also known as softbots or sitebots or customer service bots/enterprise agents to either replace or augment customer service and process repetitive tasks like an FAQ, and routine sorts, using programmed rules and logic, but non-stateful; and Linked to Facebook Messenger, Slack, and WeChat. (2) Rule based – Medical AI⁴ as illustrated in Figure 8 – chatbot architecture. (3) Challenging – whereby not all leads are created equal, and getting the right leads in front of the right representatives at the right time is a lot more challenging (ongoing development) for Roof AI – real estate marketers using Facebook.⁵ Chatbots receive NL input, sometimes interpreted through speech recognition software, and execute one or more related commands to engage in goal-directed behaviour (often on behalf of a human user).

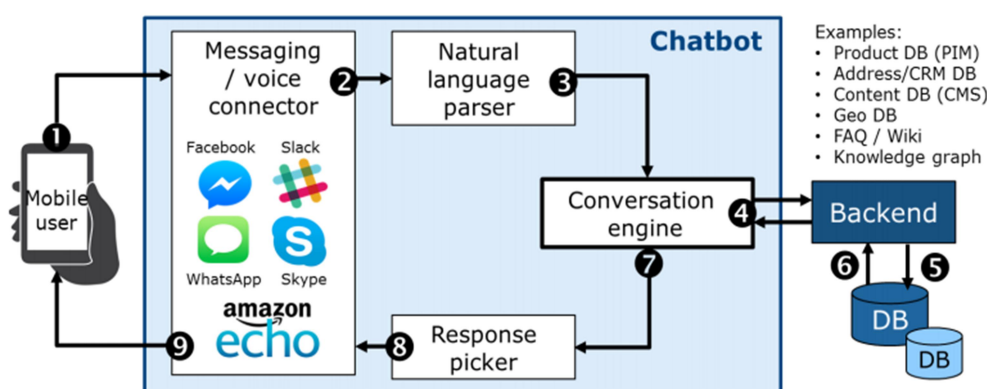


FIGURE 8. CHATBOT ARCHITECTURE (ZUMSTEIN AND HUNDERTMARK 2017, 98).

3.3. Conversational AI thinking unpacked

Over the decades, there has been the good, bad and ugly of conversation devices (Grunwitz 2017), which has inspired the Loebner prize contest for a human-like chatbot with Mitzuki being a 4-time winner (Worswick 2018). The goal here is high level cognition – not only for humans to exhibit intelligence - the capability to recognise concepts, perceive objects, or execute complex motor skills BUT more the capability to engage in multi-step reasoning to understand the meaning of NL, to design innovative artefacts, and to generate novel plans that achieve goals, and to reason about their own reasoning (Peebles 2012) .

One good example to understand the journey of conversation building is the following: (i) *I am hungry*; (ii) *Is there any food in the fridge?* (iii) *My stomach feels empty*; (iv) *I can't work without a happy belly*. (i) to (iv) means the same - *I am hungry*. But every utterance defines the hunger situation in different ways. For a human mind, it is easy to perceive the idea behind every sentence. An artificial system will not understand the hidden meaning behind these sentences. To make it understand, a process is required to define, calibrate and feed in the meanings of such words (Biswas 2019). He recommends: (i) definition modelling which is similar to a mental dictionary

⁴ <https://medwhat.com/>

⁵ <https://roof.ai/>

in our head or a lexicon in linguistic terms; (ii) contextual awareness - we tell the system what a combination of words will mean in a certain context or scenario. Contextual awareness is the understanding of defining the flow of a conversation to a certain result; (iii) understanding the domain and to judge the intention; (iv) error handling benchmarks in the system. With this in mind, there have been developments with principled and data-driven approaches to build open domain conversational systems due to the benefits from the large scale social conversation data publicly available, and the rapid progress of employing ML (often Markov chains or deep NN) to be able to adapt to new information or new requests (Yan 2018). However, how well is this being achieved for retrieval of information and conversation with multi-turns?

3.4. Challenge: retrieval based vs generation based conversational AI

Conversational AI has a general paradigm composed of a user has a query – which forms an utterance – for which the computer must respond with a response. The response can be either from a repository (retrieval based systems) or organising tokens and words to synthesis new responses – that is a generation based system, whereby the query has context, and last for multiple turns – as the context is based on previous utterances with the conversation session (Yan 2018). Note the algorithms involved are not discussed, but the concepts involved and the nature of complexity, and how to measure it. As a recap from section one, we mentioned shallow representations using one-hot representation of words – vectors, and discussed the sequence-to-sequence model - of which we will continue to look at the current state of deep neural network as automatic learning machines: “they can extract underlying abstract features of data automatically by exploring multiple layers of non-linear transformations” (Yan 2018).

3.4.1. Generation based AI

One way to build a conversational system is to use language generation techniques, as experimented by Ritter, Clark, and Etzioni (2011) in the feasibility of conducting short text conversation by using statistical machine translation (SMT) techniques, learning from millions of naturally occurring conversation data in Twitter, but demonstrating a limited guarantee of legitimate NL text at all times using DL (Yan, Song, and Wu 2016). However, these techniques have become more advanced introducing multi-turn, and trying different ways to use contexts and gradually introducing additional elements such as personas, knowledge and topics from Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18) 5521 on the generative framework. It used the encoder-decoder framework using sequence-to-sequence model for encoding inputs to hidden representation using either LSTM or GRU units, and further the attention mechanism and parameterised softmax function, as highlighted in section one.

Going back to the retrieval based responses which reflect mainstream chatbot systems, the responses are literally created by humans. The sentences are fluent and natural, and hence regarded to be reliable in practice, as there are a large number of utterance pairs, but the system cannot create new appropriate candidates. Comparative to generation based – they are flexible enough to create unlimited responses given a small vocabulary and a smaller training dataset – but limited – as it may not be natural and fluent - due to their over-generation or under-

generation (Yan, Song, and Wu 2016).

A third method is to take the advantage of both, by merging the candidates from both systems together – as an ensemble, using context, for single turn or multi-turn, and looking at the responses.

3.4.2. Working with the different types of responses in conversation

Driven by data, trivial or non-committal responses are: ‘Me too’, ‘I’m OK’ and ‘I don’t know’. Li et al. (2015) highlights 0.45% of the utterances in the conversational data referred to ‘I don’t know’ - observing a high frequency of generic and meaningless responses – and relative sparsity of more informative alternative candidates, using a range of metrics such as maximum marginal relevance (MMR) and dominate point processes (DPP), and good discourse diversity (N-best lists), and using a Stalemate Breaker concept (when the speaker does not what to say or how to say) (Li et al. 2016). The system detects when a stalemate occurs, and determines what intriguing contents to introduce by re-ranking candidate responses. The content of the candidate responses uses a backward-and-forward (B/F) language modelling algorithm – starting the sentence with the designated algorithm. This method is applied to conversations as a sequence-to-B/F (Seq2BF) model (Mou et al. 2016) After encoding the input utterance, the neural generator decodes from introduced content word (s) to respond to the input utterance.

These content words exist explicitly in the utterance using the method of generative Sequence-to-Sequence (Seq2Seq) model through neural networks, which provide trivial responses (Yao et al. 2017). Alternatively, the semantics of the words are implicitly included in the utterance, using semantic information known as dialogue acts. This implicit content –introducing conversational model combine the standard decoder, a designated content encoder and a fusion decoder fused together to generate responses – compared to explicit way – provides software, a more relaxed and better flexibility in generation.

3.4.3. How accurate is the evaluation of DL conversational AI generation?

Automatic evaluation metrics can be divided into non-learnable and learnable approaches. Automatic evaluation is crucial for language generation tasks. Retrieval based conversational systems – use techniques such a mean average precision (MAP), and normalised Discounted Cumulative Gain (nDCG) – but a challenging for measuring dialogues. Metrics such as BLEU (Papineni et al. 2002) from IBM and METEOR (Banerjee and Lavie 2005) for machine translation and/or summarisation (ROUGE) (Lin and Och 2004) are used but provided insufficient evidence to evaluate conversations with context information, one-to-many diversity and word overlap and ground truth (Liu et al. 2016). Further Lowe et al. (2017) – propose Turing test for dialogue, which learns to predict a score of a reply given its query (previous user issued utterance) and a ground truth reply. Here an automatic dialogue evaluation model (ADEM) was trained in a semi-supervised manner using a hierarchical RNN but required human-annotated scores to train the network, being less flexible and extensible.

A later unsupervised technique requires no human scoring, but a metric with (1) a referenced part to measure the overlap between the system response and the ground truth, and (2) an

unreferenced part to measure the correlation between the system response and the query utterance – as identified in Figure 9. RUBER (Referenced metric and Unreferenced metric Blended Evaluation Routine) evaluates a reply by taking into consideration both a ground-truth reply and a query (previous user-issued utterance). RUBER is flexible and extensible to different datasets and languages (Tao et al. 2018).

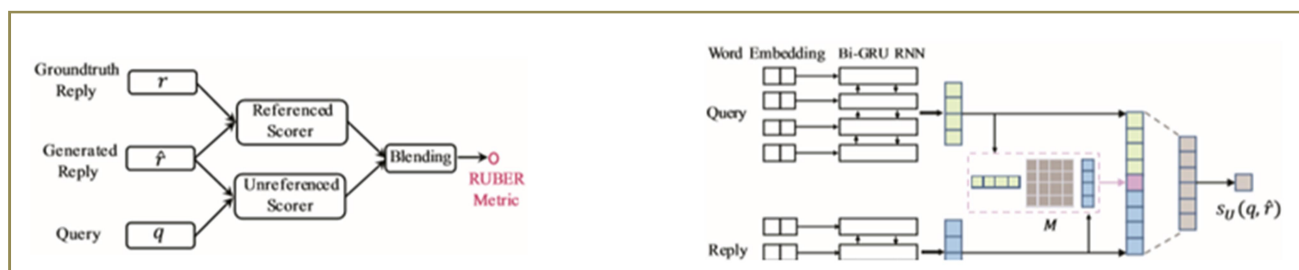


FIGURE 9. (LEFT) OVERVIEW OF RUBRIC METRIC; (RIGHT) NN PREDICTS UNREFERENCED SCORE (TAO ET AL. 2018, 3).

RUBER includes no human annotation, and based on a corpus created to train embeddings and neural scorers for the query-reply data. It compares the average human satisfactory score of a ground-truth reply between the reply and its query. Important elements: generated reply \hat{r} and a ground-truth r as a referenced metric with maximum vector pooling. Unreference metric - measures the relatedness between the generated reply \hat{r} and its query q . This metric, denoted as $s_U(q, \hat{r})$, is unreferenced because it does not refer to a ground-truth reply. Differently from the $r - \hat{r}$ metric, which mainly measures the similarity of two utterances, the $q - \hat{r}$ metric in this part involves more semantics, as highlighted in Table 2.

TABLE 2. RUBER - QUERY AND GROUND TRUTH/CANDIDATE REPLIES.

Query: Why not adopt one?
Ground truth: I hope so, but it's hard to find one
Candidate1: I'm looking for one from my friends
Candidate2: Pappilon is a rare breed
Candidate3: OK, thank you for your advice

A good response can either resemble the ground truth well or be closely related to the query utterance. Both referenced metric and unreferenced metric as a hybrid method RUBER for the evaluation of open-domain dialog systems, using various statistical analysis. For conversational AI – this focuses on single-turn conversation in Chinese, as a starting point, but RUBER suggests it can be used for more complicated context aware dialog systems – where the NN is designed taking the context into account, for the unreferenced metric.

In this section, we have seen that conversational AI design and development is very innovative, with lots of research effort and commercialised solution, but conversational AI - dialogued based has some positive results for single turn – but more challenging, in terms of response rates and evaluation metrics. This explains why we are seeing chatbots providing the customer the right journey for the right interaction – and where it is complex/challenging there is a blend of self-service and assisted service.

4. DISCUSSION OF STATISTICS NLP VS LINGUISTIC NLP SOLUTIONS

4.1. Introduction

As a recap, NLP tasks include POS tagging, parsing, named-entity recognition, semantic role labelling, sentiment classification, machine translation, question & answering, dialogue systems, and contextualised embedding (Omarsar 2018). From a DL space convolutional neural networks (CNN) have a feature function that is applied to the words or n-grams to extract high-level features. The resultant features are used for NLP tasks such as sentiment analysis, machine translation and question answering. (Omarsar 2018). The pioneering work of CNNs for NLP tasks by Collobert and Weston (2008) is illustrated in Figure 10 and Figure 11, showing the CNN framework to perform word class prediction and modelling in text. It transforms words into a vector representation via a look-up table – producing a primitive word embedding approach that learns weights during the training network.

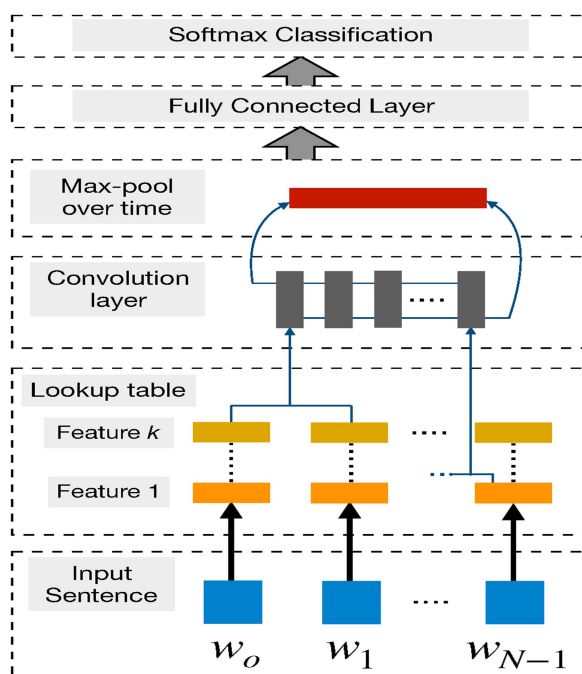


FIGURE 10. CNN FRAMEWORK WORD WISE CLASS PREDICTION (ZHANG AND WALLACE 2015).

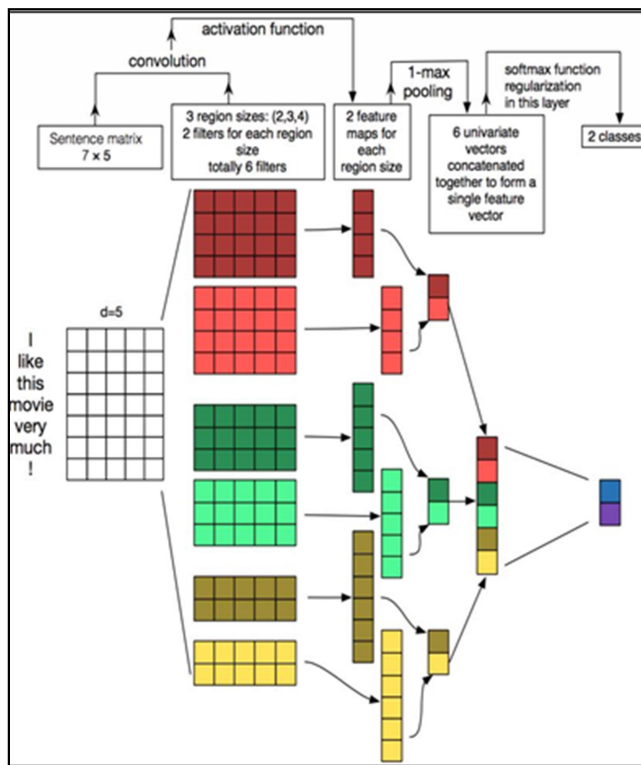


FIGURE 11. CNN MODELING ON TEXT (ZHANG AND WALLACE, 2015).

CNNs have been used for sentence classification (Kim 2014) short text categorization, and sarcasm detection via tweet sentiment classification and visualising ConvNets for text and using applied DCNN to map meanings of words that constitute a sentence to that of documents for summarization (Denil et al. 2014). One shortfall of CNNs was its inability to model long distance dependencies – as highlighted in Section 1. To solve this problem, they were coupled with time-delayed neural networks (TDNN) – which have an extended contextual range during training.

An RNN is effective for processing sequential information, and based on feedforward mechanism and models context dependencies. Key concept is that it can memorise the results of previous computations and use that information in the current computations – based on word embedding, and able to establish the difference in meaning between ‘dog’ and ‘hot dog’.

Compared to the CNN, the RNN is similarly effective, but they model different aspects of the data such natural language generation, semantic matching, word level classification (NER). Recursive neural networks are strongly motivated - as language has a recursive structure where words and sub-phrases compose other high-level phrases in a hierarchy, and consider terminal nodes. It is useful for parsing; semantic relationship, classification such as a topic and semantic relatedness. In summary, DL model utilise internal memory (bottom-up knowledge learned from the data) enriched with an external memory (top-down knowledge inherited from a KB) (Omarsar 2018).

4.2. Latest research: from NLP to the neural approach

Microsoft and Google have jointly worked on conversational AI in three categories: (1) question

feature extractors, learning capabilities – more powerful than shallow/classical ML. Statistical NLP applies linguistic knowledge to ML and DL via mathematical rules. ML’s goal is to enable computers to learn on their own. An ML’s algorithm enables it to identify patterns in observed data, build models that explain the world, and predict things without having explicit pre-programmed rules and models. Linguistic NLP – concerned with language formation, syntax, semantics, meaning, knowledge representation, named entity recognition, word sense disambiguation (WSD), co-reference resolution etc. via grammatical based computational structures.

4.4. Role and Reference Grammar (RRG) + Patom Theory (Patterns) vs Mathematical analysis via ML (PatInc)

The Patom theory postulates that all a brain can do is to store, match and use patterns. The assumption is that a pattern cannot be divided beyond a certain point, and as it is indivisible, it is called an atom. Role and Reference Grammar (RRG) (Van Valin and Foley 1980; Van Valin and LaPolla 1997) takes language to be a system of communicative social action, and accordingly, analysing the communicative functions of grammatical structures plays a vital role in grammatical description and theory from this perspective. Language is a system, and grammar is a system in the traditional structuralism sense; what distinguishes the RRG conception of language is the conviction that grammatical structure can only be understood and explained with reference to its semantic and communicative functions. That's why it has the name pattern atom, or Patom (Ball 2019). Figure 13 illustrates the consolidation and semantic set for representing a clause.

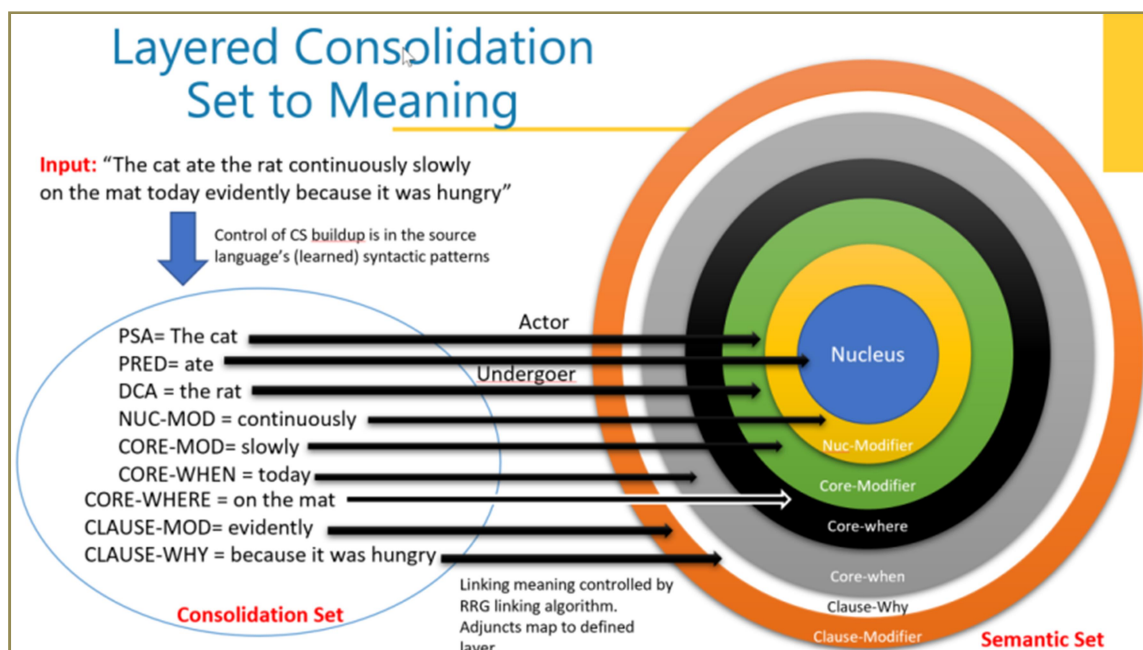


FIGURE 13. CLAUSE REPRESENTATION (BALL, 2019).

4.4.1. Linguistic phenomena

Consider the combinations based on these layers for a simple referent phrase (RP) - cat: *the cat ate the rat continuously slowly on the mat today evidently because it was hungry* in Figure 2. Limiting: the number of arguments = 2 two, for simplicity. The number of distinct word sequences in the model excluding ‘reasons’ is:

$$\text{PSA} \times \text{PRED} \times \text{DCA} \times \text{NUC-MOD} \times \text{CORE-MOD} \times \text{CLAUSE-MOD} \times \text{WHEN} \times \text{WHERE} \\ = 1.26 \times 10^{32} \times 60,000 \times 1.26 \times 10^{32} \times 2000 \times 2000 \times 1000 \times 3.2 \times 10^{12} \times 1.89 \times 10^{34} = 2.3 \times 10^{125}$$

If we change the above to a complex referent phrase – using ‘demonstrative’: *“I saw **the rat** THAT the cat ate continuously slowly on the mat today evidently”*.

Here the RP is replaced, with new PSA and DCA of 2.3×10^{125}

$$\text{WHERE} = 150 \times (1.26 \times 10^{32}) = 1.89 \times 10^{34} \rightarrow 150 \times 2.3 \times 10^{125} = 3.46 \times 10^{127}$$

$$\text{PSA} \times \text{PRED} \times \text{DCA} \times \text{NUC-MOD} \times \text{CORE-MOD} \times \text{CLAUSE-MOD} \times \text{WHEN} \times \text{WHERE} \\ = 2.3 \times 10^{125} \times 60,000 \times 2.3 \times 10^{125} \times 2000 \times 2000 \times 1000 \times 3.2 \times 10^{12} \times 3.46 \times 10^{127} = 1.4 \times 10^{405}$$

Add more complexity – with more arguments at clause level and phrase level - as in the next case. English motion predicates (predicates that describe a manner of motion) can accept and combine with 5 types of secondary predicates describing the motion and path further. Examples from Ball (2019):

- 1. Goal/destination: e.g. “She walked” + “**TO the store**”
- 2. Source: e.g. “She walked” + “**FROM the airport**”
- 3. Trajectory/target: e.g. “She walked” + “**TOWARDS the tower**”
- 4. Path: e.g. “She walked” + “**down the road**”
- 5. Directional: e.g. “She walked” + “**away**”

A visual example of this: *“The sergeant marched the soldiers slowly up **from** the garden down the road towards the church **to** the airport in the city evidently”*. In each case, the example secondary motion predicate is bolded. Technically, the words “TO”, “FROM” and the others are all 2-argument predicates, except for the directional, in which one argument is the traditional object of the preposition and the other is the actor of the first predicate “walked” i.e. “she.” The result of using all of these motion predicates at the same time is bordering on incomprehensible, but it saves showing 5 different phrases for a simple analysis. This combines the predicate-level components in the nuclear and core layers of the RRG model. The RRG layered model, of course, extends the possible constituents with clause and sentence level elements.

4.4.2. Further Mathematical Analysis (Ball 2019)

Ball (2019) further adds 284 predicates that extend motion, using categories of target, goal/destination, source, path and directional. He creates a consolidation set (CS), syntactic set and semantic set (SS) for RRG analysis. For example, *“The sergeant marched the soldiers slowly up from the garden down the road towards the church to the airport in the city evidently”*. Here are the labelled elements (RRG elements are in bold):

**PSA PRED DCA ADV-CORE DIRECTIONAL SOURCE PATH TARGET GOAL
POSITION-WHERE ADV-CLAUSE**

Looking at only the combination of words in the arguments, we utilise some metrics: PSA (like subject) and DCA (direct core argument, like object) are both RPs. The ‘where’ case above for the number of word sequences for the 4 phrases (and a value of 30 for the directional).

$$2.74 \times 10^{1215} \times 284 \times 2.74 \times 10^{1215} \times 30 \times 3.46 \times 10^{127} \times 3.46 \times 10^{127} \times 3.46 \times 10^{127} \times 3.46 \times 10^{127} = 9.16 \times 10^{2944}$$

The clause level looks like:

$$\text{PSA} \times \text{PRED} \times \text{DCA} \times \text{NUC-MOD} \times \text{CORE-MOD} \times \text{CLAUSE-MOD} \times \text{WHEN} \times \text{WHERE} \\ = 2.74 \times 10^{1215} \times 60,000 \times 2.74 \times 10^{1215} \times 2000 \times 2000 \times 1000 \times 3.2 \times 10^{12} \times 3.46 \times 10^{127} = 1.99 \times 10^{2585}$$

Here the linguistic models add exponential knowledge, but the use of ML models to process further is questionable. The NLP with Deep Learning expert Yoshua Bengio, a recent Turing Award winner, was asked about the “...main problem in current NLP systems”. He said “state-of-the-art NLP systems, e.g., the best translation systems using deep learning and attention, still make very many stupid mistakes that no human will make. The reason is that machines don’t really understand what those sentences mean, they do not comprehend anything. And the problem is that we are training current NLP deep learning systems using only texts, but it is not easy to find unconscious knowledge about the world just by reading texts” (Synced 2019). Ball (2019) states “unconscious knowledge” is somewhat undefined, and referred to as “meaning,” an element studied intensively by linguists in conjunction with discourse pragmatics and context.

The problem here is that the conversational agent requires a range of NLP qualities in a human-machine interface of understanding context, applying logic, utilising NL understanding, understanding the intent, explainable and as dialogue systems requires story comprehension, and the formulation of a response and able to learn and adapt (Panesar 2017). Further still not only from exponential knowledge, but “unconscious knowledge” which is somewhat undefined, and referred to as “meaning,” an element studied intensively by linguists in conjunction with discourse pragmatics and context (Ball 2019).

There are various popular chatbot platforms such as Microsoft Bot Framework, IBM Watson, Chatfuel, Pandabots and Wit.ai to name a few, that will enable a competent developer to create a basic bot in a few minutes. However, to create a robust conversational chatbot, it will be challenging as noted in section one that natural language is inherently ambiguous, complex, and dynamic, and conversation is not linear. There is a lot of hype about creating chatbots and conversational interfaces but creating human-like conversation is big challenge in both conversational AI and other conversational systems. Ball (2019) states “the state-of-the-art in AI today is merely a souped-up version of what machines could already do a generation ago: find hidden regularities in a large set of data.” This viewpoint aligns with the current state of AI, when discussing human intelligence and machine intelligence in another interview with Yoshua Bengio who says “but (the progress) is mostly about perception, things like computer

vision and speech recognition and synthesis of some things in natural processing. We're still far from human capabilities."

4.5. Impact of chatbots

AI ethics have also been questioned, as in 2016, when a Twitter chatbot known as Tay, developed by Microsoft, was able to "learn" the nuances of human conversation by monitoring and interacting with real people online. The bot was at the mercy of the algorithms, with a "repeat back to me" function and was eventually taken down (Shewan 2020). Here, AI can no longer be ignored, as the potential benefits far outweigh any negatives to the point we may not have much choice in the matter – and risk losing what you have built – or introduce integrity and ethics in the data collection stage, knowledge creation and inference.

5. REVIEW OF LINGUISTIC EXPERIMENT WITH LING-CSA

5.1. Introduction to LING-CSA

Just to iterate our problem: what computers find hard is processing an utterance which is implicit, high contextual, ambiguous and often imprecise (McCord, Murdock, and Boguraev 2012). The motivations of a text-based linguistically motivated conversational software agent (LING-CSA) have been addressed in (Panesar 2019a). To set the scene, for this related work, two main requirements for a CSA specified by Mao, Sansonnet, and Li (2012) and Lester, Branting, and Mott (2004) are the ability of accurate NLU from engaging in conversation and the technical integration with an application. A CSA must respond appropriately to the user's utterance via three phases: (1) interpret the utterance, (2) determine the actions (logic) that should be taken in response to the utterance (3) and present a grammatically correct response. The goals of a full-fledged NLP system are: (1) paraphrase an input text, (2) translate the text into another language (out of scope), (3) answer questions about the contents of the text, and (4) draw inferences from the text. There has been good progress on (1-3), but (4) refers to NLU – a more challenging problem which requires a deeper linguistically knowledge aware architectural solution with two goals:

1. Investigate the *integration, intersection and interface* of the language, knowledge, and speech act constructions (SAC) based on a grammatical object (Nolan 2014), and the sub-model of belief, desires and intention (BDI) (Rao and Georgeff 1995) and dialogue management (DM) for NLP.
2. A long-standing issue within NLP CSA systems is *refining the accuracy of the interpretation of meaning* to provide a realistic dialogue to support human-to-computer communication. As noted in Panesar (2019a), there are several statistical, probabilistic and linguistic based computational structures, formal language theories and grammars used in NLP and NLU, as well as links to cognitive science. For LING-CSA, a linguistic theory known as Role and Reference Grammar (RRG) by Van Valin Jr (2005) was deployed specifically focusing on the syntax-semantic interface. The reasoning behind this is that RRG's major influence on layering and its ongoing developments such as detailed rules for mapping from semantic structures to

syntactic structures (needed in modelling language production) and from syntactic structures to semantic structures (needed for comprehension), and its computational abilities and applications have provided some fruitful results (Salem, Hensman, and Nolan 2008). RRG is a mature functional grammar - that can adequately explain, describe and embed the communication-cognitive thinking in conversation, in a computational form. With this in mind, we propose a language model, knowledge model, speech act constructions (SAC) and belief, desires, and intentions (BDI) framework for a linguistically motivated text based CSA namely LING-CSA. It will further investigate how language can be comprehended and produced, to gain a deep understanding, and how it interfaces with knowledge. In order to use the RRG linking system and to create an effective parser, RRG is re-organised to facilitate the innovative use of SAC as a grammatical object (Nolan 2014) and rich lexicon, as illustrated in Figure 14 (Panesar 2019a).

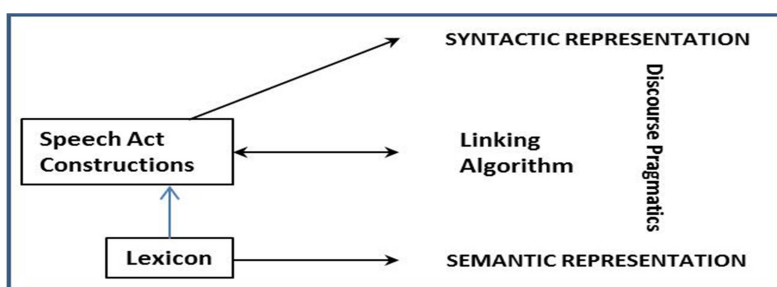


FIGURE 14. RE-ORGANISATION OF RRG FOR THE CSA (PANESAR 2019A).

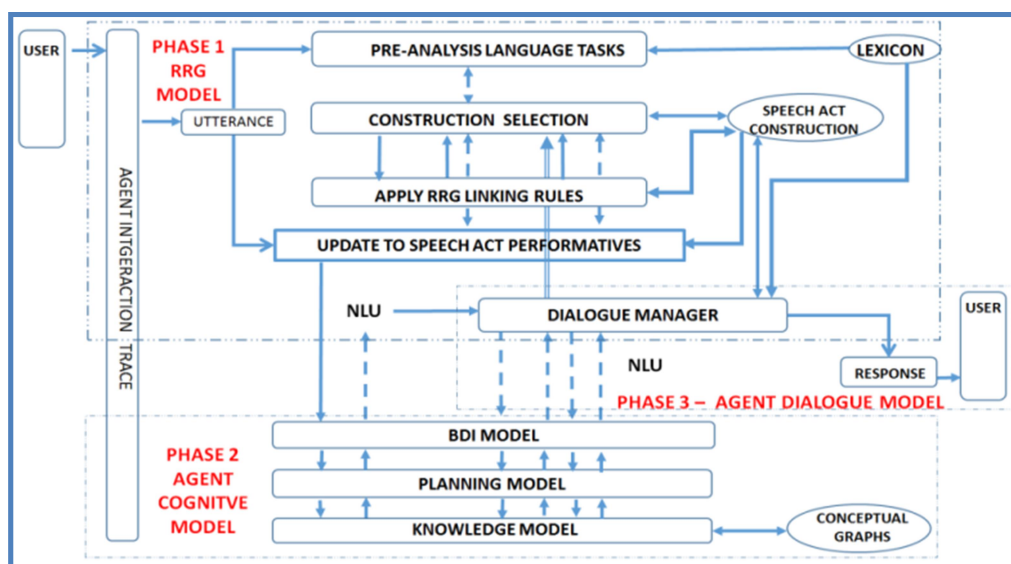


FIGURE 15. CONCEPTUAL ARCHITECTURE OF THE LING-CSA (PANESAR 2019A).

LING-CSA constitutes a three-phase model, as illustrated in Figure 15, which identifies: (1) RRG language model, (2) Agent Cognitive model (ACM), with two inner models, i.e. knowledge model and planning model, and (3) Agent Dialogue Model (ADM), where the DM is a common component of Phase 1 and Phase 3 – due to the discourse referents of an utterance and the need to create a grammatically correct response.

5.2. RRG Language Model (Phase 1)

The LING-CSA will constitute a closed domain - dialogue on food and cooking, for its richness and universality. A range of NLP tasks manipulated via the language model will include tokenisation, sentence splitting, part-of-speech-tagging, morphological analysis, syntactic and semantic parsing. It is the DM that will assist with the SA dialogue working with the syntactic parser to work effectively with the data sources, e.g. word list, lexicon, empty speech acts constructions (SACs). RRG manipulations exist for both simple and complex sentences.

The LING-CSA is limited to the manipulation of the linking system for simple sentences (active or passive) accepting transitive, intransitive and ditransitive verbs/auxiliary verbs with variable word order flexibility following SVO (subject-verb-object), and modelled via speech acts Searle (1969), applied to cognitive and rational interaction analysis. RRG's bi-directional linking algorithm and discourse-pragmatics interface will be mapped into completed SACs via computational processes such as invoking the lexicon (Table 3). Here the logical structures have been partially expanded for readability in this paper and DNA refers to 'does not apply'.

TABLE 3. SNAPSHOT LEXICON - LEXICAL ENTRIES & LEXEMES (DATA SOURCE 1).

Lexical entry 1: ate								
POS TYPE	VERB TENSE - ASPECT	DEF	PERSON TYPE	NUMBER	GENDER	CASE	ANIM	HUM
Verb	PST	DEF +/-	3	SG	M/F	DNA	ANIM	HUM
LOGICAL STRUCTURE (LS) : <tns:pst <do'(x, [eat'(x, y)]) & BECOME consumed'(y) >>								

Lexical entry 2: eat								
POS TYPE	VERB TENSE - ASPECT	DEF	PERSON TYPE	NUMBER	GENDER	CASE	ANIM	HUM
VERB	PRS/ FUT	DEF+/-	3	SG	M/F	DNA	ANIM	HUM
LOGICAL STRUCTURE (LS) : <tns:prs <do'(x, [eat'(x,y)]) & BECOME consumed'(y) >> <tns:fut <do'(x, [eat'(x,y)]) & BECOME consumed'(y)>>								

Lexical entry 3: is								
POS TYPE	VERB TENSE - ASPECT	DEF	PERSON TYPE	NUMBER	GENDER	CASE	ANIM	HUM
VERB BE	DNA	DEF+	DNA	DNA	DNA	DNA	DNA	DNA
LOGICAL STRUCTURE (LS) : be'(x,[pred'])								

The steps of the RRG model support all three LING-CSA phases, and internal BDI manipulations. In Phase 1, these include: (Step 1) pre-analysis, (Step 2) selection of the speech act construction (SAC), (Step 3) application of the syntax-to-semantics algorithms; (Step 4) application of the semantic-to-syntax algorithm; (Step 5) anaphoric resolution and dialogue management; (Step 6) create a speech act performative message to input to the agent cognitive model (Panesar, 2017). The analysis of utterance/responses will be constrained to assertive and interrogative (WH-word) speech acts. For example, empty SACs are illustrated in Figure 16: 'assertive: ATE', which is updated with all the grammatical information, i.e. (a) voice opposition,

(b) macro-role, (c) pragmatic, (d) semantic, (e) lexical, (f) the PSA (privileged syntactic argument) - similar to the 'subject' of SVO (subject-verb-object), (g) the matching signature (matching POS tags of the utterance/response), (h) syntactic morphology (rules), (i) focus, and (j) semantic information such as tense and aspect.

```

ASSERTIVE:ATE RRG [NP VERB NP], [PN VERB], [ADV PN VERB DET N], [PN VERB N ADJ], [PRP DET N PN VERB
DET N], [PN VBE VERB N], [PN PRP DET N PRP DET N], [PRO VERB DET N], [PN VERB NP], [PN VERB DET N], [NP
VERB QNT N], [DET N VERB DET N], [DET N VERB QNT N], [NP VERB (DET) (ADJ) N (ADJ)], [PN VERB DET N
ADJ], [PN VERB (DET) ADV N ADJ], [PN VERB DET N PRP DET N], [PN, VERB, N, PRP, DET, N], [PN VERB N PRP
DET N] RRG NONE RRG UTTINPUT RRG WKSPACE RRG DEFAULT ASSUMPTION (1ST NP = 'ACTOR') RRG NO
PARTICULAR SPEC RRG NONE RRG CONTAINS A NOUN PHRASE BEFORE AND AFTER THE VERB RRG
DEFAULT RRG TRUE/FALSE RRG ASSERTIVE RRG NARROW FOCUS ON THE ELEMENT RRG LOG
STRUCTURE TO ADD

```

FIGURE 16. EMPTY SPEECH ACT CONSTRUCTIONS (EXAMPLE 1).

Here the updated SAC will use the generalised lexicon and computationally work with the surface syntax to the underlying semantic forms. Any discourse referents that are generated are also checked and updated accordingly. The updated SAC will store each text and the associated complete logical structure (LS) based on the working semantic predicating element. The linking system will facilitate the syntactic parse to facilitate word order for English (SVO), and to unpack the agreement features between elements of the sentence into a semantic representation (the logical structure (LS)) and a representation of the Layered Structure of the Clause (LSC). The linking system will facilitate procedures for semantic-to-syntax and syntax-to-semantics, parsing and in the process of formulating a grammatical correct response.

5.3. Design framework - Agent Cognitive Model (Phase 2), Implementation and Evaluation

Cohen and Levesque (1988) take the viewpoint of language as action, and view utterances as events, updated to speech act performatives (SAP) messages, that change the state of the world, and hence speakers and hearer's mental state change as a result of these utterances. Moreover, it is necessary for a computer program to recognise the illocutionary act (IA) of a SA, for both the speaker (USER) utterance and hearer (AGENT) and a response. The single agent environment will constitute a language model, mental model to work with the BDI states, working model of memory (to reason with states based on current knowledge), knowledge model (world knowledge – both shared and individual beliefs) and dialogue model (for a response).

The next question is how this will integrate and function - this is achieved in the Phase 2 – Agent Cognitive Model (ACM) with an input of the Phase 1 RRG Language Model. The ACM contains two inner models namely the Agent Knowledge Model (AKM) and Agent Dialogue Model (ADM). The ACM Phase 2 model has a series of pre-agent steps and main agent steps illustrated in Figure 17. The behaviour of the agent is identified by the following basic loop where agent iterates the following two steps at regular intervals: (1) read the current message and update the mental state including the beliefs and commitments, and (2) execute the commitments for the current time, possibly resulting in a further belief change. The preliminary steps (Panesar 2017, 2019a) involve: Step Pre-Agent (1) - Create Belief Base - the agent's knowledge, Step Pre-Agent (2) - Map to Message Format - the appropriate Speech Act Performative (SAP) (from the Phase 1- RRG Language Model) is re-mapped into a data structure to reduce redundancy and enable efficiency in use of the Agent Cognitive Model.

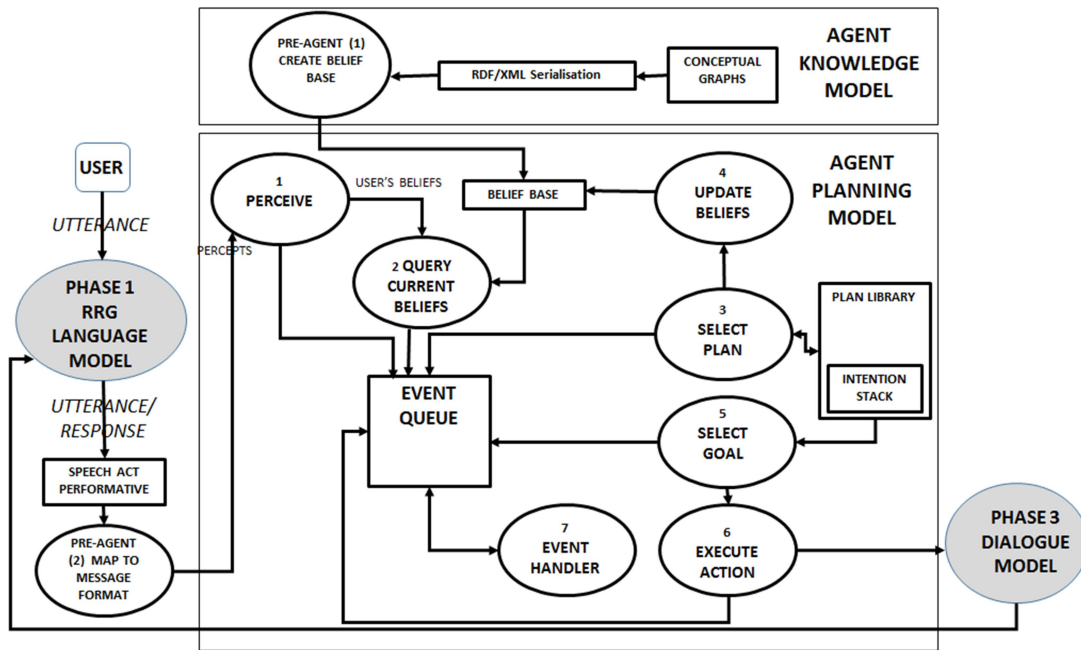


FIGURE 17. DESIGN FRAMEWORK - AGENT COGNITIVE MODEL (PANESAR 2019A).

Bratman (1987) identifies the workings of how people make decisions. LING-CSA has a planning model underpinned by BDI concepts (Wooldridge 2013) and intentionality (Searle 1983) and rational interaction (Cohen and Levesque 1990). The user instigates the utterance into the CSA framework, and it is the agent which decides on the sub-dialogue internal representations which are stored – based on theories and models such as the discourse representation theory (DRT) (Kamp, Van Genabith, and Reyle 2011). To facilitate conversation, the DM is invoked, and discourse referents in the previous utterance of the sub-dialogue are resolved employing common ground (Stalnaker 2002). This will serve two purposes: (1) to establish the NLU of the utterance, and (2) to forward to the dialogue model to ascertain a response. This response generator here will further make use of the language model, and SAC model, to formulate a grammatically correct response (Panesar 2019a).

The conceptual architecture of LING-CSA is implemented as a Java prototype, developed in Eclipse IDE platform, predominantly as POJO (plain old java objects), constituting a three-phase model. The evaluation of LING-CSA has been discussed at length in Panesar (2019b). She explains that the aim of LING-CSA is not to act as a human (Wilson 2017), but addresses the ISO 9241 concept of usability and performs the evaluation via a robust testing strategy that addresses the goals of RRG - explanatory, descriptive, cognitive and computational adequacies - with specific evaluation criteria for the three-phase models, as in Table 4 (Panesar 2017, 2019b).

TABLE 4. LING-CSA EVALUATION CRITERIA.

Criteria 1 - Could the system present a mapping of the syntactic representation to a semantic representation, for a simple utterance?
Criteria 2 - Could the system present an adequate explanation of the NLU of the utterance?
Criteria 3 - Could the system demonstrate the SAC use in the manipulation of the utterance?
Criteria 4 - Can the dialogue manager interface the language model?
Criteria 5 - Could the system demonstrate the agent BDI and knowledge representation?
Criteria 6 - Could the system represent the user's BDI states?

Criteria 7 – Could the system query the knowledge base for a fact (from the speech act performative (SAP))
Criteria 8 - Could the system devise an appropriate plan based on the BDI states?
Criteria 9 - Could the system generate a grammatically correct response in RRG based on the agent's knowledge?

```

<terminated> MainCAVersion30 [Java Application] C:\Program Files (x86)\Java\jre1.8.0_101\bin\javaw.exe (22 Jun 2
*****
Syntactic representation of this utterance >>>>>
SENTENCE ( <LDP> ( CLAUSE ( <CORE> <NP> they ( <NUC> ( <PRED> <AUX> are ) )
ADJ hungry ) )
*****
          Speech Act Performative
*****

:::Performative =SAP ASSERTIVE are :::Sender =<USER>:::Receiver<AGENT:::ontology
= :::Signature =[PRO VBE ADJ]:::Constraint =DEFAULT:::Input =they are
hungry:::Workspace =[they, PRO], [are, VBE], [hungry, ADJ]:::Syntax =SENTENCE (
<LDP> ( CLAUSE ( <CORE> <NP> they ( <NUC> ( <PRED> <AUX> are ) ) ADJ hungry
) ) ::: PSA =they:::SemanticsRRG =NONE:::Linking =CONTAINS A NOUN PHRASE BEFORE
AND AFTER THE VERB:::Morphology =DEFAULT:::Pragmatics =TRUE/FALSE:::IllForce
=ASSERTIVE:::FocusStructure=NARROW FOCUS ON THE ELEMENT:::OutputLS
=<IF>ASS<TNS><PRT> be'(they,[hungry'])

```

FIGURE 18. LSC AND LS REPRESENTATIONS OF THE UTTERANCE *THEY ARE HUNGRY* (PANESAR 2017).

Figure 18 demonstrates an example of an utterance (i.e. *They are hungry*) with both LSC and LS representations. The main evaluation point relevant to this paper is that that RRG is fit-for-purpose as the linguistic engine for LING-CSA, and it explored the gap between language and knowledge, in the agent space. Here, language is in the context of a RRG linguistic representation against the RDF Triples representing the agent belief base posed the key obstacle of interoperability of low level mapping and representation – with a lexical bridge solution proposed due to the ontology semantics demonstrating morphological, lexical and syntax gaps and connotations (Panesar 2019b). To summarise, our experiment demonstrates that language is a communicative action that references knowledge. But for effective sustained dialogue, common ground and the reformulation of this knowledge is achieved by a common conceptual lower-level bridging interface. This framework will facilitate the interactions with this knowledge (world) and to further to make changes to this world, as per conversational behaviour.

6. CONCLUSIONS

In this paper, the main goal was to explore conversational artificial intelligence (AI) by detailing statistical solutions developed through ML by firstly looking under the hood and exploring these NLP advancements that facilitate a host of exciting popular solutions. A second goal was to review the nature of conversational systems and how the solutions work and how they are evaluated. A third goal was to look at how knowledge is represented in statistical ML architectural models and the impact of increasing knowledge. A fourth goal was a discussion of statistical versus linguistic conversational solutions. A fifth goal was to focus on a linguistic experiment with a text based CSA using RRG – a strong functional grammar that has the ability to support the production and comprehension of sentences.

From this paper, we have derived a range of conclusions. (1) A chatbot can be made in 10

minutes – but will be limited on many fronts. There will be continued heavy investment in chatbot and conversational AI from key players (GAMFA) for effective customer experiences and competitive advantage. (2) Organisations are leveraging ML strategies, but acquiring the correct knowledge (and data) is critical for effective predictions of responses (unlike Microsoft Tay). (4) Human- and cognition-based thinking and ability is required to make a conversational AI effective. (5) LING-CSA is a great CSA proof-of-concept which has a rich RRG based linguistic engine having communication – cognition features and its ability to describe and explain computationally. A simple utterance is computed as a logical structure based on the layered structure on the clause, and provides an interpretation of the meaning of the simple sentence in English. LING-CSA includes multi-disciplinary efforts from cognitive science, linguistics, agent thinking, knowledge representation, software engineering and computational linguistics. It is insightful about the bridge between knowledge and language, and at an internal layer of conversational AI – the knowledge/language interface must be semantically aligned. (6) It is necessary to have a trade-off between a deeper NLP vs deeper intelligence logic via statistical exponential mathematical knowledge from DL methods which adopt inconsistent NLP evaluation metrics/benchmarks. (7) An ongoing conversational AI challenge is in the content creation, and story comprehension (Wallace 2018) which impacts knowledge inference.

REFERENCES

- AbuShawar, Bayan, and Eric Atwell. 2015. "ALICE chatbot: Trials and outputs." *Computación y Sistemas* 19 (4):625-632. <https://doi.org/10.13053/cys-19-4-2326>
- Alammar, Jay 2018. "Visualizing A Neural Machine Translation Model (Mechanics of Seq2seq Models With Attention)." accessed 12 December 2019. <https://jalammar.github.io/visualizing-neural-machine-translation-mechanics-of-seq2seq-models-with-attention/>.
- Bahdanau, Dzmitry, Jan Chorowski, Dmitriy Serdyuk, Philemon Brakel, and Yoshua Bengio. 2016. "End-to-end attention-based large vocabulary speech recognition." 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP). <https://doi.org/10.1109/ICASSP.2016.7472618>
- Ball, John. 2019. "The Problem with AI State of the Art Methodology." Pat Inc., accessed 1 June 2019. <https://medium.com/datadriveninvestor/the-problem-with-ai-state-of-the-art-methodology-db30762d4b84>.
- Banerjee, Satanjeev, and Alon Lavie. 2005. "METEOR: An automatic metric for MT evaluation with improved correlation with human judgments." Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization.
- Bhashkar, Kunal. 2019. "Conversational AI Chatbot using Deep Learning: How Bi-directional LSTM, Machine Reading Comprehension, Transfer Learning, Sequence to Sequence Model with multi-headed attention mechanism, Generative Adversarial Network, Self Learning based Sentiment Analysis and Deep Reinforcement Learning can help in Dialog Management for Conversational AI chatbot." accessed 12 November 2019.

- <https://medium.com/@BhashkarKunal/conversational-ai-chatbot-using-deep-learning-how-bi-directional-lstm-machine-reading-38dc5cf5a5a3>.
- Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python*: O'Reilly Media, Inc.
- Biswas, Sourav. 2019. "The Basics Of Conversation Designing in Chatbots." <https://medium.com/datadriveninvestor/the-basics-of-conversation-designing-in-chatbots-c15491bd3ec4>.
- Bourke, Daneil. 2018. "An In-depth Review of Andrew Ng's deeplearning.ai Specialization." *Towards Data Science*, accessed 12 October 2019. <https://towardsdatascience.com/an-in-depth-review-of-andrew-ngs-deeplearning-ai-speciliazation-99612991eb61>.
- Bratman, Michael. 1987. *Intention, plans, and practical reason*. Center for the Study of Language and Information: Stanford University.
- Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." *arXiv preprint arXiv:1406.1078*. <https://doi.org/10.3115/v1/D14-1179>
- Cohen, Philip R, and Hector J Levesque. 1988. Rational interaction as the basis for communication. DTIC Document.
- Cohen, Philip R, and Hector J Levesque. 1990. "Intention is choice with commitment." *Artificial Intelligence* 42 (2):213-261. [https://doi.org/10.1016/0004-3702\(90\)90055-5](https://doi.org/10.1016/0004-3702(90)90055-5)
- Collobert, Ronan, and Jason Weston. 2008. "A unified architecture for natural language processing: Deep neural networks with multitask learning." *Proceedings of the 25th international conference on Machine learning*. <https://doi.org/10.1145/1390156.1390177>
- Denil, Misha, Alban Demiraj, Nal Kalchbrenner, Phil Blunsom, and Nando de Freitas. 2014. "Modelling, visualising and summarising documents with a single convolutional neural network." *arXiv preprint arXiv:1406.3830*.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805*.
- Ekzarian, Daniil 2018. "Hack4th0n & RASA chatbot." 16 May 2019. <https://medium.com/reflash-programming-adventures/hack4th0n-rasa-chatbot-19c0778b4da9>.
- Feldman, Dave. 2018. "Chatbots: What happened?". <https://chatbotlife.com/chatbots-what-happened-dcc3f91a512c>.
- Gao, Jianfeng, Michel Galley, and Lihong Li. 2019. "Neural approaches to conversational AI."

Foundations and Trends® in Information Retrieval 13 (2-3):127-298.
<https://doi.org/10.1561/15000000074>.

Gartner. 2018. "Emerging Technology Trends in 2018." accessed 6 May 2019.
<https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/>.

Grunwitz, Kai. 2017. "Chatbot: the good, the bad and the ugly." 2 February 2018.
<https://twitter.com/KaiGrunwitz/status/1010025618220544000>.

Joulin, Armand, Edouard Grave, Piotr Bojanowski, Matthijs Douze, H erve J egou, and Tomas Mikolov. 2016. "Fasttext. zip: Compressing text classification models." *arXiv preprint arXiv:1612.03651*.

Kamp, Hans, Josef Van Genabith, and Uwe Reyle. 2011. "Discourse representation theory." In *Handbook of philosophical logic*, 125-394. Springer. https://doi.org/10.1007/978-94-007-0485-5_3.

Kim, Yoon. 2014. "Convolutional neural networks for sentence classification." *arXiv preprint arXiv:1408.5882*. <https://doi.org/10.3115/v1/D14-1181>

Lester, James, Karl Branting, and Bradford Mott. 2004. "Conversational agents." *The Practical Handbook of Internet Computing*, CRC Press, UK.

Li, Jiwei, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2015. "A diversity-promoting objective function for neural conversation models." *arXiv preprint arXiv:1510.03055*.
<https://doi.org/10.18653/v1/N16-1014>

Li, Xiang, Lili Mou, Rui Yan, and Ming Zhang. 2016. "Stalematebreaker: A proactive content-introducing approach to automatic human-computer conversation." *arXiv preprint arXiv:1604.04358*.

Liddy, Elizabeth D. 2001. "Natural language processing." In *Encyclopedia of Library and Information Science*. NY: Marcel Decker, Inc.

Lin, Chin-Yew, and FJ Och. 2004. "Looking for a few good metrics: ROUGE and its evaluation." Ntcir Workshop.

Liu, Chia-Wei, Ryan Lowe, Iulian V Serban, Michael Noseworthy, Laurent Charlin, and Joelle Pineau. 2016. "How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation." *arXiv preprint arXiv:1603.08023*.
<https://doi.org/10.18653/v1/D16-1230>.

Lowe, Ryan, Michael Noseworthy, Iulian V Serban, Nicolas Angelard-Gontier, Yoshua Bengio, and Joelle Pineau. 2017. "Towards an automatic turing test: Learning to evaluate dialogue responses." *arXiv preprint arXiv:1708.07149*. <https://doi.org/10.18653/v1/P17-1103>

Mao, Xuetao, Jean-Paul Sansonnet, and Lin Li. 2012. "Textual Conversation Agent for Enhancing Attraction in E-Learning." *Proceedings of International Conference on Computer Science and Information Technology*.

McTear, Michael Frederick, Zoraida Callejas, and David Griol. 2016. *The conversational interface*. Vol. 6: Springer. <https://doi.org/10.1007/978-3-319-32967-3>

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems*.

Mou, Lili, Zhengdong Lu, Hang Li, and Zhi Jin. 2017. "Coupling distributed and symbolic execution for natural language queries." *Proceedings of the 34th International Conference on Machine Learning-Volume 70*.

Mou, Lili, Yiping Song, Rui Yan, Ge Li, Lu Zhang, and Zhi Jin. 2016. "Sequence to backward and forward sequences: A content-introducing approach to generative short-text conversation." *arXiv preprint arXiv:1607.00970*.

Nolan, Brian. 2014. "Constructions as grammatical objects : A case study of prepositional ditransitive construction in Modern Irish." In *Linking Constructions into Functional Linguistics: The role of constructions in grammar*, edited by Brian Nolan and Elke Diedrichsen, 143-178. Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.145.06nol>

O'Shea, J. 2010. "A Framework for Applying Short Text Semantic Similarity in Goal Orientated Conversational Agents." PhD thesis PhD thesis, School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University.

O'Shea, Karen, Zuhair Bandar, and Keeley Crockett. 2010. "A conversational agent framework using semantic analysis." *International Journal of Intelligent Computing Research (IJICR)* 1 (1/2). <https://doi.org/10.20533/ijicr.2042.4655.2010.0002>.

Omarsar, Elvis. 2018. "Deep Learning for NLP: An Overview of Recent Trends." accessed 1 May 2019. <https://medium.com/dair-ai/deep-learning-for-nlp-an-overview-of-recent-trends-d0d8f40a776d>.

Pan, Sinno Jialin, and Qiang Yang. 2009. "A survey on transfer learning." *IEEE Transactions on knowledge and data engineering* 22 (10):1345-1359. <https://doi.org/10.1109/TKDE.2009.191>.

Panesar, Kulvinder. 2017. "A linguistically centred text-based conversational software agent." Unpublished PhD thesis, School of Computing, Creative Technologies and Engineering, Leeds Beckett University.

Panesar, Kulvinder. 2019a. "CHAPTER FIFTEEN FUNCTIONAL LINGUISTIC BASED

MOTIVATIONS FOR A CONVERSATIONAL SOFTWARE AGENT KULVINDER PANESAR." In *Linguistic Perspectives on the Construction of Meaning and Knowledge*, edited by Brian Nolan and Elke Diedrichsen, 340. Cambridge Scholars Publishing. <https://doi.org/10.4995/jclr.2019.11118>.

Panesar, Kulvinder. 2019b. "An Evaluation Of A Linguistically Motivated Conversational Software Agent Framework." *Journal of Computer-Assisted Linguistic Research* 3 (3):41-66. <https://doi.org/10.4995/jclr.2019.11118>.

Peebles, D. 2012. "Artificial Intelligence and Cognitive Systems." *The Newsletter of the Society for the Study of Artificial Intelligence and Simulation of Behaviour*, January 2012, 16.

Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. "Glove: Global vectors for word representation." Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). <https://doi.org/10.3115/v1/D14-1162>.

Radziwill, Nicole M, and Morgan C Benton. 2017. "Evaluating Quality of Chatbots and Intelligent Conversational Agents." *arXiv preprint arXiv:1704.04579*.

Ritter, Alan, Sam Clark, and Oren Etzioni. 2011. "Named entity recognition in tweets: an experimental study." Proceedings of the conference on empirical methods in natural language processing.

Ruder, Sebastian. 2016. "Highlights of NIPS 2016: Adversarial learning, Meta-learning, and more." accessed 12 December 2019. <https://ruder.io/highlights-nips-2016/index.html#thenutsandboltsofmachinelearning>.

Saifee, M. 2019a. "Recent Advancements in NLP (1/2)." *The Startup*, 1st December 2019. <https://medium.com/swlh/recent-advancements-in-nlp-1-2-192ac7eefe3c>.

Saifee, M. 2019b. "Recent Advancements in NLP (2/2)." *The Startup*, 1st December 2019. <https://medium.com/swlh/recent-advancements-in-nlp-1-2-192ac7eefe3c>.

Salem, Y., A. Hensman, and B. Nolan. 2008. "Implementing Arabic-to-English machine translation using the Role and Reference Grammar linguistic model." the 8th Annual Information Technology and Telecommunication Conference (IT&T 2008), Galway, Ireland.

Searle, John R. 1969. *Speech acts: An essay in the philosophy of language*. Vol. 626. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139173438>

Searle, John R. 1983. *Intentionality: An essay in the philosophy of mind*. Cambridge Cambridge University Press. <https://doi.org/10.1017/CBO9781139173452>

Sharda, R, Dursun Delen, and E Turban. 2019. *Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*. Eleventh ed: Pearson.

Shewan, Dan. 2020. "10 of the Most Innovative Chatbots on the Web." *WordStream*.

<https://www.wordstream.com/blog/ws/2017/10/04/chatbots>.

Singh, Seema. 2018. "Cousins of AI." Medium - Towards Data Science. <https://towardsdatascience.com/cousins-of-artificial-intelligence-dda4edc27b55>.

Stalnaker, Robert. 2002. "Common ground." *Linguistics and philosophy* 25 (5):701-721. <https://doi.org/10.1023/A:1020867916902>

Sundermeyer, Martin, Ralf Schlüter, and Hermann Ney. 2012. "LSTM neural networks for language modeling." Thirteenth annual conference of the international speech communication association.

Sutskever, I, O Vinyals, and QV Le. 2014. "Sequence to sequence learning with neural networks." *Advances in NIPS*.

Synced. 2019. "AI - Interview:Yoshua Bengio On AI Priors and Challenges." <https://syncedreview.com/2019/02/16/yoshua-bengio-on-ai-priors-and-challenges/>.

Tao, Chongyang, Lili Mou, Dongyan Zhao, and Rui Yan. 2018. "Ruber: An unsupervised method for automatic evaluation of open-domain dialog systems." Thirty-Second AAAI Conference on Artificial Intelligence.

Uszkoreit, Jakob. 2017. "Transformer: A Novel Neural Network Architecture for Language Understanding." *Google AI Blog*. <https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html>.

Van Valin Jr, Robert D. 2005. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511610578.001>.

Van Valin, Robert D, and William A Foley. 1980. "Role and reference grammar." In *Moravcik and Wirth*:329-352. https://doi.org/10.1163/9789004373105_014.

Van Valin, Robert D, and Randy J LaPolla. 1997. *Syntax: Structure, meaning, and function*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139166799>.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention is all you need." *Advances in neural information processing systems*.

Wallace, Richard. 2018. "Chatbots - a personal perspective." *Society for the study of Artificial Intelligence and Simulation of Behaviour (AISB) quarterly* (No.149):6.

Wilson, H. J., Daugherty, P. R., & Morini-Bianzino, N. . 2017. "Will AI Create as Many Jobs as it Eliminates?". MIT Sloan Management Review, accessed 13 March 2017. <http://sloanreview.mit.edu/article/will-ai-create-as-many-jobs-as-it-eliminates/>.

Wooldridge, Michael. 2013. "Intelligent Agents." In *Multiagent Systems*, edited by Gerhard Weiss, 3-50. USA: Massachusetts Institute of Technology.

Woolf, Joseph. 2019. "An introduction to Chatbots." accessed 12 March 2019. <https://josephwoolf.com/an-introduction-to-chatbots/>.

Worswick, Steve. 2018. "Mitsuku - Loebner Prize 2018." <https://www.pandorabots.com/mitsuku/>, accessed 12th September 2018. <http://www.mitsuku.com/>.

Yan, Rui. 2018. "'Chitty-Chitty-Chat Bot': Deep Learning for Conversational AI." *IJCAI*. <https://doi.org/10.24963/ijcai.2018/778>

Yan, Rui, Yiping Song, and Hua Wu. 2016. "Learning to respond with deep neural networks for retrieval-based human-computer conversation system." *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. <https://doi.org/10.1145/2911451.2911542>

Yao, Lili, Yaoyuan Zhang, Yansong Feng, Dongyan Zhao, and Rui Yan. 2017. "Towards implicit content-introducing for generative short-text conversation systems." *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. <https://doi.org/10.18653/v1/D17-1233>

Zhang, Ye, and Byron Wallace. 2015. "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification." *arXiv preprint arXiv:1510.03820*.

Zumstein, Darius, and Sophie Hundertmark. 2017. "Chatbots - an interactive technology for personalized communication, transactions and services." *IADIS International Journal on WWW/Internet* 15 (1).