Lexical access: Accessing the stored word forms

Aditya Sanjay Jagdale

Indian Institute of Technology, Kanpur

November 20, 2024

1 Abstract

In this review paper we study lexical access, which is important for understanding cognitive comprehension and recognition of words. This review paper explores various models that explain lexical access, including the Logogen Model, Frequency Ordered Bin Search (FOBS) Model, COHORT Model, TRACE Model, Distributed COHORT Model, and Simple Recurrent Network (SRN) Model. We will examine how different approach of models explain different aspects of lexical access. The key findings of the paper are that the logogen model is a simple threshold-based model. FOBS and COHORT model are more applicable as they use morphological decomposition and auditory features. TRACE and COHORT use interaction between semantic and phonological representation. SRN is neural network-based model for lexical access. This review paper also explains the limitations of various models and the scope of future improvement to build more robust models that take into context, phonological and semantic processing into consideration while addressing scalability issues.

2 Introduction

Lexical access is a fundamental cognitive process in language comprehension which is necessary for retrieval of words in memory during perception and production of language. It also plays important role in daily life as with the help of lexical access people can communicate effectively.

Understanding lexical access is a research topic in psycholinguistics which led to invention of models like FOBS, COHORT, TRACE, etc. In this paper we will review their strengths, limitations and implications. The goal of this review paper according to me is to appreciate the findings and areas of further research in lexical access for understanding language.

3 Logogen model

3.1 Explanation of the Logogen Model

Logogen model was given by John Morton in 1969 [2], is a bottom-up working model which explains the word recognition. This model takes spoken and visual input and use it to activate stored word form representations in our memory. The basic units of this model are called logogens, these are evidence-collecting devices with different thresholds. Each word is represented by a corresponding logogen, and words are recognized when its logogen's activation level crosses a specific threshold.

Logogen works by taking input from spoken words, written words, or context. The logogen system depends on these three types of inputs, and when individual logogens get activated more than their thresholds, they send signals to output buffer. If new input stops activating the logogen, a decay function reduces its activation back to baseline levels. This model assumes that the information is flowing in only one direction which is bottom-up manner, meaning that from input to activation of logogens, but not vice versa case. Also, there are no direct connections between and among the logogens themselves, so the activation levels of logogens does not affect each other.

The frequency effect[13] is also explained by the logogen model, according to which words that are more frequently encountered have lower thresholds, making them easy to recognize compared to word that are not much encountered. This model explains how words are recognized and processed in the brain, by considering the role of context and frequency in word recognition.

3.2 Computational Equation

A computational equation related to the logogen model involves the activation threshold of logogens. The activation level **A** of a logogen can be represented as:

$A = \sum_{i=1}^{n} I_i - D$

where I_i represents the level of input signals received by the logogen, and D is the decay function that returns the activation to baseline levels if no new input is received. This equation shows how the accumulation of input signals can lead to the activation of a logogen once the threshold is crossed.

3.3 Shortcomings of logogen model

The logogen model has various limitations as it is just a basic model. First one is that its assumption of strictly bottom-up information flow, which does not account for the 'interactive nature' of language processing where top-down processing also affects the lexical access. The model does not take into consideration the complexities in syntactic and grammatical structures which are present in language. The lack of connections between logogens also limits the model to explain the interactions between different words and concepts that occur in the mental lexicon.

4 Frequency order serial bin search model

4.1 Explanation of the Frequency Ordered Bin Search Model

The Frequency Ordered Bin Search (FOBS)[3] model is another model that explains how the words are recognized and accessed in the long-term memory. The key principles of the FOBS model are:

Word Form Representations: Word form representations are activated in bottom-up way by input, similar to logogen model. Inputs can be auditory or visual. Individuals search for the matching representation of the word form in their long-term memory.

Lexical Representations: The mental representations of words (i.e. lexical representations) are organized into different frequency-ordered bins. This means that words with higher frequency are stored in bins that are accessed more quickly.

Self-Terminating Search: The search process is self-terminating as once a matching representation is found, the search stops. This approach is more efficient since we don't have to search through all possible representations.

Morphological Decomposition: The FOBS model assumes that we perform morphological decomposition of word forms for lexical accessing. Words are recognized based on their shared roots (e.g., "blackboard" is recognized based on the root "black"). This allows for faster processing of words with high frequency shared root as compared to other low frequency words, as the listener only needs to find the correct bin on the basis of root, rather than the whole word form.

The evidence for the FOBS model comes from the experiment which show that root frequency (how frequently shared root occurs) better predicts response times than surface frequency (how frequently the exact word occurs). Additionally, it is observed that people find it harder to recognize words with pseudo-suffixes (e.g., "sister") compared to words with real suffixes (e.g., "grower"), suggesting morphological decomposition happens during lexical access.

4.2 Computational Equations and Research

Although the FOBS model does not rely on any specific set of computational equations, there are various research going on for computational implementation of similar models. One relevant study is by Baayen [14][1], who proposed a computational model of morphological processing based on the principles of parallel dual-route processing. In this model, lexical representations are organized into frequency-ordered bins similar to FOBS model, and word recognition involves the parallel activation of full-form representations and decomposed representations. The authors proved that this model could account for a various of findings, including the effects of surface and root frequency on response times.

Research by Seidenberg and Gonnerman [15] has explained the computational modeling of morphological processing using connectionist networks.

4.3 Shortcomings of the FOBS Model

The FOBS model simplifies how words are organized and accessed in our memory. While the FOBS model offers a basic approach, but it does not consider the factors such as semantic relationships, phonological similarity, and contextual information.

The FOBS model has been tested and developed based on single-word recognition tasks, so it might not work as well for understanding sentences or more complex language use in everyday situations.

5 COHORT model

5.1 Explanation of the COHORT Model

The COHORT model was proposed by William Marslen-Wilson in the 1970s[11]. COHORT model is a model for lexical access in spoken word recognition. According to this model the lexical access as has three stages: activation, selection, and integration. During the activation phase, word form representations are activated in response to the auditory stimulus. This activation is an independent process, influenced only by auditory stimulation. In the selection phase, we sort the activated word form representations to find the one that best matches the auditory stimulus. Selection depends on the level of bottom-up stimulus, as bottom-up information activates word choices. If there is ambiguous bottom-up input, context also considered in the process. In the integration phase, the features of the chosen words are combined to form the complete representation of the spoken sentence. The grammatical and meaning-related properties of the selected word are evaluated based on the context. In the COHORT model, recognizing words involves continuously comparing the sounds we hear with the stored word forms. It is highly incremental, as word form representations get activated as the initial sounds are heard.

5.2 Evidence

Evidence for the COHORT model comes from cross-modal priming experiments[6]. For example, primes like "captain" and "captive" were used with visual hints presented early (at or before "t") or late (during the final syllable, "ain" or "ive"). Target words were semantically related to one of the two primes, such as "ship" or "guard." Early in the word, both "ship" and "guard" were primed, but later, only "ship" was primed. This demonstrates how the COHORT model accounts for the filtering of word choices as more phonetic information becomes available.

5.3 Shortcomings

First shortcoming this model is that it depends on match at the beginning of words, that means it less flexible when there is little mismatch in initial sound, not like the TRACE model, which considers everything for similarity. Second, the COHORT model also suggests that the number of possible word choices doesn't affect how fast we recognize a word. But studies have shown that having more competing words slows the recognition, as our brain takes longer to sort through the options. These limit its effectiveness in real-world, where language processing needs to be flexible and adaptable. The COHORT model links sounds to word meanings, which makes it less flexible when understanding the natural variations in speech, like when words are stretched out (elongated pronunciations).

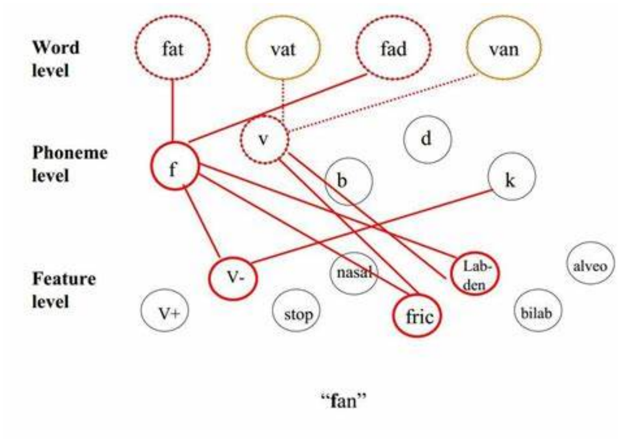

Figure 1: *Activation in different levels of TRACE model*

This rigidity limits its effectiveness in real-world situations where speech is often dynamic in nature.

6 TRACE model

6.1 Explanation of the TRACE Model

The TRACE model was developed by James McClelland and Jeffrey Elman in 1986[12]. This model is an interactive model of speech perception and spoken word recognition. The model works based on interactive activation, where units at different levels (features, phonemes, and words) influence each other. The three layers in model are: the feature layer, the phoneme layer, and the word layer. Each layer has units that represent specific features, phonemes, or words, and these units are connected through excitatory and inhibitory links.

According to TRACE model, understanding of speech starts when feature units respond to heard sound. These feature units then activate phoneme units, which then activate word units. This process helps in recognizing and understanding spoken words. The activation of units at each level is influenced by both bottom-up input (from lower levels) and top-down feedback (from higher levels). This interaction helps to explain things like speech perception, such as the word superiority effect , where letters are recognized with more accuracy when they are part of a word rather than when they are isolated.

The TRACE model also includes explanation for lateral inhibition[10], where active units suppress the activity of nearby units in the same layer. This helps to clearly differentiate between competing sounds and words, which improves the model's accuracy in recognizing speech. The model's interactive(dynamic) processing, known as to as the "trace," allows it to simulate a various research findings related to the perception of sounds and words, as well as their interactions.

6.2 Computational model based on TRACE

The TRACE model, as implemented in jTRACE at | Computational Cognitive Neuroscience of Language Lab , is a computational model of speech perception and spoken word recognition. In jTRACE, the activation $'A'$ of a unit is determined by the balance of excitatory $'E'$ and inhibitory $'I'$ inputs it receives. The activation equation can be represented as:

$A = \sum_{i=1}^{n}(E_i - I_i)$

where E_i represents the excitatory input from connected units, and I_i represents the inhibitory input from neighboring units. This equation shows that how the balance between excitatory (activating) and inhibitory (suppressing) inputs determines how active a unit is.

Research using jTRACE has demonstrated that it can effectively simulates many findings related to how we perceive speech. For example, the model explains the word superiority effect and how word knowledge influences our understanding of sounds(phoneme perception).

6.3 Shortcomings of TRACE

The TRACE model has several limitations[9]. One major shortcoming is that its computationally inefficient. TRACE model uses a large number of units and connections to represent features. As it tries to handle more sounds and words, the number of parts it needs grows at very high rate, making it highly complex. Also, the model assumes certain things about how our mental dictionary (lexicon) works, which might not match how our brains actually process speech. The TRACE model also cannot handle variations in speech, like different accents, speaking speeds, and contexts, which can affect how it recognizes speech.

7 Distributed cohort model

7.1 Explanation of the Distributed COHORT Model

The Distributed COHORT Model (DCM)[7] is a more advanced model for lexical access as compared to previous models. It processes phonetic features through a hidden layer of processing units. These units are also connected to context units, which provide additional information about the word with respect to the current context. This model helps to explain how we recognize the words even when they have similar sound or have multiple meanings. The system uses the output of these hidden units to activate two more groups of processing units: one representing phonological word forms (phonological units) and the other representing word meanings (semantic units). This separation allows auditory/phonological information and semantic information to get stored separately. Auditory information is directly and concurrently connected to both stored phonological representation and stored meanings. A word is recognized when the activation patterns in the phonological and semantic units become stable and settle down into a pattern that match to the target word. According to this model when we hear the onset of a word, both the semantic and phonological units are activated in lexical access. Phonological activity becomes clear and reinforcing because words with similar onsets share phonological representations. On the other hand, activation in the semantic space represents a mix of semantic patterns, as different phonological representations can the share meanings.

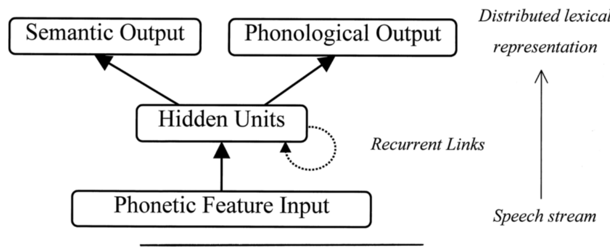

Figure 2: *Distributed COHORT Model*

7.2 Evidence

Evidence for the Distributed COHORT Model comes from experiments showing its handling of coarticulation effects[8] and words with different meanings. For example, the model demonstrates the coarticulation effect by representing the /o/ sound in "jog" with a slightly different pattern than the /o/ sound in "job."

Words with different meanings, like "bark" (a dog barks; a bark of the tree), leads to less clear activation in the semantic space, but words with different senses, like the word "twist" (give the handle a twist; can you do the twist?), leads to more clear activation. These findings support the model's ability to process complex lexical information.

8 Simple recurrent network model

8.1 Explanation of the Simple Recurrent Network Model

The Simple Recurrent Network (SRN) model was given by Jeff Elman in 1990[5].. It is a recurrent neural network that can remember information from previous steps in a sequence. It is different from most feedforward neural networks because SRNs have connections that allow them to remember information from previous steps. This helps the model in the tasks that uses sequences, like finding the next word in a sentence or recognizing patterns. The SRN has four layers: an in-

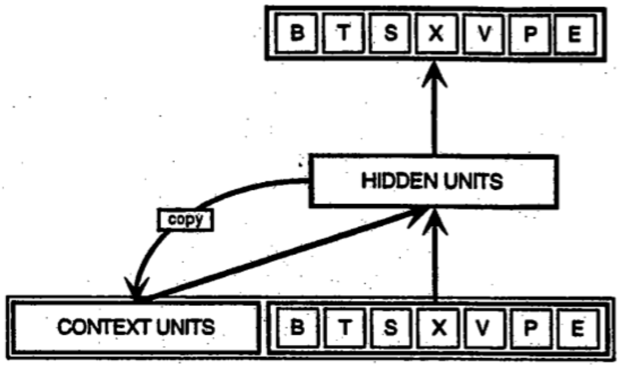

Figure 3: *Simple recurrent network model*

put layer, a hidden layer, an output layer, and a context layer which stores the hidden layer's previous state. The SRN processes the input sequences step by step with one element at a time. It changes the context layer in each step. This helps in finding the next elements in the sequence with good accuracy. These steps make SRNs effective for doing things such as language modeling, time series prediction, and sequence classification.

8.2 Evidence

Evidence supporting the SRN model comes from experiments which also demonstrate the model's ability to learn and predict sequential patterns. In the study conducted by Elman (1990)The Simple Recurrent Network: A Simple Model that Captures the Structure in Sequences trained an SRN to predict the next word in a sequence of words. The network was able to learn the structure of simple sentences and predict upcoming words based on the context given by previous words. This experiment showed that SRNs can explain the grammar of language without any need of input of grammatical rules. SRNs have also been used in many areas, including speech recognition and natural language processing, where they have handled sequential data effectively.

8.3 Shortcomings

In his 1991 paper, Jeffrey L. Elman[4] talks about the challenges associated with Simple Recurrent Networks (SRNs). These networks have trouble in remembering the long-term information because their memory is limited. They often face vanishing gradient issues (learning process becomes difficult), leading to difficulties in learning complex sequence patterns. SRNs can also overfit on training data which results in good performance on training data but poorly on new data. SRNs are sensitive to weight initialization, and it can also struggle to generalize on new data. Training can become unstable, leading to big variance in the loss function. Elman suggests that these limitations shows the need to improve recurrent network architectures to enhance their performance.

9 Appendix

Cognitive comprehension: refers to the mental process of understanding and making sense of information. It involves interpreting, processing, and integrating new information with existing knowledge to form a coherent understanding. This process is crucial for tasks such as reading, listening, problem-solving, and learning.

Lexical Access: It is the process of finding and retrieving words from our memory when we hear, read, or think about them. It's like looking up a word in a mental dictionary to understand its meaning, sound, and how to use it in a sentence. This process is essential for understanding and producing language.

Bottom-up information flow: Refers to the process where information starts from the simplest or most basic level and moves up to more complex levels. In this approach, data is first processed at the sensory level (like hearing or seeing) and then sent to higher levels of processing (like understanding and interpreting). This means that the initial input influences the activation of higher-level processes, without feedback from those higher levels.

Mental lexicon: Mental lexicon is like a mental dictionary that stores all the words we know, along with their meanings, sounds, and how to use them in sentences. It's where our brain keeps all the information about words, so we can quickly find and use them when we speak, listen, read, or write.

Semantic space: Way to represent the meanings of words in a multi-dimensional space. Each word is placed in this space based on its meaning and how it relates to other words. Words with similar meanings are placed closer together, while words with different meanings are farther apart. This helps in understanding and analyzing the relationships between different words and their meanings.

Top-down driven system: Refers to a process where higher-level cognitive functions, such as expectations, knowledge, and context, influence and guide lower-level processing. In this approach, our brain uses prior knowledge and context to interpret and understand sensory information. This means that what we already know or expect can shape how we perceive and process new information.

Feedforward neural networks: Theses are a type of artificial neural network where connections between the nodes do not form a cycle. In simple terms, information moves in one direction—from the input layer, through hidden layers, to the output layer. This type of network is used for tasks like image recognition and classification, where the goal is to map input data to the correct output without needing to remember previous inputs.

10 Conclusion

This review has explored various models of lexical access, each contributing its own unique views into how we recognize and process words. The research problem addressed was understanding the mechanisms behind lexical access and how different models explain this complex cognitive process.

These models improve our understanding of lexical access, each with its strengths and limitations. Future research should aim to combine the strengths of these models, which can lead to hybrid approaches that better imitate the dynamic and interactive nature of lexical access. Addressing current limitations, such as computational efficiency and handling variability in speech, will be necessary for further advancements in this field.

Bibliography

[1] R. Harald Baayen and Robert Schreuder. "Towards a psycholinguistic computational model for morphological parsing". In: *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 358.3 (2000), pp. 1281–1293. DOI: 10.1098/rsta.2000.0586. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2000.0586. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2000.0586.

[2] Robert Clarke and John Morton. "Cross Modality Facilitation in Tachistoscopic Word Recognition". In: *Quarterly Journal of Experimental Psychology* 35.1 (1983), Page Range. DOI: https://doi.org/10.1080/14640748308402118.

[3] David Dorfman and Murray Glanzer. "List composition effects in lexical decision and recognition memory". In: *Journal of Memory and Language* 27.4 (1988), pp. 633–648. ISSN: 0749-596X. DOI: https://doi.org/10.1016/0749-596X(88)90012-5. URL: https://www.sciencedirect.com/science/article/pii/0749596X88900125.

[4] J.L. Elman. "Distributed representations, simple recurrent networks, and grammatical structure". In: *Psychological Review.* 7.15 (1991), pp. 195–225. DOI: https://doi.org/10.1007/BF00114844.

[5] Jeffrey L Elman. "An alternative view of the mental lexicon". In: *Trends in Cognitive Sciences* 8.14 (2004), pp. 301–306. ISSN: 1364-6613. DOI: https://doi.org/10.1016/j.tics.2004.05.003. URL: https://www.sciencedirect.com/science/article/pii/S1364661304001366.

[6] M. Gareth Gaskell and William D. Marslen-Wilson. "Representation and competition in the perception of spoken words". In: *Cognitive Psychology* 45.8 (2002), pp. 220–266. ISSN: 0010-0285. DOI: https://doi.org/10.1016/S0010-0285(02)00003-8. URL: https://www.sciencedirect.com/science/article/pii/S0010028502000038.

[7] M. Gareth Gaskell and William D. Marslen-Wilson. "Integrating Form and Meaning: A Distributed Model of Speech Perception". In: *Language and Cognitive Processes* 12.5-6 (1997), pp. 613–656. DOI: 10.1080/016909697386646. eprint: https://doi.org/10.1080/016909697386646. URL: https://doi.org/10.1080/016909697386646.

[8] F. H. (1995) Guenther. "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production." In: *Psychological Review.* 3.13 (1995), pp. 594–621. DOI: https://psycnet.apa.org/doi/10.1037/0033-295X.102.3.594.

[9] Thomas Hannagan, James S. Magnuson, and Jonathan Grainger. "Spoken word recognition without a TRACE". In: *Frontiers in Psychology* 4 (2013). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00563. URL: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2013.00563.

[10] Benjamin J. Levy and Michael C. Anderson. "Inhibitory processes and the control of memory retrieval". In: *Trends in Cognitive Sciences* 6.7 (2002), pp. 299–305. ISSN: 1364-6613. DOI: https://doi.org/10.1016/S1364-6613(02)01923-X. URL: https://www.sciencedirect.com/science/article/pii/S136466130201923X.

[11] William D. Marslen-Wilson. "Functional parallelism in spoken word-recognition". In: *Cognition* 25.7 (1987). Special Issue Spoken Word Recognition, pp. 71–102. ISSN: 0010-0277. DOI: https://doi.org/10.1016/0010-0277(87)90005-9. URL: https://www.sciencedirect.com/science/article/pii/0010027787900059.

[12] James L McClelland and Jeffrey L Elman. "The TRACE model of speech perception". In: *Cognitive Psychology* 18.9 (1986), pp. 1–86. ISSN: 0010-0285. DOI: https://doi.org/10.1016/0010-0285(86)90015-0. URL: https://www.sciencedirect.com/science/article/pii/0010028586900150.

[13] Fanny Meunier and Juan Segui. "Frequency Effects in Auditory Word Recognition: The Case of Suffixed Words". In: *Journal of Memory and Language* 41.5 (1999), pp. 327–344. ISSN: 0749-596X. DOI: https://doi.org/10.

1006/jmla.1999.2642. URL: https://www.sciencedirect.com/science/article/pii/S0749596X99926426.

[14] Lee H. Wurm R. H. Baayen and Joanna Aycock. "Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities". In: *The Mental Lexicon* 2.3 (1983), 419–463 Range. DOI: https://doi.org/10.1075/ml.2.3.06baa.

[15] Mark Seidenberg and Laura Gonnerman. "Explaining derivational morphology as the convergence of code". In: *Trends in Cognitive Sciences* 4 (Oct. 2000), pp. 353–361. DOI: 10.1016/S1364-6613(00)01515-1.