

MICHAEL PACE-SIGGE
UNIVERSITY OF EASTERN FINLAND

LARGE-LANGUAGE-MODEL TOOLS AND THE THEORY OF *LEXICAL PRIMING*: WHERE TECHNOLOGY AND HUMAN COGNITION MEET AND DIVERGE

CITATION

Pace-Sigge, M. (2025). Large-Language-Model Tools and the Theory of Lexical Priming: where technology and human cognition meet and diverge. *Journal of Corpora and Discourse Studies*, 9:1-22

ABSTRACT

In this position paper, Michael Hoey's Lexical Priming Theory (2005) is being revisited in the light of recent discussions of Large Language Models as forms of machine learning (commonly referred to as AI), which have been the centre of a lot of publicity in the wake of tools like OpenAI's ChatGPT or Google's BARD/Gemini. Historically, theories of language have faced inherent difficulties, given language's exclusive use by humans and the complexities involved in studying language acquisition and processing. The intersection between Hoey's theory and Machine Learning tools, particularly those employing Large Language Models (LLMs), has been highlighted by several researchers. Hoey's theory relies on the psychological concept of priming, aligning with approaches dating back to Ross M. Quillian's 1960s proposal for a 'Teachable Language Comprehender'. The theory posits that every word is primed for discourse based on cumulative effects, a concept mirrored in how LLMs are trained on vast corpora of text data. This paper tests LLM-produced samples against naturally (human-)produced material in the light of a number of language usage situations, investigates results from A.I. research and compares the results with how Hoey describes his theory. While LLMs can display a high degree of structural integrity and coherence, they still appear to fall short of meeting human-language criteria which include grounding and the objective to meet a communicative need.

KEYWORDS

A.I.; Chat-GPT; Communicative Need; Gemini; Hoey; Large Language Models (LLM); Lexical Priming

CONTACT

Michael Pace-Sigge, Dept. of English Language and Culture, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland. michp@uef.fi

DOI

10.18573/jcads.129

ORCID

00-0002-5164-5242

ISSUE DOI

10.18573/jcads.v9

LICENSE

© The authors. Available under the terms of the CC-BY 4.0 license

Large-Language-Model Tools and the Theory of *Lexical Priming*: where technology and human cognition meet and diverge

Michael Pace-Sigge

University of Eastern Finland

1. Introduction

Were I to ask an AI chatbot to write the opening of this paper for me, then I surely would be offered two different starting points. I could either begin with the here and now and reflect on the sudden excitement and anxiety that seems to have arisen ever since companies like OpenAI offered highly-developed AI tools that create texts, images and videos within an incredibly short time-frame after receiving a prompt. Alternatively, I could go back twenty years ago and look at what was then called ‘a new theory of words and language’ – the *Lexical Priming (LP) Theory* presented by Michael Hoey (2005) and use a language-studies rather than a computational linguistics approach as my starting point. Much as I would love to be a quantum information processor (and wish the same for my readers), I doubt I can do both things at once.

What can be done, however, is to investigate material produced by LLM tools in the light of corpus-linguistic approaches. Thus, this article is designed to be a position paper, with the aim to inform the reader about how recent developments in Large-Language Model based tools have come about and can be seen as linked to linguistic theory; in particular, tools like ChatGPT and BARD are employed to replicate research undertaken by Hoey (2005) to underpin his Lexical Priming Theory (LPT). This is meant to provide a firm basis for further, more detailed research similar to, for example, Berber Sardinah (2024), who contrasts human-authored and AI-generated texts against the five dimensions set out by Biber (1988), as well as Curry and colleagues (2024), who compare ChatGPT 4’s approach to automated qualitative analysis with three previously published studies.

Theories are seen as a good approach in resolving unanswered questions. There can be good theories, which may, nevertheless, need a lot to convince people, and convincing theories, which may, nevertheless, lead us up a garden path. Importantly, a theory gains weight when other researchers can replicate what has been tested and obtain the same or similar outcomes. A theory is given credence and deemed valid if a setting is created where the theory can be tested, and the outcomes align with the behaviour as predicted by the theory. This paper contrasts naturally occurring texts with LLM-produced output, with the specific focus in how far the design of LLM tools aligns and diverges from the Lexical Priming Theory.

Previously, there have been inherent difficulties with theories of language. First of all, the only users of language as we know it are human beings, and experiments to test theories would have to be tangential and, to a very large degree, inferential. Secondly, it was seen as impossible to take into account all the elements which are in play during language acquisition, language processing and active production. Given these difficulties, the subtitle of Michael Hoey's 2005 book - 'A new theory of words and language' - seems, on the surface of it, to set the bar high when it comes to validating the theory presented here.

Over the last few years, however, an alternative approach has become available. One such example is given by Pace-Sigge & Sumakul (2022), who demonstrated that the set-up to train an AI to write a textbook can be compared to teaching student classes in *academic writing*. The machine turns into a virtual stand-in for a classroom – and the errors the machine makes, namely areas that appear non-natural and unsuitable are also points that need particular attention when teaching academic writing. In other words, in order to pursue a level of language proficiency, research in the 2020s no longer needs to apply theories and techniques in a human-based set up: training a machine can serve as a stand-in to highlight salient issues of language acquisition and usage. At the end of the day, teaching a language for a specific purpose consists not only of teaching technical terminology: the frequency of words and phrases used, the preferred grammatical forms are repeated several times. Similar to teaching a language, when training and then fine-tuning the output of a text generator, these characteristics have to be prominent as well.

It is at this point where the two phenomena of Hoey's LP theory and Machine Learning tools that make use of Large Language Models overlap. The pivotal point is the psychological concept of *priming*, which Hoey has adopted for his theory. Priming occurs when a listener or reader comes across a certain word sequence and construction with a frequency higher than random co-occurrence. As a result, a single word can then act as a 'prime' which leads to the activation of what is expected to come. As an example, in experiments in the 1960s and 1970s, it was shown that the term 'nurse' would lead to a far quicker recognition to 'doctor' than the unrelated word 'bread'. This approach was first proposed by Ross M. Quillian in the 1960s, when he proposed 'a model of language' that can serve a machine he dubbed the *Teachable Language Comprehender*, which would be trained on naturally occurring texts (books) and which could disambiguate meanings of words based on the context the words are found in (see also Quillian, 1967). While the model developed was then state-of-the-art, the technology to turn this into a viable machine was not available for another fifty years or so – while the approach to develop such a machine was revised and refined in the meantime. Quillian was not the only one to reckon that his construction, a basis for a computational machine, can also serve as a model to explain language. Some AI researchers (e.g. Brynjolfsson) go as far as to claim that human-like artificial intelligence has several benefits, amongst them 'perhaps most profoundly a better understanding of our own minds' (Brynjolfsson, 2022, 272).

As an explanation of what corpus linguistic insights present, the Lexical Priming theory builds on Quillian's work and postulates that '[e]very word is primed for use in discourse as a result of the cumulative effects of an individual's encounters with the word'

(Hoey 2005, 13). This appears to be remarkably similar when compared to the ways in which large language models (LLMs) are built.¹ The ‘cumulative effects’ are the training data, where each item (‘word’) will appear with differing frequencies in large corpora of training data, and the ‘use in discourse’ is how this item stands in relation to the other items: whether they, for example, cluster together with a mostly, often or only rarely (semi-)fixed set of items; whether they appear together or whether co-appearance is a rare occasion etc. In other words, the key issue is the relative statistical likelihood in which words co-occur (see Manin and Marcolli, 2016). While the issue of ‘representativeness’ has been used as a criticism of corpus linguistics, ever-larger data sets have mostly confirmed initial findings based on small earlier corpora. While ChatGPT and BARD/Gemini are on a large variety of texts, LLMs for specific purposes do already exist and there is, therefore, a high degree of data-quality which can be found to be in operation. The input of billions or even trillions of words become a matrix in which each item is assigned a vector representation. Mikolov et al (2013) describe how computer models mimic relationships between words in a way not dissimilar to the relationships human language users see (e.g., grammatical ones like *tall-taller*; semantic ones like *train-travel*, etc.). This computational process is, crucially, not pre-programmed: the fact that so much training material is available allows the algorithm to find ‘natural’ relationships between words. In *corpus linguistics* this is mirrored by *collocation* (the concept that a word is more frequently found in the vicinity of another word than mere coincidence would allow for) and *colligation* (the idea that lexemes fit into particular categories and their pre- and suffixes and occurrence patterns are limited to the grammatical patterns in which they are typically found).²

Consequently, it can be postulated that the claims on which the Lexical Priming Theory are based are being tested by algorithms which make use of LLMs. In fact, one of the key pillars of the theory is a process called *nesting*:

[A] combination of words will have priming separate from (though built up from) the primings of the individual words. ... [N]esting does not only take the form of the building of word sequences and items. It can also be the case that when a word or word sequence combines with a particular colligational priming. (Hoey 2005, 58)

It is at this point that a link to present day *generative pre-trained transformer* (GPT)-style machine learning tools becomes apparent. The decisive breakthrough in enabling artificially produced text to make natural-sound was the move from a forward-prediction model to the GPT text producer. While earlier systems were *neural network* based, which allowed them to predict the word likely to come after a node-word, the drawback of such a system was that the bigger the distance from the node-word, the larger the potential

-
- 1 The construction and design of the algorithm follows the principle of statistical probability for closeness and deeper imbedding. For this, see also Pace-Sigge (2018). Given that trillions of words are the training data and thousands of data points are correlated, it should come as no surprise that the outcome includes emergent properties that were not foreseen by the software developers, however.
 - 2 At this point, I want to leave out *semantic association* as this is one area which will not be specifically addressed in this paper.

words that could follow would become. In the end, that meant that such systems were quite good in producing relatively short texts without any supervision; however, the longer the required text becomes, the less natural-sounding the product becomes. GPTs, on the other hand, rely on a system first proposed in 1997, the *Long Short-term Memory* (LSTM), which allows data to flow forward as well as backward within the network. A possible comparison is a reader who comes across a sentence on page 200 which does not seem to make sense, then turns back to page 198 to see what happened there, understands now what is being referred to on page 200 and anticipates how things will develop on page 201. In this, the LSTM conforms to the model as described by Quillian (1967), where semantic disambiguation hints ‘can be found in surrounding words, sentences or paragraphs’. This is also in line with Hoey’s (1995) description of ‘bonding’, whereby the bond between one particular word with the previous and the next mention does not have to be in the next sentence or even paragraph or page, but can create a lexical chain which spans over longer sections of any given text. LSTM could therefore be seen as mirroring human cognitive behaviour, where what is being said links to early information while at the same time pointing forward to what is to come (cf. Halliday, 1993).

2. Bonding, cohesion and story grammar

It is said that the success of the current set of AI language production tools stems from the fact that they moved beyond earlier designs. Purely statistical models which were in use until around 2010 could predict the next word in a sequence. This meant that they could produce one or a set of several meaningful sentences that gave the impression of naturally produced language. However, their main drawback was that forward extension would become increasingly difficult to predict: as a result, longer passages were nearly impossible to produce. What was needed, therefore, was a tool that could not simply forward-guess based on what came before. Instead, each word, phrase or, indeed, sentence has to have the ability to link back to the web of words and concepts these words, phrases and sentences are associated with. A step in that direction was undertaken by computational linguists in the 1980s (see also Pace-Sigge, 2018: 40ff).

It is interesting to note that their work mirrors work undertaken by Hoey (1991; 1994; 1995). Here, Hoey made the case that the concept of *bonding* exists in both narrative and non-narrative texts. Crucially, ‘only a small proportion of the bonds formed in the passage given are between adjacent sentences’ (Hoey, 1991, 149). Hoey returns to this idea later and surmises that there ‘is the possibility of finding bonding across texts written between three and fourteen years apart, solely because of the mental concordances of the authors retained records of the texts they had read’ (Hoey, 1995, 90) – a working theory which was further supported by Hoey (2017b). In short, *bonding* shows that a single narrative, or set of similar narratives by one or several authors, share cohesive links which are specific to them. Furthermore, these words are nested within the wider context and co-text, thus ‘we retain access to the contexts of words previously encountered, or else

each new encounter with a word of whose meaning we were uncertain would be a fresh problem' (Hoey, 1991, 155).

This mirrors the approach by the computational linguist Peter Norvig developed in the late 1980s and early 1990s. Norvig makes the point that, in order to create a 'language comprehending machine' one needs to create 'story understanding machines' and highlights the resultant importance of *story grammar*: 'Story Analysis programs are interesting because they can provide an insight into the structures of stories, the notion of coherence, and the interaction between events, goals and plans' (Norvig, 1992, 1). How important this is in the design of LLM tools is explained by ChatGPT (Norvig, 2024).^{3,4}

'Story grammar' refers to the underlying structure or framework that organizes elements of a story, including characters, setting, plot, and resolution. In the context of ChatGPT, understanding story grammar is crucial for generating coherent and contextually relevant responses. While GPT models like ChatGPT don't explicitly possess a deep understanding of narratives, they learn patterns from vast datasets, which may include elements of story grammar.

Here are a few points to consider in relation to story grammar:

- 1) **Coherence:** Understanding story grammar aids in producing responses that follow a logical and coherent flow. It helps in generating text that maintains context and relevance throughout a conversation.
- 2) **Context Maintenance:** Story grammar assists in keeping track of the ongoing conversation. This is important for generating responses that are contextually appropriate and build upon prior inputs.
- 3) **Character Consistency:** In a conversational context, the 'characters' can be seen as the participants in the dialogue. Understanding story grammar helps maintain consistency in how these 'characters' respond and interact over the course of a conversation.
- 4) **Event Sequencing:** Story grammar involves understanding the sequencing of events. For ChatGPT, this means generating responses that logically follow the sequence of inputs in a conversation.

While GPT models, including ChatGPT, have shown impressive language generation capabilities, they may not always exhibit a perfect understanding of story grammar. The models generate responses based on statistical patterns learned from diverse data, and their performance depends on the quality and nature of the training data.

Improving story grammar comprehension in AI models is an ongoing area of research, and advancements aim to enhance the coherence and context-awareness of AI-generated text.

3 The prompt used is 'How important is 'story grammar' in the design of ChatGPT?'

4 This prompt may be seen as leading. However, when inverting it ('Why does ChatGPT not use story grammar?') ChatGPT is producing a similar answer. It does, however, give a more nuanced introduction: 'ChatGPT, like its predecessors, is primarily designed to generate coherent and contextually relevant

I quote ChatGPT's answer here in full to demonstrate how different theoretical approaches come together in forming the foundation for the design of the current crop of LLM tools. The move from *Recurrent Neural Networks* (RNN) and *Long Short-term Memory* (LSTM) which were difficult to scale up, were achieved after 2017 when Google presented their *Bidirectional Encoder Representations from Transformers* (BERT), followed by OpenAI's *Generative Pre-trained Transformer* (GPT) 1.⁵ It is noteworthy how insistent Chat-GPT here is on the word *understanding* – a quality that is succinctly described by many in the industry as 'today's NLP systems built on large language models still fall short of human-level understanding' (Merrill et al. 2021, 1047).

One of the major challenges lies in turning the four points made above into a tool that can independently generate text, rather than simply regurgitate existing content. It is, however, at this point that major doubts are expressed around the hype caused by Chat-GPT. While Shanahan (2022, 9) says that LLMs 'exhibit wisdom-of-the-crowd effects, while being able to draw on expertise from different domains' he concedes that 'a suitably prompted LLM appears to reason correctly [by] mimicking well-formed arguments in its training set and/or prompts'. Indeed, the investigation by Valmeekam and colleagues (2022), who found that tests that were meant to evaluate the reasoning capabilities and of AI systems can be passed either because the training material already consists of these tests or because these benchmarks are either too simplistic or too flawed to be useful for an objective evaluation. Crucially, '[o]ur results show that even in simple common-sense planning domains where humans could easily come up with plans, LLMs like GPT-3 and BLOOM seem to display a dismal performance' (2022, p.9).⁶ In other words, the LLMs can re-formulate, yet they seem unable to re-use or generalise from existing plans, neither can they modify any existing plans. Thus, one would see them as machines which are good at regurgitation and largely incapable of independent (forward-) planning. Yet all of these might be necessary to be creative, as is a modicum of self-awareness. Based on these, unexpected juxtapositions of concepts, novel combinations of words can come into being Hoey (2005, 153) says that it is 'original texts that refresh the language and force us to think and see things in new ways'.⁷ Yet LLMs use a finite set of text samples as their training base and these would, like old grammar books, lose their freshness and relevance in due course. The theory of Lexical Priming allows explicitly for

responses in a conversational manner. While it doesn't explicitly use story grammar, it leverages a large-scale language model that has been trained on a diverse range of internet text to understand and generate human-like text. ...'

5 See also Havlik (2023) p. 14.

6 It must be noted that all AI-tool developers appear to work very hard to address such criticism and the speed improvements seen means that 2024 versions seem to be so much more capable – which gives an idea how the 2025, 2026, 2027 etc versions might be far improved beyond recognition.

7 He continues: 'If linguistics cannot say something interesting about literary language, it is an admission that we have not yet got to the heart of our discipline'. In this, Hoey echoes John Sinclair: 'no systematic apparatus can claim to describe a language if it does not embrace the literature also; and not as a freakish development, but as a natural specialization of categories which are required in other parts of the descriptive system' (2004: 51).

language variation and change. A speaker can see the activation of a particular priming playing out differently in a formal descriptive text compared to the use of the same prime in a figurative text, whereby the con-text and co-text are aiding disambiguation. Current LLMs can simulate this process to a degree. However, a human speaker will also be able to passively comprehend where a priming gets breached or, over time, understand how a prime activates a different structure; they often actively create breakings in primes and re-assignment of primes themselves, too, depending on the communicative message they want to produce. LLMs, on the other hand, are free of intent and bound by the dictate of probability, which confine its ability for variation and context-dependent precision.

3. Chat GPT and structures

One key skill a corpus linguist acquires is the ability to identify frequent patterns and structures. These structures can be found to be salient to such a degree that Hoey (2005) identifies them as 'primed for semantic association'. For example, looking at the field of *journey* he identifies the sequence 'NUMBER-hour-JOURNEY (or NUMBER-TIME-JOURNEY)' (p.17). Another example is that the word sequence 'young men and women' has a semantic association with COMPLIMENTS' (p.19). Similarly, Chat-GPT (Jan 2024), when fed the same data⁸, identifies the following: 'Duration of Time–Mode of Transportation' for the first example, while saying about the second 'Adjective or Adjectival Phrase–Young Men and Women ... This pattern effectively highlights various positive qualities or characteristics describing the group of young men and women'.

Seeking confirmation from BARD (January 2024)^{9,10}, the following answer was produced:

Yes, the structure NUMBER-TIME-JOURNEY is commonly used in English. It is used to convey a sense of a journey that takes place over a specific period of time. This structure can be used to describe personal experiences, historical events, or fictional stories.

Here are some examples of the NUMBER-TIME-JOURNEY structure in use:

A five-year journey of self-discovery

A hundred-year journey through time

A one-day journey into the heart of darkness

The NUMBER-TIME-JOURNEY structure can also be used to create a sense of urgency or excitement. For example, the phrase 'a three-minute journey to the edge of the world' suggests a fast-paced and thrilling experience.

8 The prompt used is 'What structure to these have in common?'

9 Seeking confirmation from ChatGPT seemed to produce a circular argument as it gave the same examples that had been used in a prompt earlier. The prompt used in both cases is 'Is the structure NUMBER-TIME-JOURNEY commonly used in English?'

10 As of 08.02.2024, Alphabet has renamed BARD *Gemini*, a name it has used since October 2023 for its A.I. tools.

Overall, the NUMBER-TIME-JOURNEY structure is a versatile and effective way to convey a sense of time and place in English. It can be used to create a variety of effects, from personal reflection to historical narrative¹¹.

All in all, however, the use of LLM tools indicates that existing research can swiftly be replicated even for those who do not have access to large caches of data. Crucial, however, is the word ‘appears’, as it seems that BARD suffers from confirmation bias. This is demonstrated when the prompt is altered into ‘is the structure NUMBER-TIME-JOURNEY not commonly used in English?’, at which point the previous assertion is fully negated as BARD (2024) now says: ‘I apologize for my previous error in stating that the structure NUMBER-TIME-JOURNEY is commonly used in English. It is actually not a common structure in English.’ Intriguingly, BARD then gives three examples that seem to echo the experiment Hoey undertook when he re-wrote a sentence taken from a popular book. Thus, BARD offers now ‘a transformative journey that lasted for five years’, instead of ‘a five-year journey of self-discovery’. This is in stark contrast to Chat GPT (2024) which, while apologising for the confusion caused, nevertheless re-asserts that the ‘structure ‘NUMBER-TIME-JOURNEY’ is indeed commonly used in English’. This, however, leaves the researcher bemused: it feels like a form of triangulation, where two different researchers investigate similar data with their own tools and produce a fuzzy set of answers.¹²

Checking Hoey’s research results (which he used to demonstrate the validity of LPT) against answers from LLM tools can be seen as a demonstration that LLM tools can be used to test examples that Hoey gave in support of his lexical priming theory; it is not sufficient in itself, however, to show that the design of present-day GPT tools aligns with the premises set out by the theory.

Another test is to run results created by LLM tools side-by-side with naturally occurring language. For this, a number of random, low-frequency words from the BNC and BNC 2014 (Brezina et al, 2021) are checked for their occurrence in concordance lines, using WordSmith 8 and Lancsbox X – and the results are then compared with model sentences produced by ChatGPT and BARD. Using a rare word has the advantage that all variations of use within the corpus can be assessed, as all occurrences and not just a random sample is examined. Similarly, whereas an LLM tool can generate a potentially infinite amount of sample sentences, the natural occurring material provides a suitable number amount of material for the comparison.

11 While ChatGPT talks about ‘positive qualities or characteristics’ it must be noted that BARD gives no such evaluation. It is questionable whether the former independently judges here, it is more likely that it repeats an evaluation from a source in its training data.

12 It must be noted that, ever since virtual communicators like ELISA have appeared, human-machine interaction has become, for some users, like an interaction with a real human being. It is therefore not too far-fetched to make a comparison like this.

- | | |
|---------------------|--|
| 1) FictFan15.xml | all over again: as harem captives of a cruel Eastern potentate ; passengers sharing a lifeboat with hungry sailors, |
| 2) FictWom124.xml | Ellie was appalled. 'He's carrying on like an old-fashioned potentate with his harem.' |
| 3) FictHor14.xml | Bosphorus doomed to be killed or enslaved by the Muslim potentate . But I hope our fascination is not purely academic. |
| 4) FictMis252.xml | a ramshackle collection of warlords nominally controlled by the Cambrian potentate Tharv the Bountiful. The empire has almost no economy or |
| 5) FictMis469.xml | This was the act of a conqueror, a potentate ! 'The people, for his airs, they call him Jrundur |
| 6) NewMaDam3029.xml | working classes while living the opulent existence of a medieval potentate . |
| 7) AcaHumRa257.xml | decision to summon the people of the hundred to the potentate's gates. |

Figure 1: Concordance lines of all occurrences of 'potentate' in BNC-2014

Figure 1 shows that the word chosen for this comparison is, indeed, very rare and the majority appears in *fiction* texts. It collocates twice with *harem* while there is also a direct connection to *conqueror/warlord* as well as *Muslim/Eastern* in two each of seven cases. It is indicative that *potentate* is, in five cases, pre-modified in the human-written material. This is, however, rarely the case in the LLM-produced samples (Figure 2):

- | | |
|-----|--|
| 1.1 | The potentate , adorned in regal attire, addressed the assembled crowd with eloquence and authority. |
| 1.2 | As a potentate of the realm, his decisions held immense influence over the kingdom's affairs. |
| 1.3 | The courtiers awaited the potentate's pronouncement, knowing it would shape the destiny of the empire. |
| 1.4 | The diplomatic envoy sought an audience with the potentate to discuss matters of international importance. |
| 1.5 | The citizens celebrated the anniversary of their potentate's reign, grateful for the prosperity he had brought to the land. |
| 1.6 | The palace walls echoed with the whispers of intrigue as factions vied for favor with the potentate . |
| 1.7 | Despite being a potentate , known for strength and decisiveness, he ruled with benevolence, earning the admiration of his subjects. |
| 2.1 | The ancient potentate ruled his vast empire with an iron fist. |
| 2.2 | The media mogul was an influential potentate who wielded power behind the scenes. |
| 2.3 | The young potentate was eager to prove his mettle and establish his reign. |
| 2.4 | The potentate's lavish lifestyle was the envy of the world, but his subjects suffered in poverty. |
| 2.5 | The downfall of the potentate was swift and unexpected, leaving his empire in turmoil. |

2.6	The potentate's legacy is a mixed bag of achievements and atrocities.
2.7	The stories of the potentate's reign have become legendary, captivating audiences for centuries.

Figure 2: *Potentate generated sample sentences: 1.1-1.7: Chat-GPT, 2.1-2.7: BARD*

By contrast, Figure 2 shows that two LLM tools can create well-formed sentences, yet they do not match the collocational and colligational template provided by the – albeit British only – naturally occurring texts. There is no single collocate pointing to *harem*, *Eastern* or *conqueror*, though there are collocates like *kingdom*, *empire*, *regal* and *realm*. Interestingly, both Chat-GPT and BARD have an example of ‘potentate's reign’ (1.5 & 2.7) there is one further such possessive construction in 1.3. By contrast, in the BNC, the possessive is expressed once with the *genitive -s* and once with an *of*-construction. The most obvious difference is, however, the lack of premodification where the BARD samples produce three (out of seven), namely *ancient* (2.1), *influential* (2.2) or *young* (2.3) while Chat-GPT only uses determiners (*a*, *the*, *their*). It must be noted, however, that Concordance 6 and example 2.4 seem to describe something quite similar.

On the one hand, the LLM-produced samples are quite similar to the use of this very rare term found in the BNC-2014. On the other hand, the word *potentate* is quite specific. While in Figure 1, ‘king’ might be used in place of ‘potentate’, this does not work very well in all lines – for example, *Muslim* collocates with *potentate* rather than *king*. The 14 samples shown in Figure 2, however, are not specific enough to use either *king* or *potentate*. In a way, this comparison gives a snapshot of how LLMs seem to be hyper-priming.¹³ ‘Hyper-priming’ is similar to *priming*, but the activation of related concepts is **much stronger**. As a result, a broader field of terms and/or concepts is used and the result is that a less specific, and at times less clear, word is employed. Curry and colleagues (2024) talk in their paper of ‘categories that can appear quite generic’. Some of the collocates and some of the colligational structures befit the node word, yet, the end-result lacks the level of specificity that a human-produced utterances or sentences appear to reflect. For example, 2.6 in Figure 2 is highly generic (almost cliché-like) and ‘potentate’ here could be exchange with virtually any other type of ruler. In effect, the results provide a snapshot of what the, far more intricate, study by Berber Sardinha (2024) details. The LLM does not take in the broader information landscape that a human speaker or writer would fall back on, and thus this little test appears to point more towards ‘thoughtless imitation’, which is free of any producer intent, and appears therefore to be merely following frequency patterns based on its training.

A further exemplification can be found when looking at Hoey’s (2005, 64ff.) findings with regards to priming and co-hyponymy, where he states that ‘[t]rain is primed to collocate with *as a* in newspaper data and the nested combination of *train* as a* is typically primed to associate with SKILLED ROLE OR OCCUCATION’ (sic).

13 In December 2024, Sardinha and colleagues presented several papers at *DRD Hum 2024* where they showed that ChatGPT appears to over-emphasise and over-use text-type specific structures which could otherwise be referred to as *hyperpriming*.

For this examination, the parameters were slightly adapted. Therefore, the prompt was ‘give ten sample sentences with ‘train as’’. This allows for greater ambiguity as to whether *train* is the noun, or a transitive or intransitive verb use. Furthermore, ‘train as an actor/ auditory verbal therapist’ etc. is now feasible, as are phrases like ‘... train as hard as ...’.

For this exercise, Chat-GPT and BARD were asked to provide ten sample sentences; at the same time, the full collection of ‘train as’ concordance lines in the BNC-2014 (newspapers) were retrieved with LancsBox.

<i>train as</i> pattern	BNC-2014 news	Chat-GPT	BARD
TRAIN (n)	11	0	n/a
TRAIN (v) phrase	5	0	n/a
TRAIN (v) profession	11	14	n/a
TRAIN (v) non-prof.	1	14	n/a
total	28	28	n/a

Table 1: *train as* occurrence patterns compared (most frequent in bold)

First of all, *BARD* was not available for this. Even when re-prompted, all sample sentences used either ‘trains as’ or ‘trained as’.¹⁴ Table 1, crucially, seems to present a form of hyper-priming where Chat-GPT is concerned. For the tool, ‘train’ is exclusively meant as a form of training – and this is scrupulously divided into *train for a profession* (‘she as going to train as a pilot’) and *train to obtain a further skill/qualification* (‘investing in training employees to train as cybersecurity experts’; ‘train as a mentor’). For the latter, the BNC only has one occurrence (‘train as a part of a riot squad’), whereas ‘train to be a nurse/ doctor/teacher’ is far more prevalent. At the same time, *train* as a noun (‘killed by a Tube train as he tried to save a man’) is as likely to be found in newspapers as the verb-form. Furthermore, Chat GPT does not return any multi-word-units like ‘train as hard’, very unlike the natural occurring material.

An obvious objection here might be that this is not a like-for-like comparison, as Chat-GPT is expected to have been trained mainly on US rather than UK material. Therefore, the proportion of *train (n) as* would be expected to be lower. Furthermore, the BNC material is exclusively based on the newsprint sub-corpus, while the LLM tool is trained on a large variety of text types. Yet none of these points appears to hold up to scrutiny. The BNC-2014 has a total of 111 concordance lines for *train as*. Yet even when looking at different sources (biography, fiction, magazines, miscellaneous) the distribution of usage is similar to the one found in newsprint only. Yet, when asked to produce 111 sample sentences, the resulting distribution does not change at all in Chat-GPT. It is

14 In January 2025, the same prompt was given to *Gemini* and this time, 2/10 sentences used “train as” – one referring to a profession and the other to a non-profession.

also noteworthy that the idiomatic forms ('train as hard'; 'train as a group') have zero occurrences in the machine-generated samples.

A final comparison is based on Hoey's (2005) sampling of 'hypernyms of SKILLED ROLE OR OCCUPATION, namely *architect*, *accountant*, *actor* or *carpenter*'.

The first thing that can be noted in this experiment is how hyper-generic the examples produced by Chat-GPT and BARD are when prompted with 'give me sentences that use either actor OR architect or accountant or carpenter'. First of all, bar one exception, all sample sentences start with the requested noun – which is a feature hardly ever occurring in the natural data. The only difference is that BARD pre-modified each target word, therefore it produced: 'The struggling actor...'; 'The meticulous architect...'; 'The reliable accountant...'; 'The meticulous carpenter...'. In half the samples in Chat-GPT, the initial noun phrase is followed directly by a verb phrase (either verb or verb+adverb), for example, 'The accountant meticulously reviewed the company's financial records...'. That the results appear generic is underscored, furthermore, by the inclusion of collocates from the same semantic word field in almost each sentence.¹⁵ Thus *actor* has as co-text words like *play*, *screen*, *role* and *performance*; *accountant* sample sentences have *businesses* and *financial*; *architect* has *building*, *office space*, *cityscape* etc. Furthermore, the modifiers (adverbs) also appear to be generic rather than specific to any particular profession or trade. Thus, *diligently*, *meticulously* and *tirelessly* appear favoured by both LLM tools, yet their selection appears to be random.¹⁶ These results seem to support Granger who says that 'the fact that LLMs fail on unfamiliar or untrained prompts suggests that they use a simpler and more rigid grammar than human language, where 'even slight changes may cause the [program] to fail' (Granger, 2020, p. 27). Consequently, the results appear more like a simulation than a cognitively processed product (cf. Arunachalam et al., 2023; Berber-Sardinah, 2024).

Looking, by contrast, at the human-written data in the whole of the BNC-2014, the degree of overlap is minimal. For example, *carpenter*, in 8/435 concordance lines, collocates with *father*; there are also five occurrences of 'Birmingham carpenter'. The contrast is even starker if we search for the adverb modifiers or semantic field collocates like *furniture*, *tools* or *wood*. In the BNC-2014, none of these occur. In fact, amongst the professions looked at here, none of the modifiers selected by the LLMs seem to occur at all in the BNC-2014 real-world data. Thus, for example, the British texts speak of a *chartered* or a *certified* accountant, rather than an *astute* one. Likewise, the concordance lines refer to *American*, *Danish*, *Dutch*, *French* or *German* architects, rather than the *passionate* one. The closest the LLM-produced texts come to mirror what is found in naturally occurring material are some of the semantically related collocates. Consequently, *actor* co-occurs with *Hollywood*, *performance*, *plays* and *skills* – yet these are far less prominent in the BNC-2014 than, for example, *best*, *director*, *played* or *supporting*. These particular structural divergen-

15 It must be noted that BARD provides a notable exception as it produced sample sentences for *carpenter*: only one of three included *craft*, *wood*, and *furniture*.

16 *Diligently* -actor (BARD) -carpenter (Chat-GPT); *meticulously*-accountant, architect (BARD) -accountant (Chat-GPT); *tirelessly* -accountant (BARD) -actor (Chat-GPT).

cies can be found for the other professions as well. Interestingly, the collocates found in Chat-GPT and BARD for *architect*, namely *building*, *design*, *project* and *landscape* are typically amongst the most frequent collocates in the BNC-2014 as well. On the other hand, the difference is stark for *accountant*. Every sample produced by the LLM tools includes the word *financial*. In the 1,133 concordance lines of the BNC-2014, however, this word occurs as a collocate only twelve times – half as often as *experience*. Similarly, the term *business* (*businesses*) appears in half the samples produced by the LLM tools, and yet it appears as a collocate in the BNC-2014 only eight times.¹⁷

Hoey (2005, 65) posited that one might assume that all these words have a common set of collocates and largely similar colligational structures. As we have seen, the Chat-GPT and BARD produced samples meet such an expectation to a large degree. While the focus above has been on collocates, the grammatical realisations found for each word are discussed by Hoey (2005, 66ff.). Here, Hoey shows that *carpenter* occurs in 42% of all cases with an indefinite article, which is substantially higher than found with the other professions. This is supported by BNC-2014 evidence: in the 344 concordance lines for *carpenter*, there are 72 ‘a carpenter’ as opposed to 41 ‘the carpenter’ (21% vs. 12%).¹⁸ While Hoey’s data highlights that ‘an accountant’ seems to be prevalent compared to ‘the accountant’, in the BNC-2014 the ratio is even more pronounced, with the former appearing nearly four times as often than the latter. Furthermore, where Hoey identifies parenthesis (‘her father, a carpenter, became...’) as far more typical for *carpenter* than *accountant*, this colligational marker is mirrored in the usage shown in the BNC-2014, too. *Accountant* has been identified in Hoey’s data to appear in 1/10 cases with a ‘possessed’ construction – like, for example, *turf accountant*. In the BNC-2014 this construction is even more prominent (13/100 cases)¹⁹. A further feature Hoey highlights is the relative absence of the possessive for *accountant* (i.e. ‘of the accountant’; ‘the accountant’s’). Looking at the BNC-2014, the latter construction is not in evidence and the *of NP* occurs in 1.1% of all concordance lines. By contrast, Hoey shows that such constructions, in his corpus, are prevalent for *carpenter*. Indeed, the BNC-2014 shows both the genitive-*s* and the *of NP* constructions (the latter in 3.0% of all concordance lines). A final point Hoey makes is that *architect* is very frequently employed figuratively. Again, looking at 112 lines of *architect of* in the BNC-2014, the majority (101) are metaphorical (‘architect of her downfall’).

It is important that these collocational and colligational profiles of the target words are derived from, and cross-checked with more recent, more extensive data: on the one hand, the salience of Hoey’s claims can thus be supported. On the other hand, far more importantly, such detailed, qualitative reviews of word usage shine a light on how superficial and, in the end, imprecise the products of current LLM tools still can be. In this, the

17 As a side note, it must be mentioned that Hoey (2005) used 10 years of Guardian articles from the 1990s as his corpus. This research replicates his findings with more up-to-date material from a variety of sources. It must be stressed that the BNC-2014 allowed a close replication of Hoey’s results which makes the divergence found in generated material all the more relevant.

18 Lack of articles highlight how frequent ‘Carpenter’ as a surname appears in this corpus.

19 This is mainly due to 126 of the 1,133 concordance lines have *chartered accountant*. Sadly, there are only four *turf accountants* in this data set.

results shown here mirror Browning & LeCun's (2022) claim that using all available human knowledge

is just the wrong kind of knowledge for developing awareness or being a person. But they will undoubtedly seem to approximate it if we stick to the surface. While the sample sentences produced are coherent and follow a correct grammatical structure, they remain still very distant in their in-depth quality from the semantic and associative standard found in naturally produced sentences.

In a way, this already points towards the LLMs employing a fixed model of structures. Hoey (2005, 163) refers back to his earlier work from 1991 and makes the claim 'that lexis and text [are] organised, rather than structured'. Hoey accepts that there is a form of structure within language yet, crucially, highlights that this structure gets modelled and formed based on the *communicative need* – thus displaying a high degree of organisation. He denies, however, the existence of one single structural framework (a *grammar*):

Lexical priming does (...) assume that the grammars are never complete, because even the most thorough of grammar-creating language users must constantly encounter non-congruent usages produced by those without a fully integrated grammar (or occasionally, perhaps, without any grammar at all). (Hoey, 2005, p. 162)

As such, language is continually evolving as the user will adapt lexis and grammar for their needs and as listeners/readers may encounter novel applications. LLMs, however, seem confined to what already exists.

4. Communicative intent, Meaning-making and Priming

Pace-Sigge & Sumakul (2022) have already described how a longer AI-written text contrasts with the human-written equivalent through its over-use of certain words and constructions and through employing modifiers which seem atypical of a particular text-type. The evidence shown above seems to confirm suggestions that LLM tools have a number of fundamental flaws when it comes to producing longer texts involving staying true to text-type specific characteristics. These small-scale investigations enable to provide some answers to the question in how far LLM text generation and LPT are aligned. Where 'lexical priming' is understood in the broadest of terms – a word is primed to appear with certain other words in similar contexts and co-texts, there is clearly a match between what can be forecast based on LPT and the LLM-generated texts. However, Hoey's theory is far more multidimensional than that, and clearly postulates that primings are not only set by their probable lexical environment; amongst other things, the genre, pragmatics and intention form primings of a speaker/writer as well. Crucially, a number of key computational linguistics researchers highlight similar issues. Manning (2022, 130ff.), in presenting a study of human and machine language understanding, says that 'these [transformer neural network] models learn and represent the syntactic structure of a sentence and will learn to memorize facts of the world, since these things help the model to predict masked words more successfully'. The crux here being that the machine learns *sentence grammar* – the syntactic structure. The meaning and spe-

cific uses of any given word however, as described in Hoey's Lexical Priming Theory, exists only at the level of co-text. Manning points out how this can lead to problems: '[m]eaning is not all or nothing; in many circumstances, we partially appreciate the meaning of a linguistic form. I suggest that meaning arises from understanding the network of connections between linguistic form and other things, whether they be objects in the world or other linguistic forms' (Manning, 2022, 134). Similar sentiments are shown by Bisk et al. (2020) or Merrill and colleagues (2021) who speak of the 'limitations of acquiring meaning from an ungrounded form', whereby 'it is (real-world) experience which actually grounds language.' This appears to be in full agreement with Hoey, who claims that the primings of a (set of) word(s) is lodged with each individual user: 'Firth's notion of 'personal collocations' (1951) [as] it is an inherent quality of lexical priming that it is personal in the first place and can be modified by the language user's own chosen behaviour' (Hoey, 2005: 10).²⁰

Meaning-making as such goes beyond the purely linguistic expression and includes extra-lingual events (which are particular salient in spoken language) but also personal associations. Manning highlights that current AI tools rely fully on text data, yet, in order to improve their 'understanding' skills, these would need to be augmented with 'further sensory data': namely, visuals.²¹ We do see that this is the trajectory a number of AI-focused businesses (like Google and Open-AI) pursue. Yet, as Bisk and colleagues (2020) highlight, language is even more multidimensional. There is the *embodiment*, which is reference to metaphorical and less-literal expressions and, lastly, probably the most elusive of all qualities: namely that language is a key part of social interaction: '[m]eaning does not arise from statistical distribution of words, but from the use of people to communicate'. One has to, Bisk et al (2020) say, 'consider the contextual foundations of language: grounding, embodiment and social interaction'. Many of the assumptions and understandings on which communication relies lie outside of text. Furthermore, as Bender et al. (2021) highlight, LLMs use data which is fixed in time, in other words, static, whereas 'social movements produce new norms, language and ways of communicating' – thus, LLMs risk to reify older, value-locked, or indeed biased understandings. This links neatly with Hoey's assertion that 'grammars are never complete' (2005, p.162). Moving on from this, there are several critics, for example Bender and Koller (2020) or Hadfield (2022) who argue that LLMs lack a basic element of language, namely *communicative intent*. Furthermore, Hadfield (2022) looks at research into child language acquisition and highlights that '[i]nfants learn language by drawing on a wide range of cues, while LMs only train on the tiny slice of the world in their input texts'.²² When children are forced to use a more LLM-like learning process, limited to a single input modality and deprived

20 Hoey also points out that this aligns with Paul Hopper's notion of an *emergent grammar* (1998).

21 I put these concepts in inverted commas as the concepts are quoted from Manning. One might argue whether 'understanding' is an appropriate term when it comes to ML.

22 Bisk et al (2020) show a similar sentiment when they say that an LLM is 'no longer constrained to a single author or source, and the temptation for NLP is to believe everything that needs knowing can be learned from the written world. But a large and noisy text corpus is still a text corpus.' Havlik (2023) refers to it as 'LLMs understanding of language, while impressive, is shallow'

of social interaction, they fail to learn language’ (2020, p.6). This echoes Hoey who states that the ‘simple fact cannot be escaped that we do not think of a word and then start uttering, drawing as we progress on all the primings at our disposal. Self-evidently, we instead mostly start with a communicative need’ (2005:162). Furthermore, this is also a point raised by Bargh & Morsella (2008, 75) when they speak of the ‘acquisition of cultural knowledge’, or, as Havlik puts it: ‘mastery of natural language is a specific process that requires not only genetic prerequisites but also cultural and social conditions and, as a specifically developed ability, is attributed only to humans’ (2023, p. 2). Into this line of thinking fits Bender et al.’s (2021, p. 616) critique the idea that LLMs can produce longer coherent texts at all. Like Hadfield (2022) and Bisk et al. (2021) they contend that

[c]oherence is in fact in the eye of the beholder. Our human understanding of coherence derives from our ability to recognize interlocutors’ beliefs and intentions within context. That is, human language use takes place between individuals who share common ground and are mutually aware of that sharing (and its extent), who have communicative intents which they use language to convey, and who model each other’s mental states as they communicate. (Bender et al. 2021, p. 616)

Most damaging to the claim that LLMs can equal humans in acquiring a notion of meaning is the investigation made by AI researchers William Merrill and his international colleagues, who conclude the following:

While assertions allow a system to emulate semantic relations in simple cases where the semantics are referentially transparent, we find that linguistic constructs like variable binding bring this task in conflict with the fundamental laws of computability. it is not just intractable for an ungrounded system to emulate understanding of a formal language, but, in some cases, *impossible* [sic]. (Merrill et al., 2021, p. 1055)

Therefore, as a task is relatively straight-forward and is combined with prompts which are sufficiently unambiguous, an LLM will produce a passable representation of a coherent text. Beyond that, however, current systems still run the risk of being outmatched by the limits of their own frameworks, leaving them lost for suitable answers like the proverbial husband who finds his wife in the arms of her lover.

5. Conclusions

On the surface of it, the connections between Hoey’s Lexical Priming Theory and the technology that underlies current LLMs like Chat-GPT or Google’s BARD/Gemini are quite apparent. There are, for example, the origins of the concepts of *priming* and *spreading activation*, which were laid out as theory by R.M. Quillian, who aimed to create a machine which can comprehend human language input ‘Teachable Language Comprehender’ – an early form of AI. Quillian highlighted that he saw his model underlying this construct has to be understood as a theory of language (see Quillian 1967; Pace-Sigge, 2018). Furthermore, Hoey’s theory is grounded, like LLM models, in Firth’s dictum that ‘you shall know a word by its neighbours’, and the primings of words (sets of words) are

reflected in the collocations and colligations which are pervasive and statistically verifiable in human language, as found through corpus research. Based on the idea of spreading activation, Hoey's theory moves beyond mere sentence grammar; indeed, his description of *textual priming* extends forms of grammar to larger chunks of text. The notion, therefore, that coherent and apparently grammatically correct text can be produced by LLMs that follow a similar theoretical framework is a demonstration that Lexical Priming is a valid theory of language.

This notwithstanding, all this appears to happen on the surface of language use. Hoey contents that '[w]e have therefore to assume that the discursual impetus and the lexical priming are interconnected but not coterminous' (2005, p, 163). Critical voices have rightly pointed out that LLMs can replicate easily what is contained in its training data, while its lack of grounding, understanding of concrete meaning and communicative intent ensure that the texts produced are inherently flawed (and even a multi-modal system which delivers background knowledge beyond the merely textual may not be sufficient to address this issue). This is reflected by the insight given by Brynjolfsson (2022, p. 280) that AI systems can work extremely well when augmenting human endeavour – yet they are incapable of completing 100% of the necessary tasks. As the comparative experiments shown here have demonstrated, LLMs seem to act on too rigid a lexico-grammatical model. Consequently, the node words used here are too easily interchangeable, the collocational and colligational usage structures found in naturally occurring texts are only found to be replicated to a degree: the LLMs seem to be hyper-primed, leading to output which is coherent and structurally working only on the surface. Yet, at the same time, these lack the depth and precision, and the (relative) surefootedness of employing the right term in the right context and co-text (Hoey's nesting) that is only found in human-produced, naturally occurring texts, even if, for this paper, only a specific, small slice of such language has been employed.

Competing interests

The authors have no competing interests to declare.

References

- Arunachalam, H; Tang, X & Scott-Andrews, J (2023). Do LLMs really understand human language? *TMFormInform*.
<https://inform.tmforum.org/features-and-opinion/do-llms-really-understand-human-language> (last accessed 03/11/23).
- Bargh, J. A., & Morsella, E. (2008). The Unconscious Mind. Perspectives on *Psychological Science*, 3(1), 73-79. <https://doi.org/10.1111/j.1745-6916.2008.00064.x>
 PMid:18584056PMCid:PMC2440575

- Bender, E.M. & Koller, A. (2020). Climbing towards NLU: On Meaning, Form and Understanding in the Age of Data. *Proceedings of the 58th Annual Meeting of the Association of Computational Linguistics*. Pp. 5185-5198. <https://doi.org/10.18653/v1/2020.acl-main.463>
- Bender, E.M., McMillan-Major, A., Gebru, T. & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models be Too Big? *FAccT '21*. <https://doi.org/10.1145/3442188.3445922>
- Berber Sardinha, T. (2024). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics*, 4(1), 100083. <https://doi.org/10.1016/j.acorp.2023.100083>
- Brezina, Vaclav, Abi Hawtin & Tony McEnery. (2021). The Written British National Corpus 2014 – design and comparability. *Text & Talk* 41(5–6). 595–615. <https://doi.org/10.1515/text-2020-0052>.
- Browning, J & LeCun, Y. (2022). AI And the Limits of Language. *Noema*. <https://www.noemamag.com/ai-and-the-limits-of-language/> (last accessed 03/11/23).
- Brynjolfsson, E. (2022). The Turing Trap: The promise & peril of human-like artificial intelligence. *Daedalus, the Journal of the American Academy of Arts & Sciences*. 151(2), 272–287. https://doi.org/10.1162/daed_a_01915
- Curry, N., Baker, P., & Brookes, G. (2024). Generative AI for corpus approaches to discourse studies: a critical evaluation of ChatGPT. *Applied Corpus Linguistics*, 4(1), 100082. <https://doi.org/10.1016/j.acorp.2023.100082>
- Dickson, B. (2022). LLMs have not learned our language – we are trying to learn theirs. *VentureBeat*. <https://venturebeat.com/ai/llms-have-not-learned-our-language-were-trying-to-learn-theirs/> (last accessed 03/11/23).
- Futrell, R., Wilcox, E., Morita, T., Qian, P., Ballesteros, M., & Levy, R. (2019). Neural language models as psycholinguistic subjects: Representations of syntactic state. *arXiv preprint arXiv:1903.03260*. <https://doi.org/10.18653/v1/N19-1004>
- Granger, R. (2020). Toward the quantification of cognition. *arXiv preprint arXiv:2008.05580*.
- Hadfield, J. (2022). Why Large Language Models Will Not Understand Human Language. <https://jeremyhadfield.com/why-llms-will-not-understand-language/> (last accessed 03/11/23).
- Halliday, M.A.K. 1993. Towards a Language-Based Theory of Learning. *Linguistics and Education* 5, 93– 116. [https://doi.org/10.1016/0898-5898\(93\)90026-7](https://doi.org/10.1016/0898-5898(93)90026-7)

- Havlík, V. (2023). Meaning and understanding in large language models. *arXiv preprint arXiv:2310.17407*. <https://doi.org/10.1007/s11229-024-04878-4>
- Hoey, M. (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- Hoey, M. (1994). Patterns of Lexis in Narrative. In: Tanskanen, S-K. & Wårvik, B. (eds) *Topics and Comments. Anglicana Turkuensia.13*, pp. 1-41.
- Hoey, M. (1995). The Lexical Nature of Intertextuality: A Preliminary Study. In: Wårvik, B; Tanskanen, S-K. & Hiltunen, R.: Organisation in *Discourse. Proceedings from the Turku Conference. Anglicana Turkuensia.14*, pp. 73-94.
- Hoey, M. (2005). *Lexical Priming*. London: Routledge.
- Hoey, M. (2009). Corpus-driven approaches to grammar. In: Römer, U. & Schulze, R: *Exploring the lexis-grammar interface*. Amsterdam/Philadelphia: John Benjamins. pp. 33-47. <https://doi.org/10.1075/scl.35.04hoe>
- Hoey, M. (2017a). Foreword. In: Pace-Sigge, M. & Patterson, K.J. (eds.): *Lexical Priming. Applications and Advances*. Amsterdam/Philadelphia: John Benjamins.
- Hoey, M. (2017b). Cohesion and Coherence in a Content-specific Corpus. In: Pace-Sigge, M. & Patterson, K.J. (eds.): *Lexical Priming. Applications and Advances*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/scl.79.01hoe>
- Hopper, P. (1998). Emergent grammar. In M. Tomasello (ed.) *The New Psychology of Language*. NJ: Lawrence Erlbaum Associates, pp. 155-175. <https://doi.org/10.4324/9781315085678-6>
- Manning, C.D., Clark, K., Hewitt, J. Khandelai, U. & Levy, O. (2020). Emergent linguistic structure in artificial neural networks trained by self-supervision. *PNAS. Vol. 117 (48)*, pp. 30045-54. <https://doi.org/10.1073/pnas.1907367117> PMID:32493748 PMCID:PMC7720155
- Manning, C. D. (2022). Human Language Understanding and Reason. *Daedalus, the Journal of the American Academy of Arts & Sciences. Vol. 151 (2)*, pp. 127-138. https://doi.org/10.1162/daed_a_01905
- Merrill, W., Goldberg, Y., Schwartz, R., & Smith, N. A. (2021). Provable limitations of acquiring meaning from ungrounded form: What will future language models understand? *Transactions of the Association for Computational Linguistics*, 9, 1047-1060. https://doi.org/10.1162/tac1_a_00412
- Partington, A. (2014). Mind the gaps. The role of corpus linguistics in researching absence. *International Journal of Corpus Linguistics* 19 (1): 118-146. <https://doi.org/10.1075/ijcl.19.1.05par>

- Pace-Sigge, M. (2018). *Spreading Activation, Lexical Priming and the Semantic Web*. Abington: Palgrave Macmillan. <https://doi.org/10.1007/978-3-319-90719-2>
- Pace-Sigge, M. & Sumakul, T. (2022). What Teaching an Algorithm Teaches When Teaching Students How to Write Academic Texts. In Jantunen, Jarmo Harri, et al. *Diversity of Methods and Materials in Digital Human Sciences. Proceedings of the Digital Research Data and Human Sciences DRDHum Conference 2022*.
- Quillian, R. M. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioural Science*, 12(5):410-430. <https://doi.org/10.1002/bs.3830120511> PMID:6059773
- Shanahan, M. (2022). Talking about large language models. *arXiv preprint arXiv:2212.03551*.
- Sinclair, John (2004). *Trust the Text*. London: Routledge. <https://doi.org/10.4324/9780203594070>
- Valmeekam, K., Olmo, A., Sreedharan, S., & Kambhampati, S. (2022). Large Language Models Still Can't Plan (A Benchmark for LLMs on Planning and Reasoning about Change). *arXiv preprint arXiv:2206.10498*.
- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D. and Chi, E.H., (2022). Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.

Tools used

- Google [2023] (2024). *BARD/Gemini*. <https://BARD.google.com/chat>
- Brezina, V. & Platt, W. (2023) *#LancsBox X*. Lancaster University, <http://lancsbox.lancs.ac.uk>
- OpenAI. [2022] (2024) *ChatGPT (GPT 3.5)* <https://chat.openai.com/>
- Scott, M. (2023). *WordSmith Tools version 8*. Stroud: Lexical Analysis Software.