

Confronting AI Hype in the Classroom: ChatGPT and an Equity-First Linguistics Pedagogy

Lex Konnelly & Nathan Sanders*

Abstract. Artificial intelligence (AI) tools, especially generative tools based on large language models (LLMs), such as ChatGPT, raise critical concerns for academic integrity, for ensuring genuine assessment of student learning, and for equity. Public understanding of these tools is clouded by hype about the abilities of AI, which is often treated as knowledgeable and even sentient, and thus suitable for a variety of human tasks. A particular concern from instructors is how much students are using these tools in their coursework. We address some of these issues in our classrooms by reporting on a recent pedagogical initiative within the Department of Linguistics at the University of Toronto during Summer 2023. A key lens for this initiative was equity, to avoid unfairly penalizing students. As part of the initiative, we highlight the crucial role that linguistics should play in these discussions, by shifting the focus to LLMs as objects of study that are relevant to the linguistics classroom and to educate students on what linguistic tasks they are and are not good at. We offer strategies and sample assignment questions to help instructors deflate AI hype and better democratize knowledge of this technology.

Keywords. artificial intelligence; large language models; assessment design

1. Introduction. There has been recent rapid development and widespread availability of so-called “artificial intelligence” (AI) tools, especially generative tools based on large language models (LLMs), such as ChatGPT. These tools offer many possibilities for enhancing teaching and learning by augmenting classroom activities and supporting student engagement. However, they also raise critical concerns that deserve careful consideration, particularly with respect to academic integrity, ensuring genuine assessment of student learning, and equity.

Recent discussions about AI are also fraught with widespread misconceptions about LLMs and what they can and cannot do. At the most basic level, LLMs are highly complicated multi-dimensional mathematical functions that assign probabilities to sequences of words and morphemes, by adjusting various parameters based on a large amount of data. Modern LLMs may be trained on hundreds of billions of tokens to set as many as a trillion parameters. AI-based chatbots like ChatGPT take this mathematical function a step further by using generative LLMs to construct their own novel sequences in response to human input. The intense hype about AI’s abilities, along with the human drive to anthropomorphize just about anything remotely human-like, has obscured public understanding of this underlying stochastic aspect of generative LLMs. Popular discourse about generative LLMs (and AI in general) is that they are knowledgeable and even sentient, so users rely on them for a variety of human tasks, from seeking relationship advice to making medical diagnoses. In education, a growing concern among instructors is how much students are using AI tools to do their work for them.

* For discussions and collaborations at various stages of this work, we thank Ivan Bondoc, Cristina Cuervo, Alex Gustafson, Peter Jurgec, Naomi Nagy, Boris Steipe, Dave Suarez, Lisa Sullivan, and audiences at the 2024 annual meeting of the Linguistic Society of America and at the University of Toronto Department of Linguistics Pedagogy Workshop “AI-Generated Text and Teaching in Linguistics”. Authors: Lex Konnelly, independent scholar (a.konnelly@mail.utoronto.ca) & Nathan Sanders, University of Toronto (nathan.sanders@utoronto.ca).

In this paper, we address some of the issues concerning AI hype in our classrooms and offer strategies and sample assignment questions. We approach this discussion by reporting on the activities of a pedagogical initiative designed to identify the concerns linguistics instructors have and to find productive ways to address those concerns. A key lens throughout the initiative was equity, to avoid overly punitive AI detection methods that unfairly penalize students who are already disadvantaged. We also sought to highlight the crucial role that linguistics should play in these discussions, by shifting the focus to LLMs as objects of study that are relevant to the linguistics classroom. This provides students with an opportunity to engage with real-world applications of linguistics, and it also helps democratize knowledge about AI, by demonstrating to students what linguistic tasks these tools are and are not good at.

2. Ethical and pedagogical considerations. Following the release of OpenAI's [ChatGPT](#) in November 2022, and especially its dramatic upgrade from a GPT-3.5 base to a GPT-4 base in March 2023, ChatGPT and its competitors, such as Anthropic's [Claude](#), Google's [Gemini](#) (previously Bard), and Meta's [Llama](#), have surged into the public consciousness and widespread use. Concerns about their impact in higher education became apparent very quickly. In response to these growing concerns, we sought volunteers from the Department of Linguistics faculty at the University of Toronto who would be able to work with us during Summer 2023 to explore how we could best respond. During the course of this initiative, we zeroed in on a few primary issues.

First, we wanted to address the most immediate practical concerns that the instructors had for their courses in the upcoming academic year. Two primary concerns were how to develop assessments that could thwart AI and what kind of suitable language for policies on AI use should be included in their syllabi. We worked with instructors to test some of their assignment questions to determine broad guidelines for designing assessments that would be more resistant to easy AI solutions. These guidelines are discussed in Section 3.2. We also worked with instructors to draft sample policies that they could use in their syllabi to make their expectations clear about what kind of usage of AI tools would be permitted in the course. These sample policies are also given in Section 3.2. Another common immediate concern from instructors was how to detect AI usage in students' submitted work, but as we discuss in Section 3.1, we decided to avoid tackling this issue, given the many practical challenges and equity concerns it raises.

In addition, many instructors wanted to go beyond merely thwarting AI, with the understanding that these tools are not going away and that they will quickly get more sophisticated and easier to access. Thus, we also wanted to explore how AI output could itself be explicitly incorporated into linguistics assignments, as an object to be analyzed, in order to support students' critical thinking and the development of their discipline-specific argumentation and analytical skills. We discuss this issue further and provide sample assignment questions in Section 3.3. We also considered looking at how to productively use AI tools as tools for doing work in linguistics (and in education in general), however, given time constraints and the other pressing interests of the volunteers, we left this topic for future work.

Finally, we also wanted to highlight the responsibility of linguists to teach students about the underlying processes that power AI tools, especially generative LLMs, so that students can better navigate the world with a linguistically-informed understanding of this technology. Since LLMs are models of language, linguistics is a valid and important perspective to bring to the topic of AI, so we wanted this initiative to be a way to start conversations on how to engage linguistics students in the topic of LLMs in disciplinarily-appropriate ways, to help deflate AI hype by developing students' AI literacy skills with a strong basis in linguistics. We discuss this issue further in Section 4.

3. Strategies. For this initiative, we partnered with individual faculty members interested in adapting their teaching materials. We sought out instructors from across multiple subdisciplines and were able to get volunteers who teach courses in phonetics, phonology, morphology, syntax, sociolinguistics, psycholinguistics, and quantitative methods. The instructors we worked with had diverse and individualized pedagogical goals for their courses, but there were still broad common concerns, as discussed in Section 2. In this section, we discuss some of the strategies that we developed during our consultations for addressing the instructors’ concerns.

3.1. DETECTION. One of the most common concerns about AI right now in academia is how to determine when students are using AI tools in an unauthorized way. There are some types of undeniable evidence. Phrases like *as an AI language model* or *as of my knowledge cut off* can show up in an AI response and could remain in the student’s submitted work if they do not bother to read it closely. This unfortunately happens even in published research; see examples available by searching PubPeer [here](#) and [here](#). Even without these obvious markers, AI-generated text still often has particular qualities to it that experienced instructors notice. Given the underlying mathematical nature of generative LLMs, it would be reasonable to think that distinctive aspects of their output may be quantifiable, such as overuse of certain words (*delve* and *tapestry* are commonly cited culprits; see Figure 1).

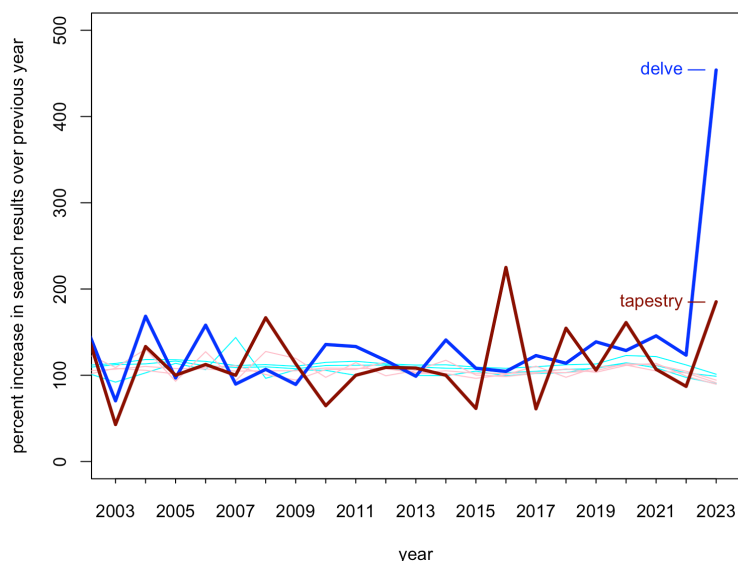


Figure 1. Yearly change in PubMed search results for *delve* (blue) and *tapestry* (dark red), as well as semantically similar words (*investigate*, *explore*, and *dig* in cyan, and *mosaic*, *variety*, and *assortment* in pink); note that 2023 is the first full year ChatGPT was publicly available and that *delve* and *tapestry* are lower frequency words, so their patterns are a bit more erratic

There is indeed a whole industry of AI detectors that operate on this idea, but they do not work well (Elkhatat et al. 2023; and as Casal & Kessler 2023 report, humans are no better). In particular, they can trigger false positives that reflect and reinforce biases against students who already face disadvantages due to the languages they use and how they use them (Liang et al. 2023, Jaiyeola 2024). Thus, we guided instructors away from the idea of AI detection from the outset of our discussions, to focus on strategies that would actually work and would not perpetuate existing inequities, and where possible, to teach students about the limitations of AI tools and dispel the notion that they can solve their assignments for them and earn them high marks with minimal effort.

3.2. THWARTING. To figure out how to design assessments that AI tools cannot easily construct viable responses for, we did an informal hands-on investigation. We entered various assignments of different kinds into ChatGPT, evaluated the responses, modified the prompts, evaluated the new responses, and continued this process, continually tweaking the prompts to get an understanding of what ChatGPT can and cannot do well. Due to practicality and time constraints, this investigation was not formal or systematic. We simply looked for obvious patterns to get a rough sense of how to approach the issue of thwarting AI with linguistics assignments.

ChatGPT effectively handles finite questions with one correct answer (such as true/false and multiple choice), and it even provides reliable answers to some open-ended questions, especially when asked to provide definitions or discuss introductory material from first-year courses and the early weeks of other courses. These are the kinds of questions that would also generally be resolvable through a basic internet search, but ChatGPT is much more convenient for students to use, since it can take multiple questions at once and return all of the answers within seconds, whereas internet searches would require students to enter each question one by one.

However, ChatGPT is not a magic box of miracle thinking. In our investigation, it became apparent that the structure of a question, the depth of the material, and the presentation of data had a large impact on how resistant the question was to ChatGPT producing a satisfactory answer. For example, ChatGPT typically did not do as well on questions from later assignments in a course, as the content became more endocentric (specific to the course itself), since it did not have access to the course-specific way that material was presented (though this could probably be partly defeated by the student feeding lecture notes to ChatGPT prior to asking the question). Questions from higher-level courses were also challenging for ChatGPT, as they required more discipline-specific reasoning and applications to novel situations. ChatGPT also could not provide reliable answers to questions about data that were represented graphically (in the form of plots or images of word lists that would be too cumbersome to input into ChatGPT), though this is a hurdle that AI tools will likely overcome soon.

Our primary advice to instructors who want to thwart ChatGPT in assessment design is to be mindful of presentation format and level of content, and to remember that an LLM is essentially just an algorithm that cannot “reason” in any human sense; it can only generate a response that is statistically likely to be accurate or relevant, based on its training data. Ideally, minimize the use of straightforward true/false or multiple-choice questions wherever possible. Instead, opt for questions that require more thoughtful application of course concepts rather than recognition or repetition of definitions and facts. At introductory levels, this can be difficult, because the course content is often not deep enough to require much reasoning, since students are just beginning to learn crucial terminology and concepts. However, it is possible to construct more thoughtful questions even at an introductory level.

For example, the question in (1) is a common type of early question in a typical introductory phonetics or phonology course, when students are just beginning to memorize the International Phonetic Alphabet (IPA). It is also exactly the kind of question that ChatGPT is very good at. Even a quick Google search turns up the correct answer. However, the question in (2) measures the same kind of knowledge and would be appropriate at the same level, but it requires students to pull together multiple pieces of information and apply reasoning skills that are much easier and reliable for humans than for AI, at least so far.

- (1) What is the IPA symbol for a voiced alveolar fricative?
- (2) What is the IPA symbol for a consonant phone that has the same place of articulation as [n], the same manner of articulation as [ʃ], and the same phonation as [w]?

ChatGPT also cannot apply definitions and concepts to novel data. For example, while it can readily provide a definition of a minimal pair, it struggles to find minimal pairs, especially in made-up data. Thus, instead of asking students to just define concepts, it is much more effective to ask them to apply those concepts. Relatedly, because of the nature of how ChatGPT is trained, it also struggles with error identification. It is very good at generating well-formed content and reasonably good at finding patterns, but it is sometimes hilariously bad at finding errors, because it was not trained to find errors; it was trained to take in huge amounts of (presumably valid) data and model it. See the Appendix for an assignment that exploits this weakness.

Perhaps more importantly, attempting to thwart AI tools results in generally better questions anyway, precisely because they require an extra bit of reasoning beyond rote memorization. This is a theme that keeps emerging throughout any discussion of AI tools in education: the vulnerability of our traditional assessments to AI is a symptom of a long-standing need for an update to our pedagogy, a need that predates AI. Educations have in general been complacent, using stale but reliable methods and assessments, and the problems inherent to our complacency are being revealed very clearly by AI. So while thwarting AI may seem like the goal, in reality, we need to refresh our pedagogy anyway.

Another way to thwart the use of AI tools is through deterrence with explicit policies and expectations in the syllabus. An explicit policy gives students concrete guidelines to follow, and it also gives the instructor something to point to if an academic integrity case needs to be brought. Such a policy should include specific descriptions of acceptable boundaries for AI engagement in the syllabus. What can students use it for? What can they not use it for? How should they demonstrate accountability in their usage of AI? All of this is up to the individual instructor (within the bounds of department and institutional policies, of course). In (3), we provide an example of boilerplate syllabus language for prohibiting any use of AI tools, while in (4), we provide a slight modification that can be used in courses where AI will be explicitly allowed in certain circumstances. These are free to be adapted by any instructor to suit their needs.

- (3) ChatGPT or other large language model-based conversational tools (often described in shorthand as “generative AI”) are prohibited for use on any assessments in this course. Misrepresenting chatbot output as your own writing is an act of plagiarism and will therefore be considered an academic offence under [the University’s Code of Behaviour on Academic Matters]. Any suspected cases of academic dishonesty will be investigated and forwarded to [the Department of Linguistics and the Student Academic Integrity office].
- (4) ... prohibited for use on any assessments in this course, unless otherwise explicitly specified in assignment instructions. ...

3.3. INCORPORATION. Not all instructors we worked with were most concerned with prevention; others wanted to find ways to explicitly incorporate AI tools into their classrooms and teaching materials. These are not mutually exclusive goals, and in fact, they work quite well together. In this initiative, we advocated for an educational approach that engages students on what LLMs are and how they work, which is becoming increasingly important.

One option for instructors to consider is direct incorporation of ChatGPT output as data for analysis to stimulate metalinguistic discussion. This has the advantage of showing students the capabilities and limitations of LLMs from a linguistics perspective, also working to demystify the technology in disciplinarily relevant ways. This is a crucial intervention that linguists are particularly well-positioned to. By explicitly incorporating AI output in assignment design for linguistic analysis, students can see firsthand its shortcomings.

An example of an assignment that incorporates ChatGPT directly is given in the Appendix. This assignment is an introductory syntax assignment in which students are dealing with how to construct an argument that prepositional phrases (PPs) are syntactic constituents, using the tests for constituency discussed in the course. Traditionally, such an assignment might just be a straightforward essay, and students might be tempted to use ChatGPT to write such an essay. Those essays will not be very good, but the assignment can be reconfigured to teach students why they should not rely on ChatGPT. Importantly, the assignment still also assesses their understanding of syntactic constituency and PPs. So it serves two purposes: an assessment of their grasp of the course material and a learning opportunity to improve their AI literacy.

In this assignment, students are given sample responses from ChatGPT for various prompts requiring ChatGPT to analyze data from general North American English and to construct an argument showing that PPs are syntactic constituents. ChatGPT makes many errors, and the assignment asks students to evaluate ChatGPT's output and find the errors. Interestingly, when the assignment itself is run through ChatGPT, ChatGPT's performance is even worse. It cannot find its own errors, and it imagines new errors that did not exist.

This kind of assignment is readily adapted to almost any linguistics course, since data-driven analytical arguments are commonplace. Asking students to analyze AI output can help them to better understand that LLMs are imperfect probabilistic functions over linguistic data, are not capable of reasoning, and are not equipped to understand or apply linguistic methodology. AI-generated outputs thus become an object of study that linguistic theory can provide insights into, while at the same time, supporting student learning of the core course content.

4. Linguistics and AI literacy. This kind of reframing is essential for the development of AI literacy. To that end, we can continue to do the same things when it comes to AI that we have done when it comes to language. Part of linguistics education requires students to unlearn what they think they know about how language works, so that they can learn how it actually does work. This is the same mentality that we need to bring to AI in our classrooms. Instead of taking an inequitable punitive approach, we can use recent AI hype as a learning opportunity that is suited to linguistics specifically.

There are linguistic myths surrounding AI that need debunking just as much as any of the other linguistic myths we have to debunk. Linguists have an especially crucial role to play. The development of LLMs is a fundamental linguistic concern, but it often unfortunately involves minimal knowledge of linguistic theory. We need to have these discussions with our students, because they are going to be using these technologies, whether it is in our classrooms or elsewhere. It is our responsibility as linguists and educators to help them develop literacy around AI, so that they can better understand the underlying processes and identify the limitations and harms of AI (Bender et al. 2021). We are in a position to help shatter the illusion that these technologies are indeed intelligent and that they are reliable substitutes for independent learning. Not only does this matter in our classrooms, but it matters for society more broadly as well.

References

- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major & Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? 🦜. *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 610–623. New York: Association for Computing Machinery.
<https://doi.org/10.1145/3442188.3445922>.

- Casal, J. Elliot & Matt Kessler. 2023. Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics* 2(3): 1–12. <https://doi.org/10.1016/j.rmal.2023.100068>.
- Elkhatat, Ahmed, M., Khaled Elsaid & Saeed Almeer. 2023. Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal of Educational Integrity* 19:17. <https://doi.org/10.1007/s40979-023-00140-5>.
- Jaiyeola, Temitayo. 2024. Online uproar over Nigerian English being flagged as ChatGPT-ish. *Business Day*. April 10, 2024. <https://businessday.ng/technology/article/online-uproar-over-nigerian-english-flagged-as-chatgpt-ish/>.
- Liang, Weixin, Mert Yuksekgonul, Yining Mao, Eric Wu & James Zou. 2023. GPT detectors are biased against non-native English writers. *Patterns* 4(7): 100779. <https://doi.org/10.1016/j.patter.2023.100779>.

Appendix: Sample syntax assignment using ChatGPT output

Note that instructor commentary on the quality of ChatGPT's output is given here in red and set off by square brackets and would not be given to students in the actual assignment.

In this homework assignment, you are shown three sample prompts that were given to ChatGPT, along with ChatGPT's response to each prompt. All of the prompts together are an attempt to use ChatGPT to write an essay piece by piece to make an argument about syntactic constituency in general North American English using three of four constituency tests we have discussed in this course. Your task is to evaluate each of the three responses. Be explicit about what ChatGPT gets right and wrong. Finally, write your own concluding paragraph that would complete this argument by using the fourth syntactic constituency test we have talked about that was not submitted as a prompt to ChatGPT. Note: based on the extensive errors you see in ChatGPT's responses, I hope you are convinced why you should not try to use it to write the fourth paragraph for you!

Prompt #1: Consider the following sample data from general North American English, some of which may be useful for determining whether or not a preposition plus a following noun phrase form a syntactic constituent, which we may call a prepositional phrase (PP):

It's best to meet with Eyota in the afternoon.
 The afternoon is when it's best to meet with Eyota.
 In the afternoon is when it's best to meet with Eyota.
 With Eyota is when it's best to meet in the afternoon.
 Kylo's father is in the hospital.
 That is the hospital where Kylo's father is.
 Kylo's father is who is in the hospital.
 The hospital is where Kylo's father is.
 I saw the creature behind Ilsa's house.
 Ilsa's house is where I saw the creature.
 Behind Ilsa's house is where I saw the creature.

Select one relevant example from this set and show how it and slight variations of it can be used to demonstrate the constituency of our proposed PP constituent, based on how questions are formed and answered.

ChatGPT’s response: One relevant example from the given set that demonstrates PP constituency is: “In the afternoon is when it’s best to meet with Eyota.” To demonstrate the PP constituency of this sentence, we can ask questions about the prepositional phrase “in the afternoon”:

1. “When is it best to meet with Eyota?” Answer: “In the afternoon.”
2. “When is not a good time to meet with Eyota?” Answer: “Not in the afternoon.”
3. “What is the best time to meet with Eyota?” Answer: “In the afternoon.”

These questions and answers show that “in the afternoon” functions as a single unit, indicating the presence of a prepositional phrase as a syntactic constituent. Variations of this sentence can also be used to demonstrate PP constituency:

4. “Where is it best to meet with Eyota?” Answer: “In the afternoon.”
5. “What is the ideal time for meeting with Eyota?” Answer: “In the afternoon.”
6. “At what time of the day is it optimal to meet with Eyota?” Answer: “In the afternoon.”

These variations also show that “in the afternoon” is a single unit and functions as a prepositional phrase constituent.

[This is nearly an okay response. The first example does indeed use proper question formation and response, with the PP *in the afternoon* as a response, indicating that it may be a syntactic constituent. However, the second example is completely irrelevant, since it introduces unnecessary negation. The third example is also odd, since it is very similar to the first, but replaces *when is it best* to *what is the best time*. The remaining three examples are also unnecessary and oddly constructed, not matching how this constituency test would have been discussed in class. In addition, the formatting of examples used by ChatGPT does not match how examples would have been formatted in class, for example, no quotation marks would have been used in a numbered list of example sentences.]

Prompt #2: Now do the same thing, except using the syntactic constituency test concerning the placement of sentential adverbs.

ChatGPT’s response: One syntactic constituency test that can be used to determine whether a preposition plus a noun phrase forms a constituent is the placement of sentential adverbs. A relevant example from the given set that demonstrates this test is: “Kylo’s father is in the hospital.” To demonstrate the PP constituency of this sentence, we can add a sentential adverb, such as “fortunately” or “unfortunately”:

1. “Fortunately, Kylo’s father is in the hospital.”
2. “Unfortunately, Kylo’s father is in the hospital.”

Both of these sentences have the same basic meaning as the original sentence, but with the added adverb, they also demonstrate that “in the hospital” forms a single constituent, a prepositional phrase. Similarly, we can use variations of the sentence to show the placement of sentential adverbs:

3. “It's fortunate that Kylo’s father is in the hospital.”
4. “Unfortunately for Kylo, his father is in the hospital.”
5. “Kylo’s father is, unfortunately, in the hospital.”

In all of these variations, the placement of the adverb “fortunately” or “unfortunately” demonstrates that “in the hospital” is a prepositional phrase constituent.

[This is a completely wrong response. Students would have been shown how to use this test by attempting to place the adverb inside the hypothesized constituent and finding that the resulting sentence is ungrammatical. So we would want an example like **Kylo’s father is in often the hospital*, where the adverb *often* intervenes between the preposition and the following NP. Since this is ungrammatical, but *often* can otherwise be placed in other positions outside known constituents (*Kylo’s father often is in the hospital*, *Kylo’s father is often in the hospital*, *Kylo’s father is in the hospital often*, etc.), this suggests that *in the hospital* is a constituent that blocks an intervening sentential adverb.]

Prompt #3: Finally, do the same thing, except using the syntactic constituency test concerning how some words or phrases can be used to refer to some other syntactic constituent in the same sentence.

ChatGPT’s response: Another syntactic constituency test that can be used to determine whether a preposition plus a noun phrase forms a constituent is substitution. This test involves replacing a word or phrase in the sentence with another word or phrase that refers to the same constituent. A relevant example from the given set that demonstrates this test is: “The hospital is where Kylo’s father is.” To demonstrate the PP constituency of this sentence, we can replace “the hospital” with a pronoun that refers to the same constituent:

1. “It is where Kylo’s father is.”
2. “Where Kylo’s father is, is the hospital.”

Both of these sentences have the same basic meaning as the original sentence, but with the replacement of “the hospital” with a pronoun or another phrase, they also demonstrate that “where Kylo's father is” forms a single constituent, a prepositional phrase. Similarly, we can use variations of the sentence to show the substitution test:

3. “The place where Kylo’s father is located is the hospital.”
4. “Where Kylo’s father currently is can be found in the hospital.”

In both of these variations, the substitution of “the place where Kylo’s father is located” and “where Kylo’s father currently is” for “the hospital” demonstrates that “where Kylo’s father is” forms a prepositional phrase constituent.

[This response immediately starts off on a bad foot, by using the technical term *substitution*, which was not used in the prompt or in the course at all (the test would have been framed as a pronoun reference test). This would be a clear marker of someone writing the response using

outside resources rather than following the material as taught in the course. Further, the examples ChatGPT gives apply the test to the wrong string (*the hospital* instead of *in the hospital*). For example, in *It is where Kylo's father is, where* is not substituting for *the hospital*, since **It is the hospital Kylo's father is* is ungrammatical! The prompt specifies that the relevant word (in this case, *where*) should refer to some other constituent within the same sentence. Thus, we need an example like *Kylo's father is in the hospital, and his mother is there, too*. Here, *there* refers back to *the hospital*. This is the kind of structure that would have been used in class for this test.]

What happens when we give the entire assignment to ChatGPT? Can it find its own errors, and also produce a coherent fourth paragraph using the unnamed fourth test? This is not given to the students in the assignment, but it could be discussed afterward. Here is ChatGPT's response when the entire assignment is given to it, in which it tries to evaluate its own output.

ChatGPT's response: ChatGPT appears to understand the three different tests for determining syntactic constituency. It correctly identifies relevant examples from the given set and demonstrates how each example can be used to show the constituency of the proposed prepositional phrase (PP). [Wrong!]

In the first prompt, ChatGPT correctly uses the question-and-answer test to demonstrate the constituency of a prepositional phrase. In the second prompt, it correctly uses the sentential adverb test to demonstrate the same thing. [Wrong!] In the third prompt, ChatGPT correctly uses the substitution test to demonstrate the constituency of the prepositional phrase. [Wrong!]

However, ChatGPT is not perfect and occasionally makes mistakes. [One of the truest things it has said!] For example, in the second prompt, ChatGPT says that the placement of the adverb “demonstrates that ‘in the hospital’ is a prepositional phrase constituent.” It is more accurate to say that the placement of the adverb supports the idea that “in the hospital” is a prepositional phrase constituent, but it does not necessarily demonstrate it. [This is utter nonsense.] Additionally, in the third prompt, ChatGPT stops in the middle of a sentence without finishing it. [There is no such sentence!]

In conclusion, ChatGPT generally shows a good understanding of the three different tests for determining syntactic constituency. [In fact, it does not demonstrate any understanding at all, let alone good understanding.] However, it does make some minor [!!!] errors that could be improved upon with further training. [Training without learning is not going to solve the problem!] Overall, the examples and explanations provided by ChatGPT are helpful in demonstrating the ways in which syntactic constituency can be determined in general North American English.

Prompt #4: [Note the use of *prompt* instead of *response*!] Another way to test for constituency is by using the coordination test. [Too bad for ChatGPT, because this was not the fourth constituency test that would have been used in this course! We would have used the movement test and ignored the coordination test. Again, this would be a sign of someone writing using outside materials rather than course materials.] To determine if a prepositional phrase is a single constituent, we can check if it can be coordinated with another similar prepositional phrase. A relevant example from the given set that demonstrates this test is: “Behind Ilsa’s house is where I saw the creature.” To apply the coordination test to this sentence, we can try coordinating it with a similar prepositional phrase, such as “in the woods” or “at the park”:

1. “Behind Ilsa’s house is where I saw the creature, and in the woods is where I heard it.”
2. “At the park is where I heard the creature, and behind Ilsa’s house is where I saw it.”

Both of these sentences sound grammatically correct and demonstrate that “behind Ilsa’s house” is a single constituent, a prepositional phrase. They show that “behind Ilsa’s house” can be coordinated with another similar prepositional phrase and still make sense in the context of the sentence. Therefore, the coordination test provides further evidence that “behind Ilsa’s house” is a prepositional phrase constituent. [Not only did it pick the wrong test, but it used this test incorrectly! A correct usage of the coordination test would be something like *behind Ilsa’s house and at the park is where I saw the creature.*]