

LEL2A: Syntax

Semester 1, 2025–26

Research Skills Assignment

1 Introduction

In addition to the Syntax-Skills that are taught and tested throughout the course, we've identified 5 big-picture cognitive skills that are useful in the study of syntax and throughout linguistics. These five skills were chosen as they represent the way in which linguists approach new data and problems. They also align with the way ideas in syntax, and other fields, are communicated. The five RESEARCH-SKILLS for the course are:

- M1 State a hypothesis clearly
- M2 Present data in support of a hypothesis
- M3 Define the conditions under which a hypothesis would be refuted
- M4 Present data to refute a hypothesis
- M5 Consider the limitations of a diagnostic test when interpreting the results

The remainder of this section gives an example of what these look like in the context if material we've already covered, followed by submission guidelines. Section 2 elaborates some more on what the skills mean in the context of LEL2A. Section 3 gives a few examples of possible topics for a *Research Skills* submission. Refer to the rubric on Learn to see how these skills would map onto your final mark for the syntax block.

1.1 Quick example

In Topic 6 we established the VP-Internal Subject Hypothesis. We can recap the steps we took, or could take, according to the five Research Skills:

- M1 **State a hypothesis clearly:** The subject of a verb starts in the specifier of VP (for selection), and then moves to the specifier of IP (for agreement, Case, or to satisfy the EPP, depending on the analysis).
- M2 **Present data in support of a hypothesis:** We saw evidence from idioms and quantifiers. Idioms showed us that only the VP-internal material is relevant to maintaining the idiomatic reading. Quantifiers showed us that the subject might leave its quantifier behind, in Spec,VP, when it raises to Spec,IP.
- M3 **Define the conditions under which a hypothesis would be refuted:** We would need to find examples where the subject doesn't raise from the VP, even though it should. Or, alternatively, a *weaker* refutation would consist of providing a different analysis of the data that supported the hypothesis. For example, if we had a different explanation for why quantifiers can be in Spec,VP when their associated noun is higher.
- M4 **Present data to refute a hypothesis:** Why should *There arrived three linguists* be grammatical when we could simply say *Three linguists arrived*? We could also look at other varieties of English; in Belfast English, imperatives might look like *Open you that door!* What's going on there? If we're looking for a slightly weaker challenge, we can ask why the quantifier can appear in the position it does in the following example, which seems to have nothing to do with the VPISH: *The children would have all been doing that.*

M5 Consider the limitations of a diagnostic test when interpreting the results: For the Belfast Irish imperatives, we'd need to think about whether the verb raises from V to I, in which case the subject might still be lower. For floating quantifiers, we'd need to say something about the conditions under which we think they can or cannot move around.

1.2 Submission guidelines

Submit a short write-up, no longer than one page, via the submission box on Learn. Explicitly flag the Research Skills you are trying to demonstrate. You may consult (and reference!) any primary literature you want, although ideally you'd want to try and think through the issues yourself, rather than provide a literature review.

The final deadline is the regular deadline for the syntax block: 12 noon on the Thursday of week “12”, 4 December. You may submit a first attempt by the Thursday of week 10, 20 November, if you'd like; you'll get feedback on that attempt within about a week and the chance to resubmit for the final deadline. It's recommended that you get in touch with Itamar before submitting anything to discuss your topic.

2 Doing syntax

We study language, including syntax, because we want to know more about how people work. Unfortunately, people's minds are black boxes; we cannot see what's going on inside them. However, language is a tangible output of the the mind and so understanding language gives us a means of understanding the unobservable.

We said in Topic 1 that we were interested in regularities relating to: the linguistic knowledge, and linguistic behaviour, of groups of humans.

The fact that all normal [*sic*] children acquire essentially comparable grammars of great complexity with remarkable rapidity suggests that human beings are somehow specially designed to do this, with data-handling or ‘hypothesis-formulating’ ability of unknown character and complexity. (Chomsky 1959)

Despite some variation between speakers, there is undeniable similarity between the language common to a group of speakers. Speakers make similar acceptability judgements and uniformly reject plausible alternative structures as ungrammatical. As the quote above argues, the acquisition of these common grammars happens quite early and without full exposure to all possible grammatical and ungrammatical sentences.¹ One of the big contentions of GENERATIVE SYNTAX is that humans have an inbuilt capacity for learning languages.

This brings us back to the black-box above; we cannot observe any inbuilt capacity for language, only its output.

The important question is: What are the initial assumptions concerning the nature of language that the child brings to language learning, and how detailed and specific is the innate schema (the general definition of “grammar”) that gradually becomes more explicit and differentiated as the child learns the language? For the present we cannot come at all close to making a hypothesis about the innate schemata that is rich, detailed, and specific enough to account for the fact of language acquisition. Consequently, the main task of linguistic theory must be to develop an account of linguistic universals that, on the one hand, will not be falsified by the actual diversity of languages and, on the other, will be sufficiently rich and explicit to account for the rapidity and uniformity of language learning, and the remarkable complexity and range of the generative grammars that are the product of language learning. (Chomsky 1965)

¹The argument hinted at here is the *Poverty of the Stimulus* argument and I haven't really done it justice here. If you want to know more, Chomsky's (1986) makes a good argument for it.

The job of syntax is to describe and account for the patterns and constraints we see in natural languages. By describing what one language can or can't do (and then comparing this against what other languages can or can't do), we can arrive at an idea of what *Language* can or can't do. When we find these *linguistic universals*, as they are referred to in the quote above, it helps us infer information about the *innate schema* inside the black-box.

In order to exemplify the RESEARCH-SKILLS above, below we will look at one of the *linguistic universals* described in the first week of the course, CONSTITUENCY, and how we would set out arguing for it.

2.1 Forming a Hypothesis and providing the positive argument

Skill: M1, M2

Often, linguistic work starts from a simple observation. From the examples below, we can make the observation: *some words seem to lump together, they can't be separated, and while the whole chunk can be replaced, partial substitution is disallowed.*

- (1) a. I think she was here yesterday.
- b. I think the woman was here yesterday.
- c. *I think the she was here yesterday.
- d. It is the dog that is chasing the cat.
- e. *It is dog that the is chasing the cat.
- f. I gave Jill a key.
- g. I did so too. (*give Jill a key*)
- h. *I did so a key (too). (*give Jill*)

The next step in an analysis is to try to propose an *explanation* for the observation we've made:

(2) **Hypothesis:**

There are structural units to language, CONSTITUENTS, that group lexical items together within the larger sentence.

If language is structured in CONSTITUENTS, we have a reason for the data above. Substitution and displacement work on certain constituents. If our hypothesis pans out, this also points to a possible *linguistic universal* as described above.

In practice, steps M1 and M2 often happen together, in that data leads to a hypothesis. Sometimes, though, theoreticians first think of an appealing idea (M1) and then look for the supporting evidence.

2.2 Identifying the alternatives & refuting them

Skill: M3, M4

We have a hypothesis (2) but we should ask if it's necessary. That is, can our observation be accounted for under an alternative or NULL HYPOTHESIS. For our example, a reasonable NULL HYPOTHESIS might be that language has no structure and whether a word can appear in a sentence is determined entirely by the word that immediately precedes it. We can actually account for some of the data above in (1) this way. If articles precede nouns but not pronouns, then the contrast in (1a-1c) is accounted for. Likewise, if articles cannot precede verbs, we have explained (1d-1e), and if *did so* cannot license an article, (1f-1h) is explained. Our current data doesn't help us.

In order to support our hypothesis, we need better examples to refute our NULL HYPOTHESIS. We could attempt to go through every possible example exhaustively, and show that there is no way to formulate rules like those above that captures every preceding word-following word pair while also excluding those that are impossible, but that would be very labour intensive. What we really need is some way of showing action over arbitrarily long distances or that there is some other distinction the NULL HYPOTHESIS fails to capture.

(3) Q: Which rakes should we buy?

A1: Those really really long ones in the black box are the best.

A2: * That really really long ones in the black box is the best.

(4) a. I fed the baby with a spoon.

b. I fed the baby with a spoon and she did so too.

c. I fed the baby with a spoon and she did so with a fork.

In (3), we see something touched on in the course, AGREEMENT. The noun *rakes* can be substituted for *ones* but the demonstrative on the left edge of the sentence must AGREE with it for number. Similarly, the verb *be*, must also reflect the number of *rakes*, despite the immediately preceding noun, *box*, being singular. This type of agreement cannot be accounted for if only the preceding word matters. We can insert arbitrarily long distances between the determiner and noun, *those really really long and really really thin rakes*, and between *ones* and *be*, *the ones in the black box with a green sticker are the best*. Likewise, in (4) we see another pattern from the course that cannot be captured by the NULL HYPOTHESIS, STRUCTURAL AMBIGUITY. The example (4a) has two interpretations, one where I use a spoon to feed a baby (*tool reading*) and another where I feed the baby that is holding a spoon (*possession reading*). In (4b), we see that *do so* substitution of *fed the baby with a spoon* preserves both readings. However, in (4c), substitution of only the string *fed the baby with do so*, causes the *possession reading* to be lost. This ambiguity and loss of readings is difficult to account for without structure. We've shown that there are problems with the NULL HYPOTHESIS.

2.3 Understanding our tools

Skill: M5

Each assumption or argument can in principle open up a number of additional questions. For example, the existence of AGREEMENT ATTRACTION indicates that there's something about the human parser that does pay attention to linear order; and if you look up CLOSEST CONJUNCT AGREEMENT, you'll find additional cases where it really looks like linear order, rather than hierarchical structure alone, determines agreement. So we always want to be aware of the limitations of both our tools and our knowledge.

3 Suggested Topics

Below are some suggested topics based on phenomena that relate to topics covered on the course. You are free to choose one of these topics or suggest your own. There is no penalty for choosing one of the suggested topics and no bonus for suggesting your own.

3.1 Double objects

In the Topic 10 course notes, you were introduced to the idea of verbal heads between V and I. In the notes this was called *vP*. It was given as a way of preserving BINARY BRANCHING, but due to a lack of space we didn't have time to motivate it. These extra functional heads are usually discussed in the context of DOUBLE OBJECT VERBS and we saw them in the context of OBJECT CONTROL.

(5) They [_{vP} persuaded [_{VP} persuaded *them* [_{CP} PRO to answer the questions]]].

Let's start again with an observation about some natural language data and how it relates to our theory.

(6) a. John [_{IP} did [_{NegP} not [_{VP} walk the dog]]]

b. The dog [_{IP?} was [_{NegP} not [_{VP} walked the dog]]]

c. The dog has not been walked

So far, we have said that given a sentence like *John did not walk the dog*, we might expect a structure like (6a). In Topic 11, we extended this to deal with passives like *the dog was not walked*, resulting in a structure like (6b). However, we then have examples like (6c). If *walk* is the verb and *has* is in the I head, where is *been*?

This may seem at first like a notation convention problem, but BINARY BRANCHING has wider implications for syntax and our understanding of Language. A BINARY BRANCHING constraint leads to a structural asymmetry between dependents of a head. It has implications for how θ -roles and Case are assigned. Additionally, it is key to a simplification of the mechanism by which elements combine into larger structures. The operation MERGE takes two syntactic objects and combines them to form one new one (Chomsky 1995). This greatly limits the things syntax can do (this is a good thing!).

In 300-500 words, write a short response to this problem that addresses the RESEARCH-SKILLS above. Propose a hypothesis, set out clearly why it's necessary, and provide evidence in support of your claims. As above, also discuss any relevant limitations. You may want to consult and cite relevant literature, but this is not necessary. It is possible to provide an answer that addresses all five skills fully relying only on the concepts developed on the course.

3.2 Split intransitivity

In the first tutorial we say verbs that alternate between transitive and intransitive readings. We also discussed the subject requirement in finite English clauses. With that as background, it has been suggested that there are two classes of intransitive verbs.

One has a range of properties, including that it allows what we call a **resultative**:

- (7) a. The winds froze the river.
- b. The river froze.
- c. The river froze **solid**.

- (8) a. Kim opened the box.
- b. The box opened.
- c. The box opened **wide**.

The other has a complementary range of properties, including that it doesn't allow resultatives:

- (9) a. I ran the race.
- b. *The race ran.
- c. *The race ran full.

- (10) a. *The clown cried the toddler.
- b. The toddler cried.
- c. # The toddler cried tired. [ungrammatical on the resultative reading]

What might be going on?

3.3 Large Language Models (ChatGPT, etc)

What are the syntactic abilities of Large Language Models? How good are they at providing explanations for syntactic phenomena when compared to linguists? Start by reading Konnelly & Sanders 2024 and Begus, Dabkowski & Rhodes 2025, then provide a critique of any of their points that allows you to showcase M1–M5 above.

If you wish to run experiments using an LLM yourself, we recommend using Edinburgh Language Models (ELM).

3.4 Other languages

Another possibility is to take any of the topics we've covered so far and examine them in another language. What differences arise? How might we account for them? If you choose to work on another language, it's particularly recommended that you get in touch with Itamar first.

References

Begus, Gasper, Maksymilian Dabkowski & Ryan Rhodes. 2025. Large linguistic models: investigating LLMs metalinguistic abilities. *IEEE Transactions on Artificial Intelligence*. 1–15. <https://doi.org/10.1109/TAI.2025.3575745>.

Chomsky, Noam. 1959. Review of verbal behavior. *Language* 35(1). Publisher: Linguistic Society of America, 26–58. <https://doi.org/10.2307/411334>.

Chomsky, Noam. 1965. *Aspects of the theory of syntax* (Special technical report / Massachusetts Institute of Technology Research Laboratory of Electronics ; no. 11). Cambridge, Massachusetts: MIT Press.

Chomsky, Noam. 1986. *Knowledge of language*. New York: Praeger.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, Massachusetts: MIT Press.

Konnelly, Lex & Nathan Sanders. 2024. AI hype in the classroom. *Proceedings of the Linguistic Society of America* 9(3). 5852. <https://doi.org/10.3765/plsa.v9i3.5852>.