LING 696G: Syntax and Computation

- Lecture 2

- Sandiway Fong
- University of Arizona
- sandiway.arizona.edu
- sandiway@arizona.edu

Course Outline

- **Today (Lecture 2):**
  - **Knowledge of Language**
  - **SMT Parser** (*696A course version, less buggy*)
  - has some English and Spanish
  - should work on macOS, Windows 11 and Linux
  - how to install, run and modify LEX: read the tutorial and manual
  - hope you have a laptop!
- **Heads up: Sept 7th is Labor Day (no classes)**

Course Outline

- **Next time (Lecture 3):**
  - **let's get deeper into background and theory!**
  - Collaborator Guest spot: Jason Ginsburg (Kyoto University)
    - *Serial verbs* (*Japanese*)
  - Background
    - Minimalism: Where Are We Now, and Where Can We Hope to Go, Chomsky N. A.
    - *GENGO KENKYU* 2021 Volume 160 Pages 1-41
    - Download and read (*as much as you can*):
      - https://www.jstage.jst.go.jp/article/gengo/160/0/160_1/_article

Lecture 1: Recap of Main Points

- **Workspace** (WS): is needed for **recursion**
- **recursion**: **SBOs** output can (later) feed **SBOs**
- **Minimal Search** (MS): a property of certain atoms of syntax, e.g. $INFL_{\phi}$, C_{Q}. (Also, relativization C_{rel} ?)
- **MS** can have **EXT** effects (the option to spell at the left edge)
- **EXT**: locus of language variation
- **Process Model**: hear something, **LEX** atoms into **WS**, **SBOs** fire, **INT** on converged **syntactic objects** (SOs), (optionally) **EXT** fires.
- **Parsing**: if what you heard is what is **EXT**'d, that **SO** is valid.
- Multiple **WS**s possible. Multiple **SO**s possible. Multiple **EXT**s possible.

Knowledge of Language

- **Knowledge** (*explicit*):
  - you know how to read and write a computer program
  - you know how to work the computer
  - **conscious propositional knowledge**
  - explicitly learned skills (*could include an implicit component*)
- **Knowledge of Language** (*implicit*):
  - *You obviously **wouldn't expect** the **native speaker of the language** to be **able to tell you anything very sensible about it**, or for that matter, to tell you anything very sensible about any other aspect of his **knowledge**. He might, on occasion, be able to, but he might very well not be able to.*

I-Language

Knowledge of Language

- From the same talk:
  - *As a speaker of English, everyone knows that there's something wrong with this sentence*:
    - **John's certainty to leave*
    - *John is certain to leave* (*that Bill will leave*)
  - *In fact, people who've **never seen either of these things** will understand them instantaneously in just the right fashion if they know English.*
  - **Linguist**: some condition on Nominalization? How is it acquired?
  - [Bruening 2018: *I for one as a native speaker of English have always found [..] perfectly acceptable.*]
  - *Whatever it is that the linguist, say, knows about is knowledge that he knows about the knowledge of the speaker, and **that knowledge is not consciously extricable**.*

Knowledge of Language

- From the same talk:
  1. *John is eager to please*
  2. *John is easy to please*
  1'. *John's eagerness to please* (Nominalization)
  2'. **John's easiness to please*
  3. *John is certain to leave*
  3'. **John's certainty to leave*
  - *[Knowledge]* ***not available to consciousness****, and there's no particular reason to suppose that it could ever be conscious. It doesn't become available through* ***introspection****.*
  - ***No*** *particular* ***reason why a person should be able to introspect*** *and ultimately sort of* ***see the principles upon which his decisions about his understanding of sentences*** *and his decisions about their character are based.*

Knowledge of Language

*There's a difference between the notion of **certain**, which is a **property of people** ...*
*and the notion of **certain**, which is a logical **property of propositions**.*

- Aside:
  - reinterpret Chomsky's *Remarks on Nominalization* (1970)
  - *John is certain to win the prize* / *It is certain that John will win the prize*
  - $\{certain_{v}, \{John \{v, \{win, prize_{the}\}\}\}$ (θ-configuration)
  - $\{INFL_{\phi}, \{be, \{certain_{v}, \{John \{v, \{win, prize_{the}\}\}\}\}\}\}$
  - *John is certain*
  - $\{certain_{\theta}, John\}$ (θ-configuration)
  - $\{INFL_{\phi}, \{be, \{certain_{\theta}, John\}\}\}$
  - *John is certain that Bill will win the prize*
  - $\{certainty_{\theta}, John\}$
  - *John's certainty that Bill will win the prize* / **John's certainty to win the prize*

Knowledge of Language

- Chomsky 1970:
  - *John is eager (for us) to please*
  - *John's being eager to please* (**Gerundive nominals**)
  - *the shooting of the hunters* (ambiguous)
  - *the growling of lions* (unambiguous)

  - *John's eagerness to please* (**Derived nominals**)
  - *the growth of tomatoes* (unambiguous)

To account for the distinction between (56c) and (57), we must restrict the feature [+cause] with respect to the feature that distinguishes derived nominals such as *growth* from forms such as *growing*, limiting it to the latter case.

Knowledge of Language

- On "***never seen before***":
  - *… each of these [cases] has exactly the same probability on any* ***empirical grounds for detecting probability****.*
  - *In other words, the probability of finding* ***probability in English is zero*** *in the really technical sense.*
  - *That is, if you look at larger and larger segments of English text, the* ***probability*** *of finding either of these* ***e reduces without limit****, clearly.*

$$P(\text{“Thank you”}) > 0 \qquad \lim_{|C|\to\infty} P_C(u) = 0$$ **"language is generative"**

where C is the corpus and u is a **particular complete utterance**.

Knowledge of Language

- From the same talk:
  - *I don't see any way to avoid the conclusion that knowledge of a language involves some sort of **internal representation**.*
  - *… a **system of rules**, of what is called nowadays a **generative grammar**, meaning a **system of principles** somehow mentally represented.*
  - *Nobody knows anything about [the] **physical representation**, but which we can describe abstractly that expresses what the speaker of the language intuitively knows, the kind of knowledge that underlies his performance, though it's **not the kind of knowledge that he can tell you about very often**.*

Knowledge of Language

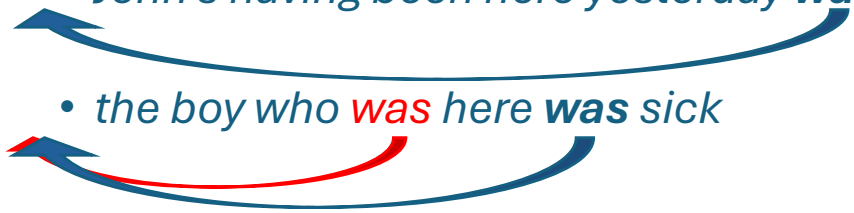
- From the same talk:
  - *So, in particular, a **generative grammar** will say for each sentence that it has such and such a phonetic shape, that it has such and such a meaning, and that it has such and such a **syntactic structure**, the latter being, again, a kind of an **abstract representation** that relates phonetic shape and meaning in some fashion.*
  - *And a generative grammar should say that for every utterance of the language.*
  - [*a person with a mental representation of a set of rules*] *may be able to detect the conclusions **without being able to have any insight** into the principles that he uses to arrive at these conclusions.*

Use of Knowledge of Language

- From the same talk:
  - *Well, if it is correct to describe knowledge of language, to offer a generative grammar as an explication of the notion of knowledge of a language, then the next question you want to ask is* ***how is that knowledge put to use****?*
  - *For example, you might want to ask the question,* ***can I construct a model****, a theory if you want, which* ***accounts for the ability of an ordinary human to accept some sort of signal written or spoken, let's say, and to produce the sort of structural description,****…*
- The **SMT Parser** is a model, a computer model of the knowledge of language (in these terms)
  - non-probabilistic (because that knowledge is non-probabilistic)

Structure dependent/independent operations

- From the same talk:
  - Language uses structure dependent operations only, e.g. for question formation
  - *It's an operation that doesn't depend on the substantive content of the string of elements in question, but rather on an abstract structure associated with it.*
  - Structure-dependence:
    - *John's having been here yesterday **was** a surprise*
    - *the boy who was here **was** sick*

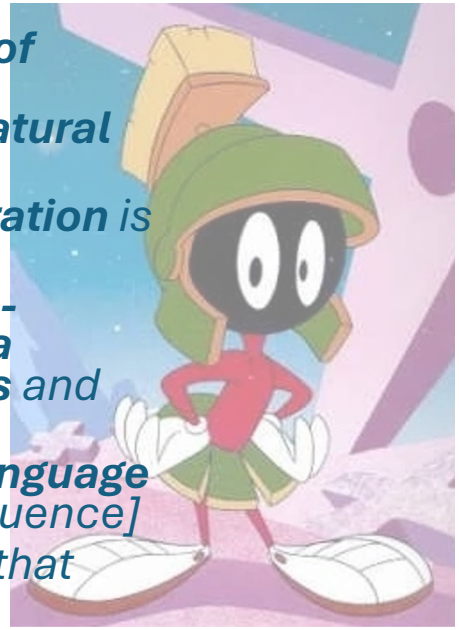


Structure dependent/independent operations

- From the same talk:
  - **Structure-independent:**
    - *you could imagine a language which formed questions by taking, let's say, by taking a declarative and **reading it backwards**,* [*e.g.* ***sick was here was who boy the***]
    - *Or you could imagine a language which has a structure-independent operation of forming a question by taking a declarative by taking, let's say, the **leftmost occurrence of the verbal element and moving that to the front**,* [*e.g.* ***was the boy who here was sick***]
    - *In fact, there's plenty of [simple] mathematics that deals with structure-independent operations like **shuffling** and **reflection** and **inversion** and all sorts of things.*

Structure dependent/independent operations

- From the same talk:
  - *And in fact, if there's any sense to the **notion of simplicity of theory**, then such a theory would be much simpler, by any considerations that we know about, **much simpler than natural language**.*
  - *Because, as I say, the notion of **structure-dependent operation** is a very **complex notion**.*
  - *If the **Martians** come along, they're likely to have **structure-independent operations**, and that'll be **perfectly fine for a communication system**, you know, **simpler mathematics** and errors and so on and so forth.*
  - *Yet, no language, you know, you can be **quite certain no language** does form questions in the latter fashion. [inverting the sequence]*
  - *And more generally, **no language has any operation** at all that operates in the latter fashion.*
  - ***All operations** in all languages […] are **structure-dependent**.*



Structure dependent/independent operations

- From the same talk:
  - *We could easily, let's say, to understand the language as **cryptanalysts**.*
  - *You know, as soon as I explain to you that in this **language** they're forming **questions by writing declarers backwards**, you could immediately carry out the appropriate cryptanalytic operations and interpret what those people are saying.*
  - *But you would be using a **different faculty of the mind** than the faculty that you use when you acquire knowledge of a language.*
  - *I'm saying that a **child** put into a situation where people are forming **questions by putting declarers backwards** would have **one hell of a time learning that language**.*
  - *Whereas when he uses the, in some sense, **much more complex rule of English**, he **learns it instantaneously**.*

Acquisition and Knowledge of Language

- From the same talk:
  - ***what kind of a system*** *must be postulated in order to account for the* ***ability of the person to blend the grammar****?*
  - *We could find out a great deal about that data if we wanted. We could* ***make recordings of everything*** *the person's ever heard and the situations they've been in and so on and so forth. And* ***it would be fairly small****, this amount of data.*
  - *But on the basis of whatever data he has available, he's come out with a* ***grammar that has certain peculiar properties****.*
  - *So, we could argue that* ***nominalization is possible in one case, but not in the other case*** *up here.*
  - *Obviously,* ***nobody's taught him that*** *because* ***nobody knew it to teach him*** *and so on and so forth in the narrow sense of know.*
  - *So, we could ask the question, what kind of a system must be presupposed, postulated, in order to* ***account for the transition from data to knowledge****?*

Induction and Knowledge of Language

- From the same talk:
  - *I think you find that many of the principles that are present in the functioning of the grammar are not many… are* ***not obtainable from the data by any imaginable induction****.*
  - *No* ***inductive process*** *that anyone can dream of or has dreamt of could lead from the kind of data that's available to* ***principles of the required abstractness****, for example, the principles that account for normalization in these cases.*
  - *If you really look at grammar, you can say that the following set of principles, which we have evidence to say are known by every speaker, normal speakers of English, are not determined by the data by any inductive process.*
  - *And furthermore, I think one can say not only that these* ***principles are common to all normal speakers of English****, but that* ***very similar principles****, if you move to the right level of abstractness, are also* ***true of every other language*** *that we know of them.*

Innateness and Knowledge of Language

- From the same talk:
  - *If, say, a biologist or an engineer were faced with such a situation, and it was not the human being that was involved, but rather some other organism or some automaton, let's say, he would immediately conclude that these principles are part of the* ***internal structure of the device****.*
  - *And that is used as the basis for developing specific knowledge on the basis of whatever data is available.*
  - *Now, if the theories are widely enough spread in the sense that not too many of them are compatible with small amounts of data, then the process of* ***acquiring a very rich theory on the basis of very degenerate and restrictive data*** *is possible.*

W. von Humboldt (1836)

- Chomsky (1964), *Logical Basis of Linguistic Theory* (LSLT), cites von Humboldt extensively
- *On Language* (1836):
  - "*Language is an **involuntary emanation of the mind**, no work of nations, but a gift fallen to them by their inner destiny*."
  - "*Every language is produced by the **original tendency** or '**original talent**' shared by all human beings*."

Peirce and Human Learning (Discovery)

Background:
Charles Sanders Peirce (abduction):
https://plato.stanford.edu/entries/abduction/peirce.html

- From the same talk:
  - *If you had an **automaton**, let's say, which had to come out with a very rich and highly articulated and specific and extensive theory and only had small degenerate data available to it, it would be able to do that **only if it had an initial characterization of the form of the theory that was quite restrictive** and that made available only a **fairly small number of possible hypotheses.***
  - *And it seems to me that the study of language leads to the, suggests very strongly, that **that's the way that the human mind works**, at least in this domain.*
  - *And that does recall very, well, almost in word for word paraphrase, some very interesting lectures of **Peirce on abduction**, where he proposed, on very weak grounds actually, that **all human learning is of this sort.***
  - *In fact, you may remember he argued even that **there's only a very small number of possible theories**, that human beings will ever be able to acquire.*
  - *he argued that if you look at the history of science, you see that when people have asked certain questions, then **almost immediately something rather like what is taken to be the right answer came out.***

Acquisition and Innateness

And it can be shown false by somebody solving the problem of constructing general techniques, heuristic techniques, or techniques of induction that don't have those two properties.

- From the same talk:
  - *On what grounds is the **data in the acquisition of language held to be so limited** that **only the theory of innate knowledge** can account for the acquisition of language?*
  - *One person has a certain, **reasonably small number of seconds of lifetime**. And in that number, there's a certain number of seconds he's had **a certain number of utterances presented to him** in particular situations.*
  - *And if you characterize all of that in any fashion that you like, **it's not the basis for induction** by any **specifiable principles of induction** that **don't incorporate the generative grammar**.*
  - *I would argue two things. First of all, that those principles of induction have **absolutely no generality**, that **they would be unique [to] language acquisition**, as far as we know.*
  - *And secondly, that the principles of induction would, in fact, be just what I described, namely, **they would list, in some manner, a set of attainable theories**.*
  - *And maybe **a technique for choosing among them**, and a **technique for judging confirmation**.*

Piantadosi (2024) vs. Chomsky

- **Modern language models refute Chomsky's approach to language** (Piantadosi, *lingbuzz* 2023; 2024)
- *The rise and success of [**LLMs**] undermines virtually every strong claim for the innateness of language that has been proposed by **generative linguistics**.*
- *Modern machine learning has subverted and bypassed the entire theoretical framework of Chomsky's approach, including its core claims to particular insights, principles, structures, and processes.*

Fong (2025) IJL Volume 37, Issue 1, pp. 59-74.

Futrell and Mahowald (2025)

- How Linguistics Learned to Stop Worrying and Love the Language Models, *Behavioral & Brain Sciences (BBS)* (2025)
  - ***Formal linguistics*** *has not presented an alternative model with the demonstrated* ***practical language-learning capabilities of neural models****.*
  - *The direction of developments in machine learning suggests that the gap between human and machine learning is more likely to be closed* ***not*** *through* ***built-in domain-specific formal restrictions****, but rather through* ***more powerful domain-general learning algorithms****.*
  - *Finding these internal mechanisms is not a prerequisite for the reality of linguistic structure —* ***the mechanisms may be irreducibly complex****.*
  - *It is rare in the history of science for a* ***scientific theory*** *to turn out as* ***disconnected*** *from a corresponding* ***engineering application*** *as formal generative linguistics has turned out to be for language models.*

?

?

Einstein's **Miracle Creed**

The Miracle Creed and SMT (Chomsky 2024)

Noam Chomsky
UNIVERSIY OF ARIZONA

Abstract. Einstein once observed that the "true theorist believes that the logically simple is also the real,...a miracle creed, but it is a miracle creed which has been borne out to an amazing extent by the development of science." The Strong Minimalist Thesis (SMT) holds that language too may satisfy the miracle creed at its core. SMT was proposed only as a guideline for research, but increasingly it seems that it may be valid in surprising ways. Recent work, reviewed and extended here, provides some credibility to the strong interpretation of SMT, with many consequences, if tenable.

Miracle Creed

LLMs: *GPT-4 1,760 billion parameters (G. Holz, cited in Wikipedia)* GPT-5.6?

APRIL 1, 1950 | 20 MIN READ

On the Generalized Theory of Gravitation

how pure a "positivist" he may fancy himself. The metaphysicist believes that the logically simple is also the real. The tamed metaphysicist believes that not all that is logically simple is embodied in experienced reality, but that the totality of all sensory experience can be "comprehended" on the basis of a conceptual system built on premises of great simplicity. The skeptic will say that this is a "miracle creed." Admittedly so, but it is a miracle creed which has been borne out to an amazing extent by the development of science.

A. Einstein in *Scientific American* Vol. 182 No. 4 (1950)

The Tamed Metaphysicist

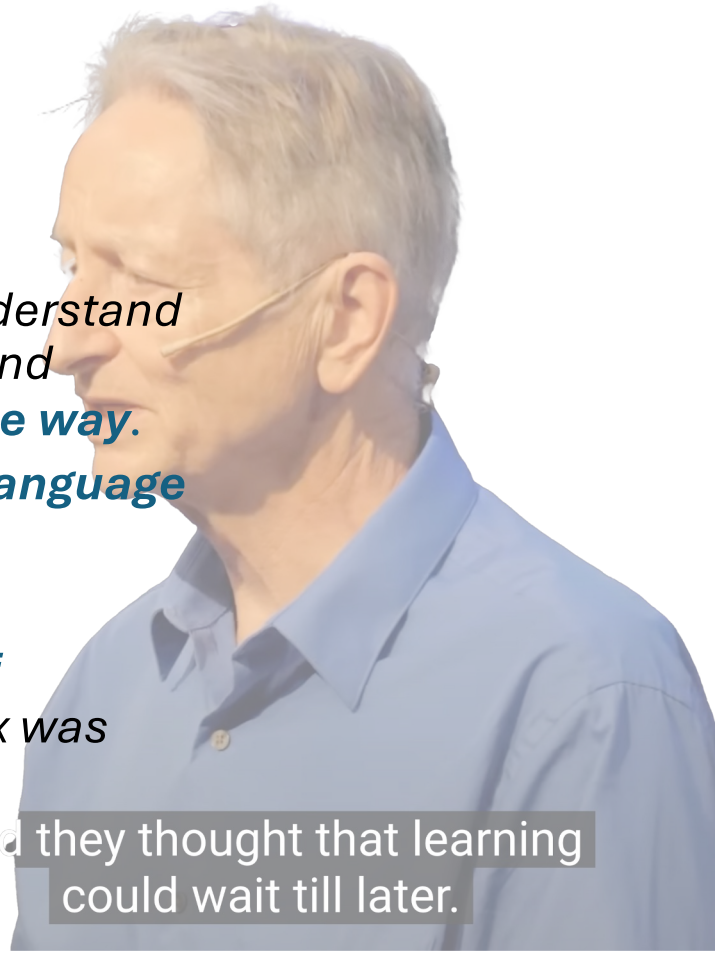
- *I believe that every true theorist is a kind of tamed metaphysicist, no matter how pure a "positivist" he may fancy himself.*

(Einstein 1950)

- positivist (e.g. Mach): empiricist – observables/measurables only.
- *[A] metaphysicist believes that the logically simple is also the real.*
  - imagine a simple, underlying model of the universe
- *[A] tamed metaphysicist believes that not all that is logically simple is embodied in experienced reality, but that the totality of all sensory experience can be "comprehended" on the basis of a conceptual system built on premises of great simplicity.*
  - there's a **testable**, simple, underlying model of the universe
- *The skeptic will say that this is a "**miracle creed**." Admittedly so, but it is a miracle creed which has been borne out to an amazing extent by the development of science.*

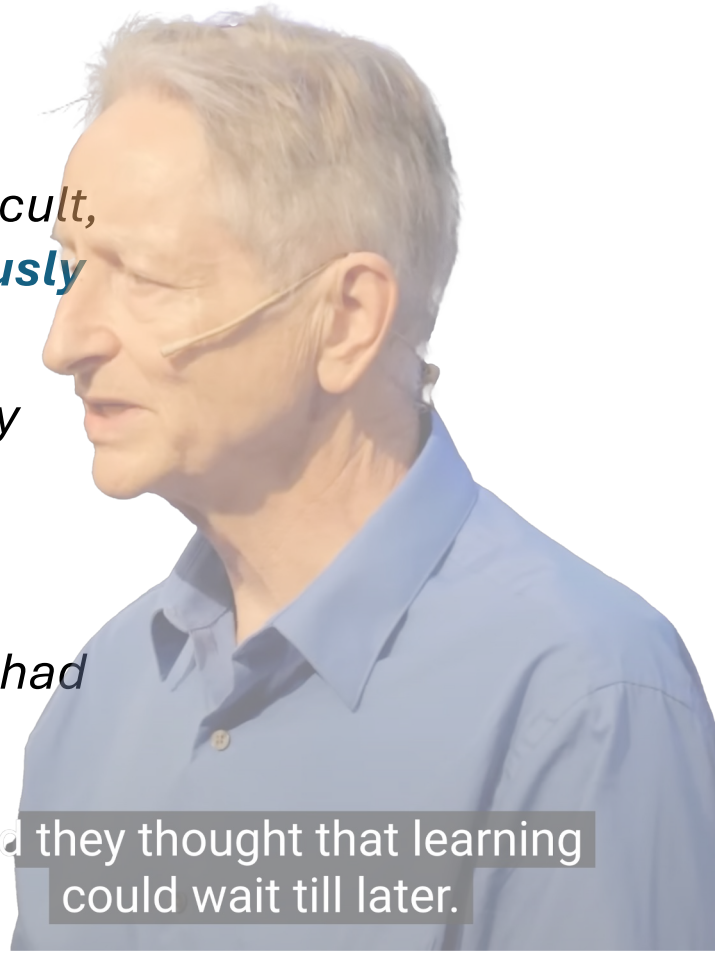
Hinton vs. Chomsky

- Hinton *Royal Institution* lecture (2025):
  - youtu.be/IkdziSLYzHw
  - *That's what understanding is when you understand language and when these [LLMs] understand language. **We understand in just the same way**.*
  - *And that's a **much better model of what language is** than anything the linguists ever had.*
  - *The linguists of course hate it.*
  - *[Linguists] also thought that **knowledge of language** was **innate**, knowledge of syntax was innate, which is just stupid.*



Hinton vs. Chomsky

- *It's the sign of a cult, that in order to join the cult, you have to believe something that is **obviously silly, like language isn't learned**.*
- *There's a guy called Chomsky, when they say things like, "These things don't understand anything, they're just a statistical trick," they don't actually have a model of what understanding is, because they never really had a model of what understanding was.*



Acquisition and Correction

- From the same talk:
  - *The problem is to understand how the child had come to know that certain expressions are correct and others are incorrect.*
  - *It's **not corrections**. It must be **something beyond corrections,** some either analogy or some principles that he used to support the correction support or something of that kind.*
  - *What is the nature of the **principles with the corrections support** that leads to this tremendous proliferation?*
  - *And incidentally, we know for a fact that the **correction from the environment is almost irrelevant**.*

Acquisition and Correction

- From the same talk:
  - *And incidentally, we know for a fact that the **correction from the environment is almost irrelevant**.*
  - *That if **children learn just as fast**, in every experiment that's been done, exactly as fast, **if they never get corrected**, they do get corrected.*
  - *So whatever kind of interaction there is, is first of all **small** because of the bound of time, and secondly, even smaller because we know that it has **almost marginal impact**.*
  - **Technically**: *you have been […] **making up a language on the basis of inadequate data** and have been **informed that your data is inadequate** and you therefore **made up the wrong grammar**.*

Empiricism and Learning

- From the same talk:
  - *It seems to me [**concrete empiricist theories**] immediately collapse as soon as they go beyond the level of few examples.*
  - *let me put it as strongly as I can, that **empiricism as an approach** to either epistemology or psychology is of purely historical interest, has **no rational basis***
  - *I really think it's just an extreme form of, well, **superstition**, basically.*
  - *And that it's inconsistent with the approach that any scientist would take towards the problem, if one was asking **how a highly structured final state emerges from a brief interaction with an environment**.*

Quine and Learning

Learn *John's certainty that Bill will leave*
Fail to learn *John's certainty to leave*

- From the same talk:
  - *If you remember, what [Quine] says is that there are **exactly three ways** of doing what he calls **learning sentences** [expressions].*
  - ***One**, you can learn them by **stimulus to expression association**.*
  - ***Two**, by a process of **expression to expression association**.*
  - *And **three**, by **analogy**.*
  - *But if it suggests that **analogy is a matter of substitution of similar items in fixed frames**, then you might as well forget about it, because it's not going to take you from the trivial data available to the potentially infinite data to which knowledge extends.*

Quine and Learning

- From the same talk:
  - *He says that **analogy serves only to speed the process up**. He seems to imply that you could really do the whole thing only by the first of the methods, but it would take much longer.*
  - *Tell me what the **basis for the analogy** is.*
  - *If you can't tell me the basis for the analogy, **you're not saying anything**.*
  - *If you can tell me the basis for the analogy, then **you're talking about the innate structure of his mind**.*

What Chomsky talk was that?

UW-Milwaukee talk,
July 16th 1968

nearly 60 years ago!

Willem Diepraam
(1968)

Communication and Thought

- Language organ is designed to construct thoughts efficiently
- Language is not designed for efficient communication
- *If that makes externalization harder to process and even makes some thoughts impossible to express without circumlocution,* ***too bad****.* ***Nature doesn't care****.* [Chomsky GK (2021), pg.11]
  - *EXT cannot have come before Merge.*
  - *The modern doctrine that language may have evolved from animal communication seems quite untenable.* [pg.10, Chomsky GK (2021)]
- It makes no sense to say that *some system evolved for* ***X***
  - "the spine evolved for keeping us upright," or "language evolved for communication"

too bad. Nature doesn't care. [Chomsky GK (2021), pg.11]

Communication and Thought

- Communicative efficiency is always sacrificed
  - *The most serious cases involve deletion of copies in accord with computational efficiency, leading to some of the **hardest parsing problems**.* [pg.10, fn.12, Chomsky (2021)]
    - *see possible solutions in the SMT Parser …*
  - Externalization: *John or the men *is/*are in the room*
    - *… unproblematic for expression of thought if feature valuation kept to late insertion so that only the bare copula reaches the thought level (as in some spoken dialects)*
  - *Note that **statistical information is irrelevant to I-language as a matter of principle**, though as has always been assumed in the generative enterprise (see Chomsky 1957), it can be highly relevant to processing and acquisition.*

696A SMT Parser

Download `english.zip` from the course website.

1. You need to install SWI Prolog
2. You need Python
3. You need nltk

- Download the *Tutorial and Manual for LEX* also.
  - Installation is covered in Section 1
  - You should try Sections 2 and 3.
  - (Sorry: Section 4 is missing.)
- (There is also a `spanish.zip`.)

696A SMT Parser

SMT Parser: Tutorial and Reference Manual

Sandiway Fong
University of Arizona
sandiway@arizona.edu
sandiway.arizona.edu

September 2024, updated April-May 2026. Draft August 2026.
[Incomplete: NOT FOR GENERAL DISTRIBUTION.]

The tutorial section covers installation, how to run the provided examples, and on how to modify **LEX** for the *LING 696A* course version of the SMT Parser, henceforth referred to as the **696A SMT Parser.** This version comes in macOS, Windows 11 and Linux flavors and is freely available for download and play.[1] Caveat: it is a limited teaching vehicle and not designed to be a general-purpose parser. The reference manual lists the **LEX** facts that must be defined for the parser to operate correctly for a particular language.

Brief History:
2024: SMT Parser using spec-INFL$_\varphi$. (*No longer available.*) First version of the manual. (*Deprecated.*)
2024-2025: NSF sub-project to add Spanish. Collaboration with PI Hilton Alers-Valentin (U. of Puerto Rico at Mayaguez).
2026: The **696A SMT Parser**. spec-INFL$_\varphi$ eliminated with respect to Merge. Language-particular Edge of INFL$_\varphi$ spellout defined in **EXT**. Revised manual (*still not quite complete*).
2027: A few examples from languages other than English and Spanish to illustrate points of theory, e.g. serial verbs.
Late 2027: Anticipated addition of the acquisition module for **EXT**. (*Perhaps an overly optimistic timeframe.*)

(The research version of the SMT Parser handles more examples, more languages, and contains a more advanced theory but currently has restricted distribution limited to collaborators. It is not freely available for obvious reasons, e.g. unclean development source code, instability due to implementation bugs, incomplete testing and outright mistakes in my tinkering with theory, etc. However, examples from the research version (unavailable in the 696A SMT Parser) may show up in papers and talks. If interested in discussing theoretic aspects or particular languages, please email me.)

Note: the linguistic theory behind the SMT Parser and how it works is documented elsewhere.

- See my BFSU talk slides for an overview.
  URL: `sandiway.arizona.edu/bfsu2026/bfsu2026.pdf`
- There is a paper in preparation (based on the BFSU talk).
- See also extensive LING 696A course notes.

[1] The author gratefully acknowledges the work of UA graduate students *Alex Tubens* (Ph.D linguistics) and *Gabriela Quinones* (M.S. Human Language Technology) for their Spanish LEX contributions.

Contents

696A SMT Parser

Brief Glossary

Important: this document does not describe the implemented linguistic theory (*described elsewhere*). Nevertheless, this alphabetical list of terms may be helpful in understanding the workings of the parser.

appl: *applicative verbal head*, written as $\texttt{appl}_{\theta}$ as it selects for a (usually dative) **IA**. appl is a head of the *propositional domain*. See discussion of *gustar* (*like* in Spanish).
aux verb: *auxiliary verb*, a non-θ-selecting head in the verb phrase. These are **v**'s that extend the core verb phrase, i.e. the θ-configuration, prior to transition to the *clausal domain*.
C: *complementizer* head, e.g. C/C_{rel}/C_{Q} for declarative/relative/interrogative C, respectively. The head of the *clausal domain*.
Clausal domain, the domain of **C** and **INFL** heads.
EA: *external argument*, e.g. the θ-relevant item selected by $\mathbf{v}_{\theta}$. (Or a_{θ}, but not used in the 696A SMT parser).
EXT: language-particular module for externalization (of a **SO**). (A parameterized theory of **EXT** is under development.)
FormCopy (FC): an operation that is automatically triggered at the interface with **EXT**. In the case of c-commanding identical inscriptions within the same Phase, only the higher occurrence is pronounced. See M-gap in (Chomsky 2021). The *SMT Parser* indicates unpronounced occurrences in **SO**s using gray coloring.
FormSet (FS): *form set*, a **SBO**, a set-forming operations that selects for 2 (or more) WS items that are *coherent* in some sense, e.g. all are θ-relevant items, predicate heads or θ-configurations. Resulting set is stored into the (next) **WS**, and is accessible for future **Merge** or **FS** operations (as is the case with all new **WS** items).
`fut`: future tense (**TNS**), a feature of highest **v** (*not used in the English implementation*).
Head: a bundle of features from a *syntactic* "atom" of **LEX**, i.e. the smallest unit that can enter the **WS** in this theory. (Sub-atomic heads are not subject to syntactic **SBO**s, we assume they are formed within **LEX**.)
I-Language: core unconscious *Knowledge of Language* that is part of our heritage as humans and inaccessible to conscious thought (or introspection). The SMT Parser implements a theory of **I-Language**.
IA: *internal argument, e.g. the θ-relevant item selected by* $\boldsymbol{R}_{\theta}$ *or* $\boldsymbol{appl}_{\theta}$.
INFL: *inflectional head*, written as INFL_{φ} in this theory. The subscript φ stands for φ-features, e.g. Person, Number and grammatical Gender. INFL_{φ} triggers **MS** and finds a θ-relevant item valuing φ. INFL_{φ} is a head of the *clausal domain*.
Initial WS: the state of the **WS** before **Merge**, but after **heads** from **LEX** have been inserted.
LEX: *lexicon*. Contains heads that enter the initial **WS**. Currently, for simplicity there are separate **LEX** files for English and Spanish.
Linear operation: an operation that applies to a sequence (not to structure). In Chomsky's terms, these are *structure-independent* operations, not part of **I-Language**.
main verb: a verb selecting for at least one θ-relevant item.
Markovian assumption: **SBO**s operate on items from the current **WS** only, cannot reach back to items from earlier stages of the **WS**. Moreover, there are no **SBO** records kept in the **WS**, e.g. we cannot tell which **SBO** operation created the higher occurrence of X in {X, {X, Y}}. See M-gap in (Chomsky 2021).

Merge: a **SBO**, *binary set Merge* of two items X and Y from a **WS**. Resulting set {X, Y} is stored into the (next) **WS**, call it WS'. X and Y are *accessible* items from the **WS** prior to WS'. For *External* **Merge**, both X and Y are distinct WS items. For Internal **Merge**, only X is a WS item and Y is a sub-term of X found by **Minimal Search**.
MS: *Minimal Search*, c-command domain search for the closest relevant item, e.g. θ or *wh*.
NP: *noun phrase*, a **SO** headed by a nominal.
φ-features: person, number (and gender), e.g. `3sg-m` for 3rd person singular masculine.
`pastp`: past participle **EXT** feature on **R**, e.g. `pastp` *comer* (*eat*) = *comido*.
`pres`: present tense (**TNS**), a feature of highest **v**.
`presp`: present participle **EXT** feature on **R**, e.g. `presp` *eat* = *eating*.
Propositional domain, see *θ-configuration*, for verb phrases the domain of (potential) θ-selecting **R**, **v** and **appl** heads.
`pst`: past tense (**TNS**), a feature of highest **v**.
Recursion: a **SBO** is *recursive* if its output can serve as input to **SBO**s. **Merge** and **FormSet** are recursive **SBO**s. Recursion requires a **WS**.
R: root, e.g. verbal root `dance` or $\texttt{eat}_{\theta}$. Written with a θ-subscript if it selects for an **IA**.
SBO: (primitive) *structure building operation*, **Merge** or **FormSet** (FS) in this theory.
SO: (set-based) syntactic object, e.g. $\{C, \{\text{INFL}_{v:3pl}, \{v_{\texttt{llegar:pres}}, \{\texttt{llegar}_{\theta}, \texttt{trenes}_{dos}\}\}\}\}$.
SMT: *Strong Minimalist Thesis*.
Structure: *syntactic structure* (*or object*), a set object formed by **Merge**(s).
Spellout: a linear process, e.g. the sequence φ-TNS-R `3sg pst eat` maps to (irregular) `ate`.
θ: *theta*; if a subscript, e.g. on $\texttt{llegar}_{\theta}$ (*arrive*), head selects for a θ-relevant item. See **R**.
θ-configuration: a **SO** containing (**WS** obligatory) arguments and required predicate heads, e.g. for transitive *eat*, a configuration $\{\texttt{EA}, \{v_{\texttt{eat:}\theta\texttt{:TNS}}, \{\texttt{eat}_{\theta}, \texttt{IA}\}\}\}$ with arguments `EA` and `IA`.
θ-relevant item: a NP (or clause), a **SO** headed by a nominal (or **C**) that may bear a θ-role.
TNS: tense, e.g. `pres`/`pst`/`fut`, a feature of (structurally) highest **v**.
v: "little v" verbal head that may select for a θ-relevant item and hosts **TNS**. E.g. $v_{\texttt{eat:}\theta\texttt{:pst}}$ indicates that this head selects for verbal root *eat*, introduces a θ-relevant item (**EA**), and has **TNS** value `pst` (*past*). Generally, only the highest **v** in a verb phrase can bear **TNS**.
WS: *workspace* for recursive **SBO**s, **Merge** and **FormSet**. *Recursion* requires a **WS**. **Heads** from **LEX** populate an initial **WS**, called WS_0. Results from **SBO**s are placed into the (next) **WS**, e.g WS_1, WS_2, etc. In this serial implementation, each **SBO** results in a transition between **WS**s. Earlier WSs are not accessible to **SBO**s, e.g. at WS_n, items from WS_{n-1} are not accessible, i.e. the *Markovian assumption* applies. (However, in principle non-interfering **SBO**s may apply at the same time and update the **WS** in parallel.)

696A SMT Parser


```
For online help and background, visit https://www.swi-prolog.org
For built-in help, use ?- help(Topic). or ?- apropos(Word).

% Interactive session; added `.` to Python `sys.path`
[3 ?- parse(['John',likes,'Mary'], SO, L).

Words: John likes Mary
Initial WS: Mary  like_θ  v_like:θ:pres  INFL_v:3sg  John
INT/EXT: {C, {INFL_v:3sg, {John, {v_like:θ:pres, {like_θ, Mary}}}}}
EXT: John 3sg pres like Mary
Spellout: John likes Mary
Parse found!
SO = {C, {INFL_v:3sg, {John, {v_like:θ:pres, {like_θ, Mary}}}}},
L = [John,likes,Mary] ;
```

```
glish.prolog — 80×30

thon `sys.path`
).

 INFL_v:3sg  John
e:θ:pres, {like_θ, Mary}}}}}

pres, {like_θ, Mary}}}}},

e:θ:pres, {like_θ, John}}}}}
EXT: Mary 3sg pres like John
Spellout: Mary likes John
Blocked: inconsistent with input!
INT/EXT: {C, {INFL_v:3sg, {{Mary, John}, {v_like:θ:pres, {like_θ, {Mary, John}}}
}}}
EXT: {Mary,[and],John} 3sg pres like {Mary,[and],John}
Blocked: EXT inconsistent!
Blocked: inconsistent with input!
false.

[4 ?- report.

REPORT LENGTH: 40
true.

5 ?-
```

696A SMT Parser

file:///Users/sandiway/Desktop/codir

[Right triangle ▸ may be expanded, down triangle ▾ to close. Button ⌃ = tree pop-up. ✕ = close pop-up.]

Terminal: run. **Derivation Graphs:** (left-to-right) compact / full; (top-down) full / compact.

▸ Help: blue = parse inside. **Abbr.:** WS: Workspace; Initial WS: initial heads for Merge after LEX lookup.

Words: John likes Mary

▾ **Initial WS 1:** Mary $like_\theta$ $v_{like:\theta:pres}$ $INFL_{v:3sg}$ John

▸ **WS 1:** {$like_\theta$, Mary} $v_{like:\theta:pres}$ $INFL_{v:3sg}$ John

INT/EXT: ⌃ {C, {$INFL_{v:3sg}$, {John, {$v_{like:\theta:pres}$, {$like_\theta$, Mary}}}}}

EXT: John 3sg pres like Mary

Spellout: John likes Mary

Parse found: John 3sg pres like Mary

▸ **WS 1:** {$like_\theta$, Mary} $v_{like:\theta:pres}$ $INFL_{v:3sg}$ John

▸ **WS 1:** {$like_\theta$, John} Mary $v_{like:\theta:pres}$ $INFL_{v:3sg}$

INT/EXT: ⌃ {C, {$INFL_{v:3sg}$, {Mary, {$v_{like:\theta:pres}$, {$like_\theta$, John}}}}}

EXT: Mary 3sg pres like John

Spellout: Mary likes John

Blocked: inconsistent with input!

▸ **WS 1:** {$like_\theta$, John} Mary $v_{like:\theta:pres}$ $INFL_{v:3sg}$

▸ **WS 1:** {Mary, John} $like_\theta$ $v_{like:\theta:pres}$ $INFL_{v:3sg}$

INT/EXT: ⌃ {C, {$INFL_{v:3sg}$, {{Mary, John}, {$v_{like:\theta:pres}$, {$like_\theta$, {Mary, John}}}}}}

EXT: {Mary,[and],John} 3sg pres like {Mary,[and],John}

Blocked: EXT inconsistent!

Blocked: inconsistent with input!