Formal methods in cognitive science: language

050.370/670, JHU, Fall 2008

Instructor: Kyle Rawlins
Email: rawlins@cogsci.jhu.edu
Office: Krieger Hall 149
Office phone: 410-516-5330
Office hours: Tuesday 10–11, Wednesday 1:30–2:30
Course times: Tuesday & Thursday 12:00-1:15
Website: http://www.cog.jhu.edu/courses/370/
TA: Christo Kirov
TA: Erin Zaroukian

Textbook:

- Barbara Partee, Alice G. B. ter Meulen, and Robert Wall. Mathematical methods in linguistics, Kluwer, 1993.

1 General description

From the registrar:

> This course will be devoted to the study of formal systems that have proven useful in the cognitive science of language. We will discuss a wide range of mathematical structures and techniques and demonstrate their applications in theories of grammatical competence and performance. A major goal of this course is bringing students to a point where they can evaluate the strengths and weaknesses of existing formal theories of cognitive capacities, as well as profitably engage in such formalization, constructing precise and coherent definitions and rigorous proofs.

Let's break this down a bit:

- **Theory-building**: (controversial statements ahead) all theories of language rely, implicitly or explicitly, on mathematical concepts. Even explicitly anti-formalist work often replies, implicitly, on such tools as set theory.
- **Theories of cognitive capacities**: You have all probably seen statements such as: "Theory/analysis X of phenomenon A is simpler than theory/analysis Y, and therefore X is better." "Theory/analysis/algorithm X is obviously so complicated or non-intuitive that the human brain just can't possibly do that." "Representation X is too complicated/large/non-finite/hard to learn to be what is actually stored in the brain." In some cases (especially in earlier work, but not limited to that), such statements are made without any kind of grounding. *All* such statements, if true, can and should be given mathematical grounding. In some cases, when you attempt to do so, it turns out that the statement isn't actually true. (E.g. non-finite structures often strike people as very counter-intuitive; after all, isn't the brain finite? But you must rule out the possibility of an efficient finite representation of a non-finite object for this argument to go through. There typically is in fact such a representation.)
- **Language as a mathematical object**: Language itself can be studied as a mathematical object. This idea is (proximately) due to Chomsky and Montague, and revolutionized the modern study of natural languages.

- **Engaging in formalization**: It *always* helps in the theory building endeavor to understand how your theories might be formalized. It is never the case that you don't learn something new, surprising, and often alarming when you try to formalize your theory, or make sure your definition is actually coherent. And you want it to be *you* that discovers the alarming thing, not someone who will use it to argue against you.
- **Precise definitions**: It is not a contribution to science if your definition is too vague to argue against (e.g. unfalsifiable in some way). Mathematical tools help avoid this situation.
- **Engaging with previous literature**: Mathematical tools are in common use in linguistics, you have to understand them to understand the literature.

Therefore, it is useful for any linguist or cognitive scientist working in related fields to understand the relevant formal tools.

Note: if you have taken or are taking a course on formal language theory (e.g. 600.271 or 600.471 in the CS department here), this class may be somewhat redundant. It will still help cement your understanding of the material and how it relates to linguistics, but you may want to come talk to me to evaluate whether you should take the course.

2 Organization

- The class involves weekly homework problem sets, consisting of short(ish) mathematical problems. The problem sets will be split between purely mathematical questions and problems applied to some specific linguistic issues. Some problems will be graduate-only.
- There will be no in-class exams; around week 6 and at the end there will be take-home exams. These will essentially be somewhat longer/more interesting homework assignments. I reserve the right to give short unannounced quizzes (though I probably won't).
- We will try to split our time evenly between phonological, syntactic, and semantic applications (but I do have some biases). We will also cover topics such as acquisition.
- Collaboration is highly encouraged. Please work together, it can really help when learning this material. But, **you must write up your answers yourself**. I will not accept group assignments; the process of learning to write clear, precise mathematical prose is a crucial part of what this class teaches, over and above simply solving the problems. Please write who you work with on your assignments.
- **Please type your work.** If you haven't learned LaTeXand would like to, this course is a perfect opportunity, and I would be happy to give you pointers. Just about any formalism you would need to write up should also be doable in Word or its free alternatives (though I'm afraid I won't be able to help you with this very much.)

3 Topics

The chart on the next page breaks down the topics of the class, and suggests applications in some of the major subfields of linguistics. We won't cover all of these applications; this chart is designed to give you a taste of what this material is good for. Note that though this course focuses on natural language, many of these tools also have applications in cognitive science in general.

There are some major topics we won't cover in much detail. In particular, systems of logic will not be the focus of this class, to avoid overlap with Semantics I. We also will focus on discrete structures, leaving probabilistic models (and the like) to 050.371/671 and 050.372/672.

Topic	Subtopic	Phonetics/phonology	Morphology/syntax	Semantics
Basic tools	Set theory	many		many; possible world semantics, plurality
	Propositional logic	many		
	Relations, functions	many		many; indefinites/choice fn.s, questions
	(In)finity		cardinality of a language	possible world semantics, attitudes/modality, the limit assumption
	Proof techniques: direct proofs, contradiction, induction, diagonalization	many		
Structures	Graphs, trees	autosegmental representations	many	
	Algebraic structures, lattices	Antilla and Cho (variation and change)	Szabolcsi/Zwarts on islands	plurality
	Model theory	model theoretic OT	model theoretic syntax	many
Formal language theory	regular languages, finite state automata	finite state transducers	finite-state morphology, agreement	quantifiers
	context free grammars, pushdown automata		many	
	Type-0 grammars, turing machines		Peters and Ritchie theorem	
	Between types 1 and 0		TAG, indexed grammars	
	General		Language computability from an empirical perspective, Gold's theorem (acquisition)	Compositionality, complexity of semantic representations
	Tree-sets	model-theoretic phonology	model-theoretic syntax	
	Complexity, P, NP	parsing, etc.		

4 Schedule

The following is a rough schedule of the class, subject to change.

Week	Topic	Reading (corresponds with lecture)
Sep. 4	Set theory	PMW ch. 1
Sep. 8–12	Set theory, functions, relations	PMW ch. 2.1–2.4,3.1,3.4
Sep. 15–19	Infinities, proof techniques	PMW ch. 3
Sep. 22–26	Graphs and trees	TBA, PMW 16.3
Sep. 29–Oct. 3	Graphs and trees, model theory (?)	TBA
Oct. 6–Oct. 10	Algebraic structures	PMW ch. 9-10
Oct 13–Oct. 17	Lattices	PMW ch. 11
Oct 20–Oct. 24	Regular languages, finite automata	PMW ch. 17, TBA
Oct. 27–Oct.31	cont'd	
Nov. 3–Nov. 7	Context free languages, pushdown automata	PMW ch. 18
Nov. 10–Nov. 14	R.e. languages, Turing machines	PMW ch. 19
Nov. 17–Nov. 21	cont'd	
Nov. 24–Nov. 28	Context sensitive languages, intermediate languages	PMW 20–21, Gold 1967
Dec. 1–Dec. 5	Tree-sets, complexity theory	TBA, TBA
Dec. 9–11	reading period	
Dec. 12–19	exam period	

Bibliography

Gold, Mark. 1967. Language identification in the limit. *Information and Control* 10:447–474.

Partee, Barbara, Alice ter Meulen, and Robert Wall. 1993. *Mathematical methods in linguistics*. Dordrecht: Kluwer.