Ling 7800-065: Sign-Based Construction Grammar

- **Instructor**: Ivan A. Sag (Stanford University)
- **Email**: sag@stanford.edu
- **URL**: http://lingo.stanford.edu/sag/LI11-SBCG

What is Generative Grammar?

- **GG_1**: Any precisely formulated set of rules whose output is all (and only) the sentences of a language, i.e. the language generated by that grammar.

- **GG_2**: Any version of TRANSFORMATIONAL Generative Grammar:

  Early Transformational Grammar (e.g. *Syntactic Structures*) ⤳ The 'Standard' Theory (e.g. *Aspects of the Theory of Syntax*) ⤳ The 'Extended Standard' Theory ⤳ REST ⤳ P&P ⤳ GB ⤳ The 'Minimalist' Program

Generative Grammar as Cognitive Science

Marr's (1982) theory of Vision

- **Computational Level**: What function is computed?
- **Algorithmic Level**: How is it computed?
- **Implementational Level**: How are those algorithms implemented?

Generative Grammar

- 'An abstract characterization' of linguistic knowledge
- Evaluated by descriptive adequacy
- Very 'weak' competence theory (cf. Bresnan and Kaplan 1982)
- And the story is never completed!

Generative Grammar

- ▶ Generative Grammars are usually regarded (certainly by Chomsky) as theories of the Computational Level.
- ▶ Not clear how to evaluate weak competence theories

  Why should we choose between two formally distinct theories that derive exactly the same sound-meaning corrspondences?
- ▶ Not clear how to evaluate theories of 'I-Language'
- ▶ Even less clear how to evaluate theories of 'Universal Grammar'

Not everyone thinks this way about grammar

- ▶ Is psycholinguistic/neurolinguistic evidence relevant?
- ▶ E.g. performance errors (Fromkin,...)?
- ▶ Systematic observations about language use/processing?
- ▶ Native speakers' intuitions about analyses (perhaps at odds with the 'simplest' analysis)?
- ▶ Diachronic data?
- ▶ Functional considerations of various kinds?

A bit of History

- **The Derivational Theory of Complexity**

- Each application of a transformation increases the psycholinguistic complexity of a given sentence.
- The overall complexity of a given sentence is determined in part by the number of steps in its transformational derivation.

Derivations (TG in the 70s)

[Kim$_i$ [we [were impressed [by t$_i$]]]] (spell out)

[Kim$_i$ [us+NOM [be+past impress+ed [by t$_i$]]]] (affix hopping)

[Kim$_i$ [us+NOM past [be ed impress [by t$_i$]]]] (case marking)

[Kim$_i$ [us past [be ed impress [by t$_i$]]]]
(topicalization)

[us past [be ed impress [by Kim]]]
(passivization)

[Kim past [impress us]] (deep structure)

Fodor et al. (1974, p. 276)

- Investigations of DTC...have generally proved equivocal. This argues against the occurrence of grammatical derivations in the computations involved in sentence recognition.
- [e]xperimental investigations of the psychological reality of linguistic structural descriptions have...proved quite successful.

A bit more History

- Chomsky and fellow derivationalists rejected the relevance of the experiments that led Fodor, Bever, and Garrett to their conclusions.

- But in the 1970s, some took these results seriously, began to look for alternatives to transformations.

- 'Realistic' Grammar (Bresnan 1978)

And...

- In the 1980s, new kinds of generative grammar began to emerge that eliminated transformations, hence transformational derivations. These approaches came to be known as **Constraint-Based Grammar**.

- Generalised Phrase Structure Grammar (GPSG)
- Lexical Functional Grammar (LFG)
- Head-Driven Phrase Structure Grammar (HPSG)
- Categorial Grammar (especially Combinatory CG (CCG))
- Tree-Adjoining Grammar
- Simpler Syntax

A final bit of History

MP is evolving into a CB-Framework. When it eliminates 'Move' and has only 'Merge', it will finally be Constraint-Based.

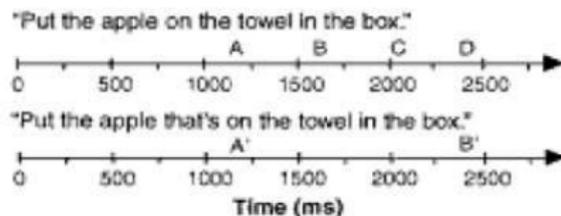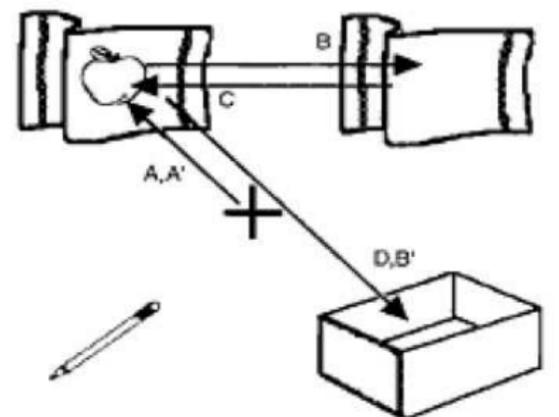$[[b_i [c [a t_i]]] d]$ (Merge)

$[b_i [c [a t_i]]$ (Move)

$[c [a b]]$ (Merge)

$[a b]$ (Merge)

Strong Theory of Linguistic Competence
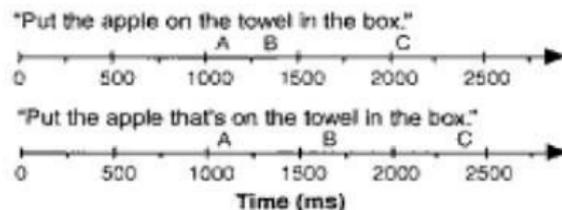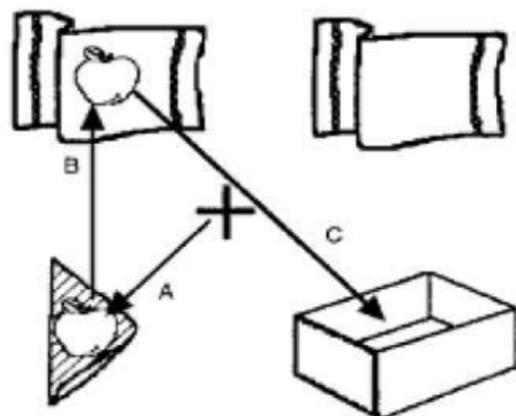
- The constructs of grammar are in part motivated by properties of language use, processing, and language change.
- The competence grammar is directly embedded in a model of performance, a model of change, etc.
- The theories of grammar and processing have to be developed in parallel.
- Evaluate grammars (and grammatical theories) in terms of their fit into this broader picture.

Tanenhaus et al. in Science (1995)

"Put the apple on the towel in the box."

	A	B	C	D	
0 500 1000 1500 2000 2500

"Put the apple that's on the towel in the box."

| | A' | | B' | |
0 500 1000 1500 2000 2500

Time (ms)

Fig. 1. Typical sequence of eye movements in the one-referent context for the ambiguous and unambiguous instructions. Letters show when in the instruction each eye movement occurred, as determined by the mean latency for that type of eye movement (A' and B' correspond to the unambiguous instruction).

"Put the apple on the towel in the box."

| | A | B | | C | |
0 500 1000 1500 2000 2500

"Put the apple that's on the towel in the box."

| | A | B | | C | |
0 500 1000 1500 2000 2500

Time (ms)

Fig. 2. Typical sequence of eye movements in the two-referent context. Note that the sequence and the timing of eye movements, relative to the nouns in the speech stream, did not differ for the ambiguous and unambiguous instructions.
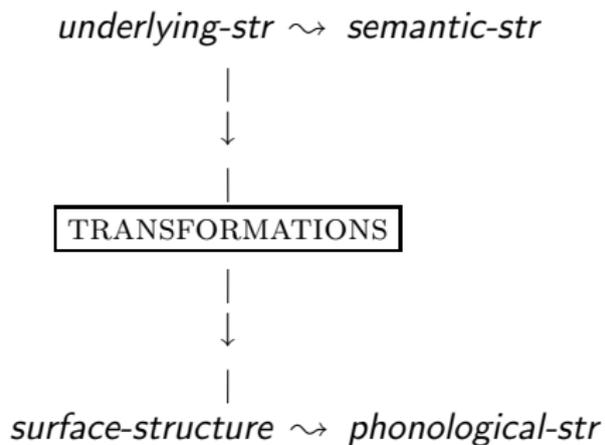
Tanenhaus et al. in Science (1995)

Our results demonstrate that in natural contexts, people seek to establish reference with respect to their behavioral goals during the earliest moments of linguistic processing. Moreover, referentially relevant nonlinguistic information immediately affects the manner in which the linguistic input is initially structured. Given these results, approaches to language comprehension that assign a central role to encapsulated linguistic subsystems are unlikely to prove fruitful. More promising are theories by which grammatical constraints are integrated into processing systems that coordinate linguistic and nonlinguistic information as the linguistic input is processed[15]

---

[15]... Jackendoff, Ray. 1992. *Languages of the Mind*...

Carl Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*.

Syntactocentric Interpretation 1

$$\textit{underlying-str} \rightsquigarrow \textit{semantic-str}$$

$$|$$
$$\downarrow$$
$$|$$

$$\boxed{\text{TRANSFORMATIONS}}$$

$$|$$
$$\downarrow$$
$$|$$

$$\textit{surface-structure} \rightsquigarrow \textit{phonological-str}$$

Syntactocentric Interpretation 2

$$
\begin{array}{c}
\textit{d-structure} \\
| \\
\downarrow \\
| \\
\boxed{\text{TRANSFORMATION}} \\
| \\
\downarrow \\
| \\
\textit{s-structure}
\end{array}
$$

```
              s-structure
            /            \
           ↓              ↓
           |              |
          PF            LF ⇝ meaning
```
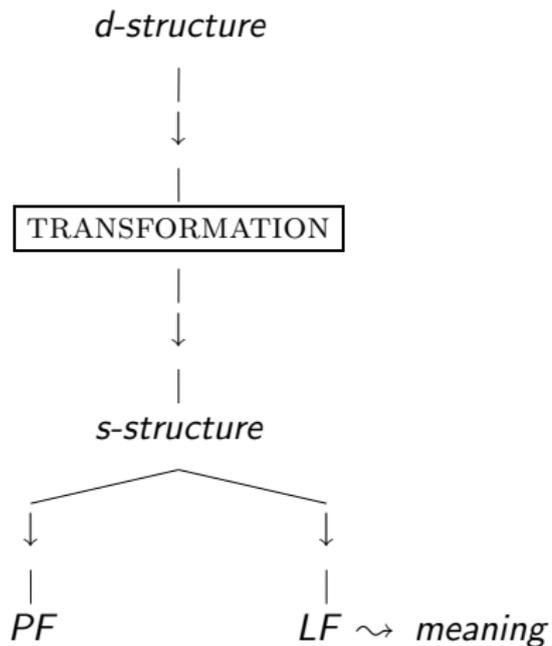
Incrementally Computed Partial Meanings

- Reject Syntactocentrism
- Surfacist analyses
- Adopt Sign-Based architecture
  (subsumes Bach's 'Rule-to-Rule' Hypothesis)

- Localized Syn-Sem interface
- Localized Phon-Syn interface
- Localized Phon-Sem interface
- Localized Contextual Inferences

Flexible Utilization of Partial Information

- Partial linguistic information is sometimes enough
- speech processing; degraded signal
- using foreign languages with imperfect knowledge
- relatively seemless integration of partial linguistic information
- integration of linguistic and nonlinguistic information

Jackendoff 2002

Because the grammar is stated simply in terms of pieces of structure, it imposes no inherent directionality: in production it is possible to start with conceptual structure and use the interfaces to develop syntactic and phonological structure; and in perception it is possible to start with phonological strings and work toward meaning.

Sag, Kaplan, Karttunen, Kay, Pollard, Shieber, and Zaenen 1986

A [unification-based] theory of grammar ... allow[s] a direct embedding of the theory of linguistic knowledge within a reasonable model of language processing. There is every reason to believe that diverse kinds of language processing - syntactic, lexical, semantic and phonological - are interleaved in language use, each making use of partial information of the relevant sort. Given that this is so, the theories of each domain of linguistic knowledge should be nothing more than a system of constraints about the relevant kind of linguistic information - constraints that are accessed by the potentially quite distinct mechanisms that are involved in the production and comprehension of language.

Fluctuation of Activation

- ► Lexical Priming
- ► Semantic Priming
- ► Phon Priming
- ► Constructional Priming
- ► Rich encoding enhances activation, facilitating processing.
- ► Relevant to the analysis of filler-gap constructions (cf. Hofmeister 2007; Hofmeister & Sag 2010)

- ► Accommodate probabilistic effects

Quantifier Scope Underspecification Resolution

- Native speakers don't struggle with the massive scope ambiguities predicted by modern theories of quantification.
- Psycholingusitic motivation for a theory of quantification that allows underspecification or partial scope resolution.

Constraint-Based Grammar

- surface-oriented,
- model-theoretic (constraint-based and monotonic), and
- strongly lexicalist.

The Competence/Performance Distinction

- ▶ The distinction isn't meaningful without some precision in developing both theories.
- ▶ Must develop explicit models of processing in which to embed explicit grammars.
- ▶ With that clarification, the C/P distinction is an extremely useful working assumption.

For Example

- ▶ Parsing with Context-Free Grammars.
- ▶ Distinguish grammar from parser.
- ▶ The operations performed by the parser consult the grammar as a resource.
- ▶ Hence the grammar simultaneously serves to specify the structures of the language and certain aspects of the processing of that language.
- ▶ E.g. Shift-Reduce Parsers (Aho and Ullman, 1972)

Shift-Reduce Parsing with a CFG

- Parser actions:

  Shift

  (go ahead to the next word without building anything new) or

  Reduce (apply a CF rule to build a tree structure)
- Consult grammar rules in performing a reduction.
- E.g. Shieber (1983) on Attachment Preferences (See also Pereira and Shieber 1985)

What's Missing?

A lot:

- ▶ Access to semantic information
- ▶ Access to world knowledge
- ▶ Access to probabilistic information
- ▶ Access to the linguistic context
- ▶ Access to the extralinguistic context
- ▶ A theory of how these factors interact

Why Do Construction Grammar?

- First reason:

  It provides uniform tools for analyzing the general patterns of language, the most idiosyncratic exceptions, and everything in between.

Kay and Fillmore 1999

One cannot analyze an idiomatic construction without simultaneously discovering and setting aside all the aspects of the data that are NOT licensed by the construction one is studying. To know what is idiomatic about a phrase one has to know what is nongeneral and to identify something as nongeneral one has to be able to identify the general. In grammar, the investigation of the idiomatic and of the general are the same; the study of the periphery is the study of the core-and vice versa. The picture that emerges from the consideration of special constructions ... is of a grammar in which the particular and the general are knit together seamlessly.

For me... Construction Families

Some Aux-Initial Constructions: (Fillmore 1999; Ginzburg & Sag 2000)

Exclamatives:	Boy, **was I stupid**!
	Wow, **can she sing**!
Conditionals:	**Were they here now**, we'd...
	Should there be a storm, we'd...
'Magic':	May they live forever!
	May all your teeth fall out!
Interrogatives:	Were they involved?
	We won't go, **will we**?
Declaratives:	So **can I**!
	Never **would I do such a thing**.
...	

- ▶ What is Construction Grammar?
- ▶ Go to a Construction Grammar conference.
- ▶ Ask Wikipedia!
- ▶ What is a construction?

- What is a construction?
- C is a CONSTRUCTION iff$_{def}$ C is is a form-meaning pair $\langle F_i, S_i \rangle$ such that some aspect of F_i or some aspect of S_i is not strictly predictable from C's component parts or from other previously established constructions. [Goldberg 1995]

Some Questions

- What does 'previously established' mean?
- What exactly are the 'component parts' of a construction?
- How do constructions define what's well-formed and what isn't?
- How do constructions interact with one another?
- Do constructions work like grammar rules?
- ...

- ▶ What is a construction?
- ▶ Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency (see Chapter 3 for discussion). [Goldberg 2005,2008]

Different Conceptions of Construction Grammar

- ▶ What Wikipedia says
- ▶ 'Cognitive Grammar', Radical CxG, Fluid CxG
- ▶ BCG (Fillmore, Kay, Goldberg, Michaelis,...)
- ▶ Constructional HPSG (Ginzburg, Sag,...)
- ▶ Simpler Syntax (Culicover, Jackendoff)
- ▶ Data Oriented Parsing (DOP; Rens Bod,...)
- ▶ SBCG (Sag, Kay, Fillmore, Michaelis,...)

The Fundamental Insight of Generative Grammar

- ▶ Language is a recursive system.
- ▶ Expressions combine in systematic ways.
- ▶ CxG Must recognize patterns of combination

Informally:

Informally

- ▶ Combine a subject and a finite VP to form a clause whose meaning is a proposition. (Subject-Predicate Construction)
- ▶ Combine a lexical head and all of its complements except its subject to form a phrase whose meaning is a predicate. (Predicational Head-Complement Construction)
- ▶ Combine an invertible (hence finite) auxiliary verb with all its valents (subject, then complements) to form an interrogative clause whose meaning is a polar question. (Polar Interrogative Construction)
- ▶ Combine a *wh*-interrogative expression (the filler) with an aux-initial clause missing an expression of the same type as the filler to form an interrogative clause whose meaning is a nonpolar question.

Misconceptions about CxG (Michaelis 2011)

- ▶ CxG is nonrigorous.
- ▶ CxG does not offer generalizations.
- ▶ CxG is obsessed with linguistic marginalia.
- ▶ CxG is opposed to compositional semantics.
- ▶ CxG is not constrained.
- ▶ CxG does not provide a universal framework for syntax.

Misconceptions about CxG

- CxG is nonrigorous.
- Not all work is 'formal', nor should it be.

Misconceptions about CxG

- CxG does not offer generalizations.
- [In a Principles-and-Parameters approach] the notion of grammatical construction is eliminated, and with it, the construction-particular rules. Constructions such as verb phrase, relative clause, and passive remain only as taxonomic artifacts, collections of phenomena explained through the interaction of the principles of UG, with the values of the parameters fixed. [Chomsky, 1986]

Langacker's Rule vs. List Fallacy

Available evidence suggests that both generalizations ('rules') and item-specific knowledge ('lists') are recorded.
Instances are represented at some level of abstraction due to selective encoding; that is, since not all features of an item are represented, the representation is necessarily partially abstract.
Moreover, generalizations across instances are also made.

Goldberg 2006

[A] similar position has been developed within the field of categorization. Most recently, categorization researchers have argued for an approach that combines exemplar-based knowledge with generalizations over that knowledge (Anderson 1991; Murphy 2002; Ross and Makin 1999).

Misconceptions about CxG

- CxG is obsessed with linguistic marginalia.

  maybe, but...

- Fillmore, Kay, Goldberg and others discuss patterns of complementation, passives, lexical representation, datives, resultatives, ...

- Ginzburg and Sag 2000, Sag 2010 provide (very) detailed accounts of *wh*-constructions

Misconceptions about CxG

- CxG is opposed to compositional semantics.
- 'Frege's Principle: the meaning of a complex expression is determined by the meanings of its constituent parts, in accordance with their syntactic combination'
- CxGrammarians take compositionality wherever they can get it.

Misconceptions about CxG

- ▶ CxG is not constrained.
- ▶ CxG does not provide a universal framework for syntax.
- ▶ This is addressed squarely in SBCG

Universals and SBCG 1

- ▶ Dryer (1997), Croft (2001), Evans and Levinson (2008), and others

  argue for theorizing about universals without a universal vocabulary.

- ▶ Most universals are probabilistic.

- ▶ Formal explanations rule out in principle what can occur with low frequency.

- ▶ E.g. SVO languages tend to be prepositional;

- ▶ Common patterns across languages have functional or cognitive motivation.

- ▶ More uniform constraints on the linearization of heads are easier to learn.

  head-final or head-initial.

Universals and SBCG 2

- ▶ But SBCG is perfectly consistent with strong nativist assumptions, including UG.
- ▶ More general types would be good candidates for principles of UG.
- ▶ In facts, computational work in HPSG has led to the development of a notion of a 'grammar matrix'.

  Rapid prototyping of fully implemented grammars of new languages.

  See the HPSG LinGO Grammar Matrix (Emily Bender and colleagues).

- ▶ But functional explanations are better explanations!

Common Themes (Analytic/Formal)

- Constructions are present and primitive in the theory and related to one another
- Variable Grain Generalizations
- No sharp distinction between Syntax and Lexicon
- Grammar is infused with Semantics (rejection of 'syntactocentrism'; Jackendoff 2002)

Common Themes (Empirical/Methodological)

- Broad Empirical Responsibility (rejection of core vs. periphery)
- Data-Based Learning (rejection of Parameter-Setting models of learning)
- Cautious approach to Universals (rejection of Chomskyan UG as a theoretical starting point)
- Explain as much as possible about language in terms of more general cognitive and/or functional considerations.
- Grammar is the residue that can't be explained without stipulation.

The History of SBCG

- A dialogue between researchers in Berkeley Construction Grammar (BCG) and Head-Driven Phrase Structure Grammar (HPSG) in the San Francisco Bay area in the late 1980s.
- led to certain refinements of BCG and to the constructional version of HPSG developed in Sag 1997 and Ginzburg and Sag 2000.
- Emergence of common framework by early 2000s.

The History of SBCG

BCG and HPSG

Common Assumptions of BCG and HPSG

1. Linguistic objects are modeled in terms of feature structures (representable as attribute-value matrices or directed graphs).

2. Feature values are sometimes complex. (Feature structures can be recursive.)

3. A language consists of a set of signs; *sign* is an abstract entity that is the locus of constraints on the interface of form and meaning.

4. A grammar is a system of constraints that work together to license and delimit the signs of a given language.

5. *Constructions*, the constraints on classes of signs and their components, are organized into a regime (a lattice-like array of types and subtypes) that allows generalizations of varying grain to be stated.

6. The distinction between lexical and grammatical entities is blurry, motivating a uniform conception of lexical and constructional constraints.

- ▶ **Construction Interaction:** How do constructions interact? Do con structions freely combine when compatible? Are some constructions optional? Are some constructions obligatory? How does a grammar guarantee that the 'right' set of constructions apply to a given example?

- ▶ **The Locality of Constructions:** Do constructions need to make reference to properties of elements embedded within phrases (or boxes) at arbitrary depth?

- ▶ **The Limits of Underspecification:** Can the various argument-structure constructions be analysed in terms of underspecification of valence in a single lexical entry? Can determinerless noun phrases (with plural or mass head nouns) be given a uniform account via feature underspecification?

- ▶ **Various Constructions:** How to analyse certain constructions (primarily in English), including passive, subcategorization, filler-gap dependencies, idioms of various kinds, genitive NPs, determiners, conditionals, control, raising, unexpressed arguments, ellipsis, reflexive binding, ...

Conclusions

The goal of SBCG is to develop a theory of grammar that is psycholinguistically responsible,

► That goal leads to an architecture where rules and principles are stated statically in terms of constraints that structures must satisfy,

► where the notions of sign and construction are central, and where lexical integrity prevails.

► In addition, explicit model(s) of processing need to be developed in tandem with the development of particular competence grammars and the competence theory.

► The desired result is a theoretically grounded theory of linguistic knowledge that is fits within a broader theory of communication.