Last few years have seen increasing synergies

still claims corpus linguistics "does not exist".

Developed almost separately from each other (philosophical differences – Chomsky appeared

Begun at about the same time (c. F. Shamoon)

From computer science / engineering (attempt to use & describe language as observable

Data-driven ("stochastical", "statistical", "empirically", "bottom-up")

Began in the 50s (c. Chomsky)

From linguistics (attempt to describe human linguistic capability

Knowledge/rule-based ("symbolic", "hand-crafted", "top-down")

Of technological: "to provide a working component of a speech or natural language system"

or psychological: "phenomenon"

Aims can be scientific: "trying to provide a computational explanation for a particular linguistic

"the scientific study of language from a computational perspective".

**What is Computational Linguistics?**
(gol a six) \(d\)

Can do this using n-grams (probabilities of words based on previous \(n-1\) words)

Although even better to use context: "I've got a ... six."

Individual word probabilities might rule out "squid".

Distinguishation requires some sort of linguistic knowledge.

Speech Recognition

Can use POS-tagging to rank words in order of importance (e.g. nouns/verbs above adjectives).

\[
\left[ \frac{(\text{words})^{\text{corr}}}{(\text{matches})^{\text{corr}}} \right] / \left[ \frac{(\text{words})^{\text{op}}} {(\text{matches})^{\text{op}}} \right] = \ \text{tffidf (term frequency) (inverse document frequency)}
\]

This based on tffidf score:

"Bag of Words" Approach: Remove common function words from search term, rank documents.

**Statistical NLP - Applications**
When is statistical NLP not enough?
<table>
<thead>
<tr>
<th>Grammatical Structure</th>
<th>Logical Meaning</th>
<th>Contextual Meaning or Impredicative</th>
<th>Pragmatics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\delta()$ ${\text{English}}$ ${x, \ldots, x}$ $\text{English} \rightarrow \text{you speak}$</td>
<td>$\delta()$ ${\text{English}}$ ${x, \ldots, x}$ $\text{English} \rightarrow \text{you speak}$</td>
<td></td>
</tr>
<tr>
<td>Syntax</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orthography</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words</td>
<td>$\text{do you speak English}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sounds</td>
<td>$/\text{spirk inglif}/$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$/\text{sprik inglif}/$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**The Nature of Natural Language**
Context-Sensitive Grammars (for later)

Agreement easily
not bad, structure available; efficient parsing techniques (chart parsing), but can’t express e.g.
Valid parse is one that covers exactly all words in string using valid rules. Expressive power

<table>
<thead>
<tr>
<th>Parse</th>
<th>( \Lambda )</th>
<th>( \Lambda )</th>
<th>( \Lambda )</th>
<th>( \Lambda )</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>( N )</td>
<td>( N )</td>
<td>( N )</td>
<td>( S )</td>
</tr>
<tr>
<td>do</td>
<td>( N )</td>
<td>( \Lambda )</td>
<td>( \Lambda )</td>
<td>( \Lambda )</td>
</tr>
<tr>
<td>sleep</td>
<td>( \Lambda )</td>
<td>( \Lambda )</td>
<td>( \Lambda )</td>
<td>( \Lambda )</td>
</tr>
<tr>
<td>sees</td>
<td>( \Lambda )</td>
<td>( \Lambda )</td>
<td>( \Lambda )</td>
<td>( \Lambda )</td>
</tr>
</tbody>
</table>

Context-Free Grammars (push-down automata)

Enough, and don’t give (easily) useful output structure.
Valid parse is one that pushes in end state and uses exactly all words in string. Not expressive

Regular Grammars (finite state automata) as commonly used in speech recognition - HMMs

Syntactic Processing

Introduction to Computational Linguistics
Still fails to capture many generalisations about rules.

\[
\begin{align*}
[d = \text{num}]_N & \iff [d = \text{num}]_P N \\
[N = \text{num}]_D & \iff [N = \text{num}]_P N \\
[N = \text{num}]_P \land [N = \text{num}]_P N & \iff S
\end{align*}
\]

We can capture generalisations with a VGC.

Ends up with many hundreds of rules.

\[
\begin{align*}
[d = \text{num}]_N & \iff [d = \text{num}]_P N \\
[N = \text{num}]_P & \iff [N = \text{num}]_P N \\
[N = \text{num}]_P \land [N = \text{num}]_P N & \iff S \\
[N = \text{num}]_P \land [N = \text{num}]_P N & \iff S
\end{align*}
\]

With a CP grammar, agreement is difficult to capture.

**Attribute-Value Grammars**
Shalom, Howard & Christian

Dynamic Predicate Logic, Discourse Representation Theory, Situational Semantics

But what to do with correspondence? And what about typification (logical omniscience)?

Model-theoretic interpretation (possible worlds, intension, to determine truth values.

\[
\begin{align*}
\{x\}p & \cup \{x\}op & \text{ } xE & \rightarrow d \alpha dN \leftarrow S \\
\{x\}d \alpha & \cup \{x\}op & \text{ } xE & \text{ } d\alpha y & \rightarrow N \text{ } \forall d & \leftarrow dN \\
(C)B & \rightarrow C & B & \leftarrow A
\end{align*}
\]

Build constituent representations:

Give each word semantic representation, and add a semantic application rule to the parser to

Compositional, easy to integrate with syntax.

First order logic, lambda calculus - familiar.

Montague Semantics

\begin{center}
\textbf{Semantic Representation}
\end{center}
Either use real-world knowledge (as we do) or (much simpler) probabilistic methods.

"Worse than you might think - (Martin et al. 1987) get 455 parses for:

\[ \text{(fruits) fish like a banana} \]

\[ \text{fish like an avocado} \]

\[ \text{time fish like an avocado} \]

Structural ambiguity: "fish like a Y"
But this wouldn’t help with “fruit flies” to help. Probabilities for rare structures may be inaccurate. Can use smoothing and back-off techniques.

Problem with Zipfian nature of NL - all unseen structures assigned zero probability, and estimated

Many need normalization on parse length (higher number of rules lower p)

(Generalization of speech recognition techniques)

Can be trained from unannotated corpora using EM and the Inside-Outside algorithm.

\[
\frac{(x \leftarrow \mathcal{A})^s u_x \mathcal{X}^s}{(\mathcal{B} \leftarrow \mathcal{A})^s u_x} = (\mathcal{B} \leftarrow \mathcal{A})^d
\]

Easy to train from hand-annotated corpora using MLE (count number of occurrences of each

Parser calculates total probability of each parse, returns most probable structure.

...`

1.0  d \Lambda dN  \leftarrow S

\text{Associate each rule with a probability: e.g.:}

Stochastic CFGs
techniques proposed (Abney 1997) in random fields for statistical sampling. Does it matter in
problem with training of AVGS - missing probability mass associated with unseen rules. Other

\[ \begin{array}{ccc}
0.5 & [d = \text{unn}] N & [d = \text{unn}] \leftarrow [d = \text{unn}] d N \\
0.5 & [s = \text{unn}] N & [d = \text{unn}] \leftarrow [s = \text{unn}] d N \\
6.0 & [s = \text{unn}] N & [s = \text{unn}] \leftarrow [s = \text{unn}] d N \\
1.0 & & [s = \text{unn}] d N
\end{array} \]

attribute values, e.g.,

- We can extend this probabilistic approach to AVGS - can even condition rule probability on

- especially with lexicized versions.

- Training methods similar to SCFGs, but sparsity of training data becomes a serious problem.

\[ \begin{array}{ccc}
0.0 & \frac{\text{time}}{\text{files}} N & \text{files} \leftarrow d N \\
0.0 & \frac{\text{time}}{\text{files}} N & \text{files} \leftarrow d N
\end{array} \]

- SCFGs, or both e.g.,

- Make rule probabilities dependent on daughter categories (SEDCFG), or words (lexically

Advanced SCFGs, Stochastic AVGS
Must also handle (amongst others) illocutionary force, grounding, hesitation.

Also need interface to phonology, and need to identify speech acts / conversational moves. Need control for anaphora, ellipsis, childbirths.

Linguistic approach required for true open-domain application.

← "sentences inducing 'route' → route planning subroutine, sentence induction." "from Statistical approach used in some applications today depends on detailed knowledge of domain"

Caller: Which is near Penitth.
Wizard: L-H-O-R-E?
Caller: Wh - L-H-O-R-E
Wizard: Yeah
Caller: B-Y
Wizard: Yeah
Caller: Yeah - K-L-R
Wizard: Can you spell that please?
Caller: to Cryptohore in Cumberland.
Wizard: Whim
Caller: From Malvern
Wizard: Where would you like to go?
Caller: Hello, um, well to plan a route really.
Wizard: Welcome to the route planning service. How may I help you?

Dialogue Systems
in an attribute-value matrix or signature in Part I of this book, the formal semantics, syntax, and computational linguistics.

Head-driven phrase structure grammar (HPSC)
schema / phrasal types - but is it learnable?

Stochastic version is (theoretically) possible ([209-1997] - associate probabilities with rule

contextual information.

Allows use of information & constraints from other levels (e.g. syntax ↔ semantics), including

Hierarchical

Default inheritance in the lexicon allows information to be specified in a type

categories); e.g.

Highly generalized rule schema over phrasal types (instead of individual rules over phrasal

HPSG (2)

Introduction to Computational Linguistics

KCL PhD Seminar
This gives us a way to calculate a participant's beliefs and the questions being discussed - i.e., a

We also need some sort of update algorithm e.g.,

\[
\begin{bmatrix}
\langle \text{chat}(X), \text{KRIBITTHORF}\rangle, \text{ask(theme)} \\
\langle \text{special}(X) \text{KRIBITTHORF}\rangle, \text{special} \text{X}
\end{bmatrix}
\]

Can include dialogue context, beliefs, future plans.


We can represent the information state for a conversational participant as an AVM (Ginzburg)

Information States
Other more complex types of ellipses include bare answers, suffixes, VP ellipses, and contraction ellipses.

SHARDS is a working implementation (http://pc-320.dcs.kcl.ac.uk:8080/).

Example: Yes, from Michael: "I-H-O-R-F?"

Ellipses are incomplete fragments which can only be fully interpreted in context.

Ellipses
But what meaningful(s) can we assign to CE? And how can we disambiguate between CE and other

•

Lexically identical moves:

But what meaningful(s) can we assign to CE? And how can we disambiguate between CE and other

•

Lexically identical moves:
That is the relation between form and meaning.

Disambiguation between meanings and from other move types could use information state (first)

There are other possible forms (e.g. reprieve; "where is X?") but what are they?

There are other possible meanings (e.g. "indexed gap"); but what are they?

Clarification Ellipsis

At least two possible meanings: a "check" type (did you really say X?) and a "reference" type.