

Natural Language Processing

Introduction to NLP

Natural Language Processing

We're going to study what goes into getting computers to perform useful and interesting tasks involving human language.

Slides by James Martin, adapted by Diana Inkpen
for CSI 5386 @ uOttawa

Natural Language Processing

More specifically, it's about the algorithms that we use process language, the formal basis for those algorithms, and the facts about human language that allow those algorithms to work.

Why Should You Care?

Three trends

1. An enormous amount of information is now available in machine readable form as natural language text (newspapers, web pages, medical records, financial filings, etc.)
2. Conversational agents are becoming an important form of human-computer communication
3. Much of human-human interaction is now mediated by computers via social media

Applications

- Let's take a quick look at three important application areas
 - ◆ Text analytics
 - ◆ Question answering
 - ◆ Machine translation

Text Analytics

- Data-mining of weblogs, microblogs, discussion forums, message boards, user groups, and other forms of user generated media
 - ◆ Product marketing information
 - ◆ Political opinion tracking
 - ◆ Social network analysis
 - ◆ Buzz analysis (what's hot, what topics are people talking about right now)

Text Analytics

canon powershot - Live Search Products

http://search.live.com/products/?q=canon%20powershot&p1=%5bCommerceService+scenario%3d%22reviews%22+doc Google


Daily Camer... Information New York Times CU Departme...ter Science James Marti... Home Page Linda's calendar

Live Search canon powershot

Products See also: [Web](#), [Images](#), [Video](#), [News](#), [Maps](#), [More](#) ▼

[Live Search cashback](#) [How this works](#) [Browse stores](#) [Your account](#)

Canon PowerShot A630 - digital camera, 8MP, 4x Optical Zoom, 4x ...

 \$174 - \$319 [Compare prices](#) (12)
★★★★☆ [User reviews](#) (260)
★★★★☆ [Expert reviews](#) (2)

Canon PowerShot A630 - Digital camera - 8.0 Mpix - optical zoom: 4 x - supported memory: MMC, SD, SDHC

[User reviews](#) | [Product details](#) | [Expert reviews](#) | [Compare prices](#)

All user reviews

Ease Of Use (90 comments)
89% positive

Photo Quality (80 comments)
89% positive

Affordability (48 comments)
81% positive

Size (38 comments)
71% positive

Screen (34 comments)
88% positive

Zoom (28 comments)
68% positive

Speed (20 comments)
90% positive

Battery Life (19 comments)
95% positive

Settings (18 comments)
61% positive

All user reviews
View by: [Most recent](#) | [Highest rating](#) | [Lowest rating](#)

★★★★☆ **canon powershot**
Like the ease of operation. Dislike the small amount of memory. Love the cannon. I bought it because consumer reports gave it their highest rating. [More...](#)
midwest-builders catalog.ebay.com 12/21/2008

★★★★☆ **Canon PowerShot A630 8.0 Megapixel**
Very good camera, it takes a nice picture. The zoom in amazing for such a small package at an affordable price. Still a little hard on batteries, but I think they all are. [More...](#)
spelhaug1 catalog.ebay.com 12/12/2008

★★★★★ **Canon digital cameras ROCK**
This camera ROCKS. I have used the Canon A620 for a couple years and LOVE it. A630 goes one better being 8MP (plus other features). And the software gives you just about everything... [More...](#)
rochs catalog.ebay.com 12/8/2008

★★★★★ **Canon A630**
This is the second Canon Powershot series camera. Not being a shutterbug, I wanted a point and shoot camera for my business (real estate appraiser). My first A80 was very reliable.... [More...](#)
nicdw catalog.ebay.com 12/3/2008

★★★★★ **fabulous camera, I have two.**
great zoom and large screen. This camera is user friendly and takes great pictures. I bought one for my daughter cuz she was always "borrowing" mine. [More...](#)
scouler827 catalog.ebay.com 11/12/2008

Sponsored sites
[Shop, Compare & Save](#)
Search now and get cashback on **canon a powershot!**
[Search.Live.com/cashback](#)
[Live Search cashback](#)

Canon Digital Camera
Large Selection of Quality **Canon** Digitals! Buy Now & Pay No Tax.
[ritzcamera.com](#)

Canon/Digital Camera
We Help Solve Your Quest! Save on **Canon/Digital Camera.**
[bizrate.com](#)

Powershot
Bid on **Powershot.** Shop eBay!
[www.eBay.com](#)

Electronics At SEARS®
Shop Electronics At SEARS.com. Buy Online, Pick-Up In-Store.
[www.SEARS.com](#)
[Live Search cashback](#)

Canon Cameras Sale
Buy **canon powershot** digital cameras at reduced price. Free shipping.
[www.amazon.com](#)

Top 10 Canon Cameras
Canon Digital Cameras Discounted. Save Money at Lowest Priced Stores.
[Canon-Cameras.ShopCompareUS.com](#)

HP @ Photo Printers
Get a Photo Printer That's Right for You. Buying Guide at hp.com.
[www.hp.com](#)

[See your message here...](#)

Text Analytics

User reviews | Product details | Expert

All user reviews

Ease Of Use (90 comments)



Photo Quality (80 comments)



Affordability (48 comments)



Size (38 comments)



Screen (34 comments)



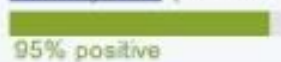
Zoom (28 comments)



Speed (20 comments)



Battery Life (19 comments)



Canon powershot - Live Search Products

Search results for Canon powershot

James Marti... Home Page Linda's calendar



Live Search cashback How this works Browse stores Your account

8MP, 4x Optical Zoom, 4x ...

8.0 Mpix - optical zoom: 4 x - supported memory:

Compare prices

Highest rating | Lowest rating

Canon powershot

Dislike the small amount of memory. Love it because consumer reports gave it their highest

og.ebay.com 12/21/2008

PowerShot A630 8.0 Megapixel

It takes a nice picture. The zoom is amazing for such an affordable price. Still a little hard on batteries, but

More...

ay.com 12/12/2008

digital cameras ROCK

I have used the Canon A620 for a couple years and it goes one better being 8MP (plus other features). And you just about everything...

More...

m 12/8/2008

A630

Canon Powershot series camera. Not being a point and shoot camera for my business (really first A80 was very reliable....

More...

m 12/3/2008

is camera, I have two.

This camera is user friendly and takes great pictures. This camera is user friendly and takes great pictures. This camera is user friendly and takes great pictures.

More...

ay.com 11/12/2008

Sponsored sites

[Shop, Compare & Save](#)

Search now and get cashback on **Canon** a powershot!

[Search.Live.com/cashback](#)

[Live Search cashback](#)

[Canon Digital Camera](#)

Large Selection of Quality **Canon** Digitals! Buy Now & Pay No Tax. [ritzcamera.com](#)

[Canon/Digital Camera](#)

We Help Solve Your Quest! Save on **Canon/Digital Camera**. [bizrate.com](#)

[Powershot](#)

Bid on **Powershot**. Shop eBay! [www.eBay.com](#)

[Electronics At SEARS®](#)

Shop Electronics At SEARS.com. Buy Online, Pick-Up In-Store. [www.SEARS.com](#)

[Live Search cashback](#)

[Canon Cameras Sale](#)

Buy **Canon powershot** digital cameras at reduced price. Free shipping. [www.amazon.com](#)

[Top 10 Canon Cameras](#)

Canon Digital Cameras Discounted. Save Money at Lowest Priced Stores. [Canon-Cameras.ShopCompareUS.com](#)

[HP @ Photo Printers](#)

Get a Photo Printer That's Right for You. Buying Guide at [hp.com](#). [www.hp.com](#)

[See your message here...](#)

Question Answering

- Traditional *information retrieval* provides documents/resources that provide users with what they need to satisfy their information needs.
- *Question answering* on the other hand directly provides an answer to information needs posed as questions.

Web Q/A

Web [Images](#) [Videos](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#) ▼

james.h.martin@gma



what's the population of Boulder

Search

[Advanced Search](#)

Web [+ Show options...](#)

Results 1 - 10 of about 6,290,000 for what's the population of Boulder. (0.24 seconds)

[Population, Boulder County, CO](#)



293,161 - Jul 2008

Source: U.S. Census Bureau, **Population** Division

[Disclaimer](#)

www.google.com/publicdata

[Boulder, Colorado - Wikipedia, the free encyclopedia](#)

The United States Census Bureau estimates that in 2008 the **population** of the city of **Boulder** was 94268, while the **population** of the **Boulder** Metropolitan ...

[Demographics](#) - [Geography and climate](#) - [History](#)

en.wikipedia.org/wiki/Boulder,_Colorado - [Cached](#) - [Similar](#) - [Comment](#) [Share](#) [Close](#)

Watson



Machine Translation

The automatic translation of texts between languages is one of the oldest non-numerical applications in Computer Science.

In the past 10 years or so, MT has gone from a niche academic curiosity to a robust commercial industry.

Google Translate

الجزيرة نت

http://www.aljazeera.net/NR/exeres/8FD54E7F-56C5-49A0-B60A-89A67426F3B3.htm

Speech and ... of Contents Book Schedule University o...ter Science James Marti... Home Page The Daily Camera The New Yor...Multimedia

نحضر لك الأخبار الساخنة أينما تكون
اشترك الآن

ملخص الخبر

انتقل الأمين العام للأمم المتحدة الذي يزور لبنان لقرعة المراقبة الدولية ببلدة الناقورة الجنوبية حيث وضع إكليلًا من الزهور على نصب لأربعة مراقبين قتلوا خلال الحرب الأخيرة، وفيما ينتظر تقعد آنان للقري المدمرة تنظم إيطاليا اليوم احتفالاً لتوديع قواتها المرسلة إلى لبنان.

أنان في الناقورة لتقعد الخط الأزرق والمناطق للمدمرة

الثلاثاء 5/8/1427 هـ - الموافق 29/8/2006 م (آخر تحديث) الساعة 17:44 (مكة المكرمة)، 14:44 (غرينتش)

مكة إسرائيلية تمدد اعتقال وزير للثلاثية الفلسطينية و13 نائباً من حماس

استشهاد فلسطينيين وإصابة تسعة في غارات بالصفقة والقطاع

أصيب تسعة فلسطينيين بينهم مدنيون في غارة جوية إسرائيلية على حي الشجاعية في قطاع غزة، يأتي ذلك مباشرة بعد استشهاد قناصين بارزين من كتائب شهداء الأقصى في عملية لقوات الاحتلال الإسرائيلي نفذها سلاح الجو وقوات المشاة في مخيم بلاطة بالصفقة الغربية.

البشير يلتقي فريزر ومجلس الأمن لن يفرض قوات بدارفور

من المقرر أن يلتقي الرئيس السوداني عمر البشير جيندائي فريزر مساعد وزير الخارجية الأميركية التي تحاول في الخرطوم إقناع المسؤولين السودانيين بنشر قوة أممية بدارفور، من جانبه قال السفير الأميركي في الأمم المتحدة إنه لا نية لمجلس الأمن بفرض قوات في الإقليم.

رمسفيدل وتشيني يصران على إبقاء القوات الأميركية بالعراق

دعا وزير الدفاع الأميركي دونالد رمسفيدل الأميركيين إلى التحلي بالصبر بخصوص العراق، وانتهد ديك تشيني نائب الرئيس دعوات الديمقراطيين لسحب القوات الأميركية من العراق للربط بين الانسحاب المبكر واحتمال وقوع هجمات داخل الولايات المتحدة.

مقتل مدنيين وإصابة ضابط بهجوم انتحاري في أفغانستان

أعلنت القوة الدولية للمساعدة على إرساء الأمن (إيساف) مقتل مدنيين وإصابة ضابط أفغاني في هجوم استهدف قافلة لقوات الأطلسي بجنوب أفغانستان. وفي العاصمة كابل انفجرت قنبلة يدوية الصنع لدى مرور عميلة فرنسية ملتحمة لنساء عن انصارات لم أخيرا.

العضوية

الأخبار الفضائية المعرفة الأعمال

روابط أخرى

خدمات الموقع

بحث

ابحث تفصيلي

انذهب

غزو غزة بالأيام

01 02 03 04 05 06

هذه الصفحة برعاية

المتفوقون الحقيقيون لا يلحقون بالأخريين بل يقودون الطريق

القطرية QATAR AIRWAYS

اسم البرنامج: بلا حدود

عنوان الحلقة: يوسف ندا، اتهامات بدعم الإرهاب الأربعاء/ مباشر 19:05 غرينتش 22:05 مكة

frontiers@aljazeera.net

التفصيلية الاقتصادية الرياضية

Google Translate

Killing Palestinians and wounding nine in the raids Sector

Nine Palestinians were wounded among civilians in an Israeli air raid in the neighborhood result in the Gaza Strip. This comes immediately after the killing of two prominent Al-Qsa Martyrs Brigades in the Israeli occupying forces carried out air and infantry forces in the Balata camp in the West Bank.



Bashir meets Fraser, the Security Council will not impose forces Darfur

Is scheduled to meet with Sudanese President Omar al-Bashir Jenday Fraser Assistant Minister for Foreign Affairs of the American attempt to persuade officials in Khartoum, Sudanese Darfur deployment of the nationalities. For his part, US Ambassador to the United Nations that it has no intention of the Security Council to impose its forces in the province.



Rmsfield and Cheney insist on keeping the American forces in Iraq

Called American Defense Minister Donald Rmsfield Americans to show patience on Iraq. I take Vice President Dick Cheney calls Democrats withdrawal of American forces from Iraq link and the possibility of early withdrawal of attacks inside the United States.



Killing civilians and wounding officer suicide attack in Afghanistan

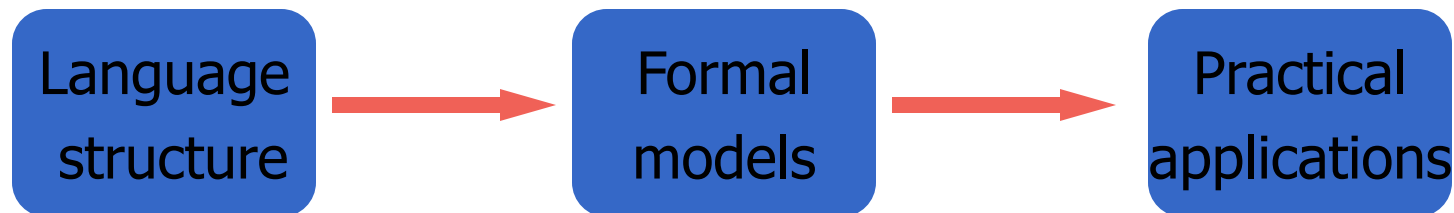
The international force to help establish security (ISAF) killed civilians and the wounding of an officer in an attack against Afghan forces convoy south Atlantic Afghanistan. In the capital Kabul, a hand grenade exploded at the passage of manufacture French patrol was not reported injuries or damage.




How?

All of these applications operate by exploiting underlying regularities inherent in human languages.

Sometimes in complex ways, sometimes in pretty trivial ways.



Major Class Topics

1. Words
 2. Syntax
 3. Meaning
(semantics)
 4. Texts
(discourse)
 5. Applications exploiting each
- 

Applications

First, what makes an application a *language processing application* (as opposed to any other piece of software)?

- ◆ An application that **requires the use of knowledge about the structure of human language**
 - Example: Is Unix `wc` (word count) an example of a language processing application?

Applications

- Word count?
 - ◆ When it counts words: **Yes**
 - To count words you need to know what a word is. That's knowledge of language.
 - Note that the definition of "word" embodied in wc doesn't work for Chinese or other languages that don't delimit words with spaces
 - ◆ When it counts lines and bytes: **No**
 - Lines and bytes are computer artifacts, not linguistic entities

Questions?

Course Material

- We' ll be intermingling discussions of:
 - ◆ Linguistic topics
 - Morphology, syntax, semantics, discourse
 - ◆ Formal systems
 - Regular languages, context-free grammars, probabilistic models
 - ◆ Applications
 - Question answering, machine translation, information extraction

Course Material

- We won't be doing speech recognition or synthesis.

Topics: Linguistics

- Word-level processing
- Syntactic processing
- Lexical and compositional semantics

Topics: Techniques

- Finite-state methods
- Context-free methods
- Probabilistic models



Supervised machine
learning methods

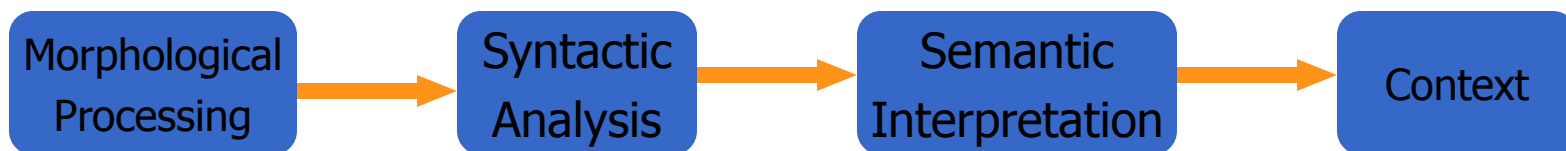
Categories of Knowledge

- Phonology
- Morphology
- Syntax
- Semantics
- Pragmatics
- Discourse

Each kind of knowledge has associated with it an encapsulated set of processes that make use of it.

Interfaces are defined that allow the various levels to communicate.

This often leads to a pipeline architecture.



Ambiguity

- Ambiguity is a fundamental problem in computational linguistics
- Hence, resolving, or managing, ambiguity is a recurrent theme

Ambiguity

- Find at least 5 meanings of this sentence:
 - ◆ I made her duck

Ambiguity

- Find at least 5 meanings of this sentence:
 - ◆ I made her duck
- I cooked waterfowl for her benefit (to eat)
- I cooked waterfowl belonging to her
- I created the (ceramic?) duck she owns
- I caused her to quickly lower her upper body
- I waved my magic wand and turned her into undifferentiated waterfowl

Ambiguity is Pervasive

- I caused her to quickly lower her head or body
 - ♦ **Lexical category:** “duck” can be a noun or verb
- I cooked waterfowl belonging to her.
 - ♦ **Lexical category:** “her” can be a possessive (“of her”) or dative (“for her”) pronoun
- I made the (ceramic) duck statue she owns
 - ♦ **Lexical Semantics:** “make” can mean “create” or “cook”, and about 100 other things as well

Ambiguity is Pervasive

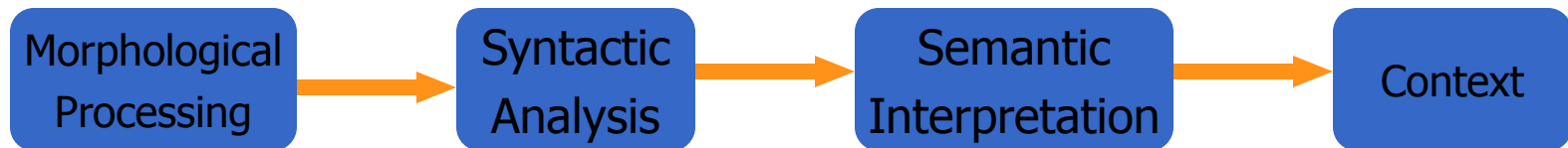
- Grammar: Make can be:
 - ◆ Transitive: (verb has a noun direct object)
 - I cooked [waterfowl belonging to her]
 - ◆ Ditransitive: (verb has 2 noun objects)
 - I made [her] (into) [undifferentiated waterfowl]
 - ◆ Action-transitive (verb has a direct object and another verb)
 - ◆ I caused [her] [to move her body]

Ambiguity is Pervasive

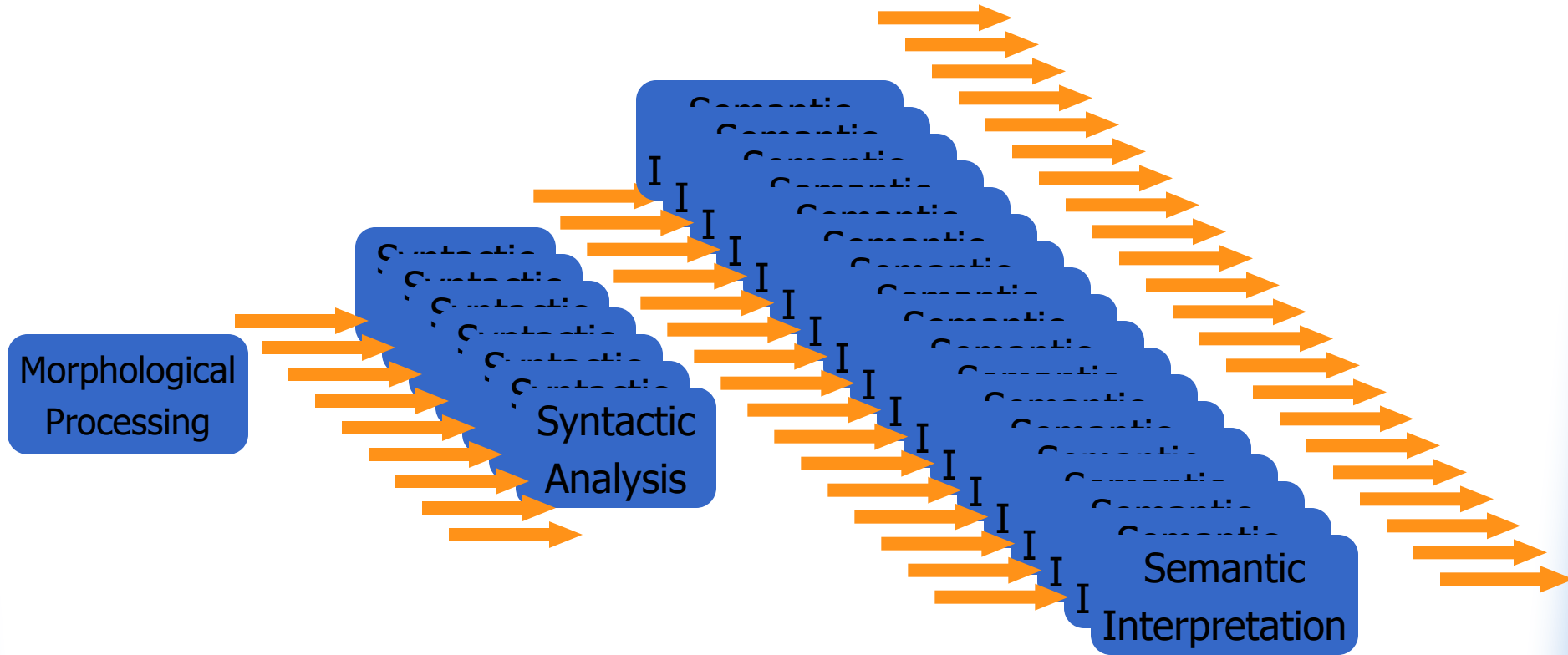
- **Phonetics!**
 - ◆ I mate or duck
 - ◆ I' m eight or duck
 - ◆ Eye maid; her duck
 - ◆ Aye mate, her duck
 - ◆ I maid her duck
 - ◆ I' m aid her duck
 - ◆ I mate her duck
 - ◆ I' m ate her duck
 - ◆ I' m ate or duck
 - ◆ I mate or duck

Problem

- Remember our pipeline...



Really it's this



Dealing with Ambiguity

- Four possible approaches:
 1. Tightly coupled interaction among processing levels; knowledge from other levels can help decide among choices at ambiguous levels.
 2. Pipeline processing that ignores ambiguity as it occurs and hopes that other levels can eliminate incorrect structures.

Dealing with Ambiguity

3. Probabilistic approaches based on making the most likely choices
 1. Or passing along n-best choices
4. Don't do anything, maybe it won't matter
 1. *We'll leave when the duck is ready to eat.*
 2. *The duck is ready to eat now.*
 - Does the “duck” ambiguity matter with respect to whether we can leave?

Models and Algorithms

- By **models** we mean the formalisms that are used to capture the various kinds of linguistic **knowledge** we need.
- **Algorithms** are then used to manipulate the knowledge representations needed to tackle the task at hand.

Models

- State machines
- Rule-based approaches
- Logical formalisms
- Probabilistic models

Algorithms

- Many of the algorithms that we'll study will turn out to be **transducers**; algorithms that take one kind of structure as input and output another.
- Unfortunately, ambiguity makes this process difficult. This leads us to employ algorithms that are designed to handle ambiguity of various kinds

Paradigms

- In particular..
 - ◆ State-space search
 - To manage the problem of making choices during processing when we lack the information needed to make the right choice
 - ◆ Dynamic programming
 - To avoid having to redo work during the course of a state-space search
 - CKY, Earley, Minimum Edit Distance, Viterbi, Baum-Welch
 - ◆ Classifiers
 - Machine learning based classifiers that are trained to make decisions based on features extracted from the local context