LECTURE 07: SYNTAX AND SEMANTICS

- Objectives:
  - Understand the role of higher level knowledge in speech recognition
  - Introduce how we can exploit knowledge about the structure of language to improve speech recognition performance
  - Gain an appreciation for the fields of linguistics and natural language processing
  - Introduce alternate choices for acoustic units

Note that this lecture is primarily based on material from the course textbook:

> X. Huang, A. Acero, and H.W. Hon, *Spoken Language Processing - A Guide to Theory, Algorithm, and System Development*, Prentice Hall, Upper Saddle River, New Jersey, USA, ISBN: 0-13-022616-5, 2001.

A good reference textbook on these topics is:

> D. Jurafsky and J.H. Martin, *SPEECH and LANGUAGE PROCESSING: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice-Hall, ISBN: 0-13-095069-6, 2000.

Return to Main

Introduction:

01: Organization
(html, pdf)

Speech Signals:

02: Production
(html, pdf)

03: Digital Models
(html, pdf)

04: Perception
(html, pdf)

05: Masking
(html, pdf)

06: Phonetics and Phonology
(html, pdf)

07: Syntax and Semantics
(html, pdf)

Signal Processing:

08: Sampling
(html, pdf)

09: Resampling
(html, pdf)

10: Acoustic Transducers
(html, pdf)

11: Temporal Analysis
(html, pdf)

12: Frequency Domain Analysis
(html, pdf)

13: Cepstral Analysis
(html, pdf)

14: **Exam No. 1**
(html, pdf)

15: Linear Prediction
(html, pdf)

16: LP-Based Representations
(html, pdf)

Parameterization:

17: Differentiation
(html, pdf)

18: Principal Components
(html, pdf)

ECE 8463: FUNDAMENTALS OF SPEECH RECOGNITION

Professor Joseph Picone
Department of Electrical and Computer Engineering
Mississippi State University

email: picone@isip.msstate.edu
phone/fax: 601-325-3149; office: 413 Simrall
URL: http://www.isip.msstate.edu/resources/courses/ece_8463

Modern speech understanding systems merge interdisciplinary technologies from Signal Processing, Pattern Recognition, Natural Language, and Linguistics into a unified statistical framework. These systems, which have applications in a wide range of signal processing problems, represent a revolution in Digital Signal Processing (DSP). Once a field dominated by vector-oriented processors and linear algebra-based mathematics, the current generation of DSP-based systems rely on sophisticated statistical models implemented using a complex software paradigm. Such systems are now capable of understanding continuous speech input for vocabularies of hundreds of thousands of words in operational environments.

In this course, we will explore the core components of modern statistically-based speech recognition systems. We will view speech recognition problem in terms of three tasks: signal modeling, network searching, and language understanding. We will conclude our discussion with an overview of state-of-the-art systems, and a review of available resources to support further research and technology development.

Tar files containing a compilation of all the notes are available. However, these files are large and will require a substantial amount of time to download. A tar file of the html version of the notes is available here. These were generated using wget:

```
wget -np -k -m http://www.isip.msstate.edu/publications/courses/ece_8463/lectures/current
```

A pdf file containing the entire set of lecture notes is available here. These were generated using Adobe Acrobat.

Questions or comments about the material presented here can be directed to help@isip.msstate.edu.

19: Linear Discriminant Analysis (html, pdf)

LECTURE 07: SYNTAX AND SEMANTICS

- Objectives:
  - Understand the role of higher level knowledge in speech recognition
  - Introduce how we can exploit knowledge about the structure of language to improve speech recognition performance
  - Gain an appreciation for the fields of linguistics and natural language processing
  - Introduce alternate choices for acoustic units

Note that this lecture is primarily based on material from the course textbook:

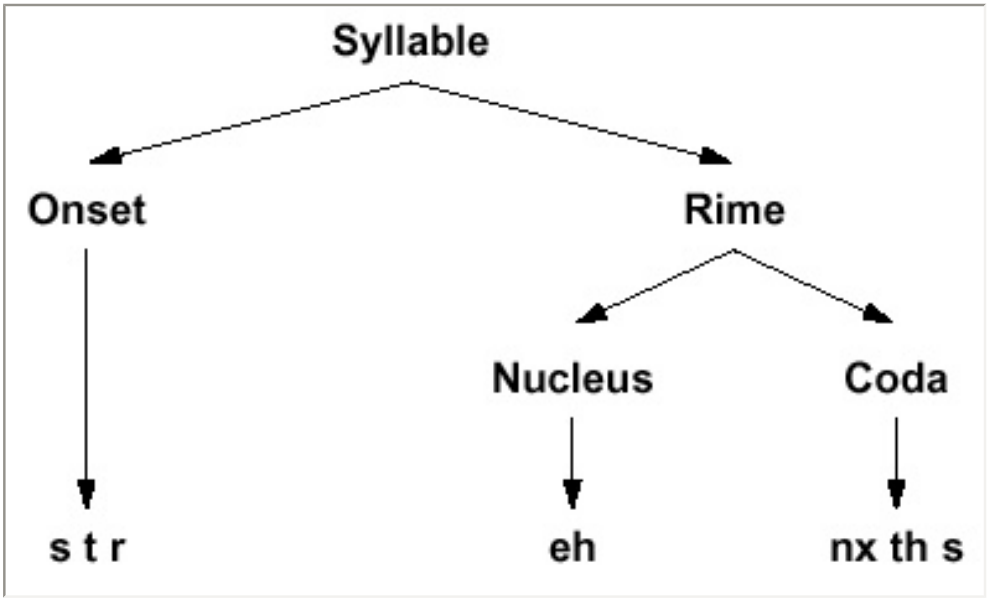
> X. Huang, A. Acero, and H.W. Hon, *Spoken Language Processing - A Guide to Theory, Algorithm, and System Development*, Prentice Hall, Upper Saddle River, New Jersey, USA, ISBN: 0-13-022616-5, 2001.

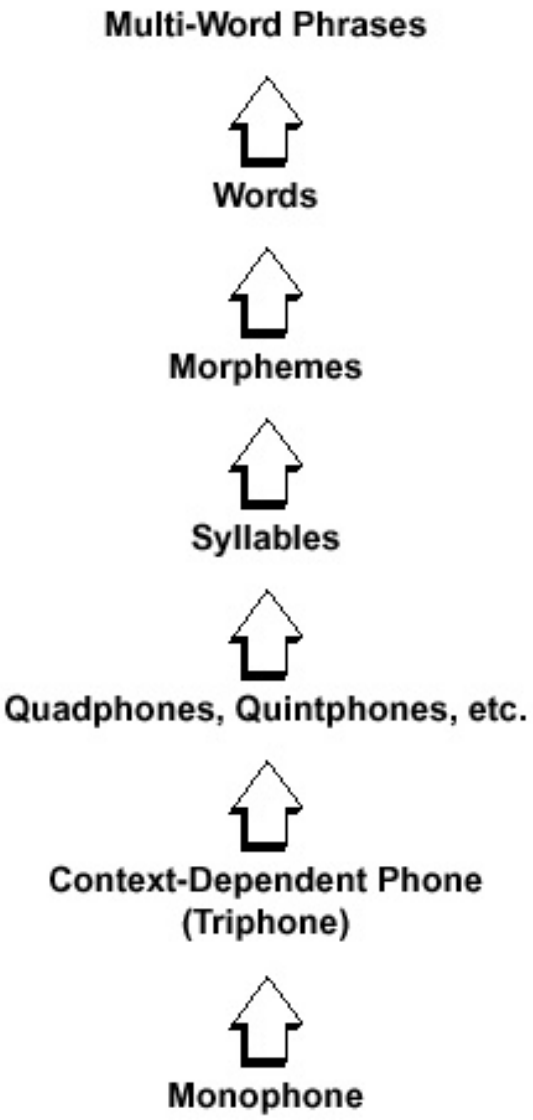
A good reference textbook on these topics is:

> D. Jurafsky and J.H. Martin, *SPEECH and LANGUAGE PROCESSING: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice-Hall, ISBN: 0-13-095069-6, 2000.

SYLLABLES: PRIMARY DOMAIN OF COARTICULATION?

- Acoustically distinct.
- There are over 10,000 syllables in English.
- There is no universal definition of a syllable.
- Can be defined from both a production and perception viewpoint.
- Centered around vowels in English.
- Consonants often span two syllables ("ambisyllabic" - "bottle").
- Three basic parts: onset (initial consonants), nucleus (vowel), and coda (consonants following the nucleus).





WORDS: OBSERVABLE UNITS OF A LANGUAGE?

- Loosely defined as a lexical unit - there is an agreed upon meaning in a given community.
- In many languages (e.g., Indo-European), easily observed in the orthographic (writing) system since it is separated by white space.
- In spoken language, however, there is a segmentation problem: words run together.
- **Syntax**: certain facts about word structure and combinatorial possibilities are evident to most native speakers.
- **Paradigmatic**: properties related to meaning.
- **Syntagmatic**: properties related to constraints imposed by word combinations (grammar).
- Word-level constraints are the most common form of "domain knowledge" in a speech recognition system.
- **N-gram** models are the most common way to implement word-level constraints.
- N-gram distributions are very interesting!

LEXICAL PART OF SPEECH:

- **Lexicon**: alphabetic arrangement of words and their definitions. A term often used to describe the list of allowable words for a speech recognition system.
- **Lexical Part of Speech**: A restricted inventory of word-type categories which capture generalizations of word forms and distributions ("dog" and "cat" are nouns and animals).
- **Part of Speech (POS)**: noun, verb, adjective, adverb, interjection, conjunction, determiner, preposition, and pronoun.
- **Proper Noun**: names such as "Velcro" or "Spandex". Pose a very challenging problem for speech recognition because of the lack of pronunciation rules (e.g., "Nyugen", "Sorbet").
- **Open POS Categories**:

Tag	Description	Function	Example
N	Noun	Named entity	*cat*
V	Verb	Event or condition	*forget*
Adj	Adjective	Descriptive	*yellow*
Adv	Adverb	Manner of action	*quickly*
Interj	Interjection	Reaction	*Oh*!

- **Closed POS Categories**: some level of universal agreement on the categories (e.g, conjunction, determiner, preposition).
- **Penn Treebank**: the LDC's Penn Treebank is one of the most ambitious projects to date in which large amounts of data have been categorized.
- **Wordnet**: Princeton's Wordnet is another very important and ambitious project to develop an on-line lexical reference system.

MORPHOLOGY: IMPORTANT IN SPECIALIZED SUB-LANGUAGES

- **Morpheme**: a distinctive collection of phonemes having no smaller meaningful parts (e.g, "pin" or "s" in "pins").
- Morphemes are often words, and in some languages (e.g., Latin), are an important sub-word unit. Some specific speech applications (e.g. medical dictation) are amenable to morpheme level acoustic units.
- **Inflectional Morphology**: variations in word form that reflect the contextual situation of a word, but do not change the fundamental meaning of the word (e.g. "cats" vs. "cat").
- **Derivational Morphology**: a given root word may serve as the source for new words (e.g., "racial" and "racist" share the morpheme "race", but have different meanings and part of speech possibilities). The baseform of a word is often called the **root**. Roots can be compounded and concatenated with derivational prefixes to form other words.

WORD CLASSES: A STATISTICAL APPROACH

- **Word Classes**: Assign words to similar classes based on their usage in real text (clustering). Can be derived automatically using statistical parsers.
- Typically more refined than POS tags (all words in a class will share the same POS tag). Based on semantics (meaning).
- Word classes are used extensively in language model probability smoothing.
- Examples:
    - {Monday, Tuesday, ..., weekends}
    - {great, big, vast, ..., gigantic}
    - {down, up, left, right, ..., sideways}

PHRASE SCHEMATA

- **Syntax**: Syntax is the study of the formation of sentences from words and the **rules** for formation of grammatical sentences.
- **Syntactic Constituents**: subdivisions of a sentence into phrase-like units that are common to many sentences. Syntactic constituents explain the word order of a language ("SOV" vs. "SVO" languages).
- **Phrase Schemata**: groups of words that have internal structure and unity (e.g., a "noun phrase" consists of a noun and its immediate modifiers).
- Example: NP -> (det) (modifier) **head-noun** (post-modifier)

NP	Det	Mod	Head Noun	Post-Mod
1	the		authority	of government
7	an	impure	one	
16	a	true	respect	for the individual

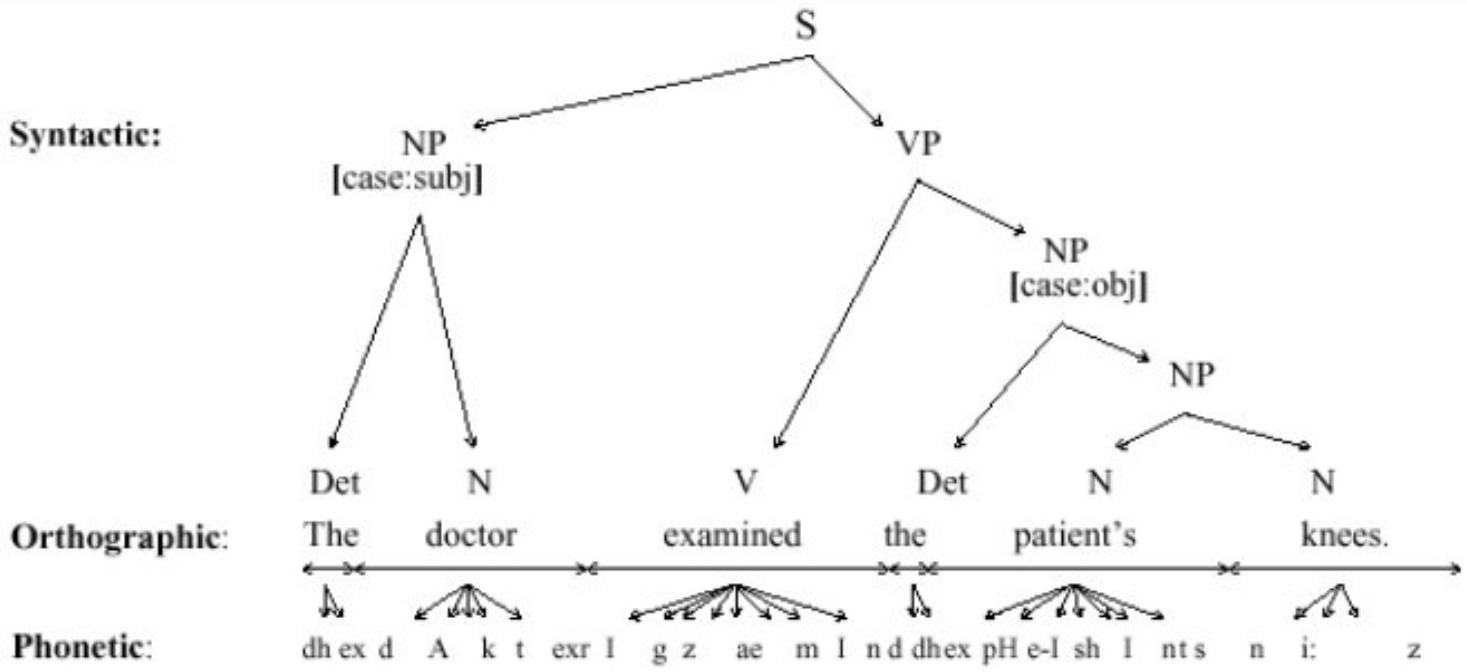
CLAUSES AND SENTENCES

- A **clause** is any phrase that has both a subject (NP) and a verb phrase (VP) that has a potentially independent interpretation.
- A **sentence** is a superset of a *clause* and can contain one or more clauses.
- Some typical types of sentences:

Type	**Example**
Declarative	I gave her a book.
Yes-No Question	Did you give her a book?
Wh-Question	What did you give her?
Alternative Question	Did you give her a book or a knife?
Tag Question	You gave it to her, didn't you?
Passive	She was given a book.
Cleft	It must have been a book that she got.
Exclamative	Hasn't this been a great birthday!
Imperative	Give me the book.

PARSE TREES

- **Parse Tree**: used to represent the structure of a sentence and the relationship between its constituents.
- Markup languages such as the standard generalized markup language (**SGML**) are often used to represent a parse tree in a textual form.
- Example:

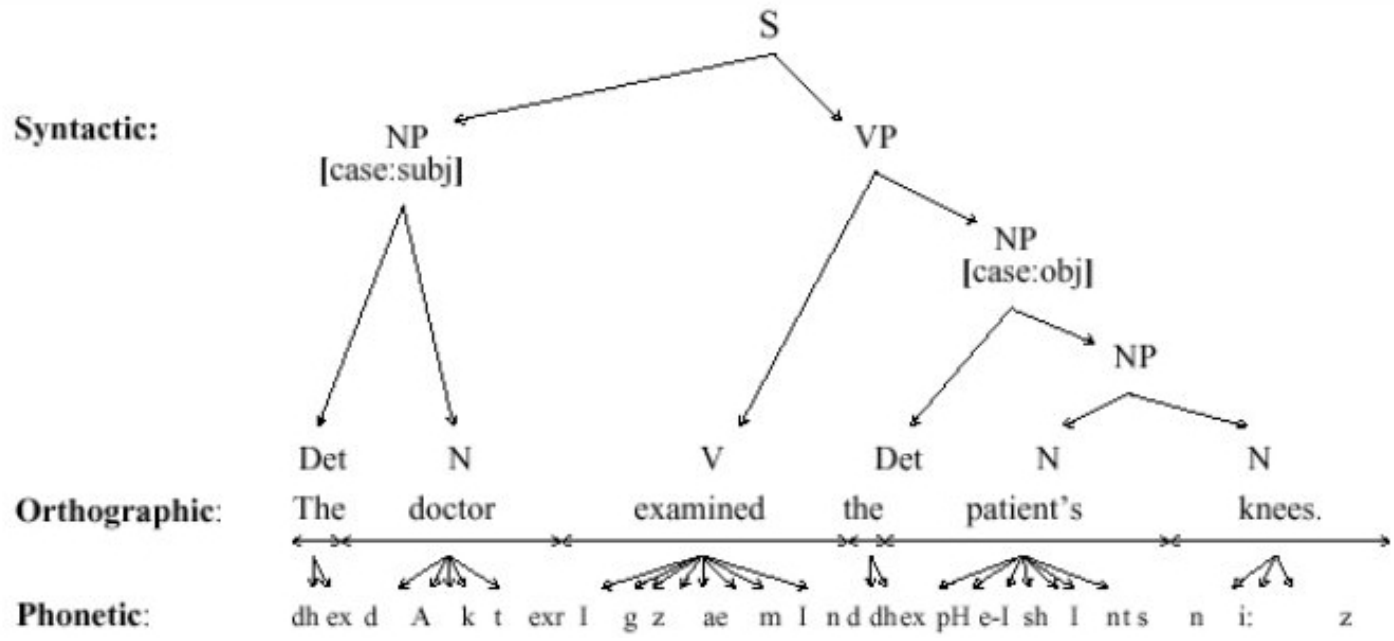


SEMANTIC ROLES

- **Grammatical roles** are often used to describe the direction of action (e.g., subject, object, indirect object).
- **Semantic roles**, also known as **case relations**, are used to make sense of the participants in an event (e.g., "who did what to whom").

Role	Description
Agent	cause or inhibitor of action
Patient/Theme	undergoer of the action
Instrument	how the action is accomplished
Goal	to whom the action is directed
Result	result or outcome of the action
Location	location or place of the action

- Example:
  - "The doctor examined the patient's knees"



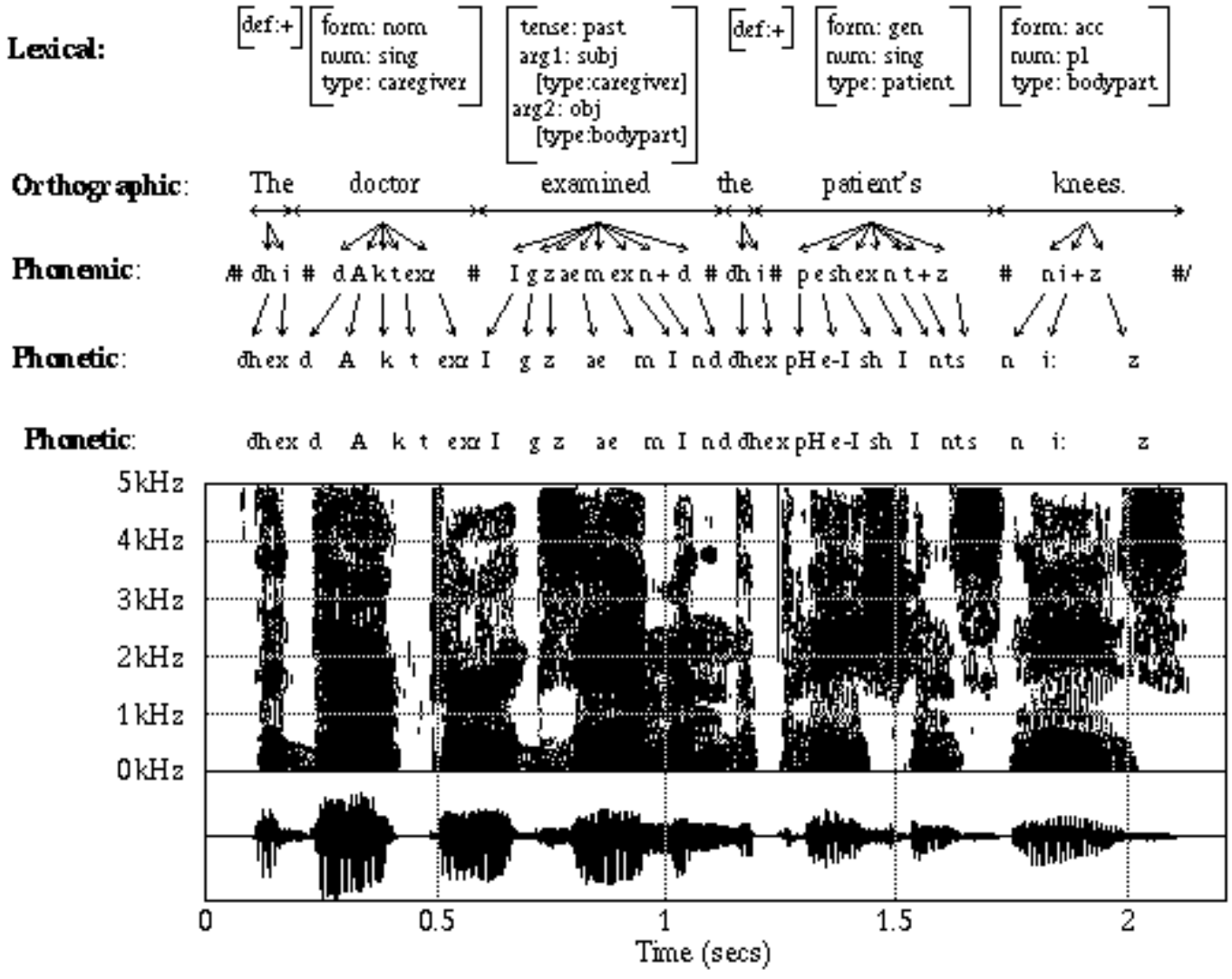
LEXICAL SEMANTICS

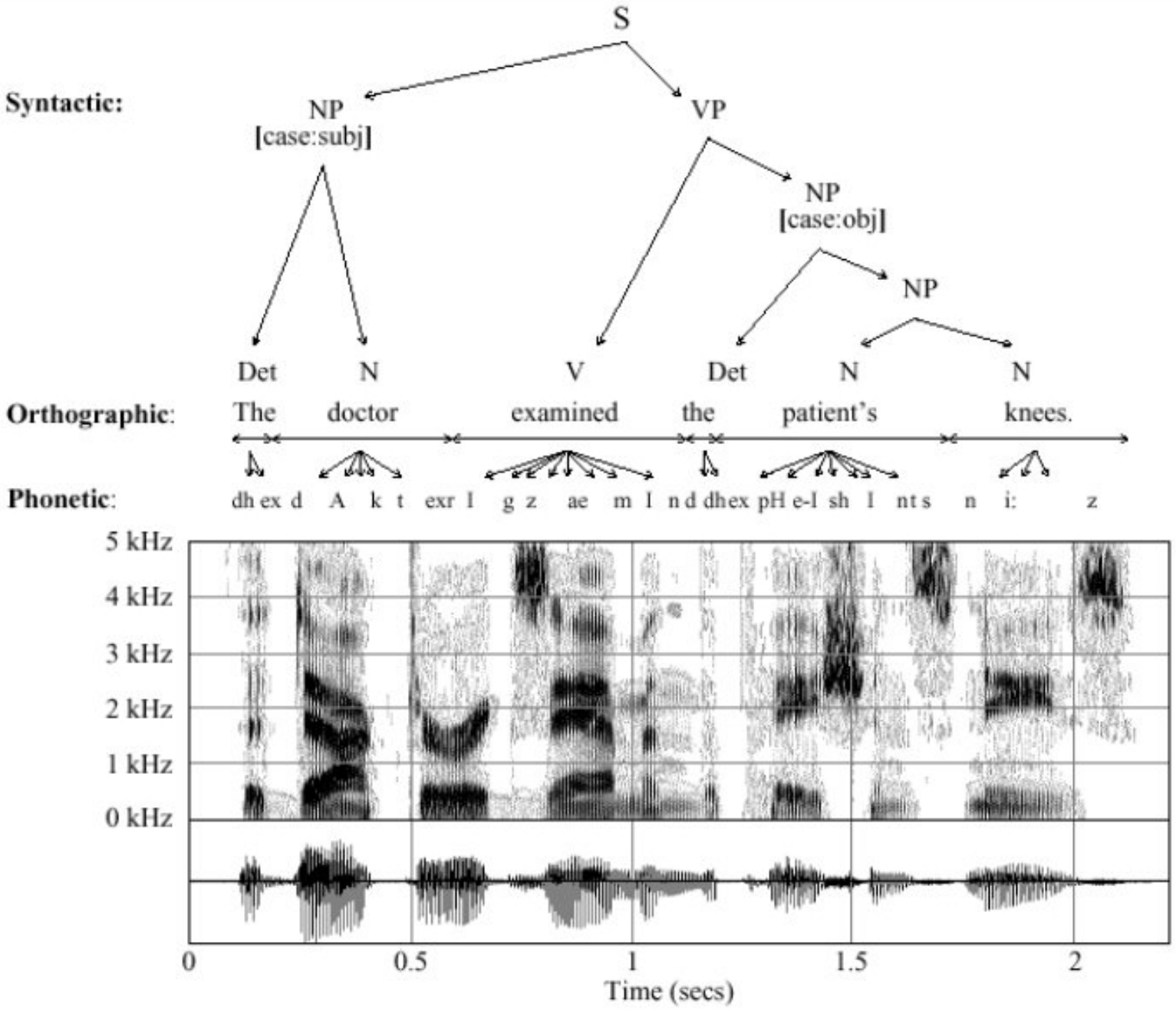
- **Lexical Semantics**: the semantic structure associated with a word, as represented in the lexicon.
- **Taxonomy**: orderly classification of words according to their presumed natural relationships.
- Examples:
    - **Is-A Taxonomy**: a *crow* is a bird.
    - **Has-a Taxonomy**: a *car* has a windshield.
    - **Action-Instrument**: a *knife* can cut.
- Words can appear in many relations and have multiple meanings and uses.
- There are no universally-accepted taxonomies:

Family	**Subtype**	**Example**
Contrasts	Contrary	old-young
	Contradictory	alive-dead
	Reverse	buy-sell
	Directional	front-back
	Incompatible	happy-morbid
	Asymmetric contrary	hot-cool
	Attribute similar	rake-fork
Case Relations	Agent-action	artist-paint
	Agent-instrument	farmer-tractor
	Agent-object	baker-bread
	Action-recipient	sit-chair
	Action-instrument	cut-knife

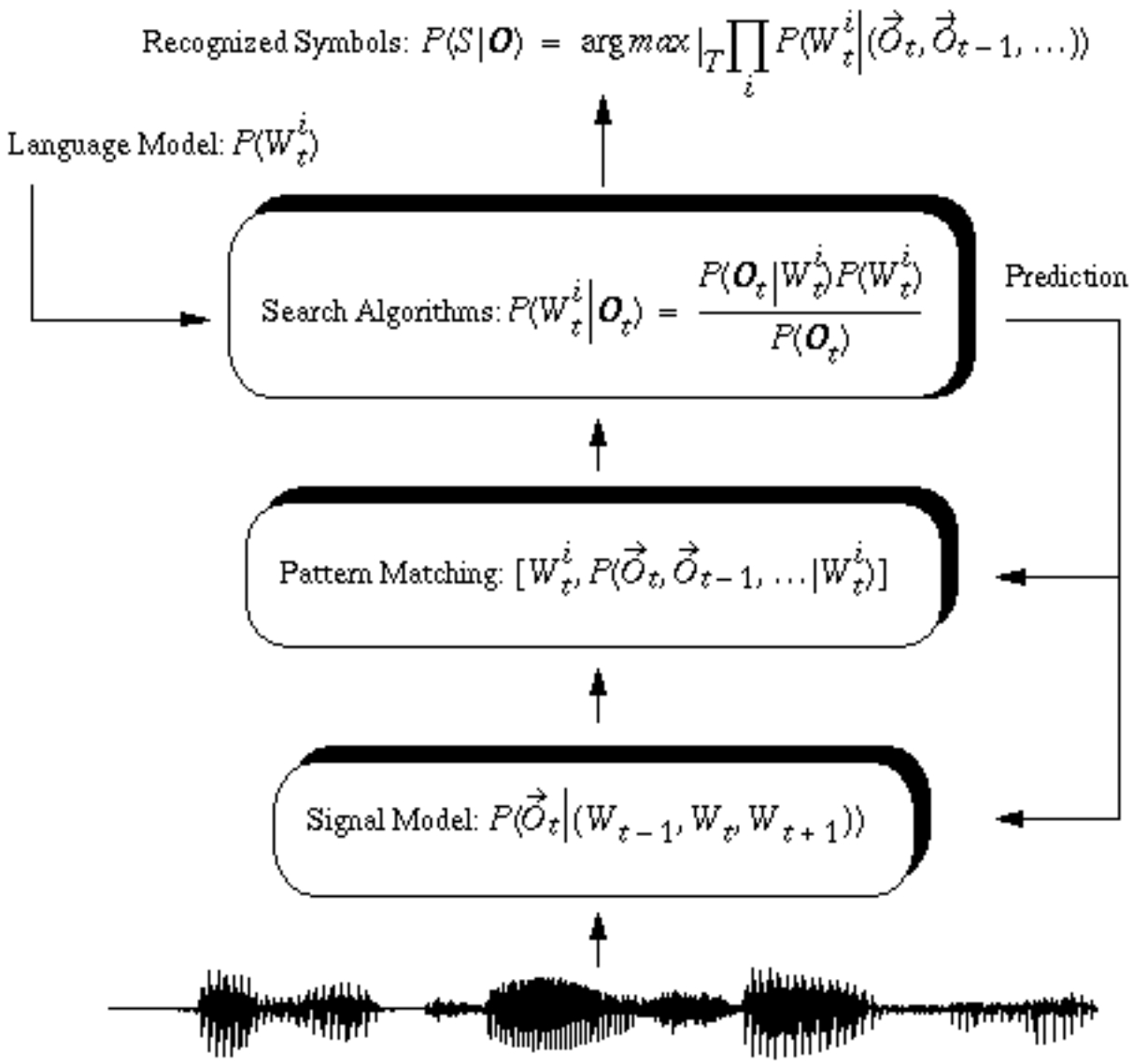
LOGICAL FORMS

- **Logical form**: a metalanguage in which we can concretely and succinctly express all linguistically possible meanings of an utterance.
- Typically used as a representation to which we can apply discourse and world knowledge to select the single-best (or N-best) alternatives.
- An attempt to bring formal logic to bear on the language understanding problem (**predicate logic**).
- Example:
  - If Romeo is happy, Juliet is happy: Happy(Romeo) -> Happy(Juliet)
  - "The doctor examined the patient's knees"




Syntactic:
S
NP
[case:subj]
VP
NP
[case:obj]
NP
Det
N
V
Det
N
N
Orthographic:
The
doctor
examined
the
patient's
knees.
Phonetic:
dh ex d A k t exr l g z ae m l n d dhex pH e-I sh l nt s n i: z
5 kHz
4 kHz
3 kHz
2 kHz
1 kHz
0 kHz
0
0.5
1
1.5
2
Time (secs)

WORD PREDICTION PLAYS A KEY ROLE



WordNet

a lexical database for the English language

cognitive science laboratory | princeton university | 221 nassau st. | princeton, nj 08542

[About WordNet](#)

[Use WordNet online](#)

[Download WordNet](#)

[Changes in version 1.7](#)

[Frequently asked questions](#)

[WordNet manuals](#)

[Glossary of terms](#)

[Current events](#)

[Publications](#)

[License & commercial use](#)

[Related projects](#)

Send comments or questions *pertaining to WordNet only* to wordnet@princeton.edu. We are not a dictionary or thesaurus service, and irrelevant mail will not be answered.

WordNet® is an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets.

WordNet was developed by the Cognitive Science Laboratory at Princeton University under the direction of Professor George A. Miller (Principal Investigator).

Over the years, many people have contributed to the success of WordNet. At the present time, the following individuals at Princeton work on the continuing development of WordNet and applying it to research:

- Dr. Christiane Fellbaum
- Professor George A. Miller
- Randee Tengi
- Pamela Wakefield

Dr. Fellbaum's work was supported in part by grant No. 9805732 from the National Science Foundation.

wordnetNEWS

- **WordNet 1.7.1** is now up on our web interface.
- **Version 1.7.1** will be released shortly for both Unix and Windows platforms. Currently available versions are 1.7 for Unix and 1.6 for Windows.
- The front page of the January 22, 2002 **Star Ledger** featured an article on WordNet.
- "WordNet: An Electronic Lexical Database" is available from MIT Press.
- **D-*lib* magazine** has a review.

SWITCHBOARD

Each release of transcription data for this project will be a superset of the previous release (in other words, you need only download the latest release). All transcriptions and segmentations developed in this project are based on the audio data from the following SWITCHBOARD release:

Switchboard-1 Telephone Speech Corpus: Release 2 August, 1997

For information regarding SWITCHBOARD, please consult the LDC web site. For more details about this project, see the project overview. A mailing list is also available to discuss progress and key issues of the project.

Transcriptions and Word Alignments:

- (10/25/01) Download manually corrected word alignments: This release contains 812 manually corrected word alignments for Switchboard conversations. Included in this release are the final transcriptions for the entire database, an updated lexicon, and automatic word alignments for the remaining conversations.
- (03/21/01) Download the ICSI Transcriptions: This release differs from the 03/15/01/release only by one utterance. Two utterances were merged to form one utterance, and the phone transcriptions were corrected.

  The original ICSI data is available from the WS97 ftp site at the Center for Language and Speech Processing (CSLP) at Johns Hopkins University. It can also be downloaded from the ISIP mirror of this data.
- (12/27/01) Download the Penn Treebank Transcriptions: This release contains an alignment of the ISIP word transcriptions to the Penn Treebank word transcriptions for all 1126 SWB conversations that are included in the Treebank. For the words which are in agreement between the two transcriptions, time marks are given. For words that don't agree, we estimate the times for the Treebank transcriptions using the ISIP transcriptions. The transcriptions also include all instances of silence, laughter and noise.

Documentation:

- Transcription FAQ: provide on-line feedback about key issues.
- Conventions: download a document describing our transcription conventions.
- SWB Statistics: download a statistical analysis of the SWB corpus.
- SWB Models: A copy of the SWB models file that we use.
- Education: an on-line educational resource for learning about the SWB corpus.
- Reports: Quarterly reports summarizing the progress made on the project.

Tools:

- Software: our transcription and segmentation tool.
- Spiker: this is a simple C program to correct Switchboard files that have been corrupted by flipping of their bits.
- HTK Tools: HTK-based tools for running SWB experiments.
- ISIP Recognizer: download a public domain speech recognition system under development in ISIP.

General Information:

- Overview: an overview of the SWITCHBOARD (SWB) resegmentation project.
- Personnel: the people that make SWB resegmentation happen.
- Job Opportunities: do you want to be a SWITCHBOARD validator?
- Timesheets: a list of due dates for timesheets.
- CLSP Workshops: summer workshops on conversational speech recognition.

Please direct questions or comments to help@isip.msstate.edu

Look here to learn more about:
Obtaining data from LDC,
Using it in your environment,
Creating your own corpora and
Providing them to others via LDC.
Or, browse our Catalog directly.

Employment at the LDC

New Corpora
Translanguage English Database (TED) Speech and **Transcripts** 188 speech files and 39 transcripts of non-native English speakers presenting academic papers for approximately 15 minutes each
Multiple-Translation Chinese Corpus 11 sets of human translations and 6 sets of automatic translations of Mandarin Chinese into English
SPINE2 Part 3 Audio and **Transcripts** ~7 hours audio of "Battleship" type game turns in English and corresponding transcript corpus.
Switchboard Cellular Part 1 Audio 1309 telephone conversations (~65 hours audio) in English, of which 250 conversations (~12 hours) represent the **Transcribed Audio**
Chinese Treebank 2.0 ~100,000 words in 325 news stories with syntactic bracketing.

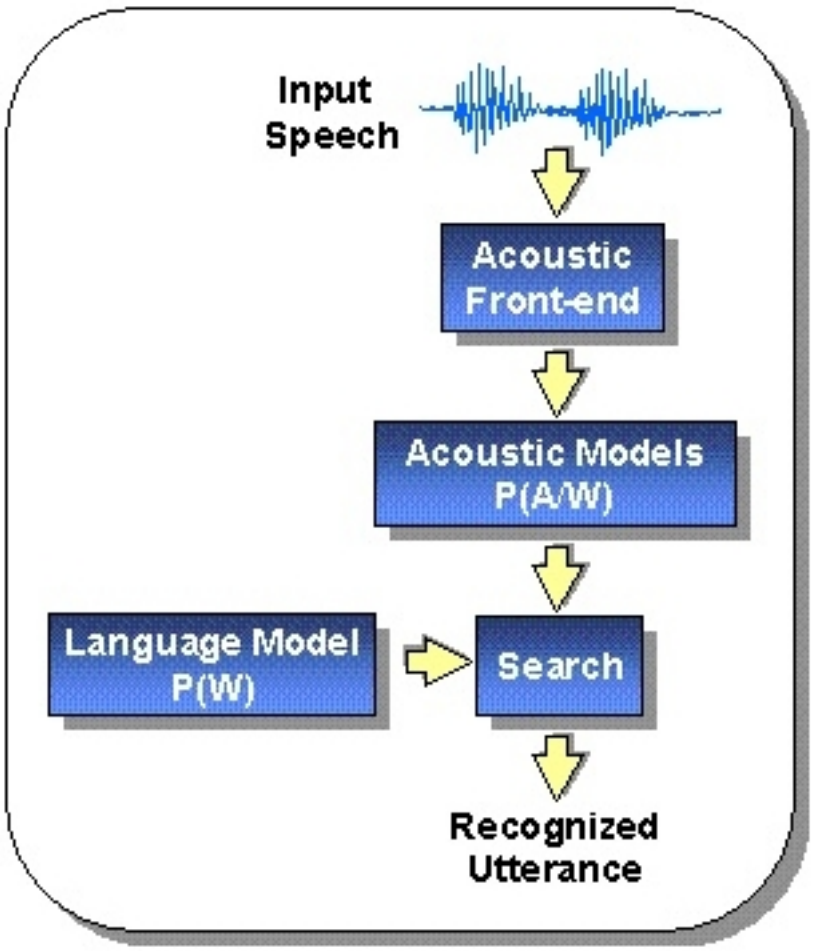
What's New! What's Free!
Updated FAQs for Members, NonMembers and for LDC Online.
Annotation Graphs - software infrastructure for linguistic annotation.
SPH Convert - tools for converting audio formats.
IRCS Workshop on Linguistic Databases
OLAC - Open Language Archives Community.
Transcriber - tool for segmenting, labeling and transcribing speech.

Linguistic Data Consortium

The Linguistic Data Consortium supports language-related education, research and technology development by creating and sharing linguistic resources: data, tools and standards.

LDC is supported in part by grant IRI-9528587 from the Information and Intelligent Systems division of the National Science Foundation. LDC's corpus creation efforts are powered in part by Academic Equipment Grant 7826-990237-US from Sun Microsystems.

HUMAN LANGUAGE TECHNOLOGY:
SPEECH RECOGNITION IS MULTIDISCIPLINARY



- Acoustic Front-End: Signal Processing
- Acoustic Models: Pattern Recognition, Linguistics
- Language Model: Natural Language Processing
- Search: Computational Linguistics, Cognitive Science