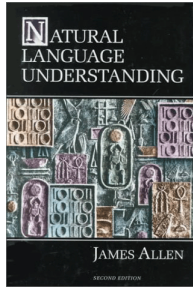


# Clinical NLP and the data dilemma

Philip Resnik, Ph.D.  
University of Maryland  
CodeRyte, Inc.

Natural Language Processing: State of the Art, Future Directions and Applications for  
Enhancing Clinical Decision-Making, National Institutes of Health, April 23-24, 2012

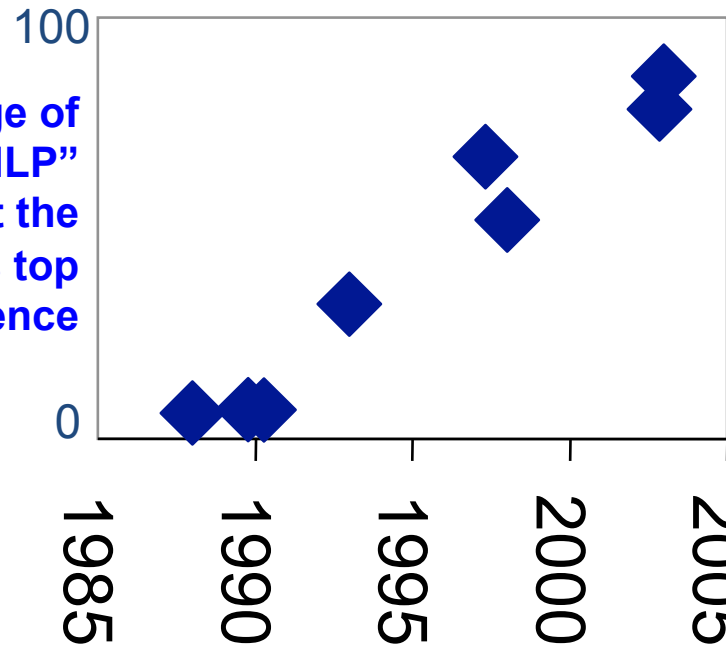


1970 — 1983  
“natural language understanding”

1970

1983

Percentage of  
“statistical NLP”  
papers at the  
field’s top  
conference



1985

1990

1995

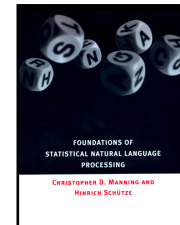
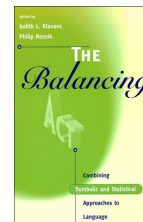
2000

2005

1983 — 1993  
“the return of empiricism...  
probabilistic models  
throughout speech and  
language processing”

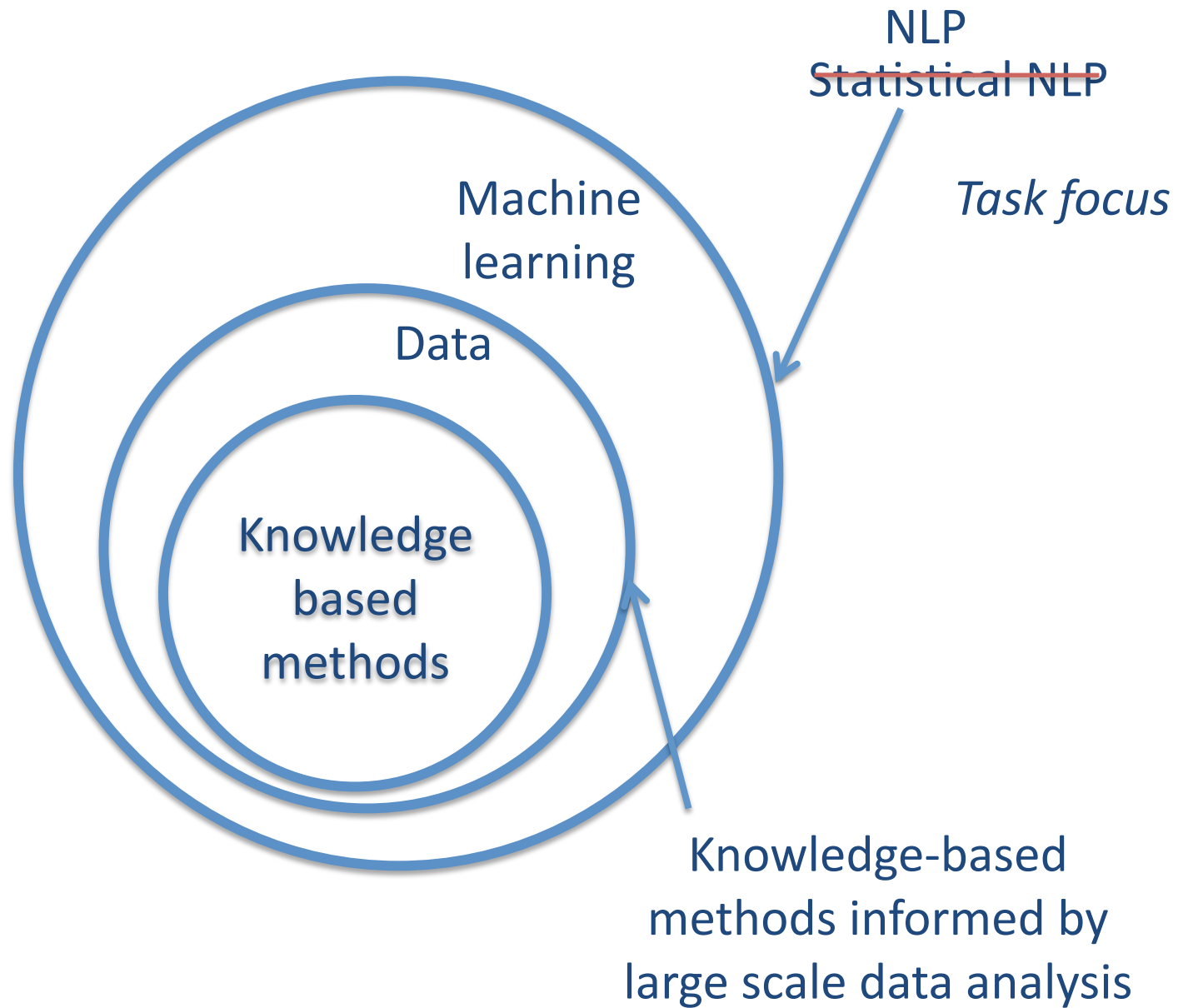
1994 — 1999  
“the field comes  
together”

2000 — 2008  
“the rise of  
machine  
learning”

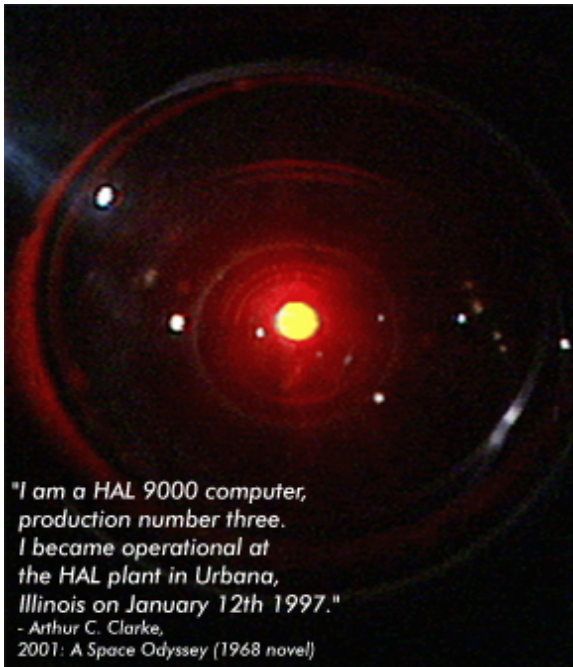


Jurafsky and Martin (2009), *Speech and Natural Language Processing*

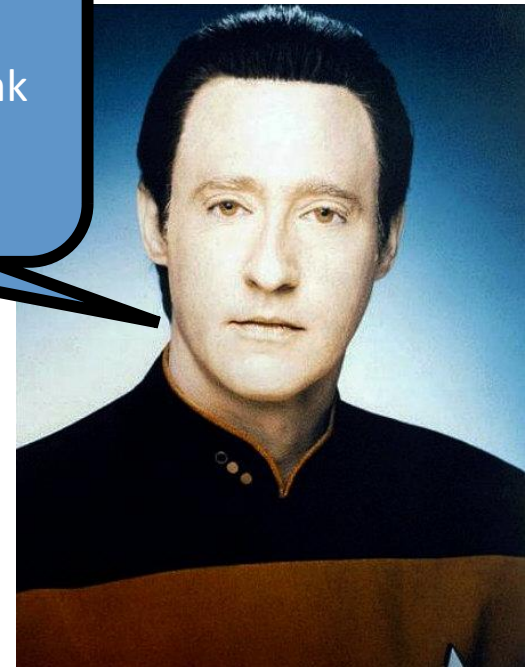
Graph adapted from Church, K. (2003) “Speech and Language Processing: Where have we been and where are we going,” Eurospeech, Geneva, Switzerland, Adding data from figures in Cardie and Mooney (1999).



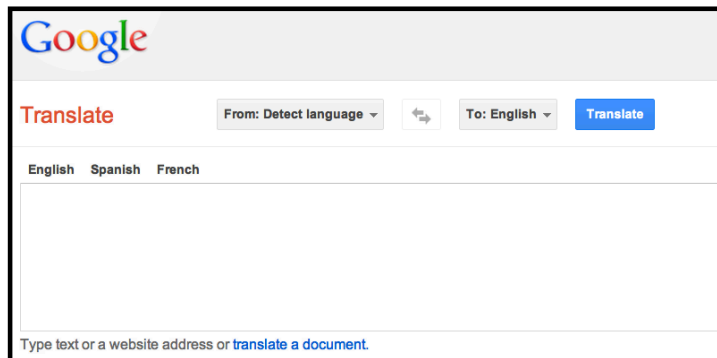
# The dream



Sure, Hal, Gingrich *does* have cool ideas about NASA. What do you think about his position on corporate tax rates?



In today's NLP,  
the question is not *whether* you  
integrate knowledge-based methods  
with statistical methods,  
but *how*.



- Features
  - Define what the analysis “pays attention to”
- Model structure
  - Defines possible relationships among variables
- Informative priors
  - “Softly” incorporate assumptions that can be overridden by enough evidence

- Features

- Define what the analysis “pays attention to”

catheter	catheter	N	3sg	CATHETER
catheters	catheter	N	3pl	CATHETER
catheterisation	catheterization	N	3sg	CATHETER
catheterisations	catheterization	N	3pl	CATHETER
catheterization	catheterization	N	3sg	CATHETER
catheterizations	catheterization	N	3pl	CATHETER
catheterise	catheterize	V	INF	CATHETER
catheterize	catheterize	V	INF	CATHETER

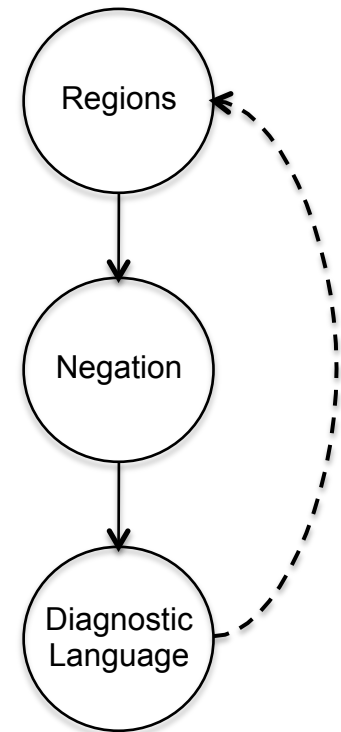
- Model structure

- Defines possible relationships among variables

**HPI:** Atrial fibrillation. This patient is a 56-year-old white gentleman who has had a history of atrial fib on and off since he had his bypass surgery. Patient was originally diagnosed with coronary artery disease as well as mitral valve problems approximately 3 years ago. Dr. Tirona used to take care of him at that time. He had a bypass surgery as well as mitral valve repair done at that time. Postop he had an episode of A-fib which then resolved spontaneously. He remembers somebody talking to him about cardioversion, but then the A-fib resolved spontaneously. So he was started on Coumadin. He would get some occasional episodes, but usually they are very brief, so he never bothered about them. Of late, over the last few months, he has been getting more frequent episodes and duration of these episodes is also prolonged for a few hours. So he saw Dr. Hagan who has referred him here for further evaluation and treatment. The patient states when he does get the A-fib, he feels very weak, tired, and short of breath. He denies any chest pain. Otherwise he is usually very active physically, he works fulltime as an electrician, and has not had any problems as far as doing his day-to-day work.

**MEDICAL HISTORY:** 1. Coronary artery disease as mentioned above. 2. Hypertension. 3. Hypercholesterolemia.

**IMPRESSION:** Paroxysmal atrial fibrillation in a patient with prior mitral valve disease, currently having more frequent breakthroughs symptoms.





- Informative priors
  - “Softly” incorporate assumptions that can be overridden by enough evidence

**COLD**



Chronic  
obstructive  
lung disease



Common  
cold

- Informative priors
  - “Softly” incorporate assumptions that can be overridden by enough evidence

**COLD**

emphysema  
obstruction  
airway  
pulmonary



Chronic  
obstructive  
lung disease



Common  
cold

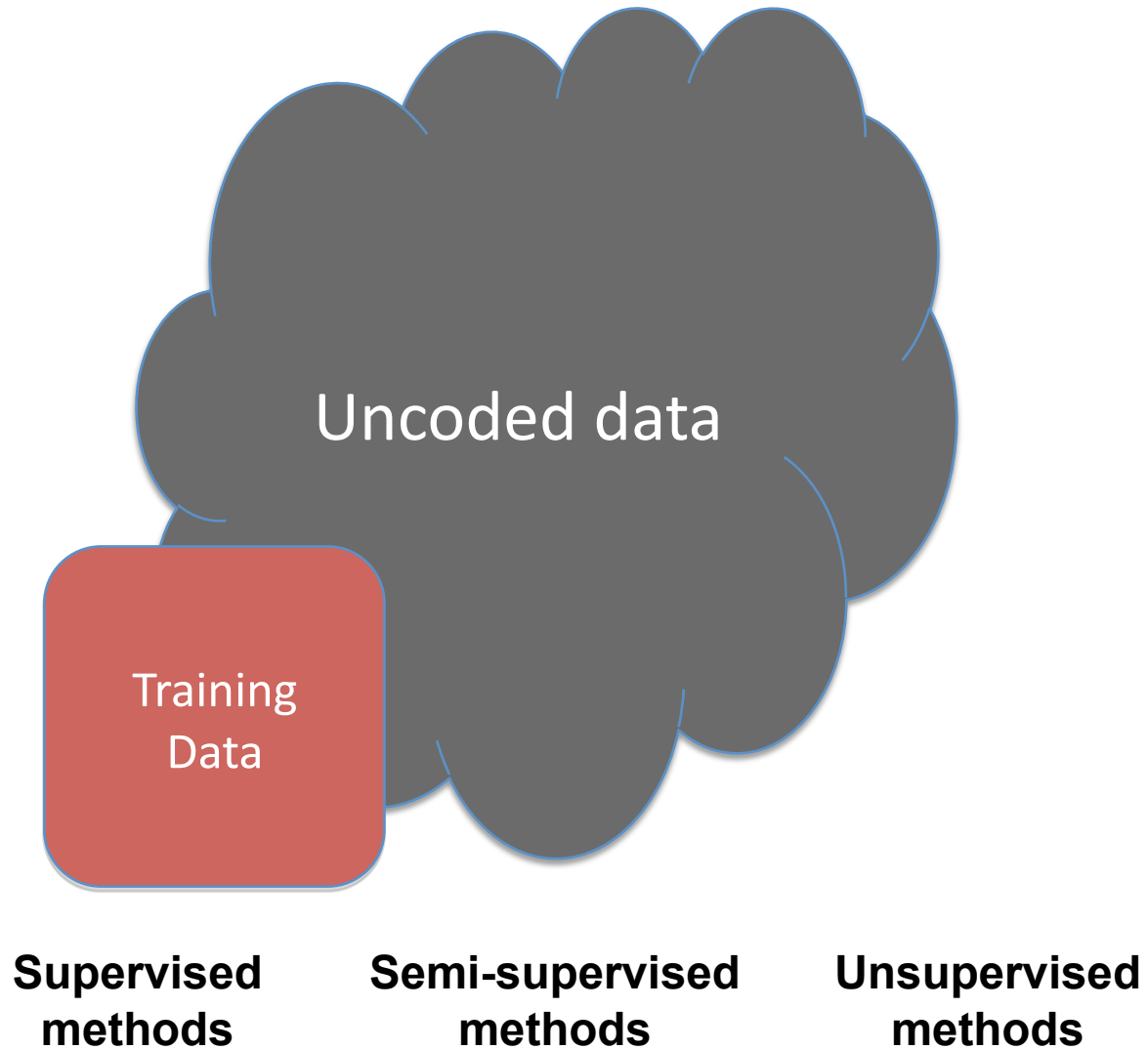
So, state-of-the-art NLP depends crucially on learning from relevant data.

That's a problem.

### SUMMARY OF THE HIPAA PRIVACY RULE

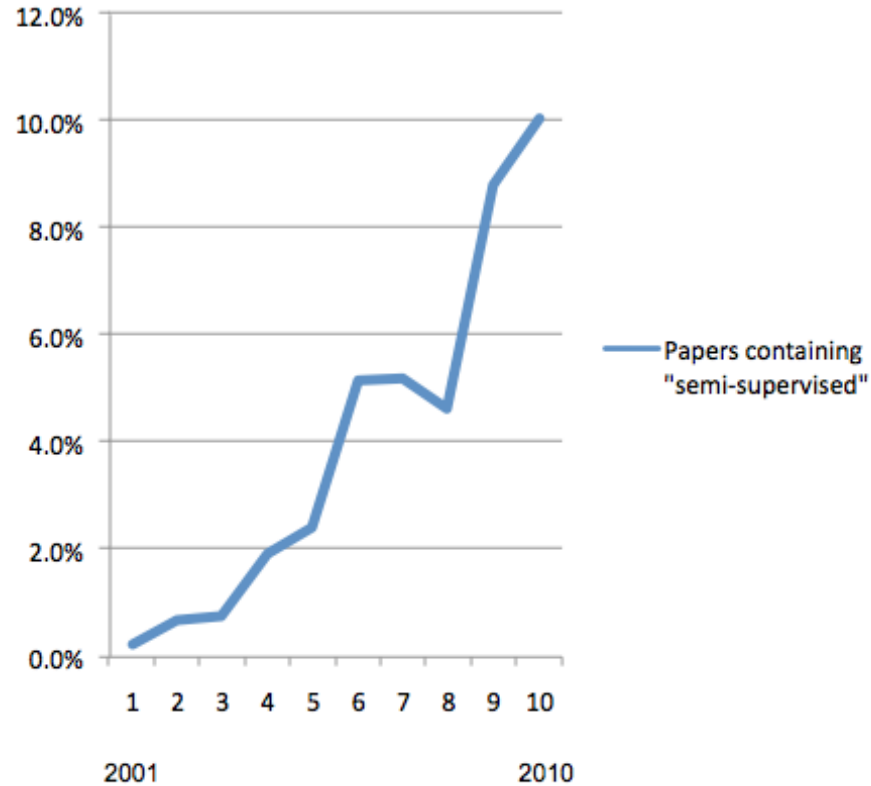
<b>Introduction</b>	<p>The <i>Standards for Privacy of Individually Identifiable Health Information</i> (“Privacy Rule”) establishes, for the first time, a set of national standards for the protection of certain health information. The U.S. Department of Health and Human Services (“HHS”) issued the Privacy Rule to implement the requirement of the Health Insurance Portability and Accountability Act of 1996 (“HIPAA”).<sup>1</sup> The Privacy Rule standards address the use and disclosure of individuals’ health information—called “protected health information” by organizations subject to the Privacy Rule — called “covered entities,” as well as standards for individuals’ privacy rights to understand and control how their health information is used. Within HHS, the Office for Civil Rights (“OCR”) has responsibility for implementation and enforcing the Privacy Rule.</p>
---------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

# One solution



# One solution

## ACL Anthology papers Containing "semi-supervised"



# Another solution

## Providers



**CHIEF COMPLAINT:**  
Shortness of breath.  
**HPI:** This is a 68-year-old female who presents to the emergency department with shortness of breath going for several days

...

Coding Tools & Services



12002  
873.0

\$\$

## Payers



**Name:** Head Injury      **DOB:** 1964-07-12  
**MRN:** 457564      **Age:** 46  
**DOS:** 2011-04-09 00:37:53      **SigningMD:** Dr. Emergency  
**OrderHistory:**  
**OrderExam:**

**HPI:**

The patient presents with headache and pt here with head injury-- jumped and hit head on beam. + LOC. no neck pain. no numbness, visual changes. no vomiting. bleeding controlled at this time. no other injuries. . The course/duration of symptoms is constant. Location: occipital. Radiating pain: none. The character of symptoms is throbbing. Associated symptoms: none.

**MEDICAL HISTORY:**

Medical history Negative. Surgical history: Negative.

**SOCIAL HISTORY:**

Social history: Alcohol use: Denies, Tobacco use: Denies, Drug use: Denies.

**ROS:**

Constitutional symptoms: Negative except as documented in HPI. Respiratory symptoms: Negative except as documented in HPI. Neurologic symptoms: Negative except as documented in HPI . Additional review of systems information: All other systems reviewed and otherwise negative.

**EXAM:**

Vital Signs. Heart Rate 73 bpm Respiratory Rate 14 breaths/min SBP NIBP 101 mmHg DBP NIBP 61 mmHg SpO2 99 % General: No acute distress. Head: 5 cm laceration over top of head to sq. does not extend to galea. Neck: Supple, trachea midline, no tenderness. Neurological: Alert and oriented to person, place, time, and situation.

CPT Code Set=2011b, ICD Code Set=2011b

RECALC SKIP EXTEND QUIT APPROVE

L1 <input checked="" type="checkbox"/> 12002 x 1 REPAIR SUPERFICIAL WOUND(S)				
1	<input checked="" type="checkbox"/>	873.0	Open wound of scalp, without mention of complication	S01.00xA Unspecified open wound of scalp, initial encounter
2	<input checked="" type="checkbox"/>			
3	<input checked="" type="checkbox"/>			
4	<input checked="" type="checkbox"/>			
5	<input checked="" type="checkbox"/>			

L2 <input type="checkbox"/> 99283-25 x 1 EMERGENCY DEPT VISIT <span style="float:right">E/M</span>				
1	<input checked="" type="checkbox"/>	873.0	Open wound of scalp, without mention of complication	S01.00xA Unspecified open wound of scalp, initial encounter
2	<input checked="" type="checkbox"/>			
3	<input checked="" type="checkbox"/>			
4	<input checked="" type="checkbox"/>			
5	<input checked="" type="checkbox"/>			

L3 <input type="checkbox"/> x 1				
1	<input checked="" type="checkbox"/>			
2	<input checked="" type="checkbox"/>			
3	<input checked="" type="checkbox"/>			
4	<input checked="" type="checkbox"/>			
5	<input checked="" type="checkbox"/>			

RECALC SKIP EXTEND QUIT APPROVE

OR ROUTE TO: Expert: CR-ED Route

COMMENTS: CLOSE ▲

## ORTHOD-132

Review: Office

CPT Encoder

ICD Encoder

PHI

NAME: [PHI Redacted] D.O.B.: 61y.o. F  
MRN: [PHI Redacted] D.O.S.: 2006-01-10 00:00:00  
Ins:

### CHIEF COMPLAINT:

Right knee pain.

**HISTORY OF PRESENT ILLNESS:** The patient is a 61-year-old white female who works part-time at Curves who said she developed sharp severe pain in her right knee beginning in November 2005. She did not have any injury that she knows of but now she is having increasing pain with walking, kneeling down and going up and down stairs. She had sharp severe pain in the medial compartment of her knee. She feels better when she is not moving at all.

### PAST MEDICAL HISTORY:

Significant for hiatal hernia and a DVT in left upper extremity.

### PAST SURGICAL HISTORY:

Significant for bilateral thoracic outlet surgery and left rib resection as well as a left knee arthroscopy.

### CURRENT MEDICATIONS:

Include Crestor.

### ALLERGIES:

Codeine, Salactic, iodine, sulfa drugs.

### FAMILY HISTORY:

Significant for heart disease, stroke, and osteoporosis.

## Overall Level Of Service Analysis

HISTORY = C HPI = Extended, ROS = Complete, FMS =  
EXAM = COM EXAM95 = COM, EXAM97 = DET  
COMPLEXITY = M MAHOPT = , RISK = Low, DATA =  
FINAL E/M = 99204 E / M Level = 4

Visit Status : Office/OP New

### History of Present Illness:

Extended(4+)

CLOSE ▲

<input checked="" type="checkbox"/> Location	right knee, medial compartment of her knee	<input checked="" type="checkbox"/> Duration	November 2005
<input checked="" type="checkbox"/> Mod. Factor	feels better when she is not moving at all	<input checked="" type="checkbox"/> Quality	sharp
<input checked="" type="checkbox"/> Timing	with walking, kneeling down and going up and down stairs	<input checked="" type="checkbox"/> Signs & Symps	increasing pain
<input checked="" type="checkbox"/> Severity	severe	<input type="checkbox"/> Context	
<input type="checkbox"/> Status 3 conds			

### Review of Systems:

Complete(10+)

CLOSE ▲

<input type="checkbox"/> Constitutional		<input checked="" type="checkbox"/> Respiratory	emphysema
<input checked="" type="checkbox"/> Genitourinary	bladder infection, incontinence	<input checked="" type="checkbox"/> Eyes	glaucoma
<input checked="" type="checkbox"/> Cardiovascular	heart disease, high blood pressure	<input checked="" type="checkbox"/> Thyroid/Endo	thyroid disease, diabetes
<input checked="" type="checkbox"/> Dermatologic	dermatitis	<input checked="" type="checkbox"/> Musculoskeletal	rheumatoid arthritis
<input checked="" type="checkbox"/> Psychiatric	anxiety, depression	<input checked="" type="checkbox"/> E/M/T	sinusitis, seasonal allergies
<input checked="" type="checkbox"/> Neurological	seizures	<input checked="" type="checkbox"/> Hemo/Lymph	hepatitis, cancer



IR D.O.B.: [PHI REDACTED] Age: 37  
 NAME: [PHI Redacted] D.O.S.: 2007-10-23 12:06:00 Gender: M  
 SSN: [PHI Redacted] Attending: null  
 Referred By: Signed By: null

[1: 203444515] Angiography Carotid Cervical Unilateral [2: 203445682] Angiography Vertebral Cervical Intracranial Unilateral [3: 203445689] Angiography Internal Carotid or Cerebral Unilateral [4: 203445699] Angiography Carotid External Unilateral

Neurointerventional procedure

Cerebral angiogram, 4/24/2006

Indication: Spontaneous subdural hematoma

Physicians: 1. Dr. John Scott.

Vessels studied: 1. Left internal carotid artery injection, PA, lateral, RAO, LAO projections, intracranial. 2. Left external carotid artery injection, PA, lateral projections, intracranial. 3. Left vertebral artery injection, PA, lateral projections, intracranial. 4. Right common carotid artery injection, PA, lateral projections, intracranial.

Discussion: Angiographic examination of the above vessels was performed after consent was obtained. A 5 French catheter was placed in the right common femoral artery, using the Seldinger technique, after Betadine prep and local anesthesia with lidocaine. After the procedure, the catheter was removed and hemostasis achieved at the puncture site.

Findings: Left internal carotid artery: There is subdural mass-effect over the left cerebral convexity manifest by displacement of pial vessels and cortical veins away from the inner table of the skull. There is rightward shift of the anterior cerebral artery complex. There are no arterial or venous vascular abnormalities.

Left external carotid artery: There are no dural vascular abnormalities.

Left vertebral artery: There is medial displacement of the left posterior cerebral artery complex. No vascular abnormalities are identified.

Right common carotid artery: There are no cerebral or dural vascular abnormalities.

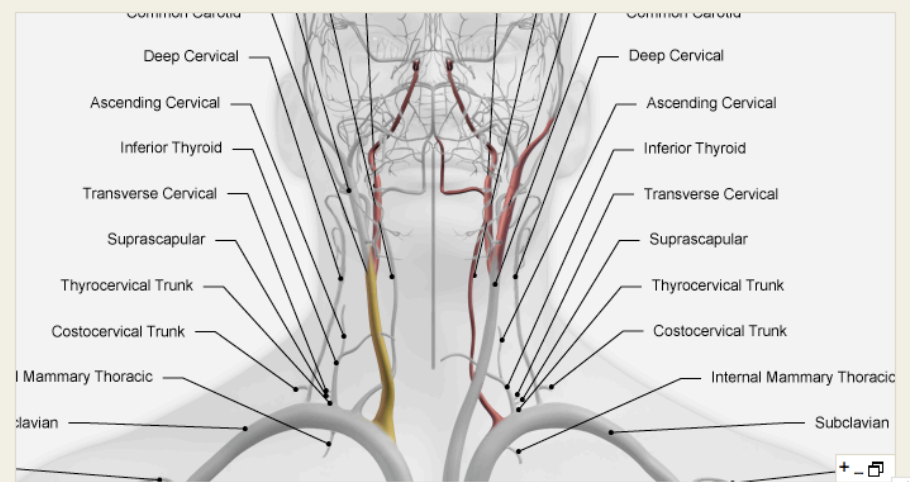
Impression: Left convexity mass-effect consistent with known subdural hematoma. Negative cerebral angiogram.

CPT CodeSet=2007d, ICD CodeSet=2007a

RECALC		SKIP		APPROVE	
ICD Code	Description				
I1 348.4	Compression of brain				
I2 432.1	Subdural hemorrhage				
+ [ ]	Add ICD Code				

APPROACH: FEMORAL

	Side	Procedure / Vessel	Type	Filter	CPT Codes	ICD Codes
L1	Left	Internal Carotid	CATH	None	36216	<input checked="" type="checkbox"/> 348.4 <input checked="" type="checkbox"/> 432.1
L2	Left	Vertebral	CATH	None	36216-59	<input checked="" type="checkbox"/> 348.4 <input checked="" type="checkbox"/> 432.1
L3	Right	Common Carotid	CATH	None	36216-59	<input checked="" type="checkbox"/> 348.4 <input checked="" type="checkbox"/> 432.1
L4	Left	External Carotid	CATH	Family	36218	<input checked="" type="checkbox"/> 348.4 <input checked="" type="checkbox"/> 432.1
L5	Bilat	Internal Carotid	S+I	None	75671	<input checked="" type="checkbox"/> 348.4 <input checked="" type="checkbox"/> 432.1



A bigger challenge for healthcare:

With the widespread adoption of EHRs, what happens to natural clinical language?

# Providers



**CHIEF COMPLAINT:**  
Shortness of breath.  
**HPI:** This is a 68-year-old female who presents to the emergency department with shortness of breath going for several days

...

Coding Tools & Services



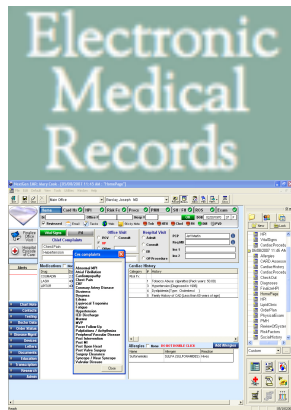
12002  
873.0

\$\$

# Payers



# Providers



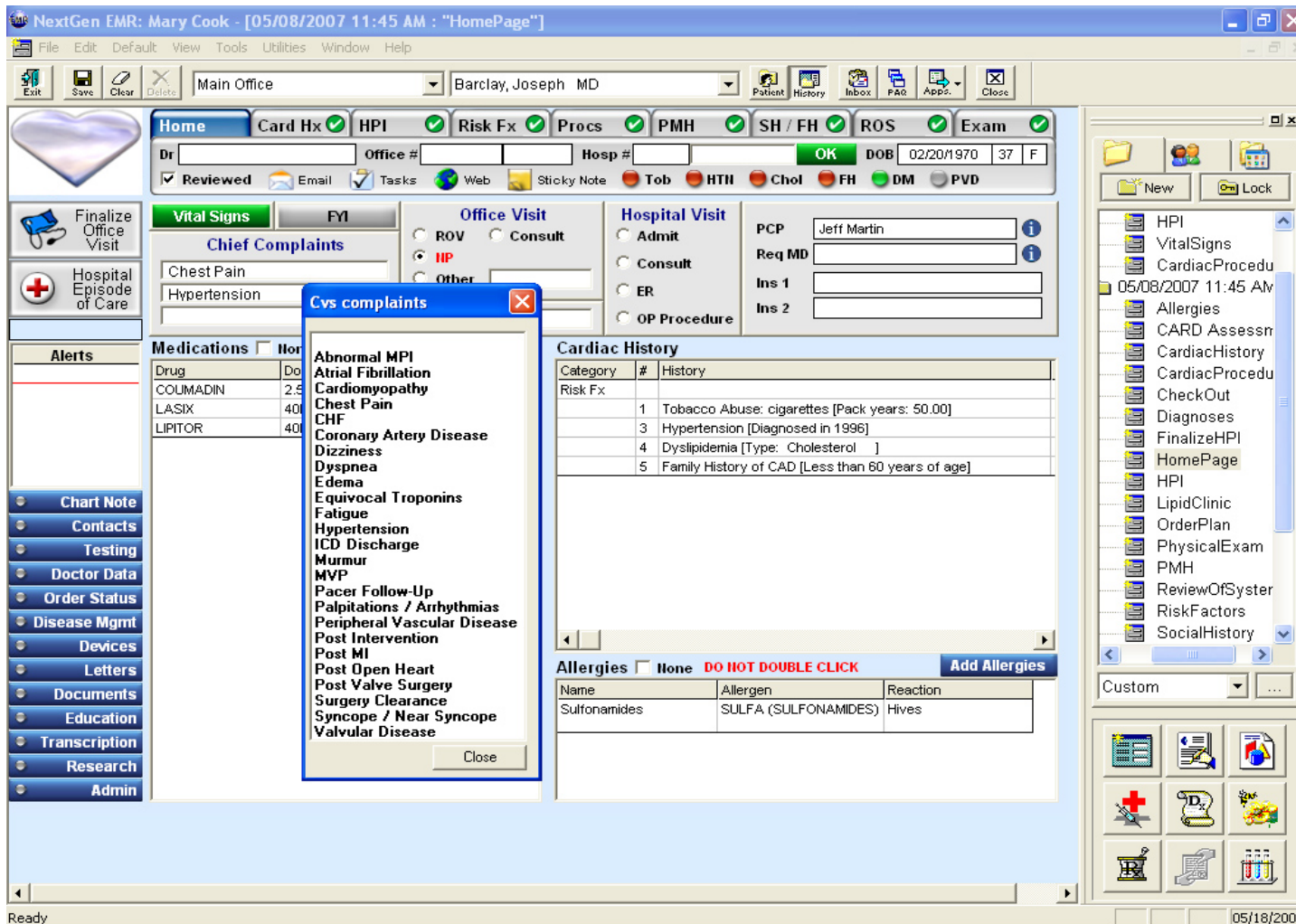
12002  
873.0



\$\$

# Payers





Source: <http://www.nextgen.com/images/screenshots/card01.jpg>

“This system is designed for physicians to point and click their way through an entire exam quickly and effortlessly.” (EMR product review)

# The clinical narrative

“...In years past, a **well-written history and physical, or progress note, would unfold like a story**, giving a vivid description of the patient’s symptoms and physical exam at the point of the encounter, as well as the synthesis of the data and the plan of care.”

- “EMRs: Finding a balance between billing efficiency and patient care”, Henry F. Smith, Jr., MD, Commentary, The Times Leader, Wilkes-Barre, PA, June 12, 2011.

**April 14, 2007**

CHIEF COMPLAINT: Shortness of breath.

HISTORY OF PRESENT ILLNESS: This 68-year-old female presents to the emergency department with shortness of breath that has gone on for 4-5 days, progressively getting worse. It comes on with any kind of activity whatsoever. She has had a nonproductive cough. She has not had any chest pain. She has had chills but no fever.

EMERGENCY DEPARTMENT COURSE: The patient was admitted. She has had intermittent episodes of severe dyspnea. Lungs were clear. These would mildly respond to breathing treatments and morphine. Her D-dimer was positive. We cannot scan her chest; therefore, a nuclear V/Q scan has been ordered. However, after consultation with Dr. C, it is felt that she is potentially too unstable to go for this. Given the positive D-dimer and her severe dyspnea, we have waved the risks and benefits of anticoagulation with her heme-positive stools. She states that she has been constipated lately and doing a lot of straining. Given the possibility of a PE, it was felt like anticoagulation was very important at this time period; therefore, she was anticoagulated. The patient will be admitted to the hospital under Dr. C.

**April 14, 2007**

CHIEF COMPLAINT: Shortness of breath.

HISTORY OF PRESENT ILLNESS: This 68-year-old female presents to the emergency department with shortness of breath that has gone on for 4-5 days, progressively getting worse. It comes on with any kind of activity whatsoever. She has had a nonproductive cough. She has not had any chest pain. She has had chills but no fever.

EMERGENCY DEPARTMENT COURSE: The patient was admitted. She has had intermittent episodes of severe dyspnea. Lungs were clear. These would mildly respond to breathing treatments and morphine. Her D-dimer was positive. We cannot scan her chest; therefore, a nuclear V/Q scan has been ordered. However, after consultation with Dr. C, it is felt that she is potentially too unstable to go for this. Given the positive D-dimer and her severe dyspnea, we have waved the risks and benefits of anticoagulation with her heme-positive stools. She states that she has been constipated lately and doing a lot of straining. Given the possibility of a PE, it was felt like anticoagulation was very important at this time period; therefore, she was anticoagulated. The patient will be admitted to the hospital under Dr. C.



**April 14, 2007**

CHIEF COMPLAINT: Shortness of breath.

HISTORY OF PRESENT ILLNESS: This 68-year-old female presents to the emergency department with shortness of breath that has gone on for 4-5 days, progressively getting worse. It comes on with any kind of activity whatsoever. She has had a nonproductive cough. She has not had any chest pain. She has had chills but no fever.

EMERGENCY DEPARTMENT COURSE: The patient was admitted. She has had intermittent episodes of severe dyspnea. Lungs were clear. These would mildly respond to breathing treatments and morphine. Her D-dimer was positive. We cannot scan her chest; therefore, a nuclear V/Q scan has been ordered. However, after consultation with Dr. C, it is felt that she is potentially too unstable to go for this. Given the positive D-dimer and her severe dyspnea, we have waved the risks and benefits of anticoagulation with her heme-positive stools. She states that she has been constipated lately and doing a lot of straining. Given the possibility of a PE, it was felt like anticoagulation was very important at this time period; therefore, she was anticoagulated. The patient will be admitted to the hospital under Dr. C.

**May 3, 2007**

EMERGENCY DEPARTMENT COURSE: The patient was admitted and nontoxic in appearance. **Blood pressure was brought down aggressively. With this combined with BiPAP, she has reversed her respiratory distress promptly.** She has improved significantly. She will not require intubation at this time period. Her family has elected to go back to M, Dr. W. I did discuss this case with Dr. G who is on call for L Cardiology. She has accepted him in transfer; however, there are no PCU or ICU beds at this time period. Will admit here for a brief period until a bed is available at M. I discussed this case with Dr. R who will admit.

**Clinicians were trying to determine whether the shortness of breath was due exclusively to her failing heart, or whether she has pneumonia.**

**Prompt response indicates that pneumonia is not the issue.**

“...As EMRs proliferate, and increased Medicare scrutiny looms, **medical documentation is evolving from its original goal of recording what actually was going on with a patient, and what the provider was actually thinking, to sterile boilerplate documents** designed to justify the highest billing codes.

- “EMRs: Finding a balance between billing efficiency and patient care”, Henry F. Smith, Jr., MD, Commentary, The Times Leader, Wilkes-Barre, PA, June 12, 2011.

***Text boxes in EMRs don't solve the problem.***

***We're at risk of losing the rich language of the clinical record.***

***And if you lose the language, you lose the story.***

# Take-aways for discussion

- Clinical NLP needs more statistical NLP
- We've got a big problem: data availability for clinical NLP R&D
- We, and everyone else, have a far bigger problem: the future of clinical language in electronic health records.

Thanks!