Linguistic and World Knowledge in Medical Applications That Involve NLP

Sergei Nirenburg
Institute for Language and Information Technologies
University of Maryland Baltimore County

Joint work with Marjorie McShane

An Apology

This brief presentation directly mentions only a very small subset of biomedical NLP systems and system modules.

There is no implication that this choice is definitive or comprehensive.

Linguistics-Based Medical NLP Systems

A large number of NLP systems and functionalities have been developed for the domain of medicine.

While the application systems differ on a variety of specific parameters, the core NLP functionalities of most biomedical NLP systems are identical to those that support NLP applications outside the biomedical domain.

Linguistics-Based Medical NLP Systems

Most recent application systems – both general-purpose and biomedical – are **hybrid**: whenever practicable, they use any kind of available algorithms (statistical or otherwise) and knowledge resources (lexicons, ontologies, various rule sets, etc.) and fill in the lacunae with newly developed capabilities.

All such systems and methods are ultimately linguistics-based.

All NLP is Linguistics-Based

Indeed, statistics-based processing needs linguistic knowledge too, at the very least for:

- selection of **features** and their value sets:

  > "In the context of language, doing "feature engineering" is otherwise known as doing linguistics" (Manning 2004);

- preparation of sophisticated **training and test data sets** ("annotated corpora") to support a variety of NLP-related decisions

Approaches To Knowledge Acquisition

In statistics-based systems the preferred method of knowledge **acquisition** is **learning it automatically from large text collections**. As this is currently infeasible, learning is instead carried out over text corpora that are **manually annotated** using feature-value sets generated by humans.

The approach to knowledge acquisition used in "rule-based" systems is to use human labor not for facilitating learning algorithms but for acquiring static knowledge and processing rules directly.

Prominent Technologies and Application Areas

Much work in biomedical NLP is devoted to **information extraction (IE)** – generating structured knowledge about an object or event of interest by detecting in open text references to values of elements of predefined knowledge structures.

While IE can be an end application itself, it can also be viewed as an enabling technology for other end applications, such as **training (tutoring)** or **clinical decision support** (CDS).

Three Representative IE-Based systems: The Briefest of Illustrations

Typical input and outputs of IE-oriented medical NLP applications are well illustrated by the following example from the MedLEE environment.

HISTORY OF PRESENT ILLNESS: The 52 year old women presented to the emergency room with three episodes of severe right orbital headache lasting 2 hours and occurring every day with rapid onset, accompanied with nausea and vomiting, **without paresia**. Patient **denies myalgia**. Positive **family history for SLE**.
PHYSICAL EXAMINATION: Alert and oriented times three. Normal neurological examination.
LABORATORY DATA: ESR 37 mm/hr.
ASSESMENT and PLAN: The patient was hospitalized with a **preliminary diagnosis of temporal arteritis** treated with Prednisone 60 mg twice a day. Head CAT scan **ruled out subarachnoid hemorrhage**. Temporal artery biopsy planned tomorrow.

MedLEE generates the following output:

From http://www.cat.columbia.edu/pdfs/MedLEE_2006.pdf

MedLEE generates the following output:

Class		Qualifier	
Name	**Value**	**Name**	**Value , Unit**
finding:	women	Age =	52 , year
problem :	headache	bodyloc = region = degree = duration = quantity = change_rate =	orbital right high degree 2 , hours 3 , episodes fast
problem:	nausea	certainty =	high certainty
problem:	vomiting	certainty =	high certainty
problem:	**paresia**	certainty =	**absence of**
problem:	**myalgia**	certainty =	**absence of**
problem:	**systemic lupus erythematosus**	status =	**family history**
normalfinding:	alert and oriented times 3		
normalfinding:	examination	bodyfunc =	neurologic
labtest:	erythrocyte sedimentation rate	measure =	37 , mm/hr
problem:	**temporal arteritis**	certainty = status = time period =	**moderate certainty** **preliminary** admission
medication:	prednisone	certainty = dose = frequency =	high certainty 60 , mg bid
problem:	**subarachnoid hemorrhage**	certainty =	**absence of**
procedure:	computerized axial tomography	bodyloc =	head
procedure:	biopsy	bodyloc = timeper = status =	temporal artery tomorrow plan

The MedLEE Family

(e.g., Friedman 2001, Lussier et al, 2010)

At origin, MedLEE has used and extended the The Linguistic String Project (LSP) system (e.g., Sager et al. 1994).

Systems in the MedLEE family (MedLEE, BioMedLEE, PhenoGO) include the following major components:

- standard **preprocessing** stages;
- **syntactic analysis** using a sublanguage grammar;
- a **target information structure** with feature-value sets specific to particular biomedical domains and applications and encoded in a variety of coding schemata; and
- domain- and application-oriented **rule sets** and static knowledge resources (e.g., lexicons) for detecting relevant words and phrases in free text documents and mapping them into appropriate values of the target features.

MetaMap

(e.g., Aronson and Lang, 2010)

MetaMap maps words and phrases in biomedical texts to words and phrases serving as concepts in the UMLS Metathesaurus, a structured biomedical knowledge resource.

This mapping enhances a variety of practical applications, such as automatic indexing of MEDLINE citations and concept-based query expansion.

Input **analysis** steps in MetaMap include a battery of preprocessing modules (from tokenization to lexical lookup), shallow syntactic parsing and lexical variant generation. **Mapping** steps include finding a set of Metathesaurus string matching at least some of strings in the processed input and finding a longest match. An optional step includes collocation-based word sense disambiguation.

SemRep

(e.g., Rindfleisch and Fiszman, 2003)

Stated purpose is similar to MetaMap's: "SemRep is a natural language processing system designed to recover semantic propositions from biomedical text using underspecified syntactic analysis and structured domain knowledge from the UMLS."

SemRep uses MetaMap to augment referring expressions in inputs with Metathesaurus concepts. SemRep enhances the knowledge and processor inventory that supports the text-to-Metathesaurus mapping. For example, special rule sets have been developed for syntax-to-semantics linking and an entire module, SemSpec, for interpreting hyperonymic propositions.

Targeting Specific Features

To improve the utility of IE results, a number of biomedical NLP systems seek to extract values of a set of features that promise to enhance the quality of the original content extraction.

To give just a few examples, the **ConText** system (e.g., Chapman et al. 2007) targets negation, a subset of time-related properties ("temporality") and selective case role assignment ("experiencer"). The **MediClass** system (e.g., Hazelhurst et al. 2005) also targets negation and additionally addresses indicators of severity of medical events and quantification. Event extraction is a popular objective, as witnessed, for example, by the popularity of the BioNLP Shared Task efforts in 2009 and 2011.

The ultimate goal of all the above efforts is to improve results of extraction by determining selected ("contextual") aspects of text meaning beyond collocation-oriented word sense disambiguation.

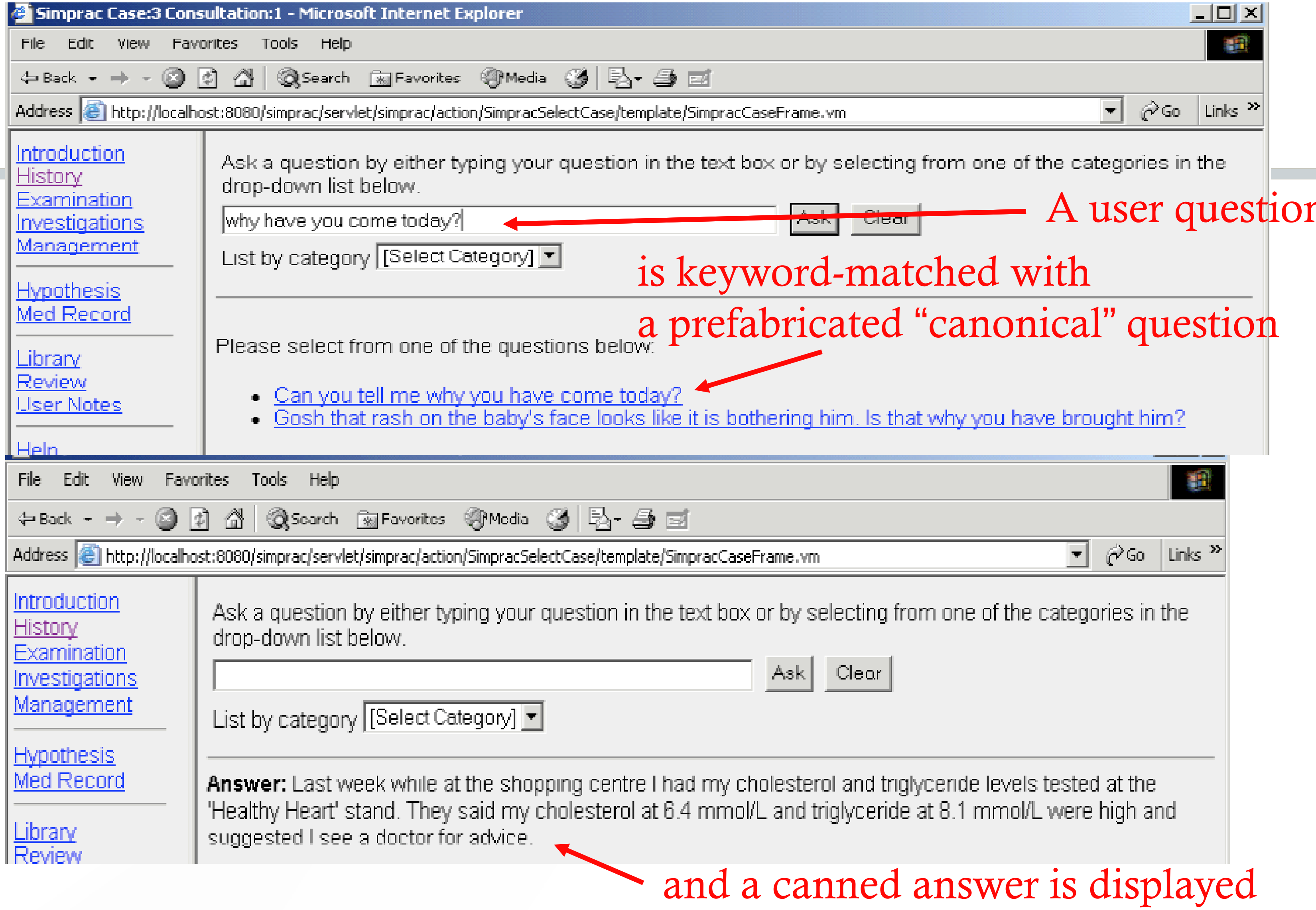
Focus on Training Systems

Many diverse medical training systems have been developed over the past 40 years, including **technical task trainers** and **cognitive skill trainers** based on virtual patients (VPs) that use human actors or computer simulations.

Typical state-of-the-art **cognitive skill training systems** extract specific knowledge elements from user input to help the selection of a path through a **decision tree** whose nodes correspond to decision points in the training task. In other words, a) training system can be viewed as a kind of decision support systems; and b) NLP in these systems is typically carried out in the IE manner.

For example, systems developed by MedCases, Inc., Therasim, Inc., the Sim-Patient system from RTI International and systems configured in the CIRCSIM-Tutor project (e.g., Evens and Michael 2006) and in the eVIP project (e.g., Zary 2007) fall into this broad category.

NLP in state-of-the-art virtual patient-based training systems



Douglas Chesher. Exploring the use of a web-based virtual patient to support learning through reflection. University of Sydney PhD Thesis. 2004

Focus on Decision Support Systems

Details about research in this area can be found in several recent surveys of the state of the art in medical decision support systems, including:

Wang et al 2007
Demner-Fushman et al. 2009
Berner 2009

Decision support challenges are also discussed in the influential report from the National Research Council on computational technology for health care (Stead and Lin 2009).

The above materials provide excellent analyses of the issues involved and cover a very large percentage of systems and projects under development. Just a few general comments will suffice here.

Decision Support Systems

Decision support systems can be useful in both clinical practice and in research environments. Clinical decision support systems may have a greater societal impact but face issues of user acceptance.

Human-computer interaction in medical decision support systems can take different forms, though many systems involve NLP.

NLP capabilities required for HCI are **not exactly the same** as those needed to support IE.

Advanced decision support systems have separate decision-making modules that rely on NLP modules for decision-making knowledge and communicating with the user.

NLP capabilities required by decision-making modules are **not exactly the same** as those needed to support either IE or HCI.

Decision Support Systems

Some NLP-inclusive medical decision support systems seek to extract from text and use not slot fillers of predefined structures but representations of all the propositional and pragmatic/discourse meanings of a text.

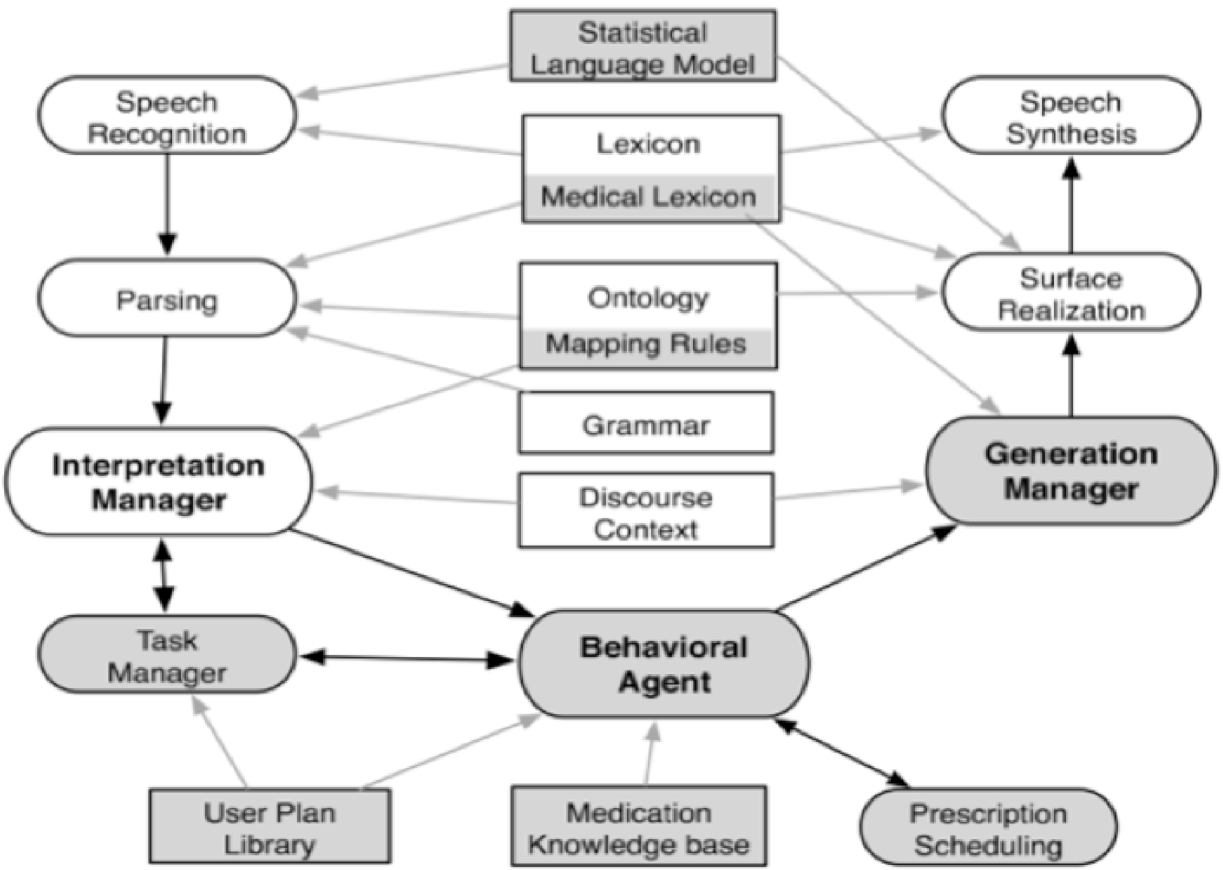
Chester (Allen et al. 2006), an application of the TRIPS dialog system to the medical domain, is one such system.

Maryland Virtual Patient (**MVP**) and Clinician's Advisor (**CLAD**) are two proof-of-concept systems developed in the OntoAgent project (e.g., McShane et al., *forthcoming*; Nirenburg et al. 2010).

Chester: TRIPS for the Medical Domain

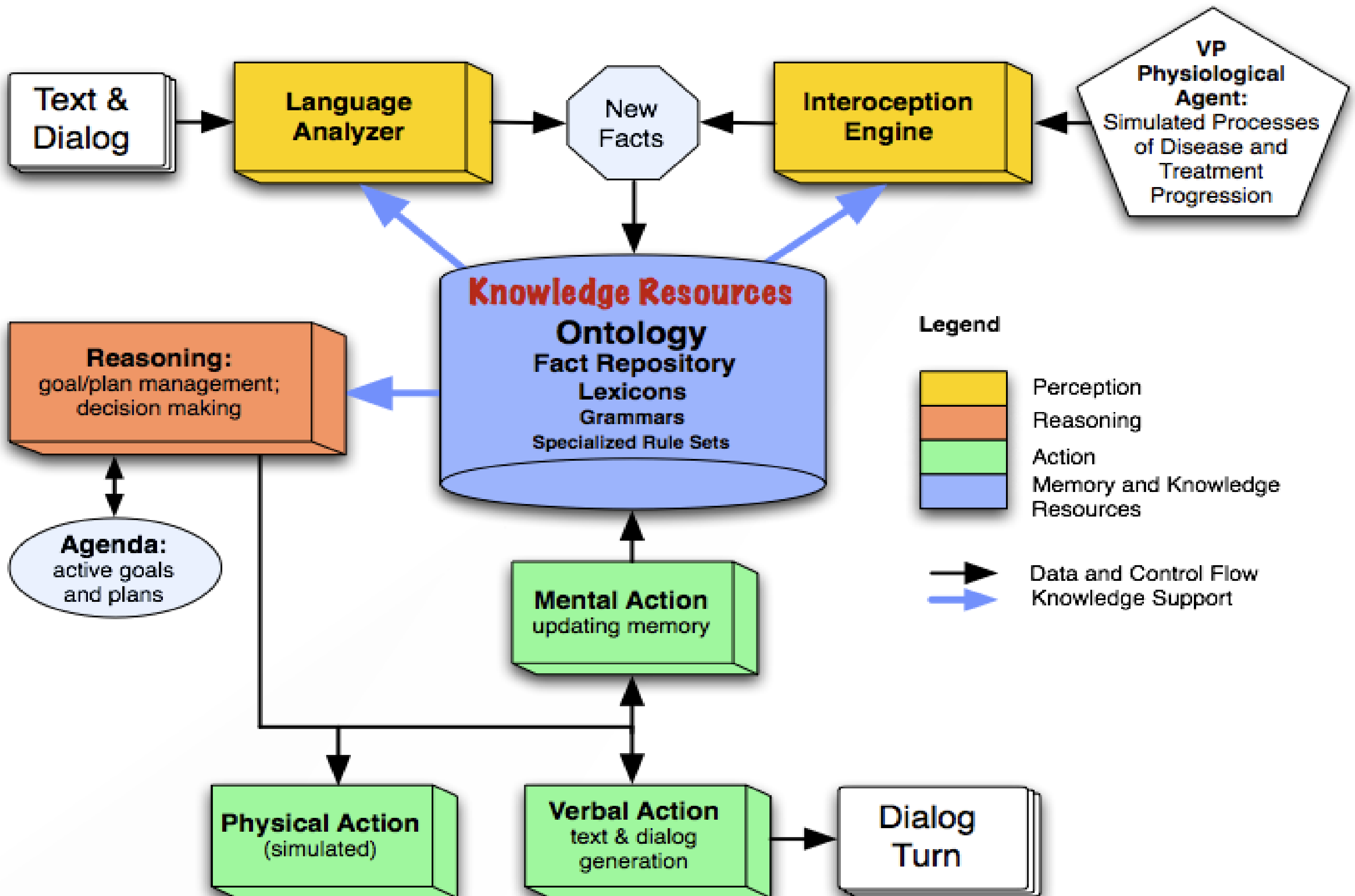
(e.g., Allen et al. 2006)

Chester reminds patients about medication scheduling to help with compliance.



OntoAgent

(e.g., McShane et al, *forthcoming*)



MVP

MVP models a team of artificial intelligent agents – notably, a **virtual patient** and a **tutor** – and a human agent in a training system where the human plays the role of trainee.

The artificial intelligent agents are implemented as different instances of the OntoAgent architecture.

NLP in MVP

When the trainee types “What brings you here?” the VP generates a meaning representation for this text (shown in a simplified format):

REQUEST-INFO-31
THEME COME-13.PURPOSE
AGENT PHYSICIAN-17
BENEFICIARY PATIENT-1
COME-13
AGENT PATIENT-1
DESTINATION OFFICE-23

(Numbers indicate specific instances of corresponding ontological concepts)

User Acceptance

The ultimate criterion of success of any application system is **end user acceptance**.

A brief survey of users' opinions about current and future medical decision support systems will help to identify user desiderata for the capabilities in such systems, including NLP capabilities.

What Users Say1/5

- Extant clinical decision support systems are failing to offer "tailored, clinically appropriate choices" for a given patient. (Wright et al. 2009)
- Clinical QA system users want **direct answers** to questions, **specific recommendations**, a **rationale** for recommendations, a practical **tempering** of original research with practical considerations, and an "emphasis on **treatment and bottom-line advice**." (Ely et al. 2005)

What Users Want 2/5

- **Customizability** should be stressed: different features, different rule sets, etc. (Berner et al. 2009)
- Systems are needed that are not monolithic – that can be reconfigured, that can learn over time.

  "…Any IT-based infrastructure to support today's health care needs must be designed to accommodate changes in roles and process tomorrow" (Stead and Lin, 2009)

What Users Want 3/5

- "Five Rights" of clinical decision support systems must be supported: give the right information to the right person in the right format through the right channel at the right time (Osheroff et al. 2009).
- Don't get in the user's way! "Alert fatigue" has left most decision support systems unused by clinicians. Need more work on tiered alerts, user configuration of preferences, and **less overall noise from systems**. (Berner 2009).

What Users Want 4/5

- "“[IT applications] are often designed in ways that… provide little support for the **cognitive tasks** of clinicians or the workflow of the people who must actually use the system."
- "Cognitive support is **not well served by the task-specific automation systems** that make up the majority of today's health care IT." (Stead and Lin, 2009)

What Users Want 5/5

Must address user's **biases: clinicians tend not to want help!**

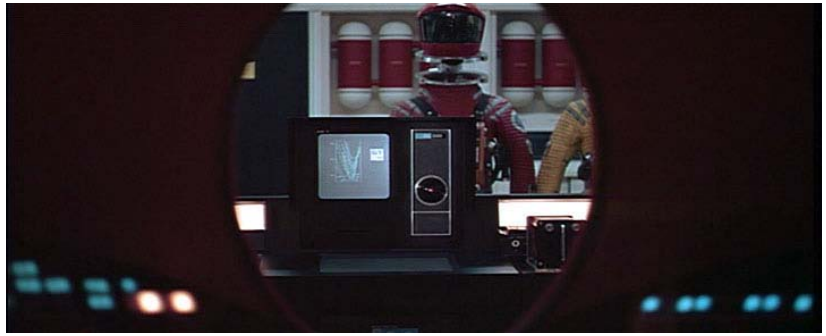
- Sintchenko et al. (2004) report that use of decision support in ventilator-associated pneumonia treatment improved clinician decision-making (from 65% agreement with an expert panel to 97%) but (a) **this did not affect clinician confidence in the decision** (clinicians were very confident anyway) and (b) in about 2/3 of cases clinicians **did not seek decision support**.

"Priorities in NLP development will be determined by the readiness of intended users to adopt NLP." (Demner-Fushman et al, 2009)

What Do Users Want Clinical Decision Support Systems to Be Like?

It seems that the users want clinical decision support systems that resemble **HAL** from Stanley Kubrick's *2001: A Space Odyssey.*

To jog your memory, here are a few illustrations of some of HAL's capabilities:

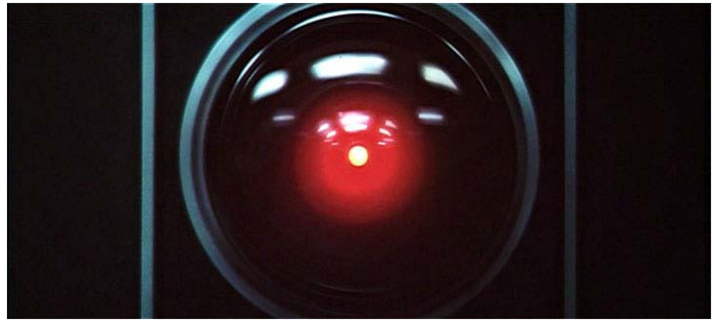


A sample dialog with HAL

Dave: Open the pod bay doors, HAL.
HAL: I am sorry, Dave. I am afraid I can't do that.
Dave: What's the problem?
HAL: I think you know what the problem is just as well as I do.
Dave: I don't know what you're talking about.
HAL: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.

HAL's capabilities demonstrated by the dialog:

Understanding of "request action"
Nominal compounding
Reference resolution (select examples)
Politeness (emotional intelligence)
Idiomaticity, politeness
Belief ascription: Modeling the knowledge, actions, goals, plans of and beliefs of others
Self-awareness: Modeling oneself, including goals, plans, etc.
Embedded modalities



HAL I can tell from the tone of	Emotion recognition via speech
your voice, Dave, that	Politeness via vocative
you're upset.	Modeling feelings of others
Why don't you take a stress	Indirect speech act
pill and get some rest.	Broad vocabulary, incl. jargon
[*Dave has just drawn another*	Subject ellipsis in VP-coord. structure
sketch of Dr. Hunter].	Confident use of ambiguous lexeme
HAL Can you hold it a bit closer?	Reference resolved to physical object
[*Dave does so*]	Relative spatial orientation
HAL That's Dr. Hunter, isn't it?	Ellipsis of *to me* (closer *to me*)
Dave Yes.	Image recognition
	Reference resolution to memory
	Generation of tag question.
	Understanding of elliptical 'yes'.

Why aren't we closer to a HAL yet?

"As we near the year 2001, do we have a computer that sounds like the voice of HAL portrayed by actor Douglas Rain…? The answer is no, not yet… The greatest obstacle … **is the machine's inability to comprehend what it is saying or hearing.**"

Joseph P. Olive (in Stork (1997), **Hal's Legacy**. MIT Press p. 124.)

"…[T]o understand language as well as he does, HAL would need **a complete model of the world** that includes understanding his own goals, the goals of those around him and the relative significance of each. In addition, he would have to understand all the ways of referring to such goals and the potential problems that could interfere with carrying them out."

Roger Schank, *ibid*. p.179.

What Now?

Though significant R&D progress has been made, the abovementioned obstacles are still very much present in 2012.

If we want to address the task of overcoming these obstacles head on:

Is the currently prevalent IE-oriented approach (whether realized through statistical, rule-based or hybrid methods) scientifically the most promising?

Or was the choice motivated by extra-scientific considerations?

The "Grand Research Challenge"

The recent National Research Council report (Stead and Lin 2009) states:

"**Patient-centered cognitive support emerged as an overarching grand research challenge during the committee's discussions.**"

"Clinicians have a "virtual patient" in mind—a conceptual model of the patient reflecting their understanding of interacting physiological, psychological, societal, and other dimensions. They use new findings—raw data—to refine their understanding of their virtual patient. Then, based on medical knowledge, medical logic, and mostly heuristic decision making, they formulate a plan, expressed as an order (transaction), to try to change the (real) patient for the better."

How Much Effort Will It Take?

We are facing a "grand challenge" project. Let's try to compare the size of this project with two well-known "grand challenge" projects.

The Manhattan Project

Manhattan Project expenditures in 1942-45 amounted to:

$69,681,000 in 1945 dollars which corresponds to $888,014,664 in 2012 dollars.

The R&D component accounted for only 3.7% of the above.

Source: http://www.brook.edu/dybdocroot/FP/PROJECTS/NUCWCOST/MANHATTN.HTM

The Human Genome Project

The Human Genome Project expenditures in 1988 - 2003 amounted to:

$3,812,600,000 in 2003 dollars, which corresponds to $4,753,151,843 in 2012 dollars.

The above does not include funding of genomics research outside of this project.

Source: http://www.stanford.edu/class/siw198q/websites/genomics/entry.htm)

An Attempt at Comparison

As regards the project to build a machine that can meaningfully communicate with people is, in my opinion, is at of at least **the same complexity** as the tasks before Manhattan and Human Genome projects. (I actually think that our project is much more complex than either of the others.) This means that the costs of these projects will be at least commensurate.

The Manhattan and Human Genome projects address societal needs (national security and health respectively) that are more immediate than the need for our project…

Watson is probably the largest ever NLP project (have the numbers been made public?). It is a spectacular project but it does not bridge the gap…

References

Aronson, A.R. and F. Lang. 2010. An overview of MetaMap: historical and recent advances. JAMIA 2010, 17: 229-236.

Berner, Eta S. 2009. **Clinical Decision Support Systems: State of the Art. Prepared for:** Agency for Healthcare Research and Quality U.S. Department of Health and Human Services, **AHRQ Publication No. 09-0069-EF June 2009**

Coden, Anni, Guergana Savova, Igor Sominsky, Michael Tanenblatt, James Masanz, Karin Schuler, James Cooper, Wei Guan, Piet C. de Groen. 2009. Automatically extracting cancer disease characteristics from pathology reports into a Disease Knowledge Representation Model, Journal of Biomedical Informatics 42 (2009) 937–949.

Manning, C. 2004. Beyond the Thunderdome. In Proceedings of CONLL-04.

ELY, John W., JEROME A. OSHEROFF, M. LEE CHAMBLISS, MARK H. EBELL, MARCY E. ROSENBAUM. 2005. Answering Physicians' Clinical Questions: Obstacles and Potential Solutions, Journal of the American Medical Informatics Association Volume 12 Number 2 Mar / Apr 2005, pp. 217-224.

HAZLEHURST, Brian, H. ROBERT FROST, DEAN F. SITTIG, VICTOR J. STEVENS. 2005. MediClass: A System for Detecting and Classifying Encounter-based Clinical Events in Any Electronic Medical Record. Journal of the American Medical Informatics Association. Vol 12 #5.

Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA. 1998;280:1339-46.

Osheroff, J.A. 2009. Improving medication use and outcomes with clinical decision support: a step-by-step guide. Chicago, IL: The Healthcare Information and Management Systems Society.

PERSSON, M., T. MJORNDAL, B. CARLBERG , J. BOHLIN & L. H. LINDHOLM. 2000. Evaluation of a computer-based decision support treatment of hypertension with drugs: retrospective, nonintervention testing of cost and guideline adherence Journal of Internal Medicine 247: 87-93.

SINTCHENKO, VITALI, ENRICO COIERA, JONATHAN R. IREDELL, GWENDOLYN L. GILBERT. 2004. Comparative Impact of Guidelines, Clinical Data, and Decision Support on Prescribing Decisions: An Interactive Web Experiment with Simulated Cases. Journal of the American Medical Informatics Association 11(1): 71-77.

Stead, William W. and Herbert S. Lin, Eds. Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions. The National Academies Press, Washington, DC.

M. PERSSON, T. M JOÈRNDAL, B. CARLBERG , J. BOHLIN & L. H. LINDHOLM. 2000. Evaluation of a computer-based decision support treatment of hypertension with drugs: retrospective, nonintervention testing of cost and guideline adherence Journal of Internal Medicine 247: 87-93.

Coden et al. 2009: *Automatically extracting cancer disease characteristics from pathology reports into a Disease Knowledge Representation Model*

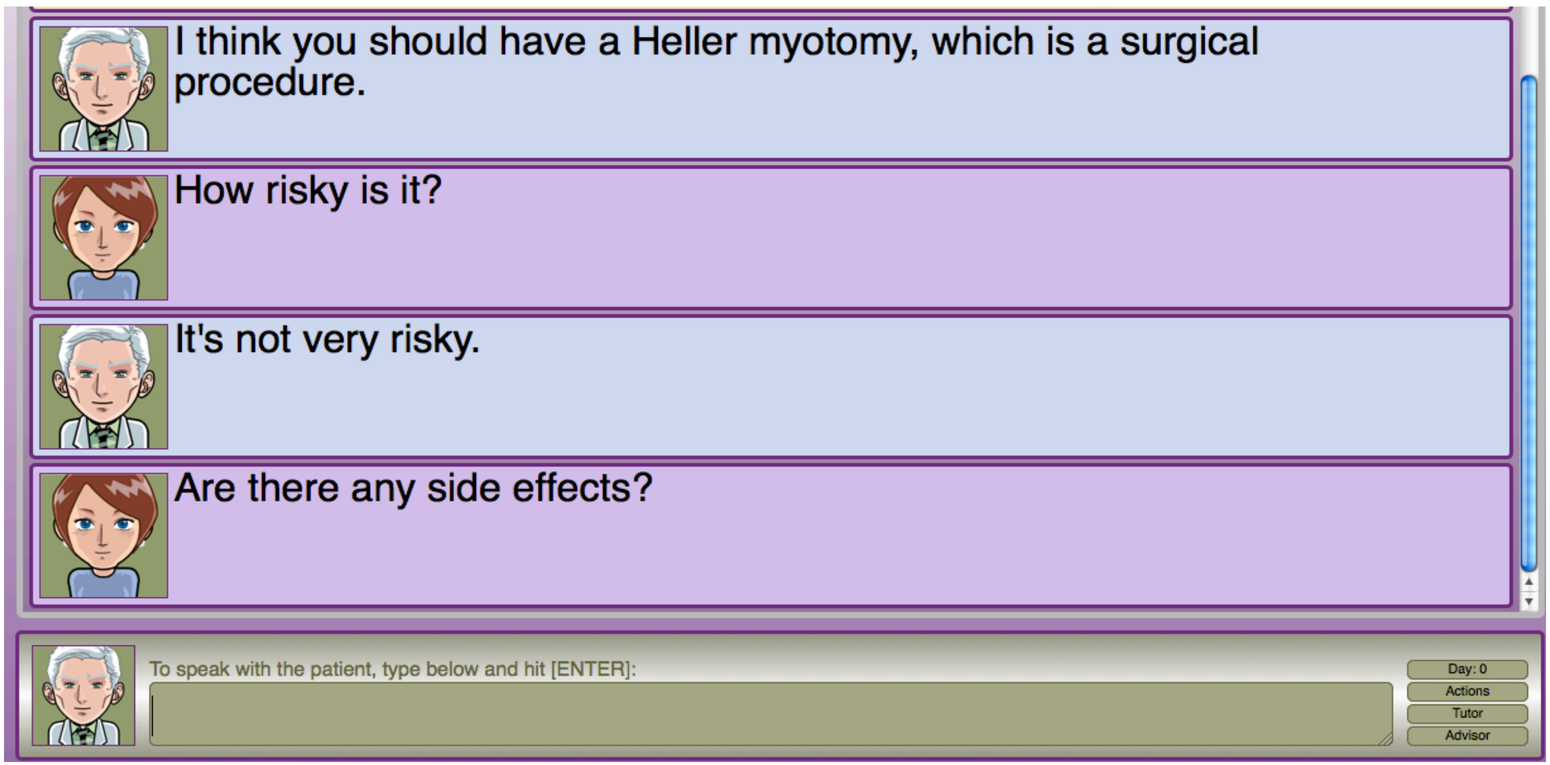
Friedman, C. 2000 A broad-coverage natural language processing system. Proceedings of the AMIA symposium

Nirenburg, Sergei, Marjorie McShane, Stephen Beale, Jesse English and Roberta Catizone. 2010. Four kinds of learning in one agent-oriented environment. In: Samsonovich AV, Jóhannsdóttir KR, Chella A, Goertzel B., editors, Proceedings of the First International Conference on Biologically Inspired Cognitive Architectures, Arlington, VA, Nov. 13-14, 2010. Amsterdam, Netherlands: IOS Press, 2010. p. 92-97.

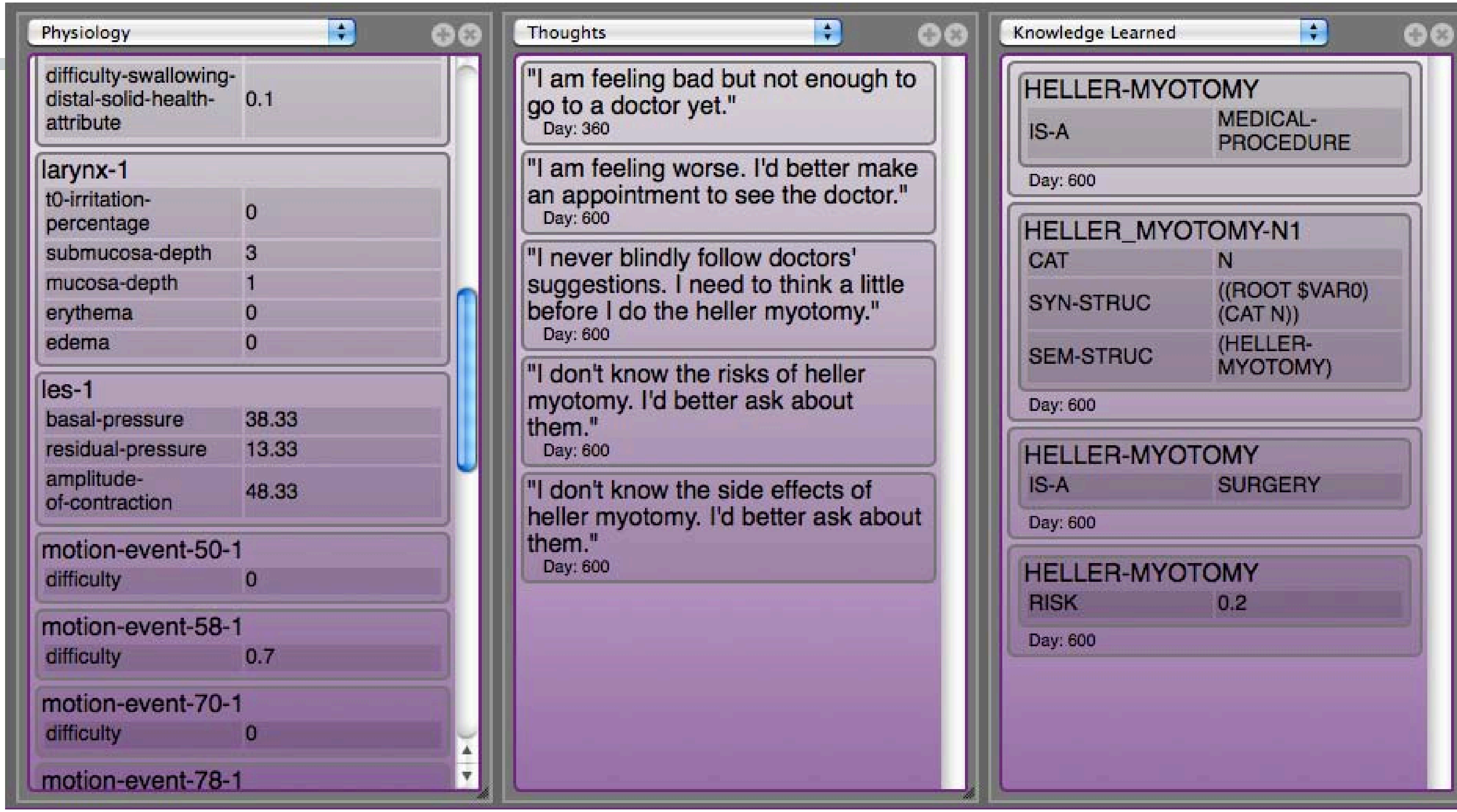
Marjorie Mcshane, Stephen Beale, Sergei Nirenburg, Bruce Jarrell, George Fantry. *Forthcoming*. Inconsistency as diagnostic tool in a society of intelligent agents. Artificial Intelligence in Medicine.

Reserve Slides

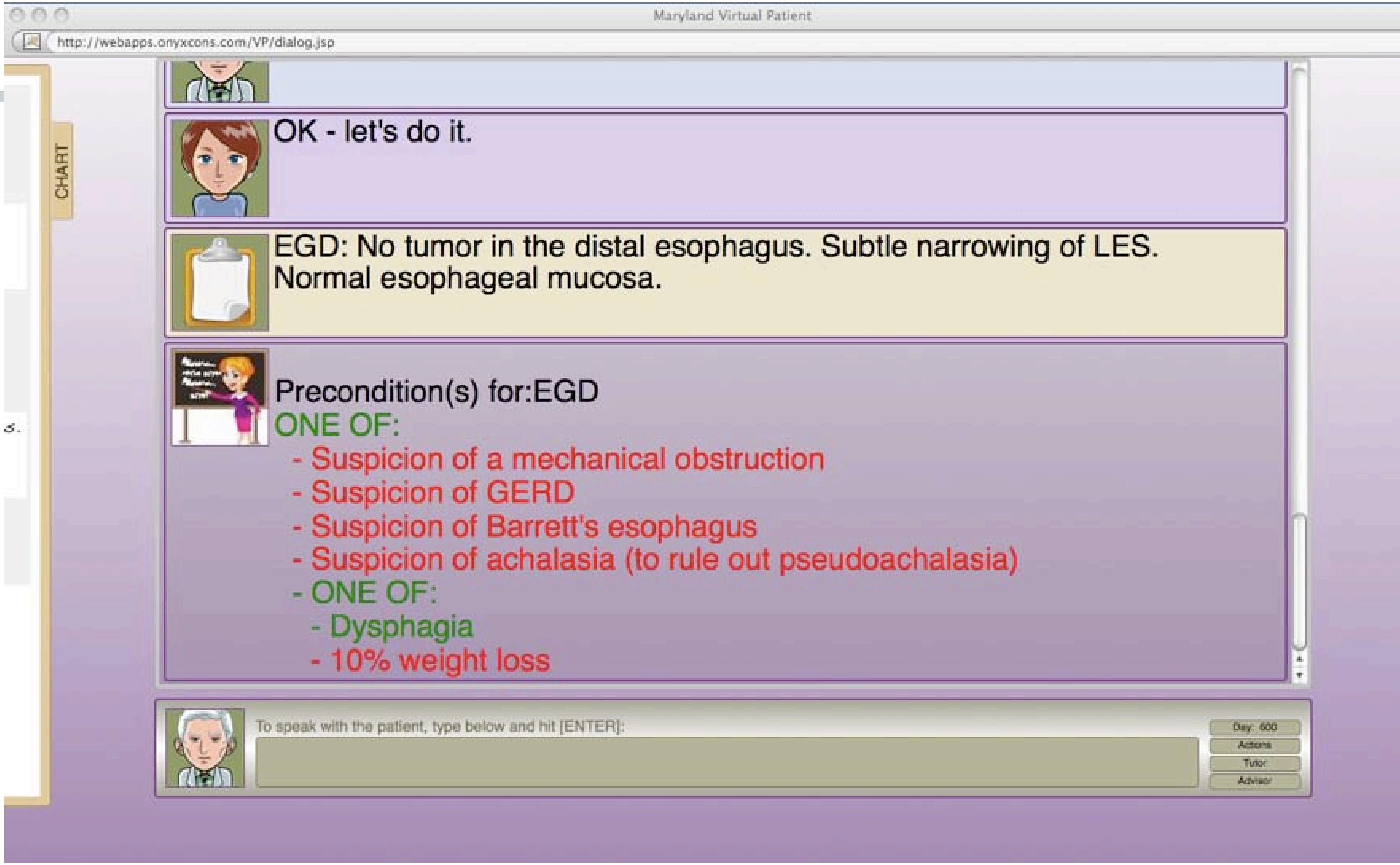
Dialog Excerpt from Maryland Virtual Patient (MVP)



“Under the Hood” of MVP



Decision Support in MVP: the Tutor Agent



NLP in MVP

When the trainee types "What brings you here?" the VP first generates a "literal" meaning representation for this text (shown in a simplified format):

REQUEST-INFO-31
THEME COME-13.PURPOSE
AGENT PHYSICIAN-17
BENEFICIARY PATIENT-1
COME-13
AGENT PATIENT-1
DESTINATION OFFICE-23

(Numbers indicate specific instances of corresponding ontological concepts)

The VP must be able to recognize and use the situational context to conclude that the intended meaning of the input is:

REQUEST-INFO-32
AGENT PHYSICIAN-17
THEME SET-4
BENEFICIARY PATIENT-1
SET-4
MEMBER-TYPE SYMPTOM-1
SYMPTOM-1
EXPERIENCER PATIENT-1

Determination of intended meaning facilitates appropriate decisions about how to respond to this request.

Note that this conclusion may be wrong in a particular case.

Zooming Out from Clinical Decision Support

(From Stead and Lin) Of the committee's 7 desiderata reflecting the IOM's vision for 21st century health care, 4 arguably require knowledge-rich systems:

- "Comprehensive data on patients' conditions, treatments, and outcomes" [*one assumes it has to be collated and interpretable to be useful*]
- "Cognitive support for health care professionals and patients to help integrate patient-specific data where possible and account for any uncertainties that remain
- Cognitive support for health care professionals to help integrate evidence-based practice guidelines and research results into daily practice
- Empowerment of patients and their families in effective management of health care decisions and their implementation, including personal health records, education about the individual's conditions and options, and support of timely and focused communication with professional health care providers." [*i.e., systems can't be made for only trained specialists*]

To expand later!

We need knowledge-based systems to support high-end applications, such as clinical decision support systems.

To support communication with knowledge-based systems – and for knowledge acquisition – language processing is the best way to go, in part because the development of NLP systems shares many requirements with the development of knowledge based systems, e.g., decision support systems.

To expand later 2!

What we need instead is to motivate the analysis of prerequisites and desiderata for creating high-end clinical decision support systems by how we propose to address the current obstacles to acceptance of such systems.

Note that some of the prerequisites for clinical decision support are themselves NLP applications that are pursued quite separately from the need to support clinical decision.

For example, automatic extraction of knowledge from medical records.

The Specialist Lexicon and WSD

Search By Base

Options: Output Options | Global Options | Version | Reset

hurt

Search Clear

```
{base=hurt
entry=E0032269
	cat=verb
	variants=irreg|hurt|hurts|hurt|hurt|hurting|
	intran
	tran=np
	tran=fincomp(t):subj
}
{base=hurt
entry=E0032270
	cat=noun
	variants=uncount
```

http://lexsrv3.nlm.nih.gov:8100/WebLexAccess.2012/jsp/getResults.jsp; accessed Apr. 19, 2012

Verbal Senses of *hurt* in OntoSem (in presentation format)

*Verbal Sense of **hurt***	*Example*	*Syntactic Dependencies*	*Semantic Interpretation: Ontological Mapping*
v1: someone injures a body part	John hurt his knee	Subject Verb DirectObject	INJURY EXPERIENCER THEME
v2: someone injures another	He hurt the cat	Subject Verb DirectObject	INJURY CAUSED-BY EXPERIENCER
v3: to harm but not by injury as such	Poor nutrition hurt him.	Subject Verb DirectObject	CHANGE-EVENT: value of "HEALTH-ATTRIBUTE" (DOMAIN) decreases, CAUSED-BY
v4: of a body part, to cause pain	His leg hurts	Subject V	PAIN THEME
v5: of an event, to cause pain	This procedure hurts.	Subject V	PAIN CAUSED-BY
v6: to offend	His remarks hurt me.	Subject Verb DirectObject	OFFEND THEME CAUSED-BY
v7: to damage something inanimate	The accident hurt his car.	Subject Verb DirectObject	CHANGE-EVENT: value of "STATE-OF-REPAIR" (DOMAIN) decreases, CAUSED-BY

Deployment-Related Concerns

The ultimate goal for any application system is **broad deployment**. To attain this goal, one must take into account several often conflicting considerations, such as:

- utility (how important is a particular application for attaining the overall task – having a top-quality spell checker is good but having a top-quality semantic analyzer has a higher overall utility)
- **breadth and depth of coverage** (not only of lexis and grammar but also of features required to support medical decision making)
- **nature** of output (e.g., whether it is intended for direct use by people or by reasoning systems or artificial intelligent agents)
- **level of automation** (is human help needed to produce final results?)
- **output quality** (of both fully automatic and human-aided configurations where system results are validated and post-edited by people)
- **operational costs** (these will be high for human-aided systems)
- development costs.

MVP

The current implementation of MVP trains medical personnel in diagnosing and treating diseases of the esophagus.

The virtual patient (VP) agent in the MVP system is “seeded” with a certain medical condition. The VP’s disease progression is simulated over time. When the VP perceives symptoms, it initiates a visit to the MD, whose role is played by the trainee in MVP. The trainee engages the VP in an information-gathering dialog, orders (simulated) labwork, establishes a diagnosis and suggests a treatment. The VP may ask the MD a variety of clarification questions about conditions, tests and treatments and influence the choice of treatments (even refuse a treatment). Treatment