

Where next?

**Why we don't care about other
fields that care about language**

Eduard Hovy

USC Information Sciences Institute

Marina del Rey, CA

hovy@isi.edu

The Statistics Revolution is over...

- ...when EMNLP 2004 called for
While reporting aggregate percentages remains essential, **instructive research should also explain a model's limitations in more meaningful ways**. This can be as simple as categorizing error statistics by finer-grained types of errors, to reveal specific areas of model limitation. Better yet, **we seek deeper insight into the models' inherent representational biases, in the form of qualitative theoretical analyses**
- ...when even Ken Church says “enough is enough”
- CL has become a field of code monkeys:
 - Excellent problem analyzers
 - Excellent large-scale data managers
 - Excellent machine learning appliers
 - Conference papers focus on **how to do something**, NOT **why it is important** or **why the method works**

So what are we going to do now?

- Large challenge problems remain:
 - **MT**: since 1950s, still not solved; statistical techniques are slowly reimplementing rule-based techniques, up the Vauquois Triangle
 - **ASR**: since 1960s, still not solved: terrible performance for open-domain vocab when untuned
 - **IR**: since 1970s, still not solved: stuck for a decade at 0.45 F-score. IR community limps along year by year searching for new problem
 - **IE**: since 1980s, static; now focusing on extraction pattern learning and web mining; terrible F-score on complex open-ended material
 - **Summarization**: since 1990s, a lot of work on news and summaries; few new ideas since early 2000s
 - **QA**: since 2000s, only factoids; still terrible performance
 - **Machine Reading**: the oldest research problem in NLP
- Smaller challenge problems:
 - **WSD**: we can do it, but we don't
 - **Subjectivity/sentiment/opinion**: ditto
 - **Textual Entailment**: ditto

To become famous, you used to
have to start a new conference
(TINAP, EMNLP, RANLP, PACLING...)
or build a useful piece of software
(Brill tagger, Charnak parser)
Now you create a new challenge
problem and corpus

What our field lacks

- Most of our energy today is spent
 - **hacking away** at applications or challenge problems
 - [and] exploring some **new technique** (SVM, Kernel methods, LSA, LDA...)
 - running some kind of **evaluation**
 - building a **new corpus**
- **Our field lacks a theory**
(or even multiple theories)
 - We really are code monkeys, typing away without a Grand Vision



So where should we turn?

- Linguistics?
- Psycholinguistics?
- Cognitive Science?
- Social Sciences?
- Information Theory?

Why we don't need Linguistics

When last did *you* find a linguistic theory helpful?

Problem 1: Linguists study different things than we do:

- We work with the average case, the typical phenomena. We leave aberrations/exceptions either for special-purpose treatment or ignore them
- Linguists study extreme/unusual/odd phenomena, to gain an insight about some specific phenomenon
 - They produce elegant theoretical accounts, representations, and rules
 - They do not produce data, numbers, or average cases
- So linguists' theories are [almost always] hard to implement, impossible to address with machine learning, and mostly simply not relevant to our problems

Problem 2: Linguists' methodology is suspect:

- Linguists annotate during theory discovery
 - Do not use multiple annotators afterwards
- Result: no verification of theoretical claims
 - (Seldom give any verification *at all*, except for example cases)

So: we can *help* the linguists by testing their theories, if we want

- but it's often a waste of time since they don't like us

Why we don't need Neurolinguistics

Learning about language by looking at firing patterns in the brain is like learning about an algorithm by looking at the flickering light patterns on old-style computer consoles

- **Problem: The architectures of the brain and computer differ**
 - There's no useful info for us in *purely structural* info about the brain
 - Can we infer *functional* lessons from the brain?
 - **Subtasks X and Y are performed in different parts** of the brain — perhaps they should be distinct code modules (e.g., syntax independent of semantics?)
 - **Subtask X is performed before/after subtask Y** — perhaps that's important for our processing (e.g., syntax before semantics?)
 - **Information of type X is stored elsewhere from info of type Y** — perhaps we should differentiate them as well (e.g., word sense granularity?)
- “We need you...You make things really *work* that the rest of us simply say in a single word” — our invited speaker John Gabrieli yesterday

Why we don't need CogSci/ Psycholinguistics



Building a plane that flies like a duck...

- Problem: **Results are limited by the crudeness of the experimental methodology.** They have essentially only three experimental procedures for people:
 - **Reaction time:** what can we infer from how long it takes people to respond. Gives hints about complexity of task. How does that help us?
 - **Confusion:** what we can infer about ‘cognitive similarity’ of objects. Perhaps useful for ontologies and/or word senses?
 - **Forgetting:** what can we infer about organization of memory. And how does this help us?
- There's no reason to believe that a psychologically / cognitively accurate computational model will tell us much about how we should do MT, IR, etc. Perhaps the opposite, even...

Why we don't need Info Theory

...And if you though the empirical approach has gone too far...

- Info Theory studies accurate transmission of info-bearing signals, and measures the quantities or loss of info conveyed
- This is **relevant to us, but not directly**:
 - The Statistics Revolution brought Machine Learning (i.e., measurements of effectiveness of notation transformations, large-scale models of language (ngram models), measures of informational similarity (pmi, etc.) into CL
 - These notions are connected to Info Theory
- But **Info Theory is too incomplete/crude/'weak'** for our needs:
 - The info content measure ($p \log p$) is a generalized score of undifferentiated 'information'
 - We need more specialized measurements of specific phenomena (parse trees built, documents returned, labels inserted, etc.)
 - (what's more, $p \log p$ is based on large-scale (averaged) entropy; no recognition of individual user's knowledge and/or inferential abilities)

Why we don't need Social Sciences

Here, finally, we're getting some connection!

- Various Social Sciences have produced results and procedures useful for us:
 - **Political Science**: Annotation of goals, attitudes, etc.
 - **Psychology** and related: Lists of emotions, attitudes, etc.
 - **Sociology**: Annotations of interpersonal behavior types (dialogue, flirting, etc.)
 - **HCI**: Studies of turn-taking and initiative, etc.
- However, the methodology is also often suspect:
 - Lists of terms and concept simply made up, annotations not carefully checked, etc.
 - We can easily read their papers and make our own lists
 - But we could enter into fruitful dialogue: we verify their theories and they give us theoretical bases for our work

What we need

1. An understanding that NLP is
 - transformation of one notation into another
 - with some insertion of additional information
 - with the purpose of modeling human communication
2. Some careful thinking about an overall computational theory of semantics, anchored on language communication
 - What it might be
 - How you might create large corpora containing
 - How the different NLP applications might use improve

(advertisement: code monkeys needed!)



What makes us unique

We are a hybrid of

- **Computer science:** methods to perform notation transformations (Finite State Automata etc.)
- **Linguistics:** phenomena of language
- **Statistics/machine learning:** obtaining necessary parameters and data
- **Philosophy / logic:** understanding the nature of meaning and communication
- **Cognitive psychology:** seeing what humans do
- **Information theory** (eventually): measuring efficiency of transformations and communication

Our theories will not look like anyone else's.

...so, it's up to us

THANK YOU

Where next?

Why we **should** take into account
other fields that **seriously** care about
language and **communication**

Michael Zock

C.N.R.S. – L.I.F.

Marseille, France

zock@free.fr

From theory to practice

After two decades of suffering from theory-driven, formal arm-chair-linguistics (symbolic approach), the computational linguistic community has moved towards data-driven corpus processing (statistical approach).

While this change has led in some areas to spectacular progress, confirming the conclusion that we had moved from toy worlds and made-up examples to the real world, it has also produced a number of negative side-effects

At least 3 consequences

- change of perspective, sometimes at odds with reality
- a certain kind of inbreeding
- overemphasis of evaluation

Change of perspectives

A subtle change of research *goals* has taken place which is somehow at odds with reality, i.e. practical needs. Many of the traditional tasks are hardly addressed anymore:

- ➡ knowledge representation,
- ➡ multimedia processing (integration of linguistic + visual information)
- ➡ user modeling,
- ➡ computational cognitive modeling,
- ➡ NLP and education,
- ➡ etc.

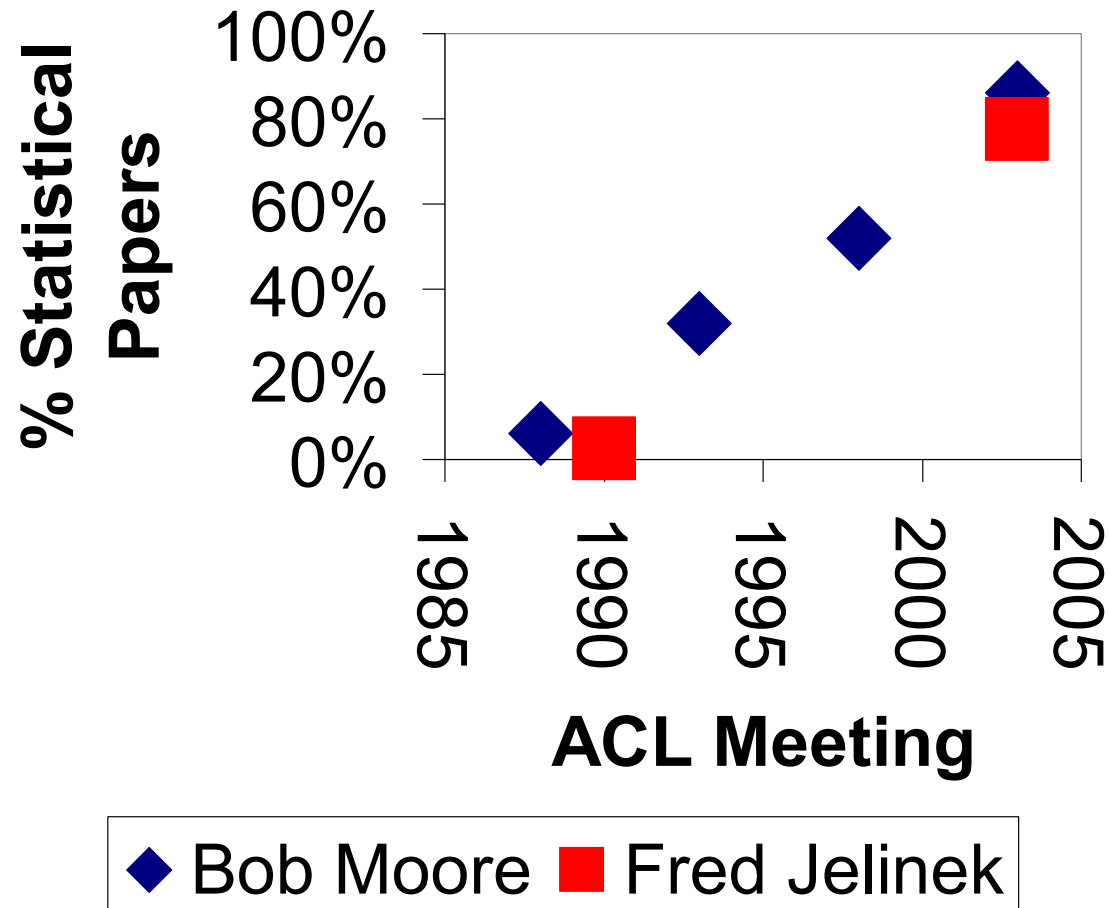
Inbreeding

Many of the traditional disciplines **don't** seem to be relevant anymore. For example, a study done by Reiter (2007) revealed that the works done by psychologists and linguists cited in major CL conferences and journals has dropped respectively from **13%** and **15%** to **5%** and **1%** in the period of 1995 and 2005.

This decrease of citing psychological research is all the more surprising, as the field has made substantial progress, partially due to a more widespread availability of new technology (eye trackers, brain scanners).

While this fact seems to be **ignored** by **our community**, it is clearly **acknowledged** in the area of **AI**. Indeed, **25%** of the work cited in the highly respected AI journal, refers to psychologists' work. In other words, the AI community shows considerably more sensitivity and awareness of the work produced by the broader language research community than the CL community (25% vs. 5%).

When will we see the **last** non-statistical paper, 2010? (slide from [Ken Church](#))



Overemphasis on evaluation schemata

Evaluation schemata have reached their limits and are not fully adequate: In many areas we seem to have come *too close to the limits* of an approach to expect *substantial gains* from it. Hence we shouldn't use it alone or as the main approach anymore.

Having noted that our community tends to forget to draw attention to the work done by those on whose shoulders they stand, Y. Wilks (2008) ends his CL lifetime achievement award by writing

« ... we need *real ideas* and *innovations*, and now may be a time for fresh ones. We have stood on the shoulders of Fred Jelinek, Ken Church, and others for nearly two decades now, and the strain is beginning to tell as papers still strive to gain that extra 1% in their scores on some small task. We know that *some change is in the air* and I have tried to hint ... to some of the places where that might be, even if that will mean a *partial return* to older, unfashionable ideas; for *there is nothing new under the sun*. »

And Yorick is not alone

« it is now essential to play down the present **narrowly-defined performance** measures in order to address the **task context**, and specifically the **role** of the **human participant** in the task, so that **new measures**, of larger value, can be developed and applied.’

Spärck Jones K. *Automatic language and information processing : **rethinking evaluation***. Natural Language Engineering, chapter 7, p. 1-18, 2001

So **we need to be alert**. It's not just that we may find ourselves putting the **cart** before the **horse**. We can get obsessed with the *wheels*, and finish up with *uncritically reinvented*, or *square*, or *over-refined* or otherwise *unsatisfactory wheels*, or even just *unicycles*.

Spärck Jones, K. *Computational linguistics: **What about the linguistics?***
Comput. Linguist., 33(3):437–441, 2007.

So **where** should we turn to?

- ⇒ Linguistics?
- ⇒ Psycholinguistics?
- ⇒ Cognitive Science?
- ⇒ Social Sciences?
- ⇒ Information Theory?

Ed's conclusions:

- ⇒ We **don't need any** of these.
- ⇒ We lack vision and theory

Some gut reactions



condescending attitude?

Chinese room?

If you see this shape,

"什麼"

followed by this shape,

"帶來"

followed by this shape,

"快樂"

then produce this shape,

"爲天"

followed by this shape,

"下式".



Linguistics

Ed is *right* to some extent, but possibly for the *wrong* reasons.

He is right

- ➡ Linguists have indeed often dealt with **borderline** cases.
- ➡ Their theories are hardly ever based on **corpora**, i.e. real world data
- ➡ Dealing only with **competency**, linguists have **never** cared to take human constraints into account (short-term memory, etc.). Hence they fail to provide a *psychologically adequate* model.
- ➡ Linguists don't care for the **needs** of other disciplines (problem of metalanguage) and they have often been *syntacto-centric*.

Linguistics

I'd say, he is wrong

- ➡ We do need linguistics, as at least half of the language is rule-based, which makes it learnable within a reasonable time frame. Of course, this is also relevant for CL.
- ➡ They have not addressed but extreme cases, as otherwise they wouldn't have been able to write grammars, which account for regularities of a substantial subset of language
- ➡ There are different kinds of linguistics. Hence the question, what kind of linguistic description do we need (procedural)?
- ➡ Also, there are some linguists who do take psycholinguistic and neurolinguistic factors into account (Jackendoff, Lamb, Hudson), and of course, there are cognitive linguists (Fillmore) and lexicographers, who, working on corpora, do cognitive linguistics (Pustejovsky and Hanks: *Concept Pattern Analysis*).

Psycholinguistics

Indeed, psychology does not have to say much about MT and IR, since these topics have been outside of their scope, but they do have to say quite a bit on sentence processing (alas, much less though, about *discourse*).

Psycholinguistics is definitely useful, especially if you are interested in man-machine communication (support people to process language). In this case it helps a lot to know how people process information, how they represent, store (organisation) and access information (dictionaries)

WN is the brainchild of a psychologist (G.A. Miller). It is clearly based on information concerning the human mind and its functioning (associations), but so is the web (Berner Lee) which is inspired by notions such as association and navigation, the latter corresponding to spreading-activation.

Notions like patterns, omnipresent in CL have long been used in psychology (Gestalt psychology). And what we know about memory is crucial for information processing (why do we chunk graphs into sentence frames?).

Neurosciences and neurolinguistics

If we didn't have a theory, we wouldn't be able to **interpret** properly **patterns** and **correlations** between them.

This information often comes from psychology and linguistics, be it for the different levels or units (meaning-form-sound) or the general architecture and the information flow.

Ed's question: can we infer functional lessons from the brain?

Yes, concerning modularity, **flexibility** of processing, possible recovery despite incomplete information.

Sociolinguistics

Contrary to what one might think, Ed provides more reasons in favor of sociolinguistics than against it, and I couldn't but agree more with him.

It has always struck me that so little attention has been paid by CL to this kind of work. Yet, if we want to deal not only with language, but also with communication (language being only a means), then we have to take a closer look at the context in which communication takes place, in order to determine the different social parameters (place, context, people) affecting the choice of a specific linguistic resource. For example, when shall we highlight or downplay some information (active voice vs. passive voice; independant vs. subordinate clause, etc.), and if so, how is this done in a given language?

What **we** may get from them and what **they** may need take into account?

What **we** can get from them

- ➡ inspiration, no doubt;
- ➡ ways to interpret our data;
- ➡ information allowing us to build our own theory;
- ➡ **functionality**: it would be absurd to simulate on a computer what has been devised for a completely different hardware, with different constraints (paper vs. brain, or, birds vs. airplanes). Nevertheless, at a functional level, there are clearly equivalences (brain, i.e. **neuronal networks** and **semantic networks**, **spreading-activation** and **search/navigation**), since we pursue similar goals.

What **they** may need take into account

- ➡ agree on annotations
- ➡ work on real world data (corpus linguistics)

Mutual benefits, i.e. added value

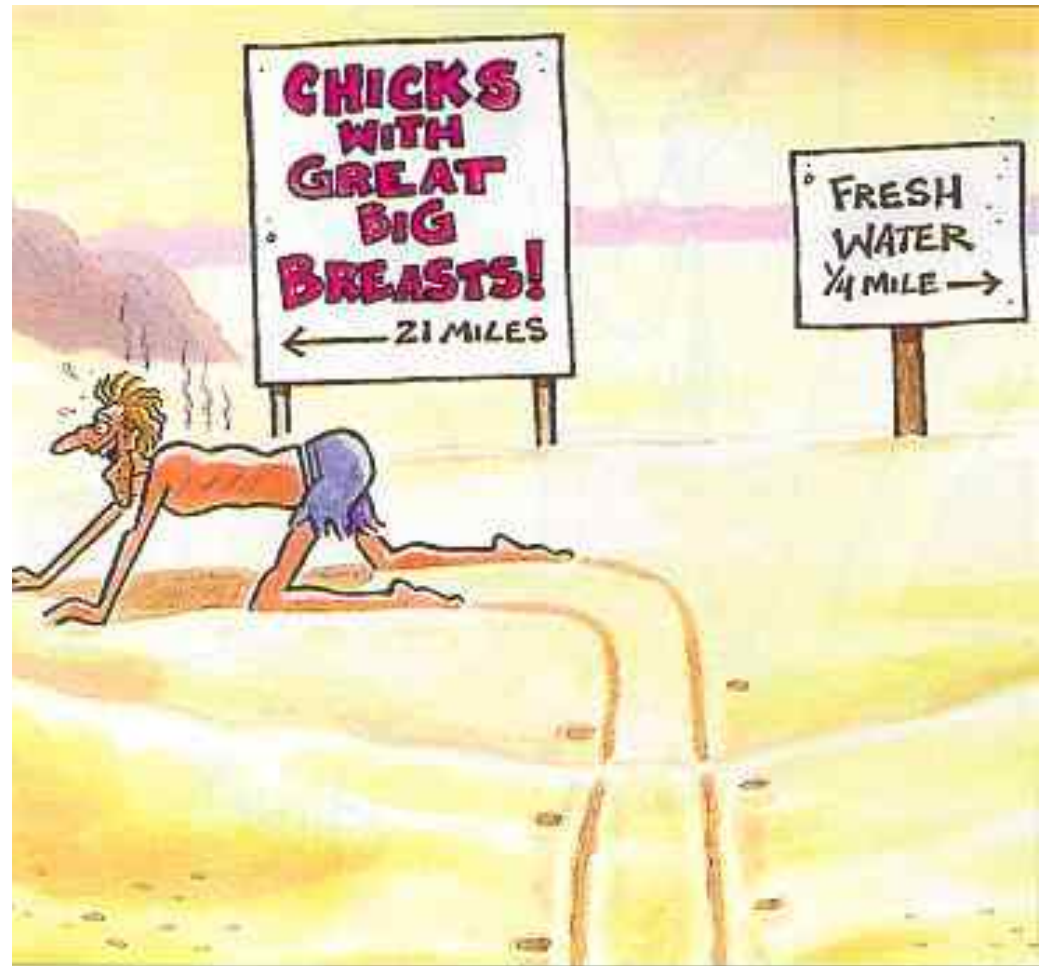
Build truly **semiotic extensions**, i.e. tools supporting people to

- ⇒ **think** (problem solving),
- ⇒ **process language** and
- ⇒ **communicate** (both in L1 and L2).

How come that so few applications have been built to support NLP (authoring tools) and education?

You've always got a choice, but it may be in your own interest to step back and think first about its adequacy, price, and consequences.

- 1° goals, i.e. needs
- 2° methods
- 3° results
- 4° adequacy/evaluation
- 5° price/consequences



Conclusion + possible recommendations

We **do need** the knowledge accumulated in other domains

Some possible **recommendations** of how to take advantage of it

- **broaden** the **range** of **papers** in ACL conferences and Computational Linguistics ;
- put more emphasis on the **poster track** ;
- include a **broader range** of **people** in journal editorial boards and conference PCs ;
- include a **broader range** of **workshops**, **tutorials**, and **invited talks** at ACL conferences;
- have **survey papers** in Computational Linguistics on **other areas** of **language-related** research ;
- offer a **hands-on approach** in smaller events like **summerschools** ;
- organize **real WORKshops**, possibly led by invited speakers ;
- support **brainstorming** by allowing the **presentation** of **ideas** (possibly even if they are **not implemented** or **evaluated**) and work describing **real world applications**.

Thanks for their **inspiration**

- Reiter, E. (2007) *The Shrinking Horizons of Computational Linguistics*. Computational Linguistics Vol. 33, No. 2, Pages 283—287
- Spärck Jones, K. *Computational linguistics: What about the linguistics?* Comput. Linguist., 33(3):437–441, 2007.
- Spärck Jones K. *Automatic language and information processing : rethinking evaluation*. Natural Language Engineering, chapter 7, p. 1-18, 2001
- Wilks, Y. (2008) *On Whose Shoulders?* Computational Linguistics 34 (4) 471-486.
- Wintner, S. (2009), *What Science Underlies Natural Language Engineering?* Computational Linguistics December 2009, Vol. 35, No. 4: 641–644