

**DESI III at ICAIL 2009**



# **Linguistic, Cultural, and Behavioral Dimensions of E-Discovery**

**DESI III Workshop  
Barcelona, Spain**

**David A. Evans  
JustSystems Evans Research, Inc.**

**June 8, 2009**

**JSE-PR-09-02**



# What Do We Do When We Do E-Discovery?



# The Work of (E-)Discovery

---

- **We collect documents and other records**
- **We try to establish for a moment of time (statically) the information that will be potentially relevant, including**
  - **The subject matter / material of the investigation**
  - **The sources (locations, scope, etc.) of the material**
  - **The people (including certain roles) of interest**
- **We analyze the material**
- **We produce lists & items of interest**
- **(We support a legal team...)**

# Sample “Production Request” (TREC)



```
<ProductionRequest>
<RequestNumber>82</RequestNumber>
<RequestText>All documents discussing the color of the paper used to make cigarettes in connection
with increasing sales.</RequestText>
<BooleanQuery>
<FinalQuery>((color! OR shade! OR pastel! OR tint!) w/5 paper) AND (increas! w/15 (sale! OR
sell!))</FinalQuery>
<NegotiationHistory>
<ProposalByDefendant>(color! w/2 paper) AND (increas! w/3 sales)</ProposalByDefendant>
<RejoinderByPlaintiff>(color! OR shade! OR pastel! OR tint!) AND paper AND (sale! OR
sell!)</RejoinderByPlaintiff>
</NegotiationHistory>
</BooleanQuery>
<FinalB>888</FinalB>
<RequestSource>2007-C-4</RequestSource>
<Instruction>
<P>1. These requests require the production of all responsive documents within the sole or joint
possession, custody or control of the Defendants, including their agents, departments, attorneys,
directors, officers, employees, consultants, investigators, insurance companies, or other persons
subject to Defendants' custody or control.</P>
<P>2. All documents that respond, in whole or in part, to any portion of these Requests must be
produced in their entirety, including all attachments and enclosures.</P> ...
```

# The Challenge of Strict Compliance...



## Judgment of Documents

- Relevant
  - Hot
    - ⇒ “Smoking Gun”
  - Highly Relevant
    - ⇒ Deposition
  - Responsive
    - ⇒ Within Scope of Order
- Not Relevant
  - ⇒ Suggests Thoroughness
- Inaccurate
- Unknown / Unavailable

***The Court only wants to see  
Relevant Documents***



“Off with her head!”

Illustration by Charles Robinson from Carroll 1906, Front'spiece.

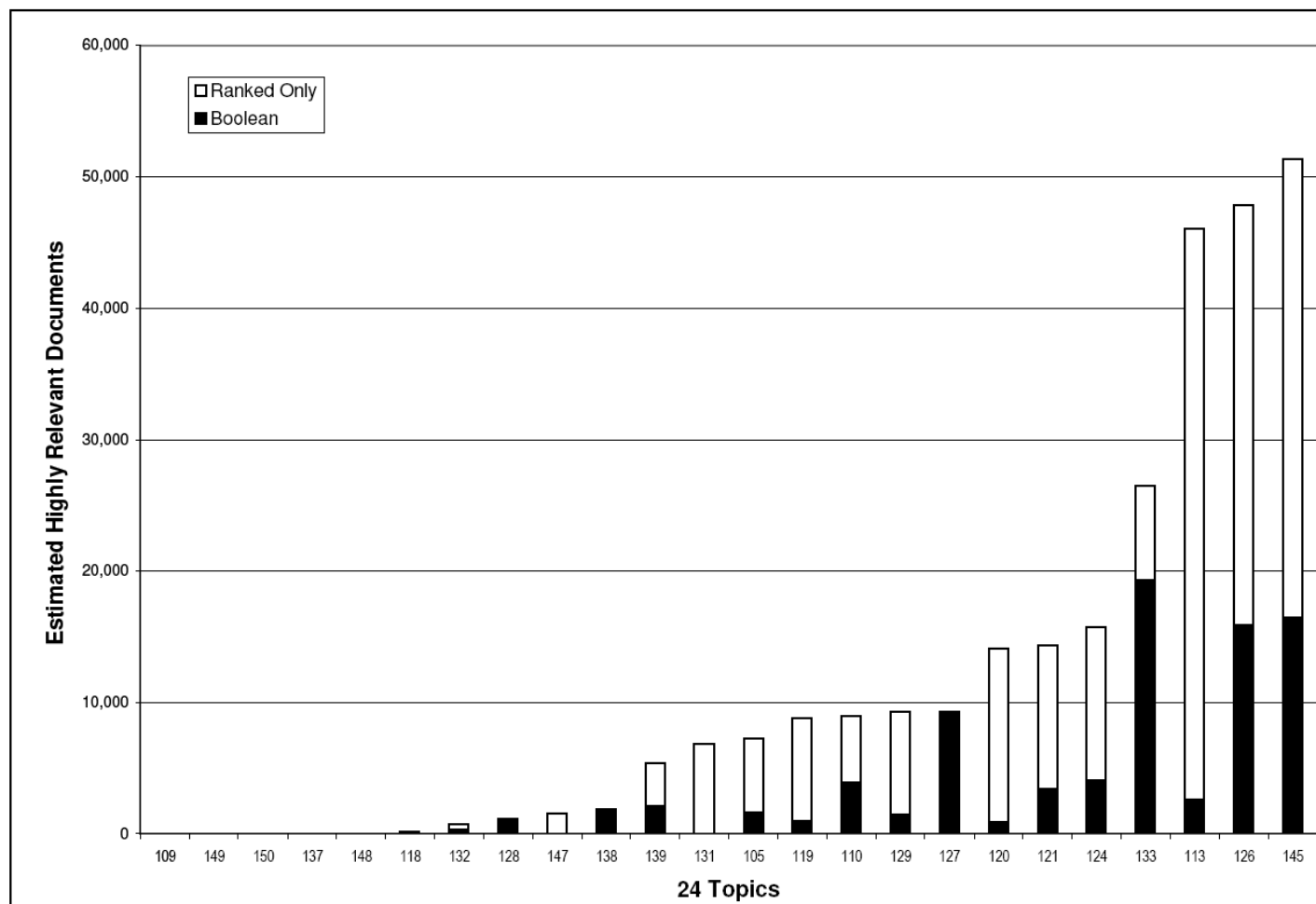


# The Boolean Trade-Off

---

- **Boolean is desirable because of...**
  - Tradition
  - “Precision”
  - **The Ability to Account for the Results**
  
- **And yet...**
  - Performance in Precision  $\Rightarrow$  *Cost of Review*
  - Performance in Recall  $\Rightarrow$  *Cost of Compliance*

# The Boolean Shortfall...



**Highly Relevant Documents not Found by the Consensus Boolean Run**  
**From: Oard et al. 2009**

# Retrieval Results



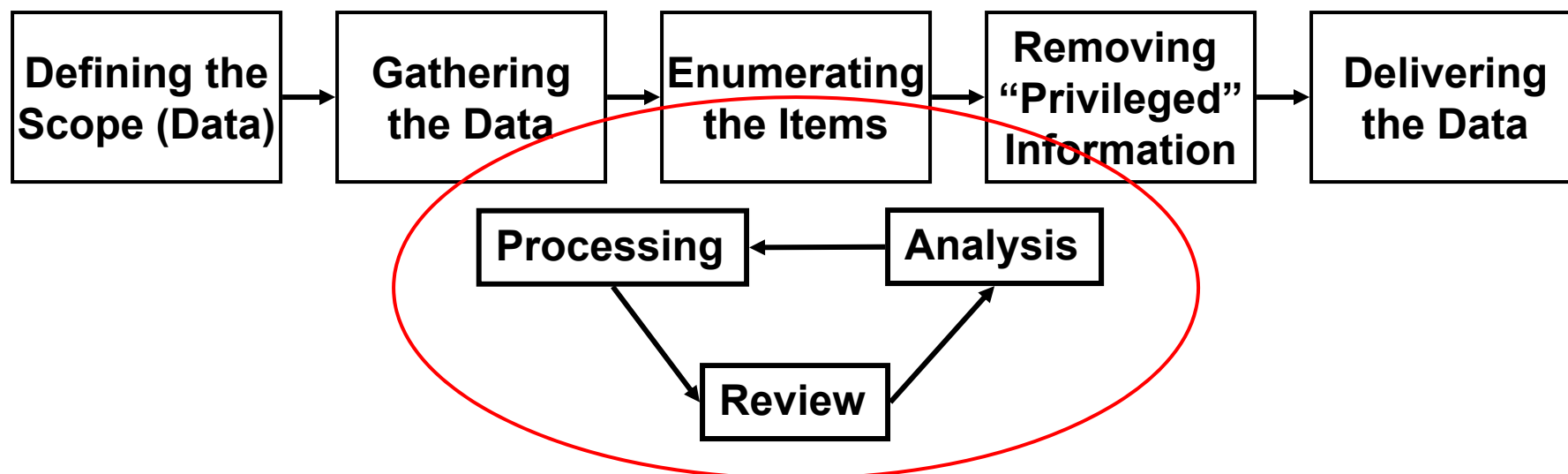
All Relevant (26 topics)	Retrieved	Precision	Recall	$F_1$	$F_1 \approx 2RP / (R+P)$
Defendant	3,180	0.41	0.04	0.05	
Plaintiff	219,606	0.23	0.43	0.19	
Consensus1	93,190	0.24	0.33	0.20	
Final	40,402	0.28	0.24	0.16	
	Avg. K				
Median (23 request runs)	14,363	0.26	0.12	0.10	
Median (41 other runs)	40,402	0.28	0.25	0.16	
<b>Highly Relevant only (24 topics)</b>					
Defendant	3,445	0.14	0.06	0.06	
Plaintiff	234,016	0.08	0.57	0.09	
Consensus1	97,259	0.07	0.42	0.09	
Final	39,930	0.08	0.33	0.09	
	Avg. $K_h$				
Median (23 request runs)	5,838	0.10	0.22	0.05	
Median (41 other runs)	19,965	0.09	0.34	0.08	

**Mean Scores of the Negotiated Boolean Queries and Median Mean Scores of the Participant Runs**  
**From: Oard et al. 2009**

# Ingredients of a Solution



## Process Flow and Techniques...



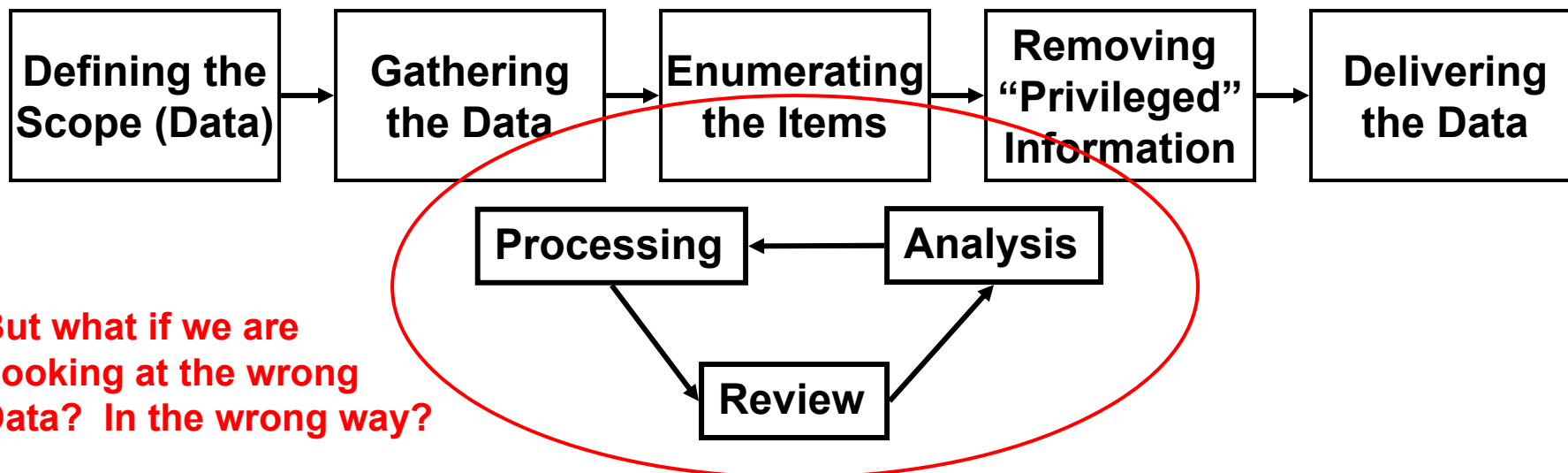
- code normalization
- unzipping compressed data
- language ID
- lexical-atom discovery
- NLP (multi-lingual)
- term EQ-class discovery
- person identification
- indexing (term/feature selection)
- duplicate/near-duplicate ID
- enumeration/individuation
- cross-linking related items
- social network analysis
- clustering (for topic threads)
- filtering
- classification (P/~P)
- topic mapping
- time series analysis
- pseudo-causal modeling

Adapted from: Evans, D.A. "Why E-Discovery is a CIKM-Hard Problem." ACM CIKM 2008.

# Ingredients of a Solution



## Process Flow and Techniques...



**But what if we are Looking at the wrong Data? In the wrong way?**

- code normalization
- unzipping compressed data
- language ID
- lexical-atom discovery
- NLP (multi-lingual)
- term EQ-class discovery
- person identification
- indexing (term/feature selection)
- duplicate/near-duplicate ID
- enumeration/individuation
- cross-linking related items
- social network analysis
- clustering (for topic threads)
- filtering
- classification (P/~P)
- topic mapping
- time series analysis
- pseudo-causal modeling

Adapted from: Evans, D.A. "Why E-Discovery is a CIKM-Hard Problem." ACM CIKM 2008.

# In Practice, Typically...

---



- **Questionable Recall**
- **Poor Enumeration of Items  
(and only that – no other context)**
- **Incomplete Classification of “Privileged”**
- **...**
- **No Attempt to Analyze Foreign-Language  
Material, except quite superficially**



# What Do We Do When We Work?

# When We Work...

---



**Dynamically over time, we**

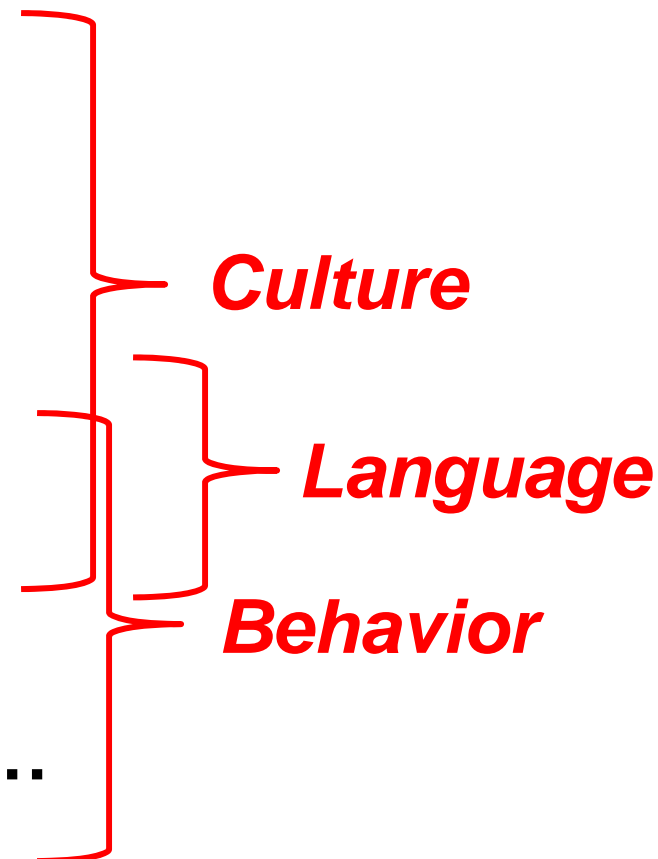
- **Observe**
- **Feel**
- **Learn**
- **Develop Knowledge & Skills**
- **Relate to Coworkers**
- **Interact & Communicate**
- **Cooperate**
- **Produce Documents, Goods, ...**

# When We Work...



**Dynamically over time, we**

- **Observe**
- **Feel**
- **Learn**
- **Develop Knowledge & Skills**
- **Relate to Coworkers**
- **Interact & Communicate**
- **Cooperate**
- **Produce Documents, Goods, ...**





# “Work” & Its Artifacts

**Meetings (Face-to-Face)**

**Videoconferences**

**Teleconferences**

**Presentations / Lectures**

**Conferences**

**Visits & “Occasions”**

**Supervision**

**Collaboration**

**Consultation**

**Conversation**



**Reports**

**Spreadsheets**

**Graphics**

**Design Documents**

**Forms**

**E-Mail**

**Images**

**Timesheets/Logs**

**Text Messages**

# Work Products in Multinational Orgs

---



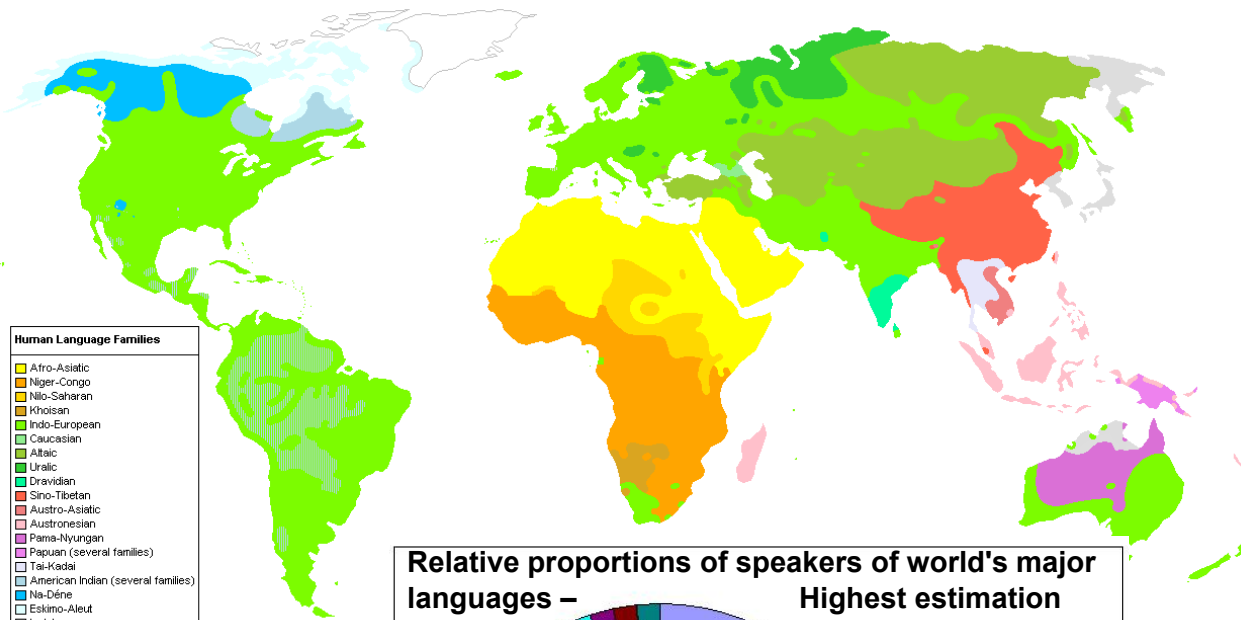
## Problems Legal Teams Have with E-Discovery Material

- **Non-English Texts**
  - “We asked for the <<foreign-language>> stuff, but we could never get anything out of it.”
- **Non-Useful / Unusual Document Types**
  - “All we got was some e-mail and a huge amount of repetitious forms.”
- **Different Boundaries for Trusted Communication**
  - “We tried to keep all the privileged stuff out, but there were copies all over the place.”



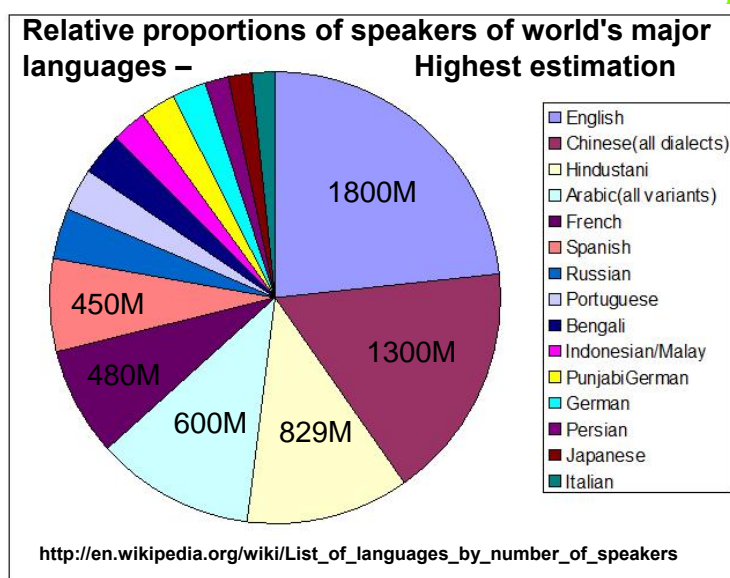
# Linguistic Dimensions

# Multilingual World



- Human Language Families**
- Afro-Asiatic
  - Niger-Congo
  - Nilo-Saharan
  - Khoisan
  - Indo-European
  - Caucasian
  - Altaic
  - Uralic
  - Dravidian
  - Sino-Tibetan
  - Austro-Asiatic
  - Austronesian
  - Pama-Nyungan
  - Papuan (several families)
  - Tai-Kadai
  - American Indian (several families)
  - Na-Déne
  - Eskimo-Aleut
  - Isolate

[http://upload.wikimedia.org/wikipedia/commons/b/b4/Human\\_Language\\_Families\\_%28wikicollors%29.png](http://upload.wikimedia.org/wikipedia/commons/b/b4/Human_Language_Families_%28wikicollors%29.png)



Language	Ethnologue (2005 Estimate)	Encarta Estimate
Mandarin	873,000,000	1,210,000,000
English	508,350,000	341,000,000
Hindi	497,000,000	366,000,000
Spanish	438,300,000	322,200,000
Arabic	206,000,000	246,000,000
Portuguese	177,500,000	176,000,000
Bengali	171,000,000	207,000,000
Russian	145,000,000	167,000,000
Japanese	122,400,000	125,000,000
German	95,400,000	100,100,000
Punjabi	88,000,000	57,000,000
French	78,000,000	78,000,000
Wu	77,200,000	—
Javanese	75,500,000	75,600,000
Tamil	78,000,000	78,000,000
Telugu	74,002,856	74,002,856
Marathi	68,000,000	68,000,000
Vietnamese	67,400,000	68,000,000
Korean	67,000,000	77,000,000
Italian	61,500,000	62,000,000

[http://en.wikipedia.org/wiki/List\\_of\\_languages\\_by\\_number\\_of\\_native\\_speakers](http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers)

# How to Say It in Japanese



## To Say:

言う	いう; ゆう	iu ; yuu
云う	いう; ゆう	iu ; yuu
謂う	いう; ゆう	iu ; yuu
話す	はなす	hanasu
語る	かたる	kataru
述べる	のべる	noberu
漏らす	もらす	mochisu
申します	もうします	mooshimasu
申し上げます	もうしあげます	mooshiagemasu
仰っしゃる	おっしゃる	ossyaru
仰る	おっしゃる	ossyaru
仰しゃる	おっしゃる	ossyaru
仰有る	おっしゃる	ossyaru

**Very polite –  
Shows humility  
in the speaker**

**Very polite –  
exalts the  
addressee or  
subject**

# Sample Japanese Text



確かに、大阪方面の方は、電化製品を購入するとき、よく値引き交渉をしますよね。実際に、大阪では**値引き**交渉がけっこう当たり前！**と**言**っ**てもいいぐらいですが、そのとき店員さんが「わかりました！では、これぐらい、**勉強**させていただきます！」  
...なんて**言**っている風景を見かけることがあります。

それにしても、**値段を負ける**ことを、どうして「**勉強**する」って**言**うんでしょうか？そこで、番組では、この「**勉強**」の語源を調査してみたところ、漢字の**オリジナル**となった中国語に、その秘密がある！ということが判明しました！

そこで今回は、中国唯一の日本語総合月刊誌「人民中国」の東京支局長、林崇珍（りん・すうちん）さんにお**話**を伺いました。

**Q.**「**勉強**」という**言**葉は、古来の中国語では**ど**う**い**う**意**味だったんですか？

**A.**「**勉**」には「**励む**」という**意**味が、現代中国語にも残っていますが、古い中国語では「一生懸命学ぶ、力を出し尽くす」という**意**味です。

**Q.**ちなみに現代の中国語では、「**勉強**」は**ど**う**い**う**意**味なんですか？

**A.**無理に強いる、強制する、という**意**味で使用します。

現代の中国語だと、「**勉強**」は、「無理を強いる」という**意**味なんですね！

同じ「**勉強**」でも、日本語と中国語では、ずいぶん違いがあるものなんですねー！

お父さん、お母さん、先生方もあんまり「**勉強**しなさい！！」なんて**言**ったら可哀想ですよ。

無理強いにならない程度に♪

From: <http://www.c-player.com/ac56998/thread/1100045039501>

# “Discount (A Price)”

---



値段を負ける (ねだんをまける) **nedanokakeru**  
**reduce the price, discount**

値引き (ねびき) **nebiki**  
**discount**

勉強 (べんきょう) **benkyo**  
**study; diligence; discount, reduction**



# “Study”

- 勉強させていただきます  
(benyosaseteitadakimasu)
- You study  
([http://babelfish.yahoo.com/translate\\_txt](http://babelfish.yahoo.com/translate_txt))
- I will study  
(<http://www.excite.co.jp/world/english/>)
- I will study  
(<http://translation.babylon.com/Japanese/to-English>)
- 勉強 させて いただきます  
{ study ; diligence ; **discount** ; **reduction** } + < cause >  
+ < I receive (humbly) >  
= I'll let you have the discount!

## Paragraph 2 – Version 1

---



- **Still, therefore there is the secret for Chinese that became original of the kanji why [I study] will say that I give a it discount on a price when I investigate the etymology of this [study] by the program! ということが became clear! Therefore I heard a story from the Tokyo Branch Chief of the Chinese only Japanese general monthly [People&apos;s China], Takashi Hayashi funny (Rin breathes it heats it) this time.**
- **<http://translation.babylon.com/Japanese/to-English>**

## Paragraph 2 – Version 2

---



- Nevertheless, why “it studies that it is defeated,” the [tsu] [te] it is price, it is probably will be? Then, with program, when it tries investigating the origin of a word of this “study”, in Chinese where Chinese character has become original, there is the secret! With what is said was ascertained. There this time, Tokyo branch manager of the Chinese only Japanese comprehensive monthly periodical “people China”, forest 崇 distinctiveness (while phosphorus doing, is story was asked). ]]>
- <http://babelfish.yahoo.com/translate txt>

## Paragraph 2 – Version 3

---



- The price is said, and however, is it said to be defeated and why is said "Study"? The secret is in the Chinese that became original of the Chinese character in [kode] and the program when the etymology of this "Study" is investigated. It turned out. Then, my did listen to of the story this time to head of the Tokyo branch and [hayashitakashi] [\*\*] (Rin and [uchin] that had it) of only Japanese, integrated monthly magazine "People China" of China.
- <http://www.excite.co.jp/world/english/>

## Paragraph 2 – Version 4

---



- **Still, why would anybody say “study” in the context of giving a discount? Then, on a television program, there was an investigation of the etymology of the word “study” and the secret is in the original Chinese characters that became used in the Chinese language! This was confirmation! Accordingly, I recently had a talk with Mr. Rin Suu-Chan, the Head of the Tokyo Branch Office of “People’s China,” the only Chinese comprehensive monthly magazine in Japanese.**
- **By DAE**

# The Whole Text in Translation



Certainly, the Osaka direction, when purchasing the electrification product, does discount negotiation well, don't you think?. Actually, in Osaka discount negotiation fairly well natural! With it is the extent which it is possible to say, but that time the salesman was recognized “! So, this extent, you study!” ... There are times when the scenery which how is said is happened to see.

Nevertheless, why “it studies that it is defeated,” the [tsu] [te] it is price, it is probably will be? Then, with program, when it tries investigating the origin of a word of this “study”, in Chinese where Chinese character has become original, there is the secret! With what is said was ascertained! Then this time, Tokyo branch manager of the Chinese only Japanese comprehensive monthly periodical “people China”, forest 崇 distinctiveness (while phosphorus doing, is story was asked).

**Q. the word, “study” with Chinese of ancient times in what sense, it is?**

**A.in the meaning working hard” ““of striving”, today remains even in Chinese, but with old Chinese “it learns with utmost effort, in the sense that power is emptied out”.**

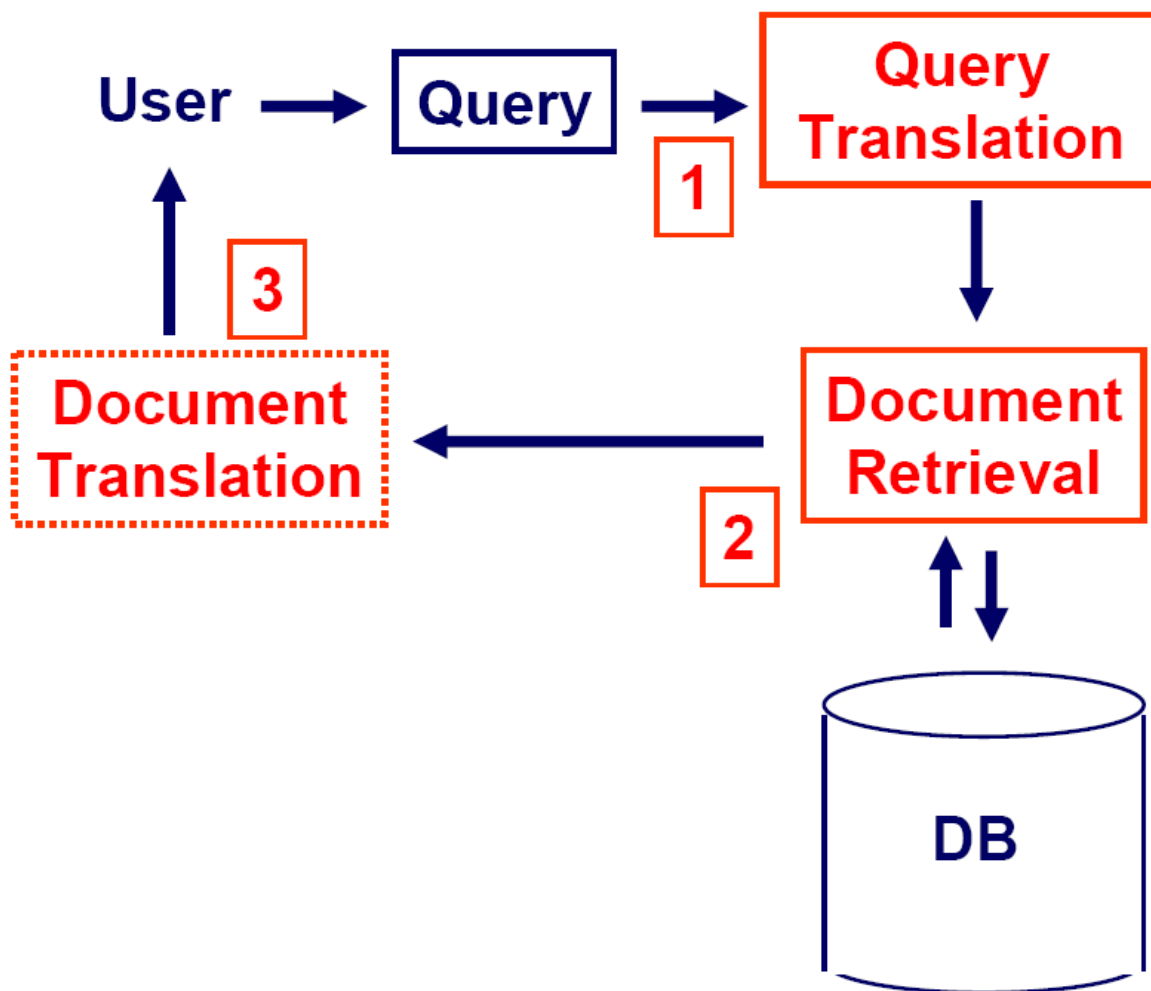
**Q. by the way with Chinese of today, as for “study” in what sense?**

**A. it forces unreasonably, you use in the sense that you force. When it is Chinese of today, with semantic something where “study” “forces the excessiveness”, the shank!**

**Even with the same “study”, with Japanese and Chinese, extremely with thing something which is the difference the shank -! Your father, the mother and the teachers are left over and “study!!”How when you say, it is yes 哀 thought. It does not become in forcing the ♪]]>**

[http://babelfish.yahoo.com/translate\\_txt](http://babelfish.yahoo.com/translate_txt)

# CLIR Functional Architecture

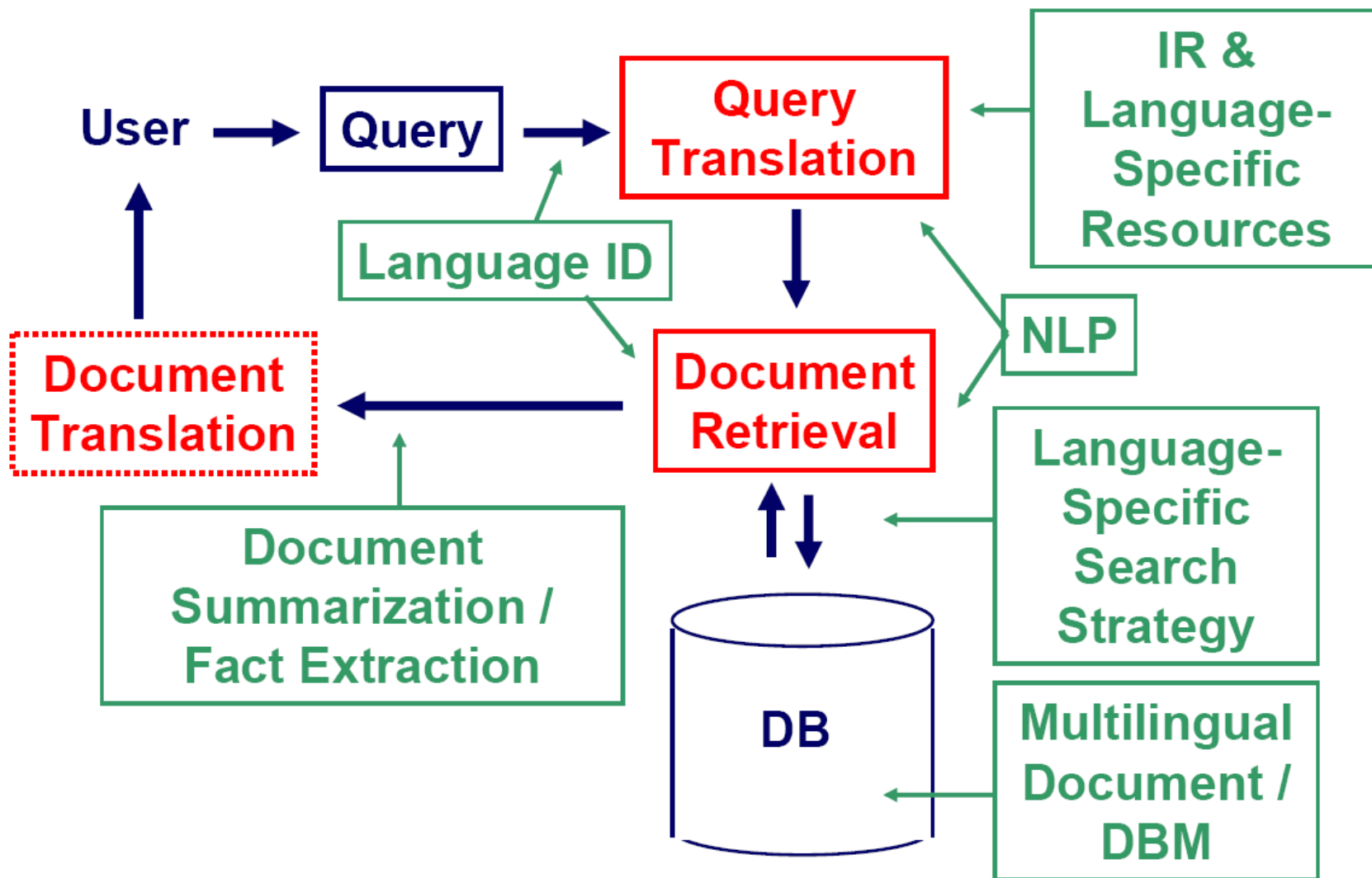


A complete CLIR system will do 1+2+3.

A minimal CLIR system will do 1+2.

Other combinations of functions do not yield CLIR systems.

# CLIR-Focused Architecture



# CLIR Performance



## Ad Hoc: Results



### Comparing bilingual results with monolingual baselines:

- TREC-6, 1997:
  - EN→FR: 49% of best monolingual French system
  - EN→DE: 64% of best monolingual German system
- CLEF 2002:
  - EN→FR: 83,4% of best monolingual French system
  - EN→DE: 85,6% of best monolingual German system
- CLEF 2003 enforced the use of “unusual” language pairs:
  - IT→ES: 83% of best monolingual Spanish IR system
  - DE→IT: 87% of best monolingual Italian IR system
  - FR→NL: 82% of best monolingual Dutch IR system
- CLEF2005 :
  - X -> FR: 85% of best monolingual French IR system
  - X -> PT: 88% of best monolingual Portuguese IR system
  - X -> BG: 74% of best monolingual Bulgarian IR system
  - X -> HU: 73% of best monolingual Hungarian IR system

Figures for FR and PT reflect state-of-the-art  
Room for improvement for “new” languages

From: Peters, C. & Ferro, N.  
“From CLEF to TrebleCLEF: the  
Evolution of the Cross-Language  
Evaluation Forum.”  
NTCIR-7 Meeting, 2008.

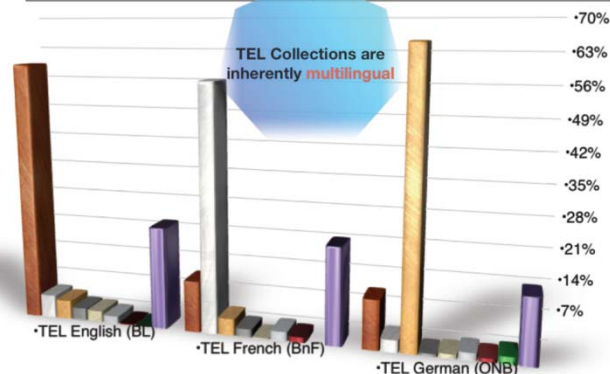
NTCIR-7 Meeting  
Tokyo, 16-19 December, 2008

# CLIR Performance



## TEL Collections: Distribution of the Languages

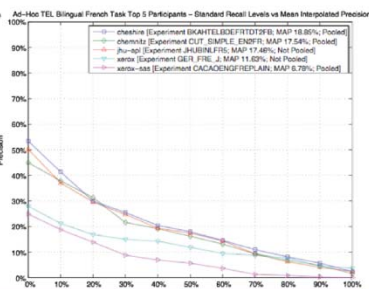
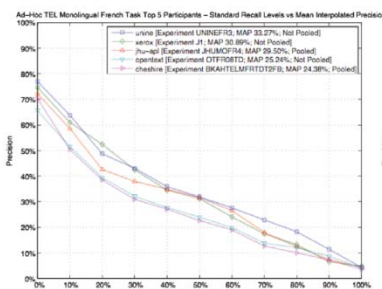
- English
- French
- German
- Spanish
- Russian
- Italian
- Latin
- Esperanto
- Other



NTCIR-7 Meeting  
Tokyo, 16-19 December, 2008

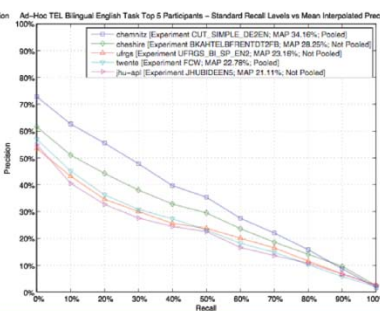
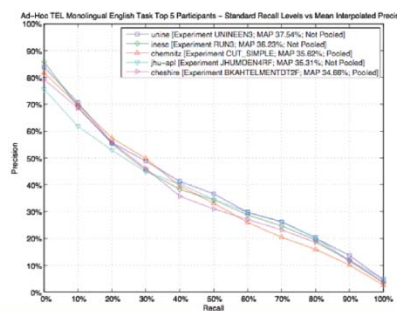
## TEL French

Bilingual is  
57%  
of monlingual



## TEL English

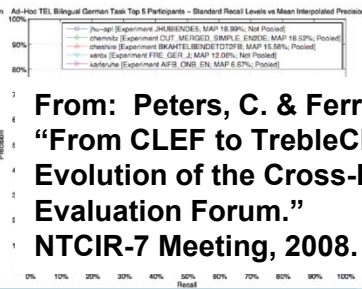
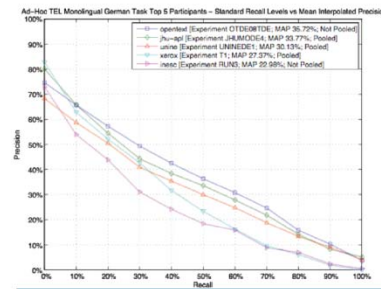
Bilingual is  
91%  
of monlingual



NTCIR-7 Meeting  
Tokyo, 16-19 December, 2008

## TEL German

Bilingual is  
53%  
of monlingual

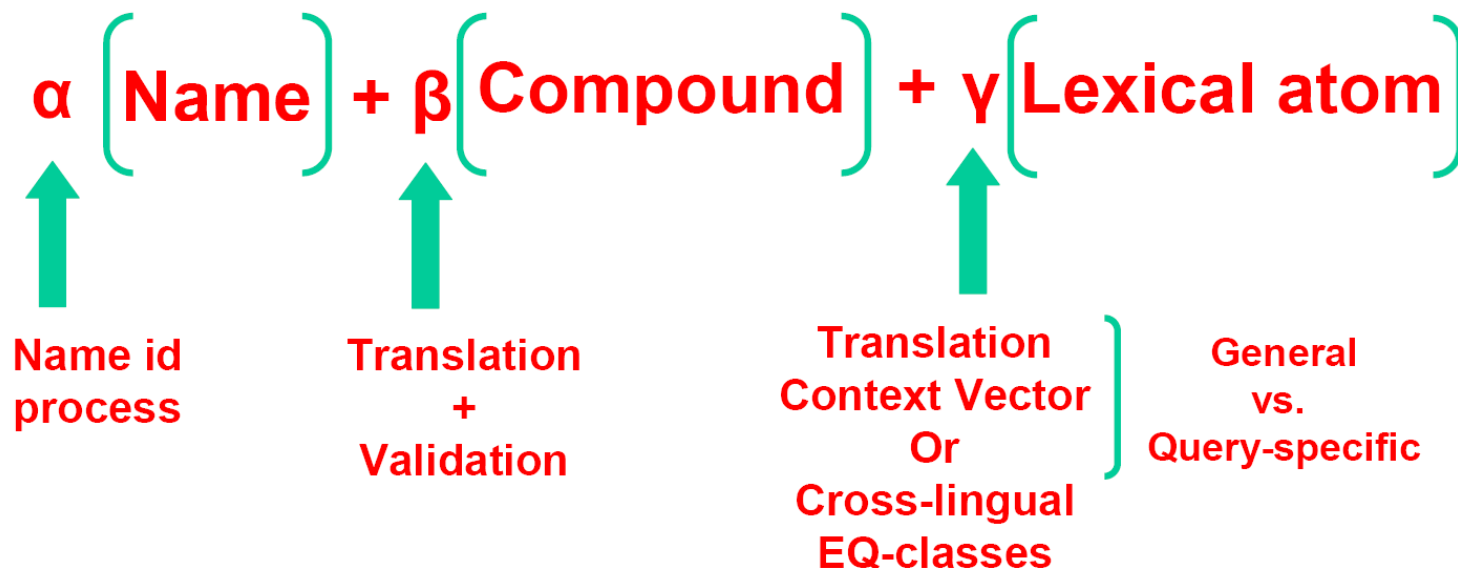


From: Peters, C. & Ferro, N.  
“From CLEF to TrebleCLEF: the Evolution of the Cross-Language Evaluation Forum.”  
NTCIR-7 Meeting, 2008.

# Special, Persistent CLIR Challenges



- Names
- Multi-word terms
- General unknown words (lexical atoms)



# One Example of the Problem



## Finding Translation of “White House”

White:

- 白
- 白色
- 纨
- 皎
- 皓
- 皙
- 皤



House:

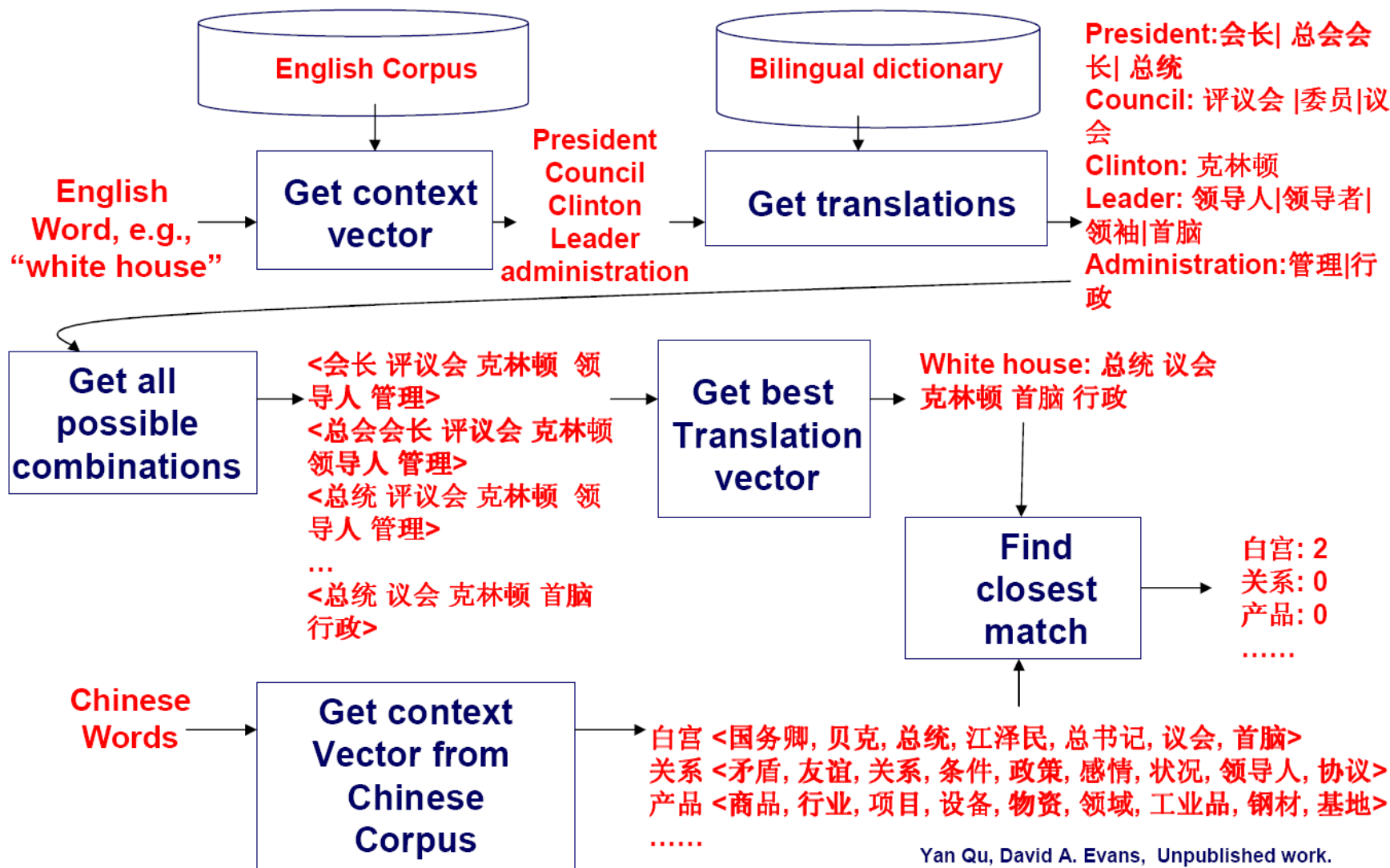
- 房
- 房屋
- 房子
- 馆
- 屋
- 屋子

Possible Translations for White House

- 白房
- 白房屋
- 白房子
- 白馆
- 白屋
- 白屋子
- 白色房
- 白色房屋
- 白色房子
- 白色馆
- 白色屋
- 白色屋子
- 纨房
- .....

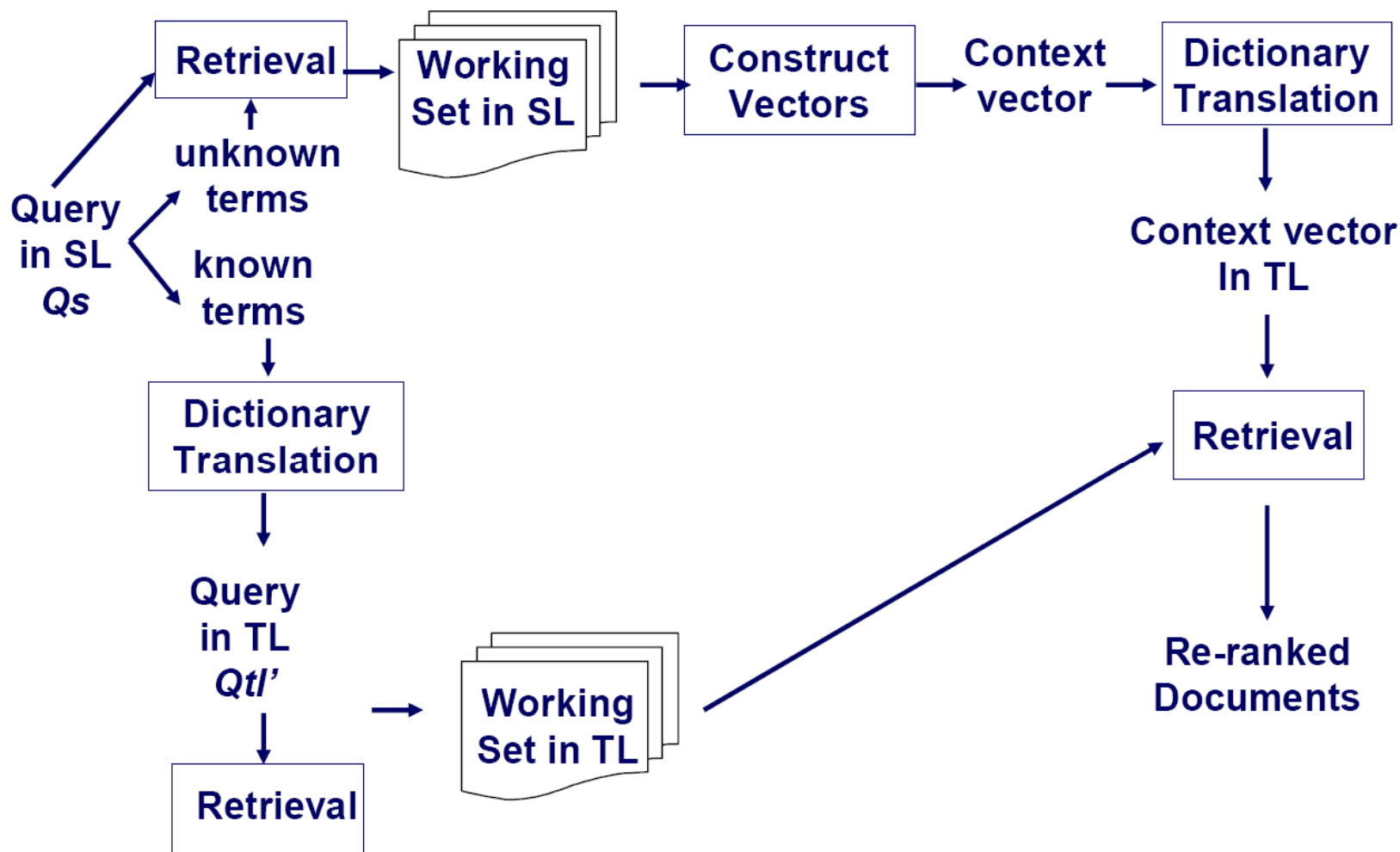
**The correct translation 白宫 cannot be constructed based on simple glossing.**

# Context Vector Approach – General



Yan Qu, David A. Evans, Unpublished work.

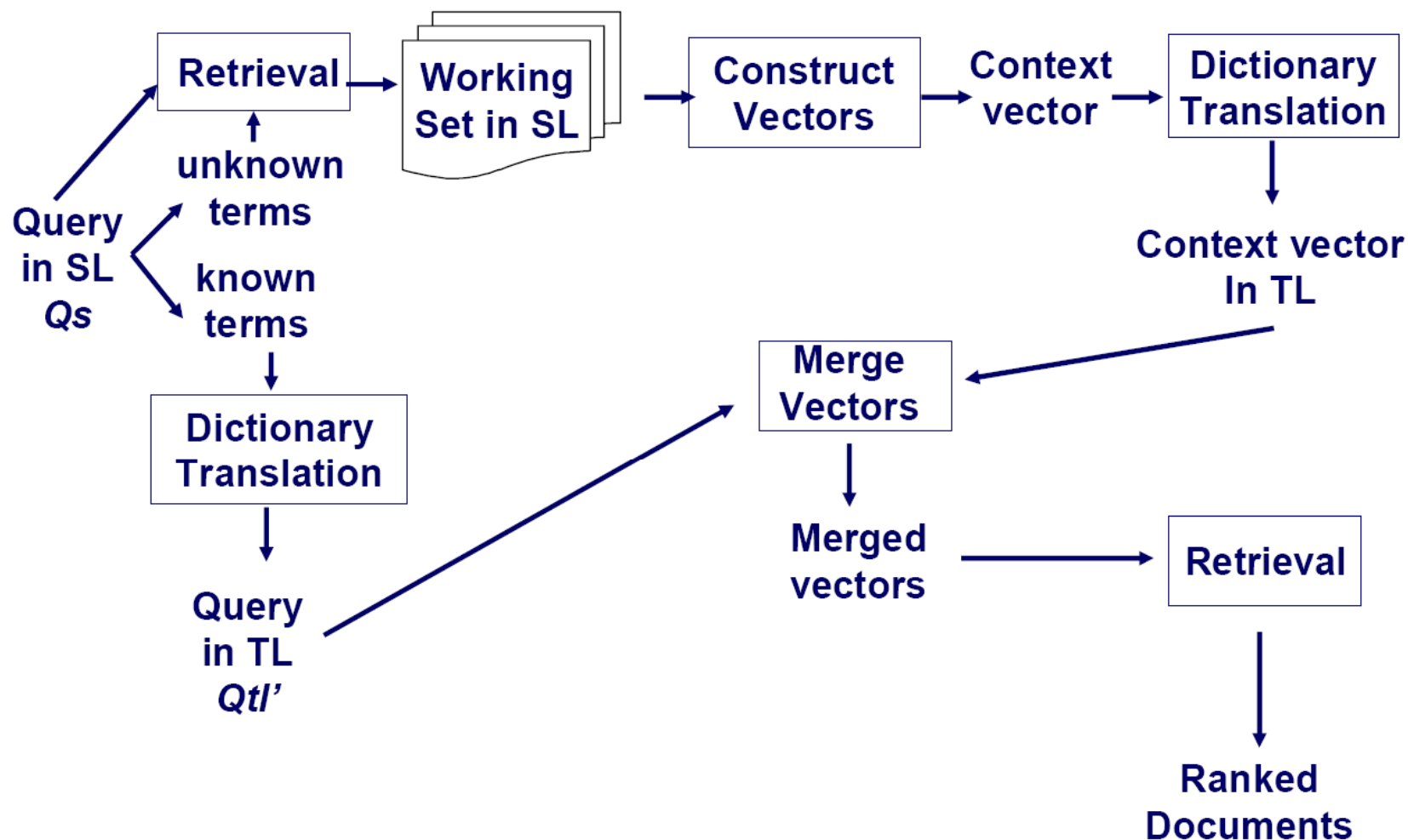
# Query-Dependent Context Vector (1)



SL: source language TL: target language

Yan Qu, David A. Evans, Unpublished work.

# Query-Dependent Context Vector (2)



SL: source language TL: target language

Yan Qu, David A. Evans, Unpublished work.



# Observations

---

- **This is not Boolean search!!**
- **The resources required to do a credible job include (at least) gloss translation lexicons, lists of the multi-word terms of importance in the domain, proper names (including people) of relevance, ... but these are insufficient!**
- **Reference, comparable corpora in the source and target languages are also needed.**



# Cultural & Behavioral Dimensions



# Culture & Behavior in Interactions

---

- **In general...**
  - We internalize culture
  - We externalize behavior
- **Culture is the “software” that encodes our sensibilities, values, expectations, default orientation and attitudes towards people in degrees of relationship to us, ...**
- **Behavior is the “application” running in social contexts, through which we manifest degrees of trust, sharing, following, leading, accepting responsibility, ...**

# Characterization of Cultural Dimensions



## Geert Hofstede, *Cultures and Organizations*, 1991

- **Power Distance**

- How a society handles inequalities – “the extent to which the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally.”

- **Individualism / Collectivism**

- Behavior towards the group – “Individualism pertains to societies in which the ties between individuals are loose: everyone is expected to look after himself or herself and his or her immediate family. Collectivism as its opposite pertains to societies in which people from birth onwards are integrated into strong, cohesive in groups, which throughout people's lifetime continue to protect them in exchange for unquestioning loyalty.”

- **Masculinity / Femininity**

- Behavior according to gender – “Masculinity pertains to societies in which social gender roles are clearly distinct; femininity pertains to societies in which social gender roles overlap.”

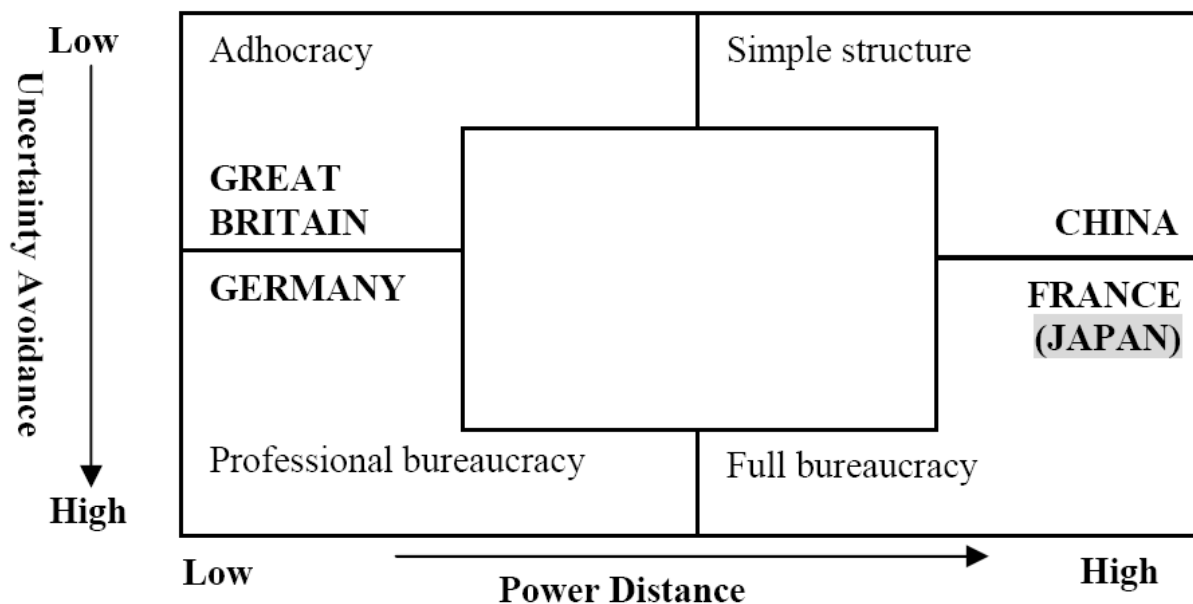
- **Uncertainty Avoidance**

- The need for Structure – “the extent to which the members of a culture feel threatened by uncertain or unknown situations.”

- **Long-Term Orientation**

- “Long-term orientation” includes values such as thrift and perseverance. “Short-term orientation” includes respect for precedence and tradition, fulfilling social obligations, saving “face.”

# Example Differences by Country



**Figure 10:** Preferred Coordination Mechanism (Adapted from Hofstede, 1991, p.152)

**Note:** This figure illustrates the typical organization structure predicted by “power distance index” and “uncertainty avoidance index.” “Power distance index” refers to the extent to which the less powerful members of organizations and institutions accept and expect that power is distributed unequally. “Uncertainty avoidance index” indicates the extent to which a culture programs its members to feel comfortable in unstructured situations such as unknown, surprising, and different from the usual. Uncertainty-avoiding cultures try to minimize the possibility of such situations by using strict laws and rules, and safety and security measures

**From Horii, Jin & Levitt 2005b**

# Parameterizing Work-Group Cultures



		<b>Culture A (American)</b>	<b>Culture J (Japanese)</b>
<b>Practices</b>	Centralization	Decentralized authority	Centralized authority
	Formalization	Medium level of formalization	High level of formalization
	Org. hierarchy	Flatter hierarchy	Multiple Hierarchy
<b>Values</b>	Decision Making	Individual decision making	Consensual decision making
	Communication	Individually-based	Group-based

**Figure 1: Summary of Cultural Differences**

**From Horii, Jin & Levitt 2004**

# Simulating Project Work

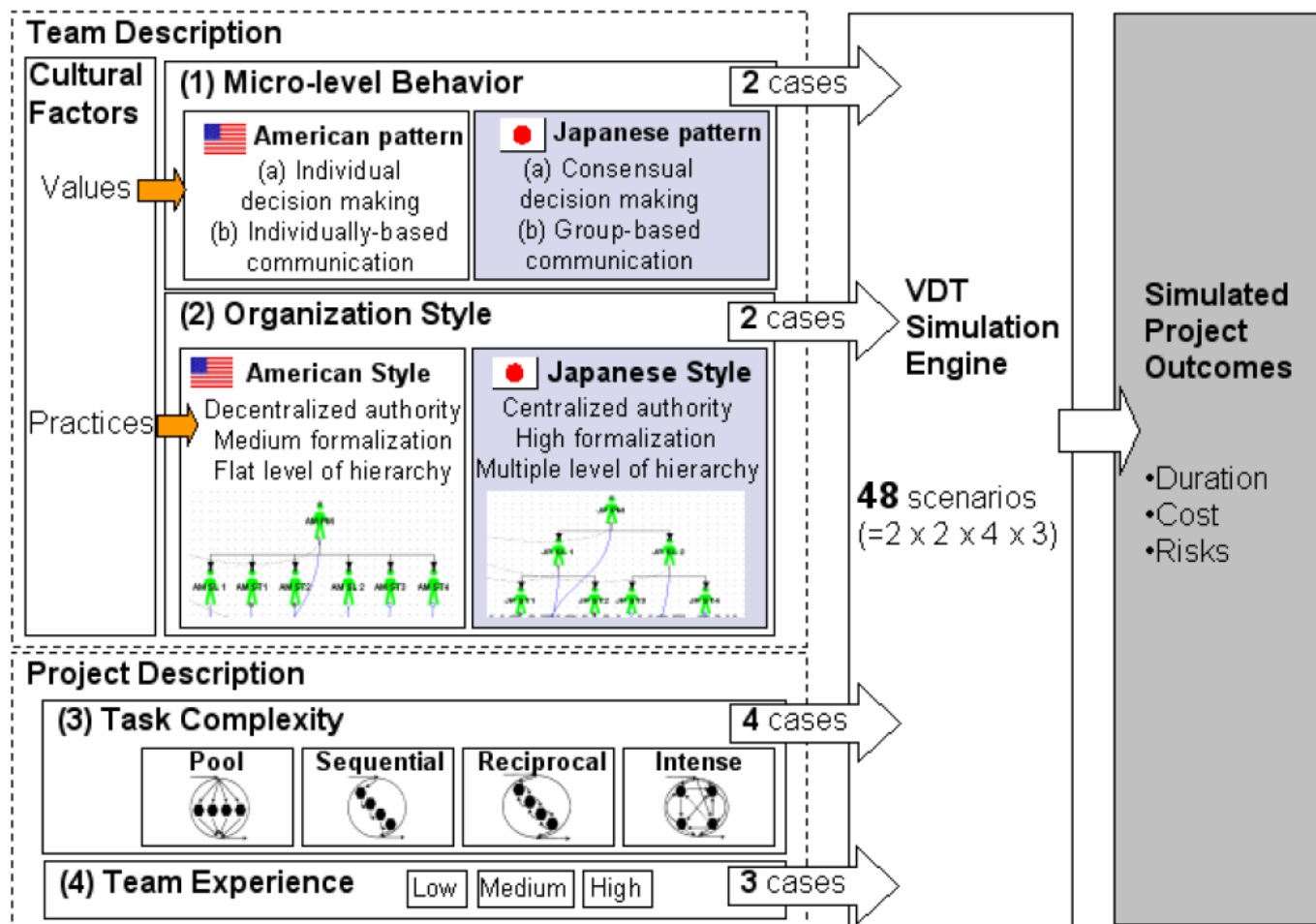
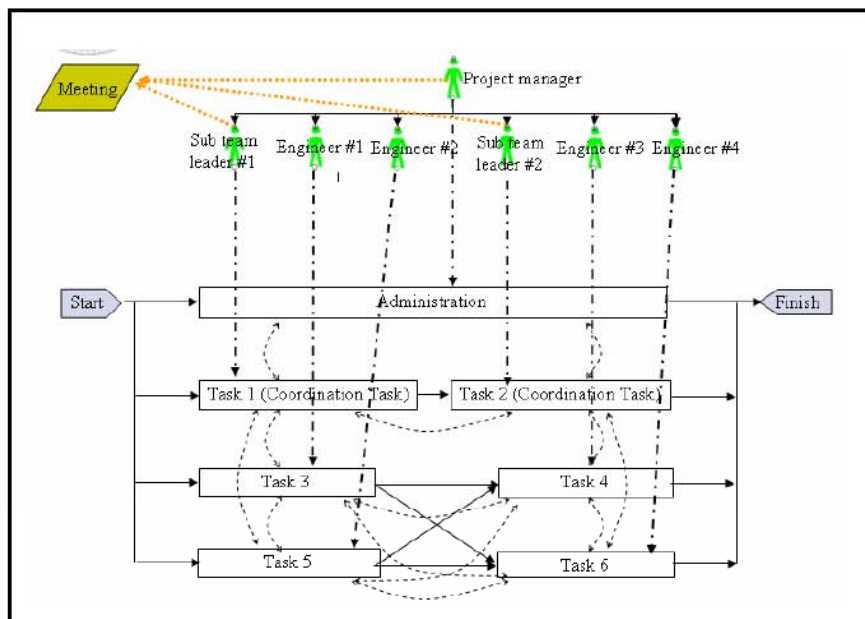


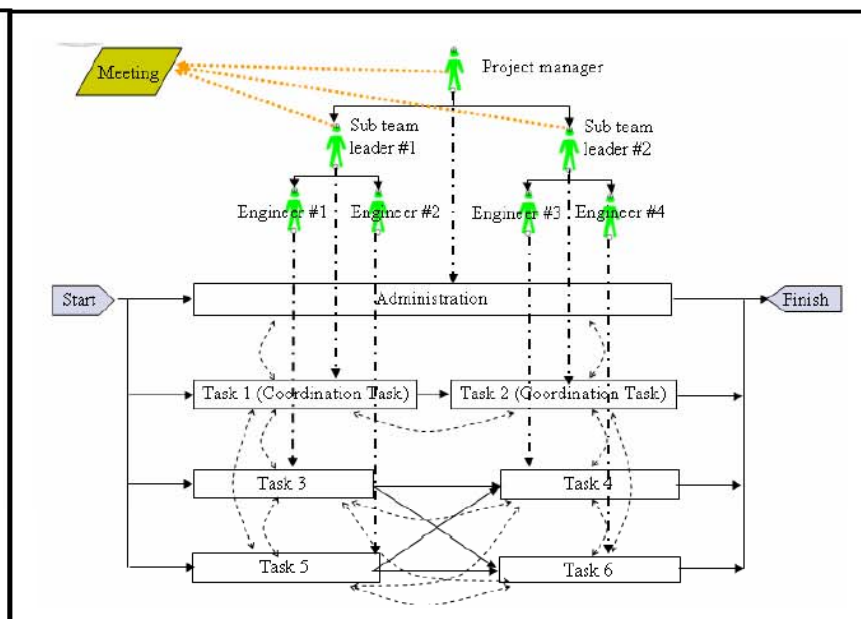
Figure 1: Modeling Framework

From Horii, Jin & Levitt 2005b

# Modeling Cultural Differences in Teams



**Figure 2:** Example of American Organization Structure Type with Intense Complexity



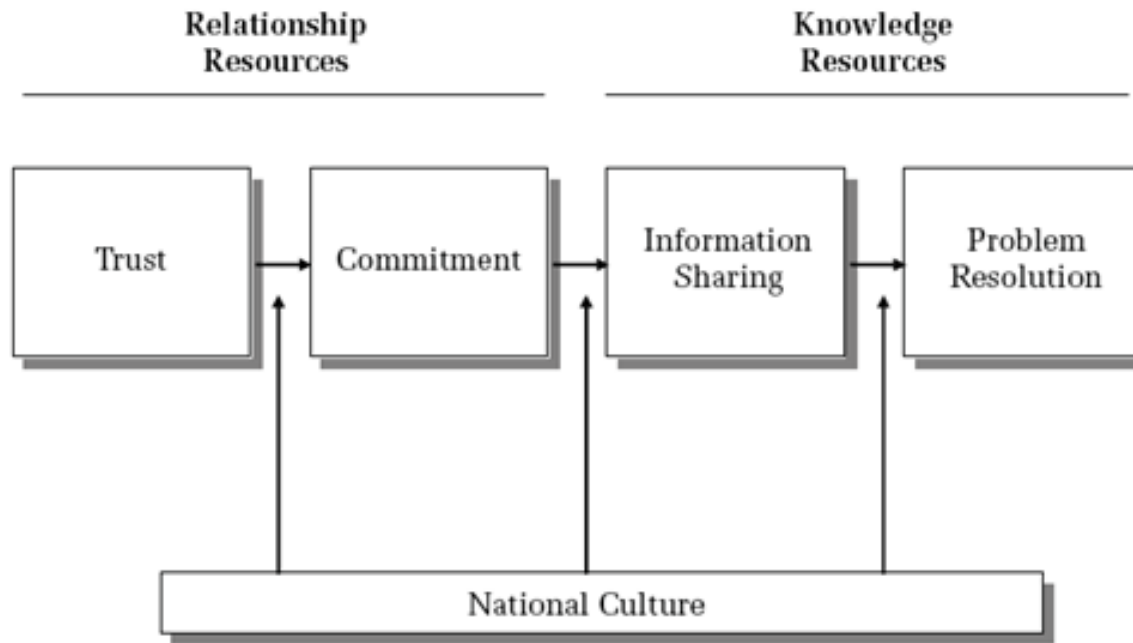
**Figure 3:** Example of Japanese Organization Structure Type with Intense Complexity

**Note:** These figures illustrate examples of the *intense* coordination complexity cases. As shown in Figures 2 and 3, both organizations have the exactly same workflow and required work volume as the intense complexity cases. All teams are composed of seven members, including one project manager, two sub-team leaders, and four sub-team members. We change only either structure types or micro-level behavior patterns actors possess.

- ▶ shows precedence links among tasks
- .....▶ shows rework and communication links among tasks
- - -▶ shows work assignment between team members and tasks

**From Horii, Jin & Levitt 2005b**

# Other Cultural Influences on Behavior



**Associated factors:**

- **Societal Monitoring**
- **Social Exchange**
- **Individual Responsibility**

From: Griffith, Myers & Harvey 2006

# Observed Differences in Behavior



H <sub>1</sub> : Japanese firms perceive a stronger positive association between trust and commitment in intracultural, interorganizational relationships than U.S. firms.
H <sub>2</sub> : Japanese firms perceive a stronger positive association between trust and commitment in intercultural, interorganizational relationships with U.S. firms than U.S. firms perceive with Japanese firms.
H <sub>3</sub> : Japanese firms exhibit a stronger positive association between commitment and information sharing in intracultural, interorganizational relationships than U.S. firms.
H <sub>4</sub> : Japanese firms exhibit a stronger positive association between commitment and information sharing in intercultural, interorganizational relationships with U.S. firms than U.S. firms exhibit with Japanese firms.
H <sub>5</sub> : Japanese firms exhibit a weaker positive association between information sharing and problem resolution in intracultural, interorganizational relationships than U.S. firms.
H <sub>6</sub> : Japanese firms exhibit a weaker positive association between information sharing and problem resolution in intercultural, interorganizational relationships with U.S. firms than U.S. firms exhibit with Japanese firms.

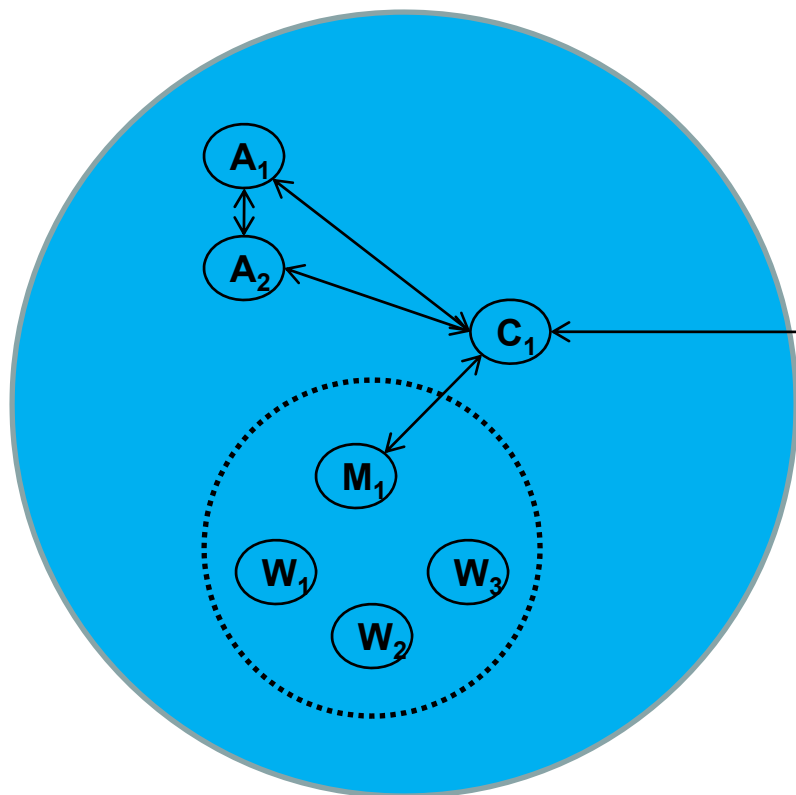
Hypotheses	Japan	United States	Assessment (p < .05)
H <sub>1</sub> : Trust → commitment (intracultural, interorganizational)	.339 t = 3.39**	.042 t = 2.36*	Supported
H <sub>2</sub> : Trust → commitment (intercultural, interorganizational)	.404 t = 3.54**	.018 t = .90	Supported
H <sub>3</sub> : Commitment → information sharing (intracultural, interorganizational)	.592 t = 4.78**	.184 t = 2.90*	Supported
H <sub>4</sub> : Commitment → information sharing (intercultural, interorganizational)	.611 t = 5.33**	.102 t = 2.14*	Supported
H <sub>5</sub> : Information sharing → problem resolution (intracultural, interorganizational)	.855 t = 4.92**	.851 t = 5.43**	n.s.
H <sub>6</sub> : Information sharing → problem resolution (intercultural, interorganizational)	.743 t = 4.71**	.980 t = 5.00**	n.s.
<b>Stacked Model Assessment</b>			
Intracultural: $\chi^2 = 515.74$ , d.f. = 174; NFI = .95; CFI = .97.			
Intercultural: $\chi^2 = 418.39$ , d.f. = 174; NFI = .96; CFI = .97.			
*p < .05.			
**p < .01.			
Notes: We report path coefficients and t-values. n.s. = not significant.			

Adapted from: Griffith, Myers & Harvey 2006

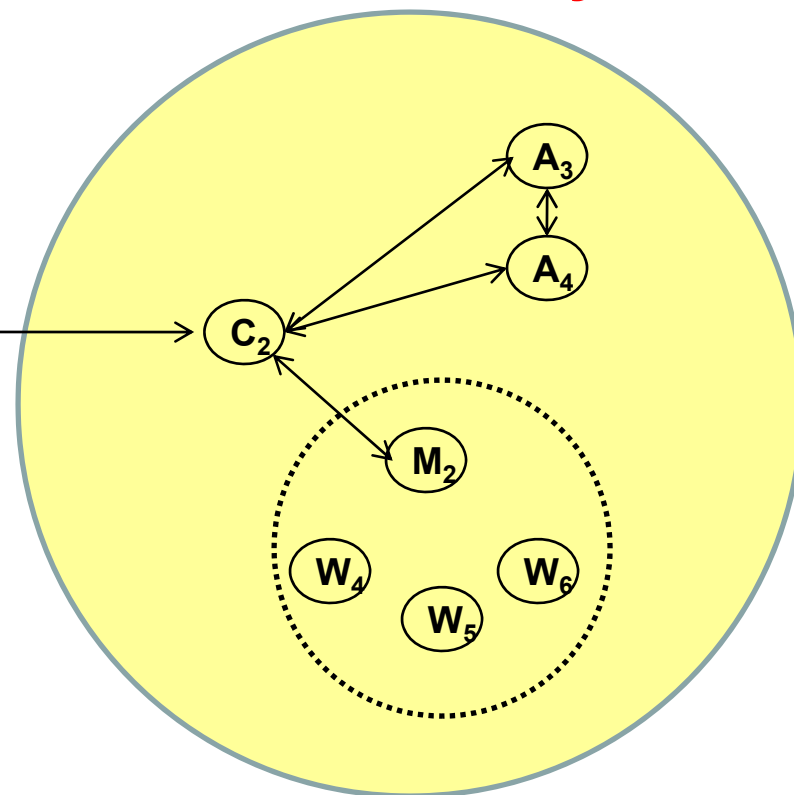
# Multinational Organization (1)



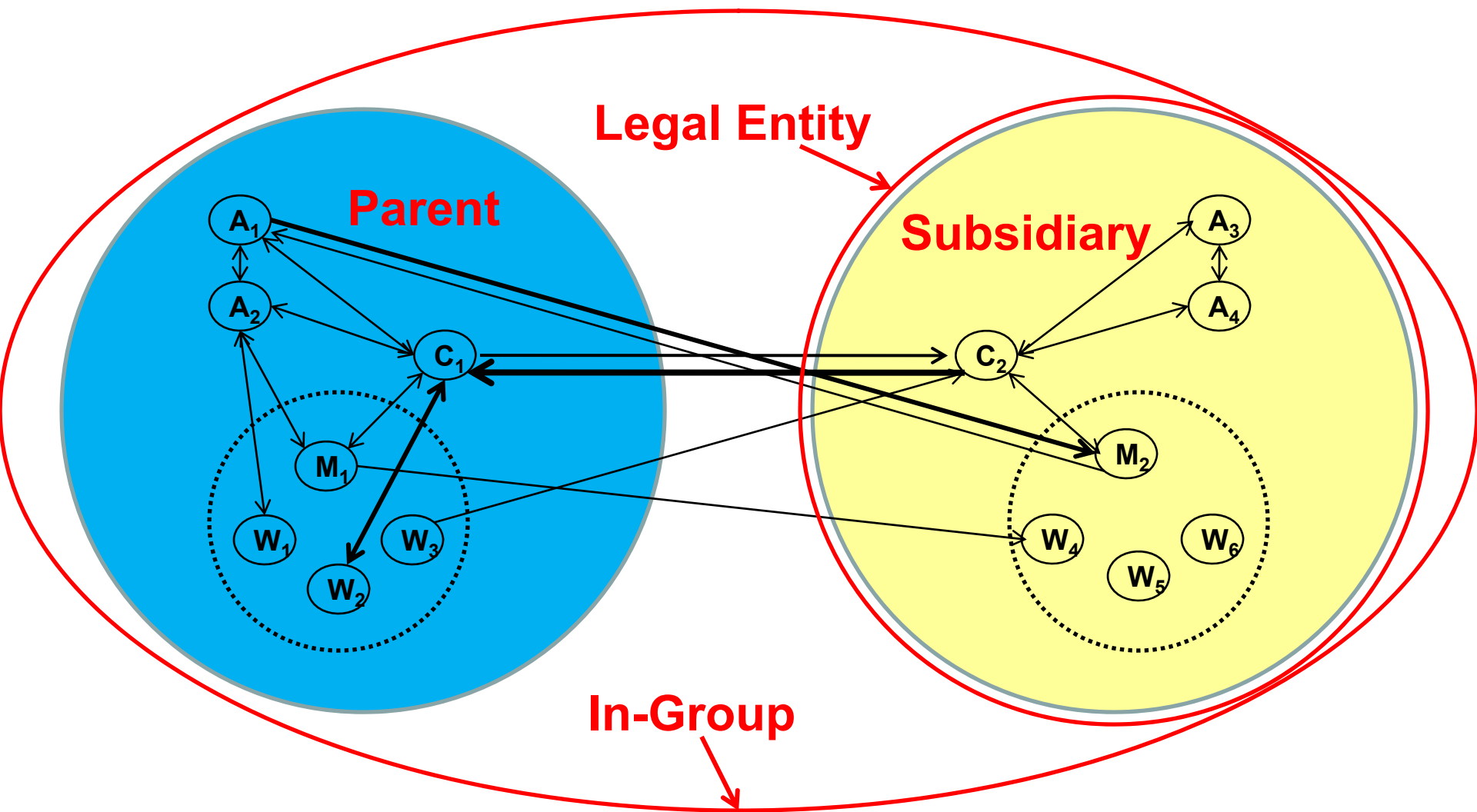
**Parent**



**Subsidiary**



# Multinational Organization (2)



# Examples of Observed Differences



## Work & Artifacts

- **Face-to-Face Meetings**      **Japan > U.S.**
- **Video/Teleconferences**      **Japan > U.S.**
- **Tape Recordings**      **Japan > U.S.**
- **Reports**      **Japan < U.S.**
- **Slides/Presentations**      **Japan > U.S.**
- **E-Mail**      **Japan < U.S.**
- **Hand-Written Notes**      **Japan > U.S.**
- **Forms**      **Japan >> U.S.**
- **Text Messages**      **Japan < U.S.**
- **Spreadsheets**      **Japan > U.S.**

# Examples of Observed Differences

---



## Sources & Information

- **Sharing of Information\*** Japan > U.S.
- **Cross-Group Knowledge** Japan > U.S.
- **Legal Boundaries** Japan << U.S.
- **Repetition / Redundancy** Japan > U.S.



# Observations

---

- **Culture Affects Work and Work Products**
- **Search Strategies May Need to Vary According to Record Type – High- vs. Low-Densities of Information**
- **Organizational Titles May Not Reflect Roles; Roles May Not Reflect Expertise or Knowledge**
- **Boundaries May Not Be Visible or Effective**



# Summary



# Lessons Learned

---

## Foreign-Language Information Processing

- **We need to use realistic approaches (as developed in advanced CLIR systems), especially to accommodate uncertainty in translation.**
- **The discovery request cannot be effectively stated in a Boolean expression.**
- **Legal teams and officers of the court should be educated about the special needs in CLIR.**



# Lessons Learned

---

## Culture & Behavior

- **We should change search strategies to mirror the cultural types and patterns of document generation.**
- **Weighting of “terms” should reflect the expected information-density of the target documents.**
- **We should anticipate “unusual” modes and patterns of information sharing, including the crossing of “legal” boundaries, when looking for sources.**



# Summary Thoughts

---

- **Languages are surprisingly “different” (!)**
- **Culture affects behavior; in work groups, this may be reflected in different patterns of communication and different types of “document” artifacts.**
- **E-Discovery practices should take account of the linguistic and cultural-behavioral contexts of companies and individual workers.**



# References

- Carroll, Lewis. *Alice's Adventures in Wonderland*. New York: Cassell and Company, Ltd. 1906.
- Griffith, David A.; Myers, Matthew B.; Harvey, Michael G. An Investigation of National Culture's Influence on Relationship and Knowledge Resources in Interorganizational Relationships Between Japan and the United States. *Journal of International Marketing*, Vol. 14, No. 2, 2006. pp. 1-32.
- Hofstede, G. (1991). *Culture and Organizations: Software of the Mind, Intercultural Cooperation and its Importance for Survival*. McGraw-Hill, New York.
- Horii, T.; Jin, Y.; Levitt, R.E. Modeling and Analyzing Cultural Influences on Team Performance through Virtual Experiments. *Proceeding of the NAACSOS Conference 2004*, Pittsburgh, PA.
- Horii, T.; Jin, Y.; Levitt, R.E. Modeling and Analyzing Cultural Influences on Project Team Performance. *Comput. Math. Organ. Theory* 10, 4 (Jan. 2005), 305-321. [2005a]  
 DOI= <http://dx.doi.org/10.1007/s10588-005-6283-1>
- Horii, T.; Jin, Y.; Levitt, R.E. Impact of Multiple Normative Systems on Organization Performance of International Joint Ventures. Symposium on Normative Multi-Agent Systems, NORMAS 2005, University of Hertfordshire, Hatfield, UK, 12-15 April 2005, 54-64. [2005b] URL: <http://dblp.uni-trier.de/db/conf/normas/normas2005.html#HoriiJL05>
- Najork, Marc; Zaragosa, Hugo; Taylor, Michael. HITS on the Web: How does it Compare? *SIGIR 2007 Proceedings*. ACM Press. 2007. 471-478.
- Oard, Douglas W.; Hedin, Bruce; Tomlinson, Stephen; Baron, Jason R. Overview of the TREC 2008 Legal Track. In E.M. Voorhees and L. Bucklin (Editors), *The Seventeenth Text Retrieval Conference Proceedings (TREC 2008)*. NIST Special Publication 500-277. 2009.  
 (Cf. [http://trec.nist.gov/pubs/trec17/t17\\_proceedings.html](http://trec.nist.gov/pubs/trec17/t17_proceedings.html))



**The End**  
*Thanks!*



# Appendix

## Notes on Najork et al. 2007

# A Modern Evaluation of Web Search

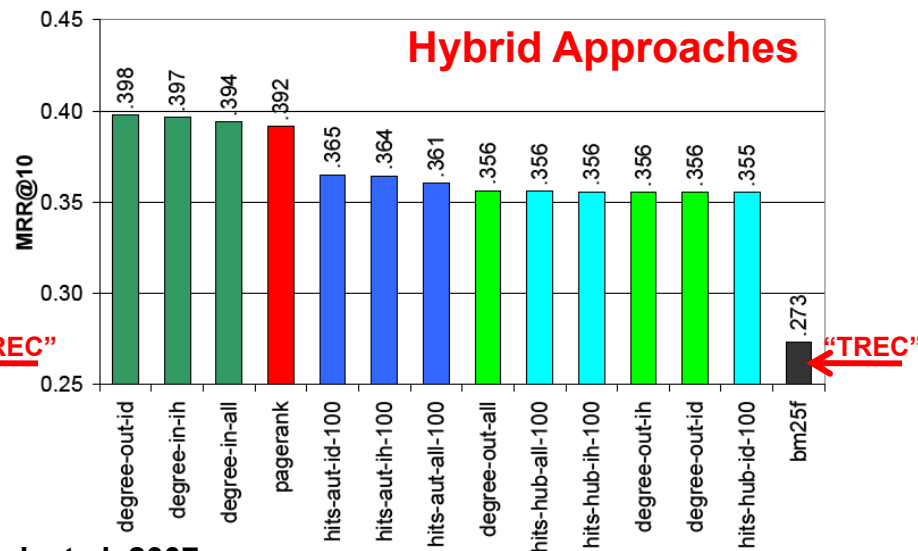
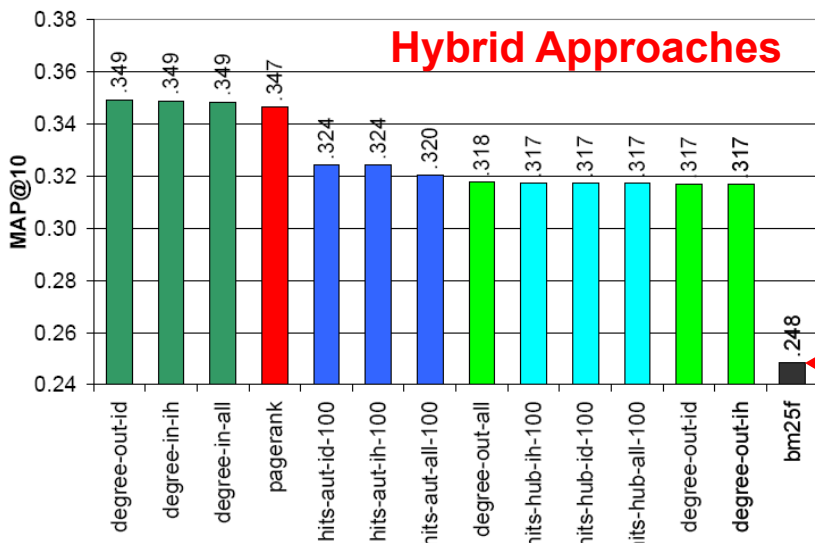
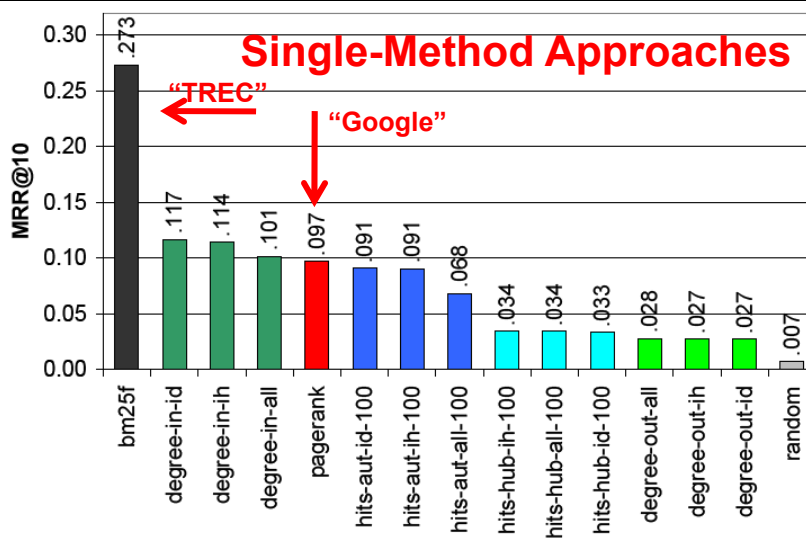
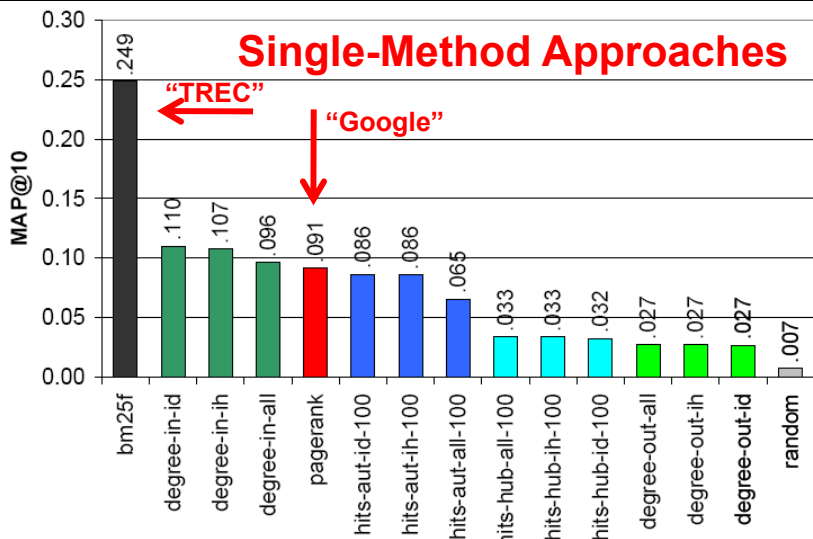
---



## **Najork et al. 2007**

- **463,685,607 HTML Pages**
- **17,672,011,890 Non-Duplicate Hyperlinks**
- **2,897,671,002 URLs – 2,433,985,395 in Frontier**
- **28,043 Queries (Sampled from User Logs)**
- **66,846,214 Result URLs for Queries (2,838/Q)**
- **485,656 Results Evaluated for Relevance by Humans (on a six-point scale)**

# State-of-the-Art Search Performance?



From Najork et al. 2007.