Full Machine Translation for Factoid Question Answering

Cristina España-Bonet and Pere R. Comas
Universitat Politècnica de Catalunya
TALP Research Center

Joint ESIRMT and HyTra Workshop
Avignon, April 23rd, 2012
Full MT for factoid QA

The Noisy Channel
Full MT for factoid QA

The Noisy Channel
Full MT for factoid QA

The Noisy Channel

Translation
Full MT for factoid QA

The Noisy Channel

Question Answering!
Qui va escriure Quarantena?

Who wrote Quarantine?

Translation
Who wrote Quarantine?

Greg Egan wrote Quarantine

Question Answering
Mathematically,

\[ P(O|I) = \frac{P(O) P(I|O)}{P(I)} \]

**SMT:**

\[ T(f) = \arg\max_e P(e|f) = \arg\max_e P(f|e) P(e) \]

**QA:**

\[ A(Q) = \arg\max_A P(A|Q) = \arg\max_A P(Q|A) P(A) \]
How old was Greg Egan when he wrote Quarantine?

**SMT:** Divide and conquer

- How old ||| Quina edat ||| prob1
- How old ||| Quants ||| prob2
- How old ||| Quants anys ||| prob3
- ... old ||| Quina edat ||| prob4
- old ||| vell ||| prob5
- old ||| gran ||| prob6
- ...

Quina edat tenia en Greg Egan quan va escriure Quarantena?
How old was Greg Egan when he wrote Quarantine?

**QA:** Divide and conquer?

How old is Johnny Depp? Johnny Depp is 49
When was *Quarantine* written? Quarantine was written in 1992
When did he write his first novel? He published his first work in 1983
How old was Greg Egan when he wrote *Permutation City*? 33 years old
How old was Greg Egan when he wrote Quarantine?

**QA**: Divide and conquer?

How old is Johnny Depp?  Johnny Depp is 49
When was Quarantine written?  Quarantine was written in 1992
When did he write his first novel?  He published his first work in 1983
How old was Greg Egan when he wrote Permutation City?  33 years old

Alignments depend on concrete questions
Overview

1. The QA system
2. Experiments
3. Final thoughts
**The QA system**

*Definition*

**Question Answering** /ˈkwestʃən ˈɑːnsərɪŋ/ n. Task of extracting short, relevant textual answers from a given document collection in response to natural language questions.
The QA system

Architecture

Question Processing → Keywords → Passage Retrieval

Documents → Passages

Expected Answer Type

Answer Extraction → Answer

Q

A
The QA system

SMT within the architecture

Q

Question Processing

Keywords

Passage Retrieval

Passages

Documents

Expected Answer Type

Answer Extraction

A
Question processing

**Annotation** with
PoS, chunks, NERC,
most frequent WordNet sense
The QA system

The Question Processing Module

**Question processing**

**Annotation** with
PoS, chunks, NERC,
most frequent WordNet sense

**Keywords extraction**

According to the *salience* of words

**Expected Answer Type**

*ME classifier* for Li and Roth (2005) answer types
The QA system

The Passage Retrieval Module

**Document retrieval**

Keywords query with Lucene IR engine

**Passage building**

Segments with two keywords separated less than $t$ words
The QA system

The Passage Retrieval Module

**Document retrieval**

Keywords query with Lucene IR engine

**Passage building**

Segments with two keywords separated less than $t$ words

**Passages processing**

Split into sentences; annotate with PoS, chunks, NERC and most frequent WordNet sense
Answer candidates

NEs and phrases with a noun within the passages
The QA system

The Answer Extraction Module

Answer candidates

NEs and phrases with a noun within the passages

Answer Ranking

Candidate answer sentences are ranked according to their similarity to SMT Question-to-Answer translations

Scoring

MT related scores; EAT score
The QA system

Question-to-Answer translation

Log-linear model (generalisation of the Noisy Channel)

\[ A(Q) = \arg\max_A \sum_m \lambda_m h_m(Q, A) \]

Level 1 Q: What is Karl Malone’s nickname?

Level 1 A: Malone, whose overall consistency has earned him the nickname ANSWER, missed both of them with nine seconds remaining.
The QA system

Question-to-Answer translation

**Log-linear model** (generalisation of the Noisy Channel)

\[ \mathcal{A}(Q) = \arg\max_A \sum_m \lambda_m h_m(Q, A) \]

**Level 1 Q:** What is Karl Malone’s nickname?

**Level 1 A:** Malone, whose overall consistency has earned him the nickname **ANSWER**, missed both of them with nine seconds remaining.
Level 1 Q: What is Karl Malone’s nickname?

Level 1 A: Malone, whose overall consistency has earned him the nickname ANSWER, missed both of them with nine seconds remaining.

Context generalisation, patterns

Level 2 Q: What STATIVE B-PERSON’s COMMUNICATION?

Level 2 A: B-PERSON, whose overall ATTRIBUTE POSSESSION POSSESSION him the COMMUNICATION ANSWER, PERCEPTION both of them with B-NUM TIME CHANGE
Level 1 Q: What is Karl Malone’s nickname?

Level 1 A: Malone, whose overall consistency has earned him the nickname ANSWER, missed both of them with nine seconds remaining.

Context generalisation, patterns for NEs

Level 2 Q: What STATIVE B-PERSON’s COMMUNICATION?

Level 2 A: B-PERSON, whose overall ATTRIBUTE POSSESSION POSSESSION him the COMMUNICATION ANSWER, PERCEPTION both of them with B-NUM TIME CHANGE
Level 1 Q: What is Karl Malone’s nickname?

Level 1 A: Malone, whose overall consistency has earned him the nickname ANSWER, missed both of them with nine seconds remaining.

Context generalisation, patterns for verbs & nouns

Level 2 Q: What STATIVE B-PERSON’s COMMUNICATION?

Level 2 A: B-PERSON, whose overall ATTRIBUTE POSSESSION POSSESSION him the COMMUNICATION ANSWER, PERCEPTION both of them with B-NUM TIME CHANGE
Level 1 Q: What is Karl Malone’s nickname?

Level 1 A: Malone, whose overall consistency has earned him the nickname ANSWER, missed both of them with nine seconds remaining.

Context generalisation, patterns for remaining words

Level 2 Q: What STATIVE B-PERSON’s COMMUNICATION?

Level 2 A: B-PERSON, whose overall ATTRIBUTE POSSESSION POSSESSION him the COMMUNICATION ANSWER, PERCEPTION both of them with B-NUM TIME CHANGE
Answering… What is Karl Malone’s nickname?
The QA system

Answering What is Karl Malone’s nickname?

Abstraction... What STATIVE B-PERSON ’S COMMUNICATION?
Answering    What is Karl Malone’s nickname?

Abstraction    What STATIVE B-PERSO N ’ S COMMUNICATION ?

Translating...    The B-ORGANIZATION B-LOCATION , B-DATE ( B-ORGANIZATION ) - B-PERSON , whose COMMUNICATION STATIV E ” ANSWER . ”
Answering: What is Karl Malone’s nickname?

Abstraction: What stative B-person’s communication?

Translating...

1st best: The B-organization B-location, B-date (B-organization) - B-person, whose communication stative ”ANSWER.”

...  

50th best: The ANSWER ANSWER, B-date (B-organization) - B-person, the person of ANSWER, the most popular artifact, serenely cognition communication.
Translation/Answer

The B-ORGANIZATION, B-LOCATION, B-DATE ( B-ORGANIZATION ) - B-PERSON, whose COMMUNICATION STATIVE ” ANSWER ”

Not a real answer!
Translation/Answer

The B-ORGANIZATION B-LOCATION, B-DATE (B-ORGANIZATION) - B-PERSON, whose COMMUNICATION STATIVE "ANSWER."

Not a real answer!

The ANSWER is found in the document collection:

- Search for the most similar candidate sentence obtained with the Passage Retrieval Module
The QA system

Answer scoring

Ranking of candidate answer sentences done by a combination of scores.

Context scores (B, R)

- *n*-gram matching metrics: BLEU & ROUGE
- Scores the similarity between translations and candidates in L2 representation

Are they enough?
**The QA system**

*Answer scoring*

**Level 1 Qa:** Where was C.S. Lewis born?

**Level 1 Qb:** Where did Hans Christian Anderson die?

**Level 2 Qx:** Where STATIVE PERSON STATIVE?
The QA system

Answer scoring

**Level 1 Qa:** Where was C.S. Lewis born?
**Level 1 Qb:** Where did Hans Christian Anderson die?
**Level 2 Qx:** Where STATIVE PERSON STATIVE?

**Language scores** \((L_b, L_r, L_f)\)

- Similarity between translations and candidates in L1 representation \((L_b, L_r)\)
- Scores candidate’s words according to their frequency in the translations \((L_f)\)
Level 2 Aa: The **B-ORG B-LOCATION, B-DATE ( B-ORGANIZATION ) - B-PERSON**, whose **COMMUNICATION STATIVE ANSWER**.

**Level 2 Ab:** The **ANSWER B-LOCATION, B-DATE ( B-ORGANIZATION ) - B-PERSON**, whose **COMMUNICATION STATIVE B-PERSON**.
The QA system

Answer scoring

**Level 2 Aa:** The **B-ORG** **B-LOCATION**, **B-PERSON**, **B-DATE** ( **B-ORGANIZATION** ) - **B-PERSON**, whose **COMMUNICATION** **STATIVE** **ANSWER**.

**Level 2 Ab:** The **ANSWER** **B-LOCATION**, **B-PERSON**, **B-DATE** ( **B-ORGANIZATION** ) - **B-PERSON**, whose **COMMUNICATION** **STATIVE** **B-PERSON**.

**Expected Answer Type score (E)**

- EAT mapped to NE or WN supersenses
- Candidates scored with the normalised probability of the ME classifier
1 The QA system
2 Experiments
3 Final thoughts
**Experiments**

**Corpora**

**TREC evaluation campaigns** [TREC9, TREC16]

**Document collection.** Newspapers (Tipster, Acquaint, Acquaint2)

**Question sets.** Questions and answer keys
Experiments

Corpora

TREC evaluation campaigns [TREC9, TREC16]

Document collection. Newspapers (Tipster, Acquaint, Acquaint2)

Question sets. Questions and answer keys

<table>
<thead>
<tr>
<th></th>
<th>Q</th>
<th>A</th>
<th>TRECts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>2264</td>
<td>12116</td>
<td>9,10,12,13,14,15,16</td>
</tr>
<tr>
<td>Dev</td>
<td>219</td>
<td>219</td>
<td>9,10,12</td>
</tr>
<tr>
<td>Test</td>
<td>500</td>
<td>2551</td>
<td>11</td>
</tr>
</tbody>
</table>
Experiments

SMT system

- **Language model**: 5-gram interpolated Kneser-Ney discounting, SRILM Toolkit

- **Alignments**: GIZA++ Toolkit

- **Translation model**: Moses package

- **Weights optimization**: MERT against BLEU

- **Decoder**: Moses
Experiments

SMT model

Characteristics
(experiments detailed in the paper)

- 8 standard features: \( P(A) \), \( \text{lex}(Q|A) \) and \( \text{lex}(A|Q) \), \( P_t(Q|A) \) and \( P_t(A|Q) \), \( P_d(A, Q) \), \( ph(A) \) and \( w(A) \)

- 5-gram language model

- 100-best list of translations
Experiments

QA system

- Applied to **Factoid questions**

- The **Question Analysis** module has been adapted from the QA system **Sibyl** (Comas, 2012)

- **Passage Retrieval** module (**Sibyl**)
  
  - 500 questions Q
  - 373,323 candidate answer sentences (747 per Q)
  - 2,866,098 candidate answers (5,732 per Q)
  - Upper bound: 66.7%
### Experiments

#### QA & SR results

<table>
<thead>
<tr>
<th>Metric</th>
<th>QA</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T50</td>
</tr>
<tr>
<td>B</td>
<td>0.018</td>
<td>0.292</td>
</tr>
<tr>
<td>R</td>
<td>0.018</td>
<td>0.283</td>
</tr>
<tr>
<td>B+R</td>
<td>0.022</td>
<td>0.294</td>
</tr>
<tr>
<td>BR</td>
<td>0.027</td>
<td>0.294</td>
</tr>
<tr>
<td>Lf</td>
<td>0.016</td>
<td>0.286</td>
</tr>
<tr>
<td>Lb</td>
<td>0.022</td>
<td>0.304</td>
</tr>
<tr>
<td>Lr</td>
<td>0.018</td>
<td>0.326</td>
</tr>
<tr>
<td>Lbrf</td>
<td>0.038</td>
<td>0.330</td>
</tr>
<tr>
<td>E</td>
<td>0.044</td>
<td>0.373</td>
</tr>
<tr>
<td>ELbrf</td>
<td>0.018</td>
<td>0.293</td>
</tr>
<tr>
<td>BLbrf</td>
<td>0.051</td>
<td>0.337</td>
</tr>
<tr>
<td>RLbrf</td>
<td>0.033</td>
<td>0.346</td>
</tr>
<tr>
<td>BRLbrf</td>
<td>0.042</td>
<td>0.350</td>
</tr>
<tr>
<td>(B+R)Lbrf</td>
<td>0.044</td>
<td>0.346</td>
</tr>
<tr>
<td>BE</td>
<td>0.035</td>
<td>0.384</td>
</tr>
<tr>
<td>RE</td>
<td>0.035</td>
<td>0.377</td>
</tr>
<tr>
<td>BRE</td>
<td>0.049</td>
<td>0.377</td>
</tr>
<tr>
<td>(B+R)E</td>
<td>0.040</td>
<td><strong>0.386</strong></td>
</tr>
<tr>
<td>BELbrf</td>
<td>0.093</td>
<td>0.379</td>
</tr>
<tr>
<td>RELbrf</td>
<td>0.071</td>
<td>0.377</td>
</tr>
<tr>
<td>BRELbrf</td>
<td>0.091</td>
<td>0.379</td>
</tr>
<tr>
<td>(B+R)ELbrf</td>
<td><strong>0.100</strong></td>
<td>0.377</td>
</tr>
</tbody>
</table>
Experiments

QA & SR results

<table>
<thead>
<tr>
<th>Metric</th>
<th>QA</th>
<th></th>
<th>SR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T50</td>
<td>MRR</td>
<td>T1</td>
</tr>
<tr>
<td>B</td>
<td>0.018</td>
<td>0.292</td>
<td>0.049</td>
<td>0.084</td>
</tr>
<tr>
<td>R</td>
<td>0.018</td>
<td>0.283</td>
<td>0.045</td>
<td>0.119</td>
</tr>
<tr>
<td>B+R</td>
<td>0.022</td>
<td>0.294</td>
<td>0.053</td>
<td>0.097</td>
</tr>
<tr>
<td>BR</td>
<td>0.027</td>
<td>0.294</td>
<td>0.057</td>
<td>0.137</td>
</tr>
<tr>
<td>L_f</td>
<td>0.016</td>
<td>0.286</td>
<td>0.046</td>
<td>0.137</td>
</tr>
<tr>
<td>L_b</td>
<td>0.022</td>
<td>0.304</td>
<td>0.054</td>
<td>0.100</td>
</tr>
<tr>
<td>L_r</td>
<td>0.018</td>
<td>0.326</td>
<td>0.060</td>
<td>0.131</td>
</tr>
<tr>
<td>L_brf</td>
<td>0.038</td>
<td>0.330</td>
<td>0.079</td>
<td>0.147</td>
</tr>
<tr>
<td>E</td>
<td>0.044</td>
<td>0.373</td>
<td>0.096</td>
<td>0.058</td>
</tr>
<tr>
<td>EL_brf</td>
<td>0.018</td>
<td>0.293</td>
<td>0.048</td>
<td>0.118</td>
</tr>
<tr>
<td>BL_brf</td>
<td>0.051</td>
<td>0.337</td>
<td>0.091</td>
<td>0.184</td>
</tr>
<tr>
<td>RL_brf</td>
<td>0.033</td>
<td>0.346</td>
<td>0.069</td>
<td>0.191</td>
</tr>
<tr>
<td>BRL_brf</td>
<td>0.042</td>
<td>0.350</td>
<td>0.082</td>
<td>0.182</td>
</tr>
<tr>
<td>(B+R)L_brf</td>
<td>0.044</td>
<td>0.346</td>
<td>0.085</td>
<td>0.187</td>
</tr>
<tr>
<td>BE</td>
<td>0.035</td>
<td>0.384</td>
<td>0.084</td>
<td>0.086</td>
</tr>
<tr>
<td>RE</td>
<td>0.035</td>
<td>0.377</td>
<td>0.086</td>
<td>0.131</td>
</tr>
<tr>
<td>BRE</td>
<td>0.049</td>
<td>0.377</td>
<td>0.098</td>
<td>0.135</td>
</tr>
<tr>
<td>(B+R)E</td>
<td>0.040</td>
<td><strong>0.386</strong></td>
<td>0.091</td>
<td>0.102</td>
</tr>
<tr>
<td>BEL_brf</td>
<td>0.093</td>
<td>0.379</td>
<td>0.137</td>
<td>0.200</td>
</tr>
<tr>
<td>REL_brf</td>
<td>0.071</td>
<td>0.377</td>
<td>0.123</td>
<td><strong>0.208</strong></td>
</tr>
<tr>
<td>BREL_brf</td>
<td>0.091</td>
<td>0.379</td>
<td>0.132</td>
<td>0.200</td>
</tr>
<tr>
<td>(B+R)EL_brf</td>
<td><strong>0.100</strong></td>
<td>0.377</td>
<td><strong>0.141</strong></td>
<td>0.204</td>
</tr>
</tbody>
</table>

Buf!
Experiments

QA & SR results

Buf!

Note that with QA, SR comes at the same price.

<table>
<thead>
<tr>
<th>Metric</th>
<th>QA T1</th>
<th>QA T50</th>
<th>QA MRR</th>
<th>SR T1</th>
<th>SR T50</th>
<th>SR MRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0.018</td>
<td>0.292</td>
<td>0.049</td>
<td>0.084</td>
<td>0.540</td>
<td>0.164</td>
</tr>
<tr>
<td>R</td>
<td>0.018</td>
<td>0.283</td>
<td>0.045</td>
<td>0.119</td>
<td>0.608</td>
<td>0.209</td>
</tr>
<tr>
<td>B+R</td>
<td>0.022</td>
<td>0.294</td>
<td>0.053</td>
<td>0.097</td>
<td>0.573</td>
<td>0.180</td>
</tr>
<tr>
<td>BR</td>
<td>0.027</td>
<td>0.294</td>
<td>0.057</td>
<td>0.137</td>
<td>0.591</td>
<td>0.211</td>
</tr>
<tr>
<td>Lf</td>
<td>0.016</td>
<td>0.286</td>
<td>0.046</td>
<td>0.137</td>
<td>0.605</td>
<td>0.236</td>
</tr>
<tr>
<td>Lb</td>
<td>0.022</td>
<td>0.304</td>
<td>0.054</td>
<td>0.100</td>
<td>0.581</td>
<td>0.192</td>
</tr>
<tr>
<td>Lr</td>
<td>0.018</td>
<td>0.326</td>
<td>0.060</td>
<td>0.131</td>
<td>0.627</td>
<td>0.225</td>
</tr>
<tr>
<td>Lbrf</td>
<td>0.038</td>
<td>0.330</td>
<td>0.079</td>
<td>0.147</td>
<td>0.622</td>
<td>0.238</td>
</tr>
<tr>
<td>E</td>
<td>0.044</td>
<td>0.373</td>
<td>0.096</td>
<td>0.058</td>
<td>0.579</td>
<td>0.142</td>
</tr>
<tr>
<td>ELbrf</td>
<td>0.018</td>
<td>0.293</td>
<td>0.048</td>
<td>0.118</td>
<td>0.623</td>
<td>0.214</td>
</tr>
<tr>
<td>BLbrf</td>
<td>0.051</td>
<td>0.337</td>
<td>0.091</td>
<td>0.184</td>
<td>0.616</td>
<td>0.271</td>
</tr>
<tr>
<td>RLbrf</td>
<td>0.033</td>
<td>0.346</td>
<td>0.069</td>
<td>0.191</td>
<td>0.618</td>
<td>0.279</td>
</tr>
<tr>
<td>BRLbrf</td>
<td>0.042</td>
<td>0.350</td>
<td>0.082</td>
<td>0.182</td>
<td>0.616</td>
<td>0.273</td>
</tr>
<tr>
<td>(B+R)Lbrf</td>
<td>0.044</td>
<td>0.346</td>
<td>0.085</td>
<td>0.187</td>
<td>0.618</td>
<td>0.273</td>
</tr>
<tr>
<td>BE</td>
<td>0.035</td>
<td>0.384</td>
<td>0.084</td>
<td>0.086</td>
<td>0.579</td>
<td>0.179</td>
</tr>
<tr>
<td>RE</td>
<td>0.035</td>
<td>0.377</td>
<td>0.086</td>
<td>0.131</td>
<td>0.630</td>
<td>0.228</td>
</tr>
<tr>
<td>BRE</td>
<td>0.049</td>
<td>0.377</td>
<td>0.098</td>
<td>0.135</td>
<td>0.608</td>
<td>0.220</td>
</tr>
<tr>
<td>(B+R)E</td>
<td>0.040</td>
<td>0.386</td>
<td>0.091</td>
<td>0.102</td>
<td>0.596</td>
<td>0.196</td>
</tr>
<tr>
<td>BELbrf</td>
<td>0.093</td>
<td>0.379</td>
<td>0.137</td>
<td>0.200</td>
<td>0.621</td>
<td>0.283</td>
</tr>
<tr>
<td>RELbrf</td>
<td>0.071</td>
<td>0.377</td>
<td>0.123</td>
<td>0.208</td>
<td>0.619</td>
<td>0.294</td>
</tr>
<tr>
<td>BRELbrf</td>
<td>0.091</td>
<td>0.379</td>
<td>0.132</td>
<td>0.200</td>
<td>0.622</td>
<td>0.287</td>
</tr>
<tr>
<td>(B+R)ELbrf</td>
<td>0.100</td>
<td>0.377</td>
<td>0.141</td>
<td>0.204</td>
<td>0.621</td>
<td>0.286</td>
</tr>
</tbody>
</table>
Individual metrics

**Weak** because:

- B and R do not take into account the lexical realisation
- $L_x$ gives the same score to all candidates in the same sentence (better in SR)
- E gives the same score to all candidates of the same type (better in QA)
## Combination of metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>QA</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T50</td>
</tr>
<tr>
<td><strong>BEL\textsubscript{brf}</strong></td>
<td>0.093</td>
<td>0.379</td>
</tr>
<tr>
<td><strong>REL\textsubscript{brf}</strong></td>
<td>0.071</td>
<td>0.377</td>
</tr>
<tr>
<td><strong>BREL\textsubscript{brf}</strong></td>
<td>0.091</td>
<td>0.379</td>
</tr>
<tr>
<td><strong>(B+R)EL\textsubscript{brf}</strong></td>
<td><strong>0.100</strong></td>
<td>0.377</td>
</tr>
</tbody>
</table>
# Combination of metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>QA</th>
<th></th>
<th></th>
<th>SR</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T50</td>
<td>MRR</td>
<td>T1</td>
<td>T50</td>
<td>MRR</td>
</tr>
<tr>
<td>BEL&lt;sub&gt;brf&lt;/sub&gt;</td>
<td>0.093</td>
<td>0.379</td>
<td>0.137</td>
<td>0.200</td>
<td>0.621</td>
<td>0.283</td>
</tr>
<tr>
<td>REL&lt;sub&gt;brf&lt;/sub&gt;</td>
<td>0.071</td>
<td>0.377</td>
<td>0.123</td>
<td>0.208</td>
<td>0.619</td>
<td>0.294</td>
</tr>
<tr>
<td>BREL&lt;sub&gt;brf&lt;/sub&gt;</td>
<td>0.091</td>
<td>0.379</td>
<td>0.132</td>
<td>0.200</td>
<td>0.622</td>
<td>0.287</td>
</tr>
<tr>
<td>(B+R)EL&lt;sub&gt;brf&lt;/sub&gt;</td>
<td><strong>0.100</strong></td>
<td>0.377</td>
<td><strong>0.141</strong></td>
<td>0.204</td>
<td>0.621</td>
<td>0.286</td>
</tr>
</tbody>
</table>
1 The QA system

2 Experiments

3 Final thoughts
approximation to QA as an MT problem

- T1 ~ 10% is in the lowest part of TREC11 evaluation.
- Other approaches that use translation probabilities (Echihabi and Marcu, 2003) are better ranked.
- This approach is more similar to Ravichandran and Hovy (2002) who learn patterns to find answer contexts.
Sentence retrieval as a complement to QA

- T50 in SR is close to the upper bound given by the Passage Retrieval module.

- **E** is not discriminative enough: T50 drops almost to a half in QA.
Final thoughts

Comments, conclusions & summary

Ranking the candidates, the key point

- Substitute $E$ for an MT-based metric.

- Introduce new scoring metrics such as the score given by the decoder in translation.

- Improve the retrieval and therefore the upper bound (query expansion with SMT?).
Thank you!
Full Machine Translation for Factoid Question Answering

Cristina España-Bonet and Pere R. Comas
Universitat Politècnica de Catalunya
TALP Research Center

Joint ESIRMT and HyTra Workshop
Avignon, April 23rd, 2012
Saliences

9  Words within quotes
8  Named entities
7  Sequences of nouns and adjectives
6  Sequences of nouns
5  Adjectives
4  Nouns
3  Verbs and adverbs
2  Question focus word
1  Any non-stop word
### Answer types, Li and Roth (2005)

<table>
<thead>
<tr>
<th>ABBREVIATION: abb</th>
<th>ENTITY: other</th>
<th>LOCATION: mount</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABBREVIATION: exp</td>
<td>ENTITY: plant</td>
<td>LOCATION: other</td>
</tr>
<tr>
<td>DESCRIPTION: def</td>
<td>ENTITY: product</td>
<td>LOCATION: state</td>
</tr>
<tr>
<td>DESCRIPTION: desc</td>
<td>ENTITY: religion</td>
<td>NUMBER: code</td>
</tr>
<tr>
<td>DESCRIPTION: manner</td>
<td>ENTITY: sport</td>
<td>NUMBER: count</td>
</tr>
<tr>
<td>DESCRIPTION: reason</td>
<td>ENTITY: substance</td>
<td>NUMBER: date</td>
</tr>
<tr>
<td>ENTITY: animal</td>
<td>ENTITY: symbol</td>
<td>NUMBER: distance</td>
</tr>
<tr>
<td>ENTITY: body</td>
<td>ENTITY: techmeth</td>
<td>NUMBER: money</td>
</tr>
<tr>
<td>ENTITY: color</td>
<td>ENTITY: termeq</td>
<td>NUMBER: order</td>
</tr>
<tr>
<td>ENTITY: cremat</td>
<td>ENTITY: veh</td>
<td>NUMBER: other</td>
</tr>
<tr>
<td>ENTITY: currency</td>
<td>ENTITY: word</td>
<td>NUMBER: perc</td>
</tr>
<tr>
<td>ENTITY: dismed</td>
<td>HUMAN: description</td>
<td>NUMBER: period</td>
</tr>
<tr>
<td>ENTITY: event</td>
<td>HUMAN: group</td>
<td>NUMBER: speed</td>
</tr>
<tr>
<td>ENTITY: food</td>
<td>HUMAN: individual</td>
<td>NUMBER: temp</td>
</tr>
<tr>
<td>ENTITY: instrument</td>
<td>HUMAN: title</td>
<td>NUMBER: volsize</td>
</tr>
<tr>
<td>ENTITY: lang</td>
<td>LOCATION: city</td>
<td>NUMBER: weight</td>
</tr>
<tr>
<td>ENTITY: letter</td>
<td>LOCATION: country</td>
<td></td>
</tr>
</tbody>
</table>
Passage building

Query Keywords: relevant, documents, process

Passage:
The log-linear model

\[ A(Q) = \hat{A} = \arg\max_A \log P(A|Q) = \]
\[ + \lambda_{lm} \log P(A) + \lambda_d \log P_d(A, Q) \]
\[ + \lambda_{lg} \log \text{lex}(Q|A) + \lambda_{ld} \log \text{lex}(A|Q) \]
\[ + \lambda_g \log P_t(Q|A) + \lambda_d \log P_t(A|Q) \]
\[ + \lambda_{ph} \log \text{ph}(A) + \lambda_w \log w(A) \]