Integrating Spoken Dialog and Question Answering: the Ritel Project

Sophie Rosset, Olivier Galibert, Gabriel Illouz, Aurélien Max

{rosset, galibert, amax, gabrieli}@limsi.fr

Human-Machine Communication Dept.
LIMSI-CNRS, France
Talk Outline

- Ritel Project - Definition and Platform
- Architecture
- Speech detection and recognition
- Non-Contextual Analysis
- General Information Retrieval
- IR on databases and Dialogic interactions
- Natural Language Generation
- Conclusions and Perspectives
The Ritel Project

*Recherche d’Informations par TELéphone* - Information Retrieval by Phone.

Limited-domain spoken dialog systems $\rightarrow$ open domain.
TREC-like Question Answering $\rightarrow$ interactive refinement.
Application-specific one-way natural language generation $\rightarrow$ natural interactivity.

Result is the Ritel Project: an Oral Open-Domain Interactive Information Retrieval System.
The Ritel Platform

Want a real system in order to collect real user data:
- Use the phone as input since everybody has one.
- Speed constraints everywhere to enable highly-reactive natural interaction.
- Fully integrate Dialogue and Information Retrieval.

First version used some databases and hand-crafted reactive answers to incite the user to speak. It was used to collect around 8 hours of user speech. Current version really integrates Information Retrieval.
Speech Activity Detection

System built for conversational telephonic speech independently of, and before, Ritel. Relatively simple:

- Two gaussian mixture models, one for silence/noise, one for voice.
- Viterbi beam decoding of the audio with symbol output on back-of-the-beam closure

End-of-speech to symbol output delay is 0 to 0.2s (not user-perceptible). No speech loss when not overlapping, 10% excess. Not tested when overlapping.
Speech Recognition

Constraints were:
- 0.85 RT max.
- Fully streamed.
- Small amount of audio for training.

Large, but context-independent phonetic acoustic models on non-normalized cepstrum. Language models built on interpolation between the Ritel corpus, broadcast news, newspapers and written queries (madwin).
Speech Recognition

Results:
- Around 28% error rate.
- 0.5RT.

Perspectives:
- Increase the amount of training data.
- Add the kitchen sink: MLLT, MMIE, dynamic VTLN and CMLLR.
- Feedback from the dialog to the language model.
Non-Contextual Analysis

- Extract “pertinent” information from a sentence.
- Must work on both written and spoken data.
- As fast as possible.
- Use existing linguistic resources when available.

What is pertinent information?

_in_ _org_ NIST _NN_ metadata evaluations
_action reported _NN_ speaker tracking
_score error rates _aux are _prep about
_val_score 15 %
Non-Contextual Analysis - Entities

Typed entities at several linguistic levels:

- Named entities
  - \texttt{\_org} NIST
  - \texttt{\_eve} 2006 Cannes festival
  - who said \texttt{\_cit} veni vidi vici

- Non-specific entities
  - \texttt{\_Eve} Cannes festival
  - the \texttt{\_Pers} president said ...
Non-Contextual Analysis - Entities

- Multi-level extended entities
  - Functions, titles (president, bishop...), ...
  - Colors, animals...
  - Hierarchical superclassing
    (bishop $\rightarrow$ religious function $\rightarrow$ function)

- Topic markers
  - I’m interested in __litterature novels__ by ...
  - who won the __sport Mundial__ in 1998
Non-Contextual Analysis - Entities

Question markers
- \_Qwho who wrote that book
- \_Qmeasure how many hours of transcription do you need

Interaction markers
- \_DA_close goodbye
- \_DA_yes yes please
Non-Contextual Analysis - Entities

- Information chunks
  - Compound nouns
    - the \_NN local farmers are ...
    - south Africa’s first \_NN multiracial elections
  - Action verbs and composites
    - Thomas Mann \_action won the Nobel prize
    - half the competitors \_action gave up on the race
  - Linguistic entities
    - the \_adj\_comp highest prize
    - the phenomena happens \_adv often when ...
Non-Contextual Analysis - Methods

Named, non-specific and extended entities:
- Lists for initial detection
- Rough word morphosyntactic (noun, verb, adjective...) and semantic (movement verbs, declarative verbs...) categories in the context for disambiguation
- Trigger contexts for out-of-list detection (who wrote X, Y university...)

The DELA dictionary from Unitex provides non-disambiguated morphosyntactic classes.
Non-Contextual Analysis - Methods

Topic detection:
- Some topics directly linked to specific extended entity superclasses
  I’m interested in literature, ...
- Other topics as co-occurences of superclasses
  _function + _org (∈ _country) → geopolitics

Question and interaction markers:
- Lists for initial detection
  how, what, yes...
- Specific verbal and prepositional structures
  How many + hours → _Qmeasure_phys
Non-Contextual Analysis - Methods

Information chunks (high level light analysis):
- Strong structures detection (noun preposition noun...)
- High-probability disambiguation (adverbs)
- Auxiliary, modal and state verbs (to be, to want, to become)
- Infinitive and -ing form inducing verbs (to start to)
- Weak structures detection (adjective noun...)
- Final remaining words classification
Currently rule-based system. Needs to be able to manage:

- Lists for initial detections
- Local contexts
- Easy categorizations

Given these constrains, a word-based Regular Expression engine was implemented with some added NLP-specific features.
NC Analysis - Implementation

RE language includes:

- Positive and negative lookaheads.
- Shy and greedy groupings.
- Named classes and macros.
- Strategies for prioritizing rule application.
- Recursive substitution modes.
- Word tagging (for tags like noun, verb...).
- Word categorization (number, acronym, proper noun...).
NC Analysis - Current system

- Lists sizes: 92 600 names, 500 countries, 185 000 towns, 300 languages...
- Around 40 analysis steps.
- Sentence analysis in the millisecond range.

Evaluation of named entities and extended entities detection gave a F-measure from 82% on broadcast news to 88% on spoken queries. Perspectives include using statistical approaches for disambiguation, semi-supervised automatic lists and structure learning from large sets of documents.
Topic change detection

Do we want to complete the request with elements from the previous exchanges?

- Use the topic markers to detect the topic of the user utterance, if any
- On topic change, flush the history
No specific Dialog Management: it’s a completely integrated approach.

Using the question and interaction markers, classify the type of utterance:

- General Information retrieval
- Information Retrieval on Databases
- Pure dialogic interaction or unclassifiable utterance
1. Detect the answer type using Question Marker, Named, Non-specific and Extended Entities co-occurences.

   - \_Qwho → \_pers or \_pers_def
   - who sold Manhattan
   - \_Qwhat + \_function → \_pers
   - what is the name of the pope

2. Decide on the relative importance of the different entities in the utterance.

   - Named entity > NN > adj_comp > action > subs ...
3. Build a primary document query with all the entities and then a series of backoff queries relaxing some of the constraints:

- Increase the allowable distance between the entities.
- Allow type changes for some entities (_loc → _org).
- Allow including or included values in the document (Bush → Georges Bush).
- Drop some of the entities.

At this point, all these queries lists are handcrafted.
General Information Retrieval

4. Send the queries to the indexation server and get snippets (sentence or groups of sentences) back.

5. Find the entities in the snippets with the expected type of the answer and cluster them.

Most frequent answers win, and the distribution of the counts gives an idea of the confidence of the system in the answers.

This is a baseline system at that point that fits the time constraint.
GIR - Perspectives

- We have a baseline to collect corpus and validate the following hypotheses:
  - H1: question or question type will occur again
  - H2: users want more information on the same topic
- Scoring the answer pattern in the snippet
- Learning and scoring the extraction and backoff patterns
GIR - Perspectives

Scoring and clusterization of answers is multidimensional

Confidence:
- syntactic (does the pattern allow to answer the question)
- in the information (computed on redundancy and source)

Clusterization:
- by existing ontology
- by emergent one
IR on Databases

Specific handling for some limited domains where accurate database-like information is available (local databases, TV programs, IMDB...). Handling is similar to traditional limited-domain dialogue systems:

1. Pick the latest query type in the dialogue history if one is not present in the user utterance.
2. Complete the query slots needed for the query type from the dialogue history.
3. Do the database lookup

Next versions should also complete the answer with results from the General Information Retrieval.
Dialogic Interaction

Management of the **pure dialogic interaction** such *please repeat, I don’t understand, goodbye ...*
Some of the reactions of that module include:

- Send the guide.
- Send a goodbye sentence then hang up.
- Reformulate and repeat.

Next system should be able to reply questions on the identity of the sources and give citations of the documents.
Natural Language Generation

Produce language output for a natural interaction

- make the user feel that the system is cooperative
- sound natural

Implications on the system design

- determine the different kinds of interaction that must take place with the other parts of the system
- identify what generation technique is appropriate for what purpose
Current status

Use of simple strategies:

1. canned text: *Welcome to the LIMSI Ritel system.*

2. canned text with slots: *The <function> of <country> is <answer>,* e.g. *The president of France is Jacques Chirac*
   - templates have to be associated with question types (specialisation Vs generalisation)
   - linguistic adaptation can be necessary, e.g.: *John Hume et David Trimble ont reçu le Prix Nobel de la Paix en 1998*

3. simple answers, e.g. *Het Wilhelmus* (10 answers)
This approach has several limitations:
- limited expressivity
- templates need to be handcrafted and maintained
- hard to maintain a natural interaction

Identify interesting research directions:
- more flexibility
- interaction with dialog state and information extraction
- reuse of document formulations
NLG - Enhanced interaction

- Communicating system confidence
- Cobuilding the user query
- Phatic functions
Communicating system confidence

- Appropriate self-assessing formulations using system confidence
  - U: *What is Noam Chomsky’s office phone number?*
  - S: *According to Noam Chomsky’s homepage at MIT, his office phone number is 1-617-253-7819*

- In case of uncertainty, decide which pieces of information to give
  - U: *Who sang the song Yesterday?*
  - S: *It seems to be the Beatles, but there are many other possible answers.*
Cobuilding the user query (1/2)

- Allow for confirmation and user barge-in
  - S: The president of Fra...
  - U: No, I’m interested in the Prime Minister of France.

- No referring expressions when uncertain of the input
  - S: It was proclaimed in...
  - S: The declaration of independence of the Dutch Republic was proclaimed in 1581.

- ... but use them once confirmed
  - S: It became independent from Spain.
Cobuilding the user query (2/2)

Resolve potential ambiguities by formulating discriminating questions, e.g.

U: *When was George Bush elected?*
S: *Do you mean George W. Bush, the current president, or George H.W. Bush?*

Decide whether a candidate is salient enough

U: *How tall is the Eiffel Tower?*
S: *Several Eiffel Towers exist in the world. Are you interested in the one in Paris, or another one?*
S: *The Eiffel Tower in Paris stands 300 meters.*
Phatic functions

- Feedback to the user
  - S: *I have more than 2000 documents to analyse, so that might take a little while.*

- Backchannel during user pauses
  - S: *Hum-hum... Yes?* (humming)

- Generation strategies that save time for IR
  - S: *Well, the countries directly to the South of the Netherlands, after processing documents from reliable sources, are... well only Belgium.*
NLG - Enhanced search

- Estimation of the granularity of the questions
- Complete answers
- Suggestive answers
Dealing with granularity

If several embedded answers

U: When was Henry IV murdered?
S: Henry IV was murdered {in 1610 | in May 1610 | on 14 May 1610}

If several possible answers

U: Where was Henry IV murdered?
S: Henry IV was murdered {in Paris | on rue de la Ferronnerie | in his carriage}
S: Are you interested in a city, a street, or a type of vehicle?
Completive answers (1/2)

- Identify pieces of information that might be of interest to the user and give them in the answers:
  - without increasing significantly the length
  - without postponing the request answer

- This extra-information should be significantly cooccurring with the sought information:
  - find recurring adverbial complements
  - use hyponyms that are more specific
  - produce the answer by modifying existing sentences
Completive answers (2/2)

When was Henry IV assassinated?
... François Ravaillac, who stabbed to death Henry IV on 14 May 1610...
Henry IV was assassinated in Paris on May 14, 1610, by a fanatical Roman Catholic named François Ravaillac.

Possible answers:

- Henry IV was assassinated on 14 May 1610.
- Henry IV was stabbed to death on 14 May 1610.
- On 14 May 1610, Henry IV was stabbed to death by a fanatical Roman Catholic named François Ravaillac.
Suggestive answers

Identifying patterns

U: Who is the president of Mexico? ...
U: And of Finland? ...
S: Are there other countries for which you would like to know who the presidents are?

Propose theme-oriented exploration

U: Who is the French president? ...
U: And the Prime Minister? ...
U: (faint noise of intense neural processing)
S: Are there other things you would like to know about the French political system?
Conclusions and Perspectives

- No dialog manager, everything is integrated and interconnected.
- System is snappy, answers are instantaneous.
- Most of the work is corpus-based.

- Increase the corpus size by an order of magnitude.
- Start work towards real interactive refinement.
- Speaker adaptation at all levels.