Natural Language Input for Voice Search

Melvyn Hunt
Novauris Technologies
The Voice Search Company
What is this young British gentleman saying?
What is this young British gentleman saying?

Could he be saying?

– Excuse me, old chap, but if you’re not too busy and if you don’t mind awfully, do you think you could perhaps come over here, please?

No way! — but he might be saying:

– Over here!

He probably isn’t even swearing
...and what might this young man be saying?

Could he be saying?

– *Excuse me ma’m, do you happen to know where Castle Street is, please?*

- Such indirect language would signal:
  - I’m a civilized person
  - I’m not going to:
    - beg for money
    - proposition you
    - rob you
    - molest you
Sadly, he didn’t use such careful language — and this is what happened….

Language must be matched to the situation

- It’s dangerous to use too many words in battle
- It’s dangerous to use too few words to a stranger in the street
- Using too many words with someone you know well can be irritating
Overview of the talk

- What spoken “natural language” really means
- How speakers skillfully adjust their speaking style
- How speakers address voice search systems
  - Observations from real-world deployments
- Usually, it’s very terse
- But there are exceptions…
  - e.g. translation systems as voice search
- It’s essential to anticipate the “natural” style for the particular application
Natural spoken language isn’t…
– speaking in complete “grammatical” sentences as though writing an essay

And it isn’t…
– using a wordy, frilly or polite style
  • unless the situation demands it
    (the soldier was definitely using natural language!)

So what is it?...
Natural spoken language is…

– speaking in the appropriate style when you take into account:
  – the situation
  – the communications channel
  – the listener’s capacities and likely reaction

Fortunately, all this comes… naturally
Some other factors that affect natural speaking style

- **Acoustic conditions**
  - *e.g.* we speak louder in noise (Lombard effect)

- **Our model of the listener**
  - We speak more carefully when we think:
    - a) the listener has hearing difficulties
    - b) the listener can’t speak our language well
    - c) the listener is mentally challenged
  - If we are a native speaker with normal hearing and the speaker overarticulates we feel patronized.
The story so far...

- Natural speech is how we spontaneously speak.
- It means using the fewest words – and the minimum articulatory effort – needed to enable the listener to get the message easily.
  - including any side messages about your respect for the listener, *etc.*
- Humans are skilled at doing this.
So how about Speaking to Machines?

- Computers don’t have emotions
- So there’s no need to use frills to assure them of our good intentions, respect, etc.
  - AND PEOPLE DON’T!
- The natural language for voice search is the fewest possible number of words.
Natural Language Input to a Very Large Voice Search Application

- Verizon Wireless Get It Now® Search:
  - an extremely large music/games mobile search application
  - published by Medio Systems
  - incorporating Novauris voice search

An ideal source of insight into real-world voice search behavior
Experience Shows: Speakers Behave Rationally

- They use the minimum number of words needed to give a clear message
  - no instances of “please”
  - no instances of “I’d like”, etc.

- The occasional exceptions are attempts at increasing clarity:
  - Repeats, e.g. “Shakira … Shakira”
  - Spelling, e.g. “Pink, … P. I. N. K.”
  - Pausing between syllables, e.g. “Ma - do - nna”

- Only potential applications providers do otherwise!
Does ASR Never Need to Cope with Wordy Language?

Yes it does, when:

a) The speaker thinks of the machine as a conscious entity:
   – a lifelike robot
   – young Japanese woman even bowing to a speech-enabled coffee machine!

b) The speaker is talking “through” the ASR to another human being:
   – automatic speech-to-speech translation
But Do These Exceptions Have Anything to Do with Voice Search?

- They aren’t what we normally think of as voice search
- But Novauris has developed techniques specifically for voice search, and...
  - surprisingly, the techniques are well suited to searching the many possible variations in spontaneous inputs to speech-to-speech translation systems
  - dealing with tech-support questions
  - and in communicating with a quasi-human robot.
Wording Variations in Speech-to-Speech Translation

- Prototype English ↔ Korean scenario-specific translator
  - A few hundred basic phrases
  - “Can you repeat that, please?” in Korean has 50,200 variations!
  - “I’m afraid the hotel is full” in English has 67,500 variations!

- Our fast search methods allow rapid, accurate identification of the appropriate basic phrase
It’s Vital to Anticipate Speaker Behavior

- One could allow a large number of variations for inputs to every application
  - But that increases computation and lowers accuracy for a given set of basic requests
  - In applications where variations hardly ever occur (e.g. destination entry, music selection…), allowing for variations is harmful.

- This talk has tried to distinguish between applications that need variations and those that don’t.
Natural speech is not a single specific style of spoken language
   - it is the most appropriate style for the situation
   - and it’s the style that speakers adopt when not given specific instructions

For most voice search applications, speakers naturally use direct language with the fewest possible number of words
   - the exception is when the speaker thinks that a conscious entity is being addressed

It’s essential to anticipate the speaking style that is “natural” for the particular application.
Thank You!

www.novauris.com