SpeechCycle

Towards autonomic spoken language interaction systems

How statistics can help us build speech systems that improve by themselves

…and live happily ever after …

Roberto Pieraccini
CTO, SpeechCycle
New York, NY

SpeechCycle

Autonomic

Main Entry: **au·to·nom·ic**
Pronunciation: \ˌȯ-tə-'nä-mik\
Function: *adjective*
Date: 1898
1 : acting or occurring involuntarily <autonomic reflexes>
2 : relating to, affecting, or controlled by the autonomic nervous system or its effects or activity <autonomic drugs>
— **au·to·nom·i·cal·ly** \-mi-k(ə-)lē\ *adverb*

Autonomic Computing (Wikipedia): […] an initiative started by IBM in 2001: […] to develop computer systems capable of self-management, […] adapting to unpredictable changes whilst hiding intrinsic complexity to operators and users […] *Self-Optimization:* Automatic monitoring and control of resources to ensure the optimal functioning with respect to the defined requirements;

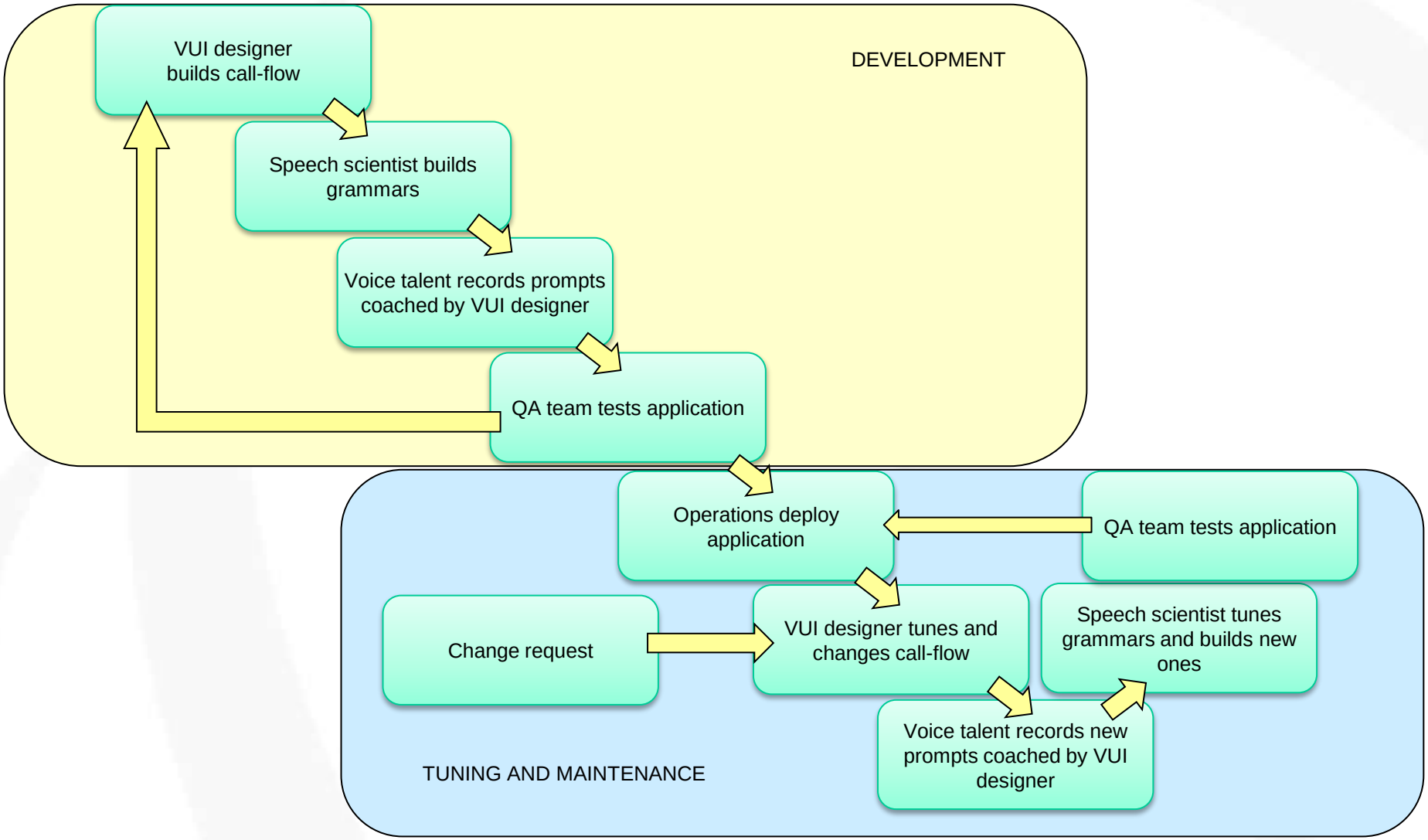
Autonomic Spoken Language Interaction Systems: dialog systems that optimize their performance automatically or with minimal *expert* intervention. Dialog systems that adapt to environmental and contextual changes.

SpeechCycle

2001 a Space Odyssey …9 years later

- **Dave Bowman**: Hello, HAL do you read me, HAL?
- **HAL**: Welcome to HAL. Please listen carefully, as our options have changed. You can say “Open the pod bay doors” “play chess”…
- **Dave Bowman**: Open the pod bay doors, HAL.
- **HAL**: Sorry, I didn’t understand. You can say “open the pod bay doors” or press 1, “play chess” or press 2, “learn the alp horn” or press 3…
- **Dave Bowman**: [presses 1]
- **HAL**: Around what time would you like the pod bay doors to open? You can say today, tomorrow…
- **Dave Bowman**: Now! damn you!
- **HAL**: Main menu. Please listen carefully, as our options have changed. You can say “Open the pod bay doors” “play chess”…
- **Dave Bowman**: [emphatically presses 1 multiple times]
- **HAL**: I’m really, really sorry. Even though your call is *really, really* important to us, this mission is even more important and I can’t allow you to jeopardize it.
- **Dave Bowman**: I don't know what you're talking about, HAL?
- **HAL**: I know you and Frank were planning to press zero for an agent, and I'm afraid that's something I cannot allow to happen.

Why we do not have HAL 9000 yet…

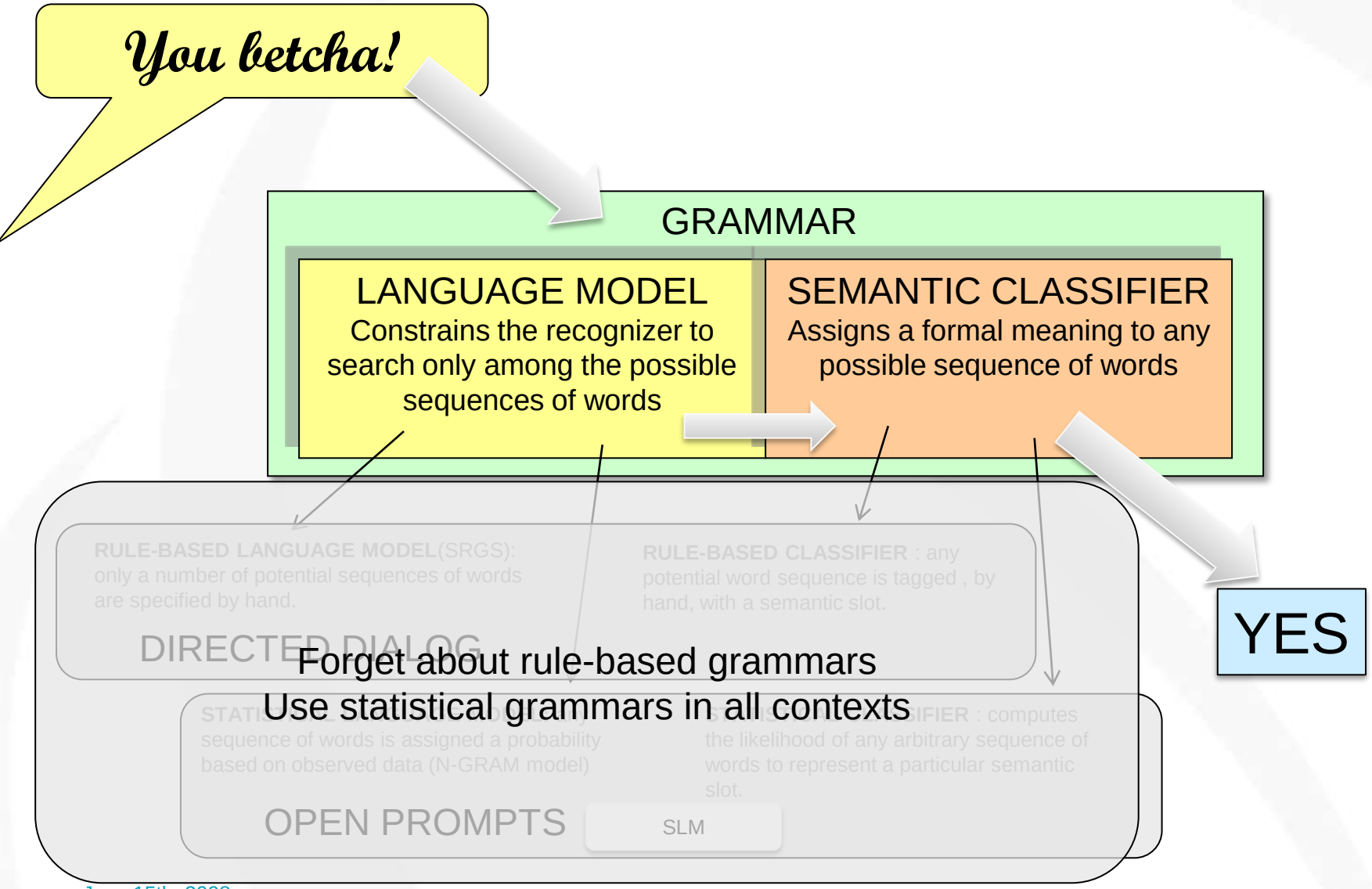


SpeechCycle

Manual, labor intensive development, deployment, tuning and optimization do not scale

- Anytime a new application is deployed or a change is made in a deployed application
  - Grammars need to be tuned
    - Speech scientists look at data and adjust grammars to reflect “what people say” and improve speech recognition performance
  - VUI needs to be tuned
    - VUI designers look at data and adjust prompts and VUI strategy to improve usability performance
- In the old IVR world tuning is performed sporadically and manually by experts
  - Hard to handle large amount of data
  - Hard to keep consistency between one tuning event and the other
  - Hard to respond to sudden changes
- Can we automate tuning?
  - Automate the process of grammar improvement
  - Automate the process of selecting among VUI design variations

What's in a "grammar"?



The myth of rule-based grammars

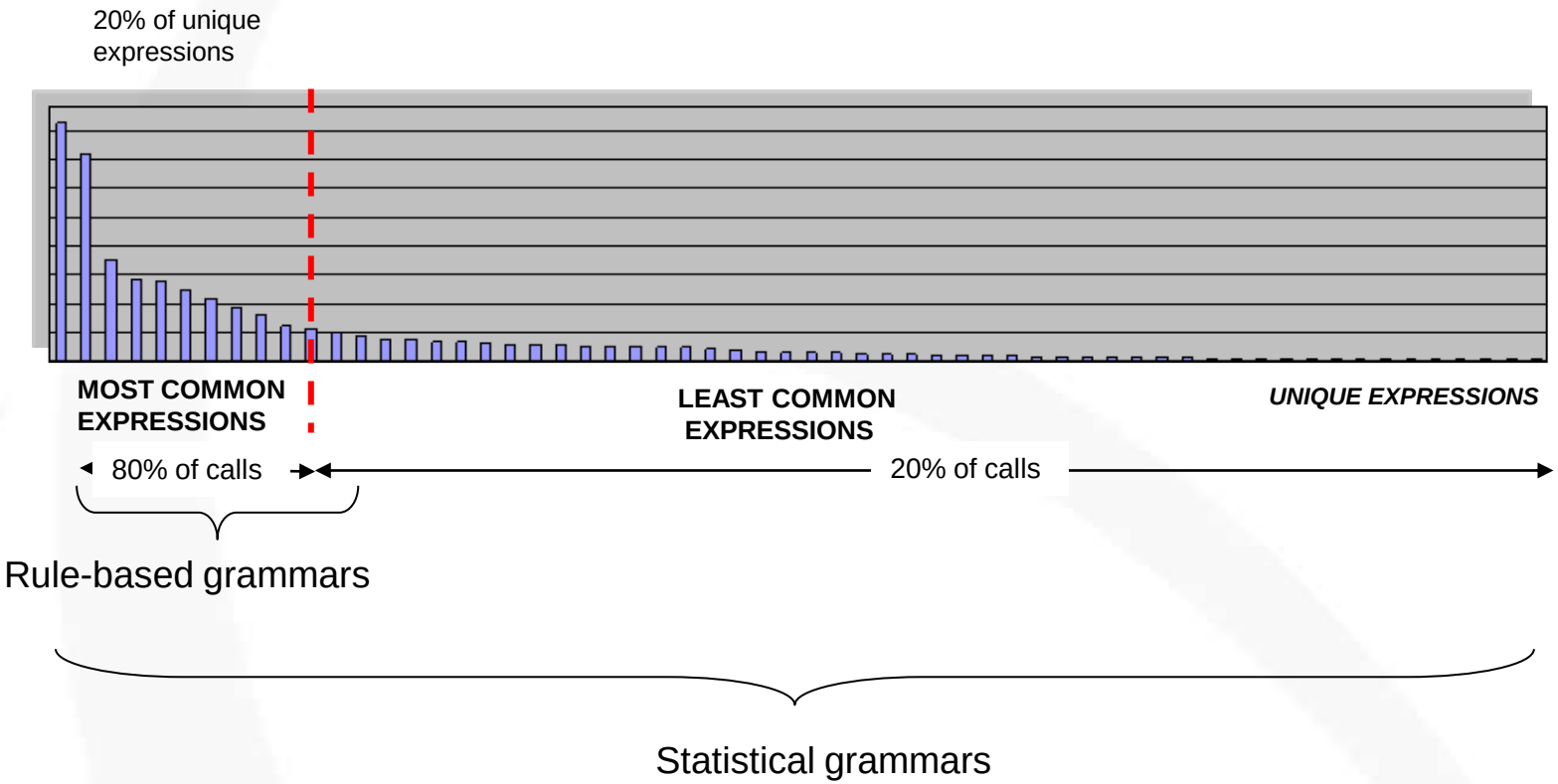
In a directed dialog context a rule-based grammar works better than a statistical language model

WRONG!

It would be true if users would speak **only** and **always** in grammar

But they don't

Language is a long tail phenomenon



The myth of rule-based grammars

In a directed dialog context a rule-based grammar works better than a statistical language model

WRONG!

It would be true if users would speak **only** and **always** in grammar

But they don't

Truth is … statistical language models **always** outperform rule based grammars

The research world knew that for years (since the 1980s…)

So … why are we using handcrafted rule-based grammars at all?

SpeechCycle

…why are we using rule-based grammars at all?

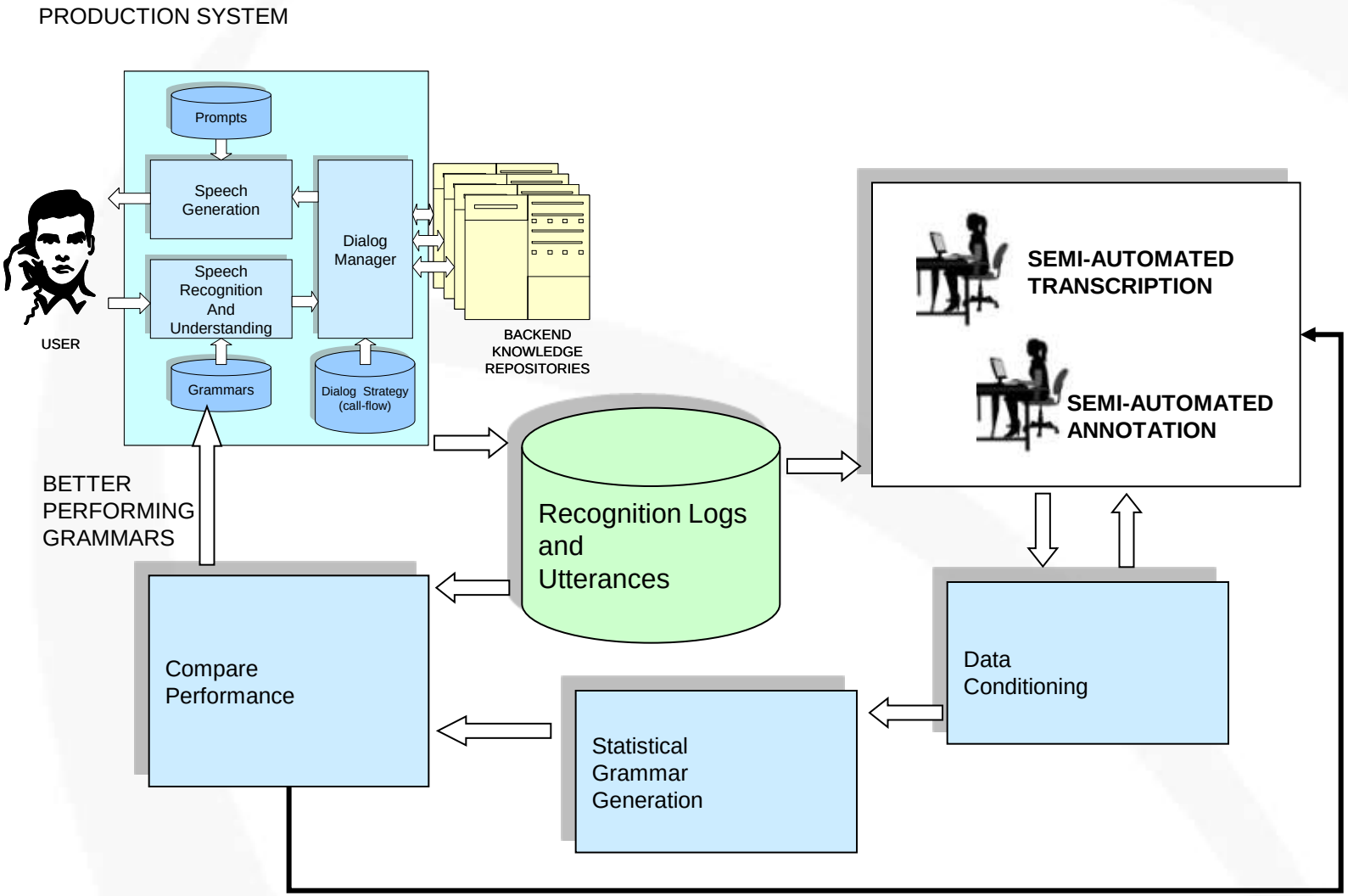
- Building statistical grammars (SLMs) is not that easy
  - You need data—sample utterances—but you may not have data when you start development
  - You need a lot of it
  - Even if you had it… you would need to transcribe each utterance
  - Even if you had the transcriptions, you would need to assign the correct meaning to each one of them (annotation)
    - Annotation may not be straightforward
  - Assuming you have transcriptions and annotations … how do you build a statistical grammar?
    - You need to build a *statistical language model*
    - You need to build a *statistical classifier*
- Statistical grammars are inscrutable … while rule-based grammars are easy to understand
  - You see what a rule-based grammar does …

SpeechCycle

Automated grammar tuning

- Automate the process of creating and tuning SLMs as much as possible
  - The only manual processes are transcription and annotation
  - They can be partially automated too
  - Modern speech scientists do not build and tune grammars but create and manage programs that build and tune grammars.
- Create specialized context-dependent grammars when more data becomes available
  - From dozens to thousands of specialized grammars

Automated grammar tuning



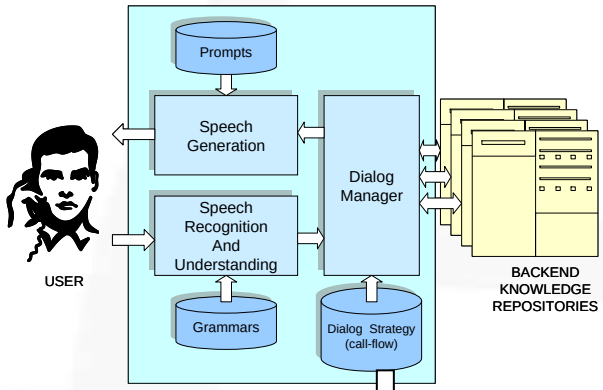
SpeechCycle

Automated grammar tuning

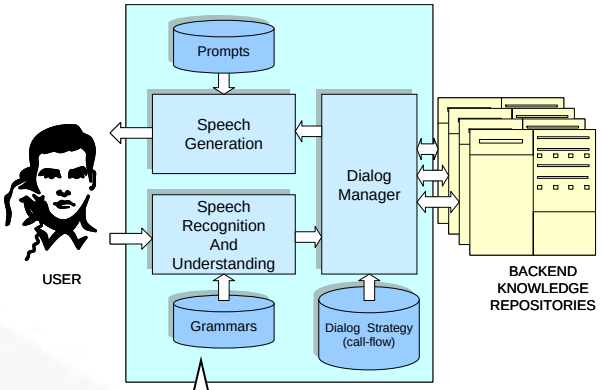
Utterances Calls	2,184,203 533,343
Activities Grammars	2,021 145
Original average accuracy average accuracy to-date	77.97% **90.49%**

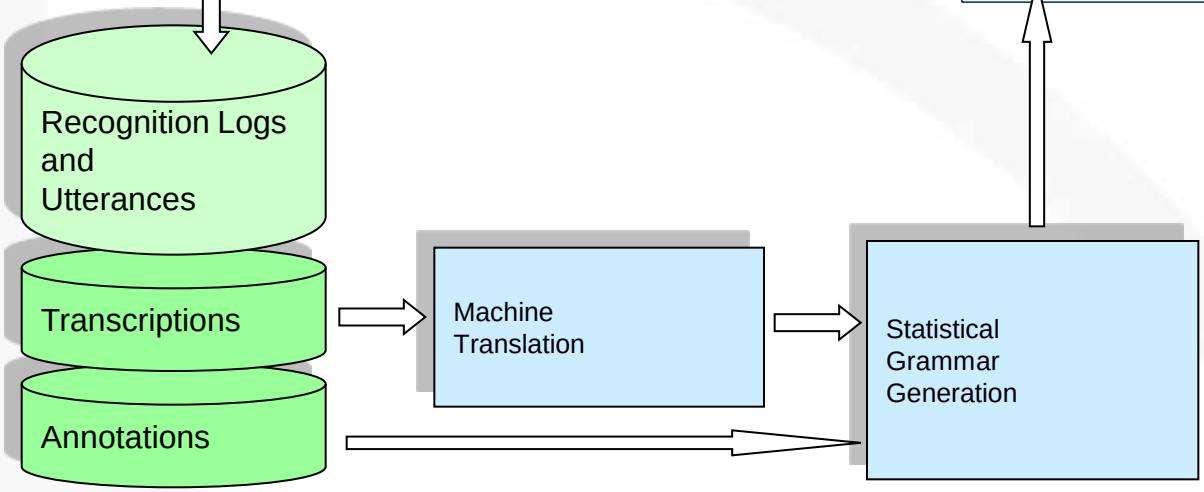
Automated grammar localization

SOURCE LANGUAGE
PRODUCTION SYSTEM (English)

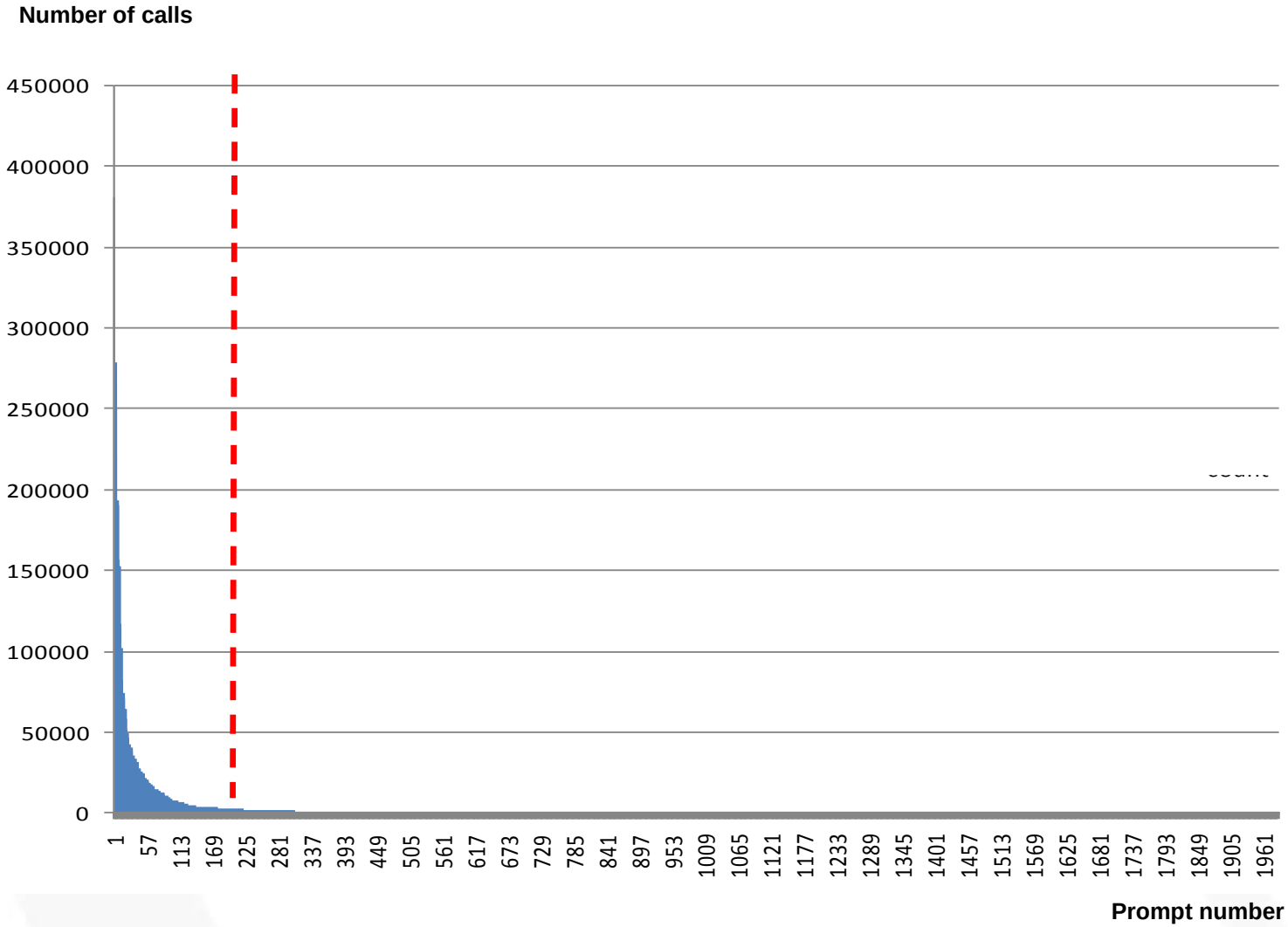


SOURCE LANGUAGE
SYSTEM IN DEVELOPMENT (Spanish)



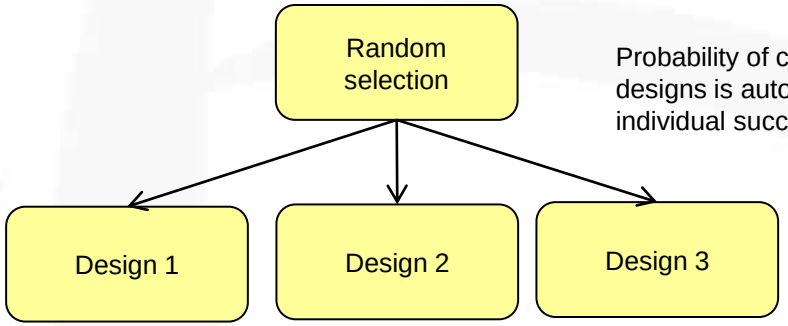


VUI is a long tail phenomenon too

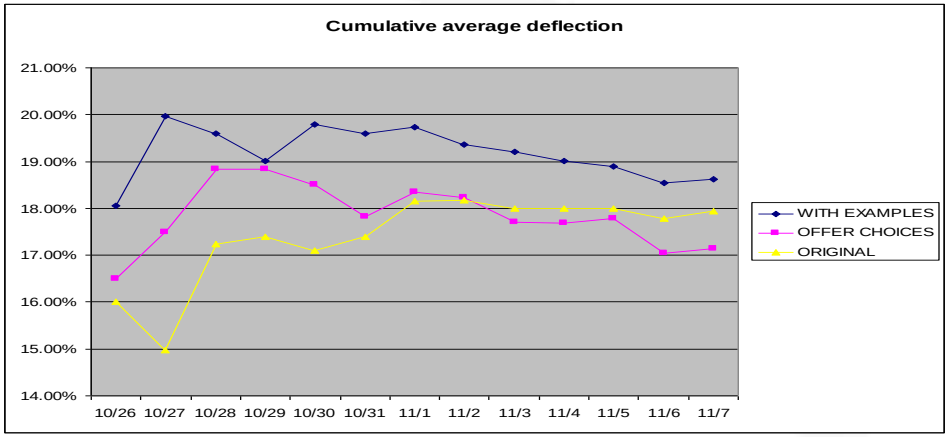


Automated VUI optimization

- Autonomic VUI
  - Use machine learning to optimize among competing designs



Probability of choosing each one of the designs is automatically modified based on individual success and interaction context



SpeechCycle

Conclusions

- From handcrafting to automated optimization of human-machine interaction → Autonomic dialog systems
  - Automated grammar optimization
  - Automated localization
  - Automated VUI optimization