2. History of Conversational Systems

Toyoaki Nishida
Kyoto University

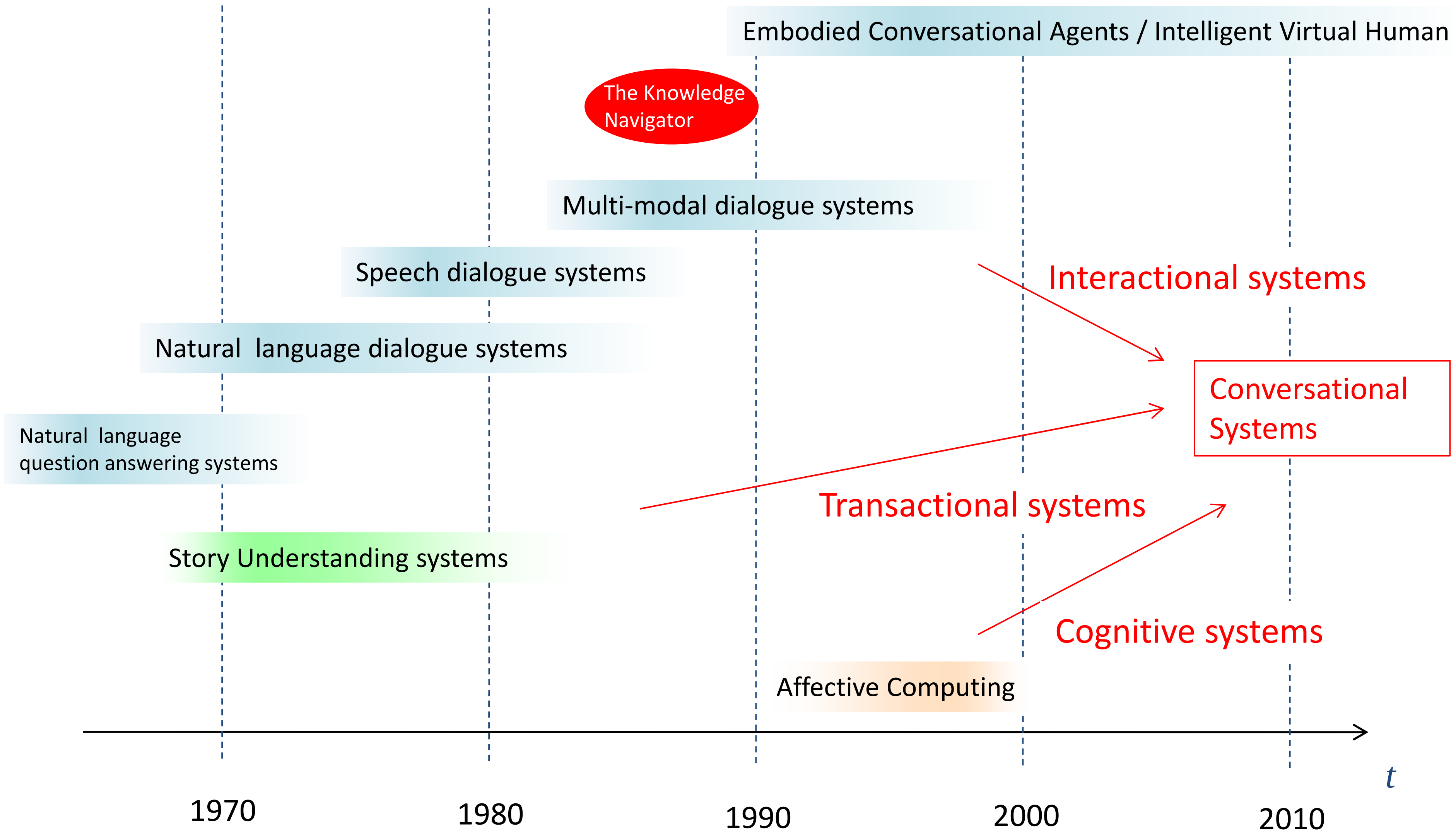
Introduction

Embodied Conversational Agents (ECAs)

"computer interfaces that can hold up their end of the conversation, interfaces that realize conversational behaviors as a function of the demands of dialogue and also as a function of emotion, personality, and social conversation." [Cassell, Sullivan, et al., 2000]

Justine Cassell Women of Vision Award Winner Leadership
http://www.youtube.com/watch?v=eQqDsf2gqYU&feature=channel

History of conversational systems development



Early Natural Language Dialogue Systems

Baseball

- One of the earliest natural language question answering system.
- Answers such questions about baseball games.

Spec[ification] list:

"Where did the Red Sox play on July 7?"
-> Place = ?
Team = Red Sox
Month = July
Day = 7

"What teams won 10 games in July?"
-> Team(winning) = ?
Game(number of) = 10
Month = July

Dictionary:
"team" -> meaning: Team=(blank)
"Red Sox" -> meaning: Team=Red Sox
"who" -> meaning: Team=?
"winning" -> meaning: subroutine

Data:
Month = July
Place= Boston
Day = 7
Game Serial No. = 96
(Team = Red Sox, Score= 5)
(Team = Yankees, Score = 3)

[Green 1961]

Early Natural Language Dialogue Systems

Modules of Baseball

1. Question Read-in
2. Dictionary Look-up
3. Syntax

(1) scan for ambiguities in part of speech: in some cases, resolved by looking at adjoining words; in other cases, resolved by inspecting the entire question.

(2) locates and brackets the noun phrases, [], and the prepositional and adverbial phrases, (). The verb is left unbracketed. E.g.,

"How many games did the Yankees play in July?"
-> [How many games] did [the Yankees] play (in [July])?

Any unbracketed preposition is attached to the first noun phrase in the sentence, and prepositional brackets added. E.g.,

"Who did the Red Sox lose <u>to</u> on July 5?"
-> (<u>To</u> [who]) did [the Red Sox] lose (on [July 5])?

The syntax routine determines whether the verb is active or passive and locates its subject and object.

The syntax analysis checks to see if any of the words is marked as a question word. If not, a signal is set to indicate that the question requires a yes/no answer

4. Content Analysis

The content analysis uses the dictionary meanings and the results of the syntactic analysis to set up a specification list for the processing program. E.g.,

"each team": Team=(blank) -> Team=each
"what team" :Team=(blank)-> Team=?
"Who beat the Yankees on July 4": Team=(blank)
-> Team=?, Team(winning)=? Team(losing)=Yankees
"six games" : Game=(blank)->Game(number of)=6
"how many games":
Game=(blank)->Game(number of)=?
"Who was the winning team…":
Team=? and Team(winning)=(blank)
-> Team(winning)=?

5. Processing

The specification list indicates to the processor what part of the stored data is relevant for answering the input question. The processor extracts the matching information from the data and procedures, for the responder, the answer to the question in the form of a list structure.

[Weizenbaum 1966]

Early Natural Language Dialogue Systems

LUNAR

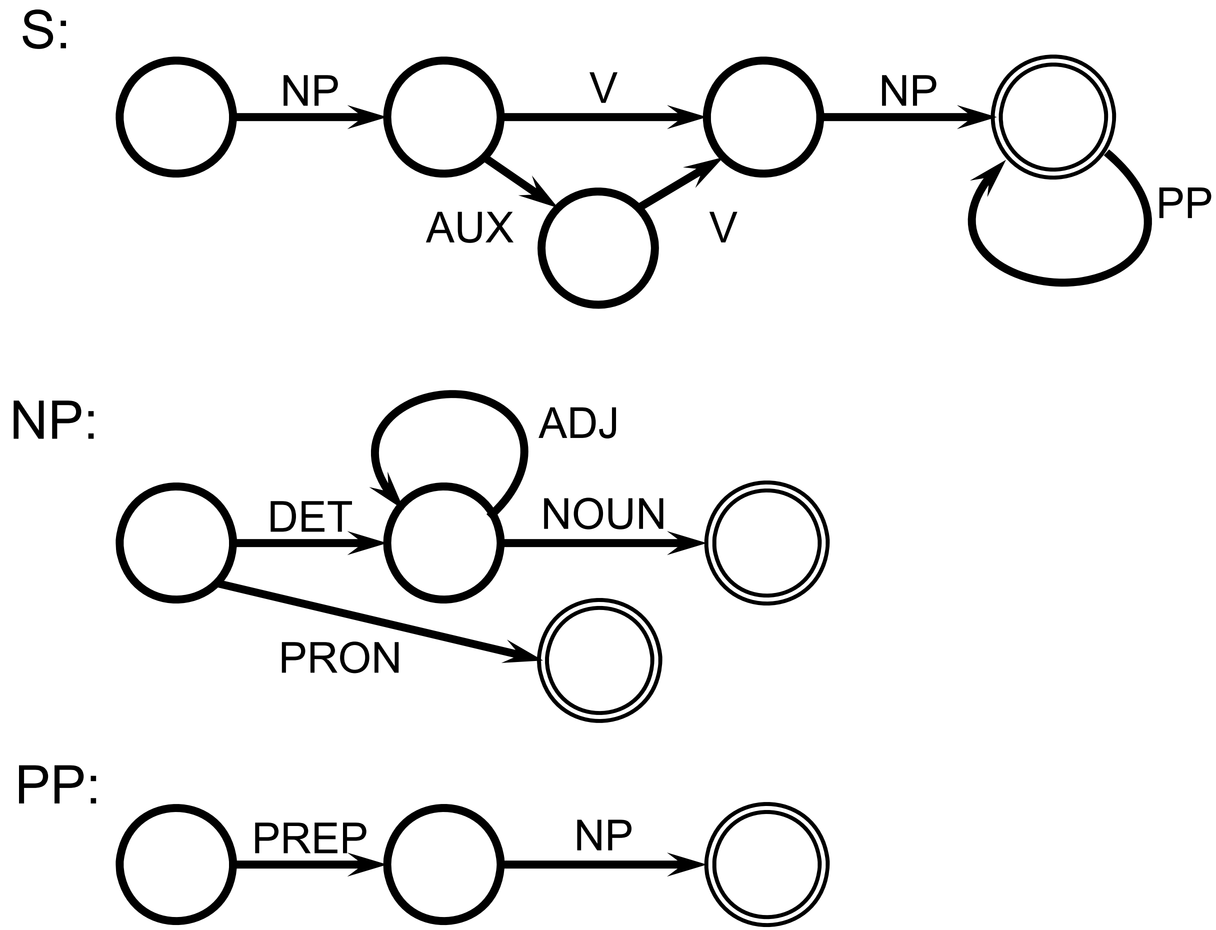
Prototype natural language question answering system that helps lunar geologists access chemical analysis data on lunar rock and soil composition.

(i)Syntactic analysis using heuristic/semantic information to choose the most likely parsing (Augmented Transition Network Grammar was used)
(ii) Semantic interpretation: produce a formal representation for queries
(iii) Execution of this formal expression in the retrieval

[Woods 1973]

Early Natural Language Dialogue Systems

Augmented Transition Network Grammar



[Woods 1973]

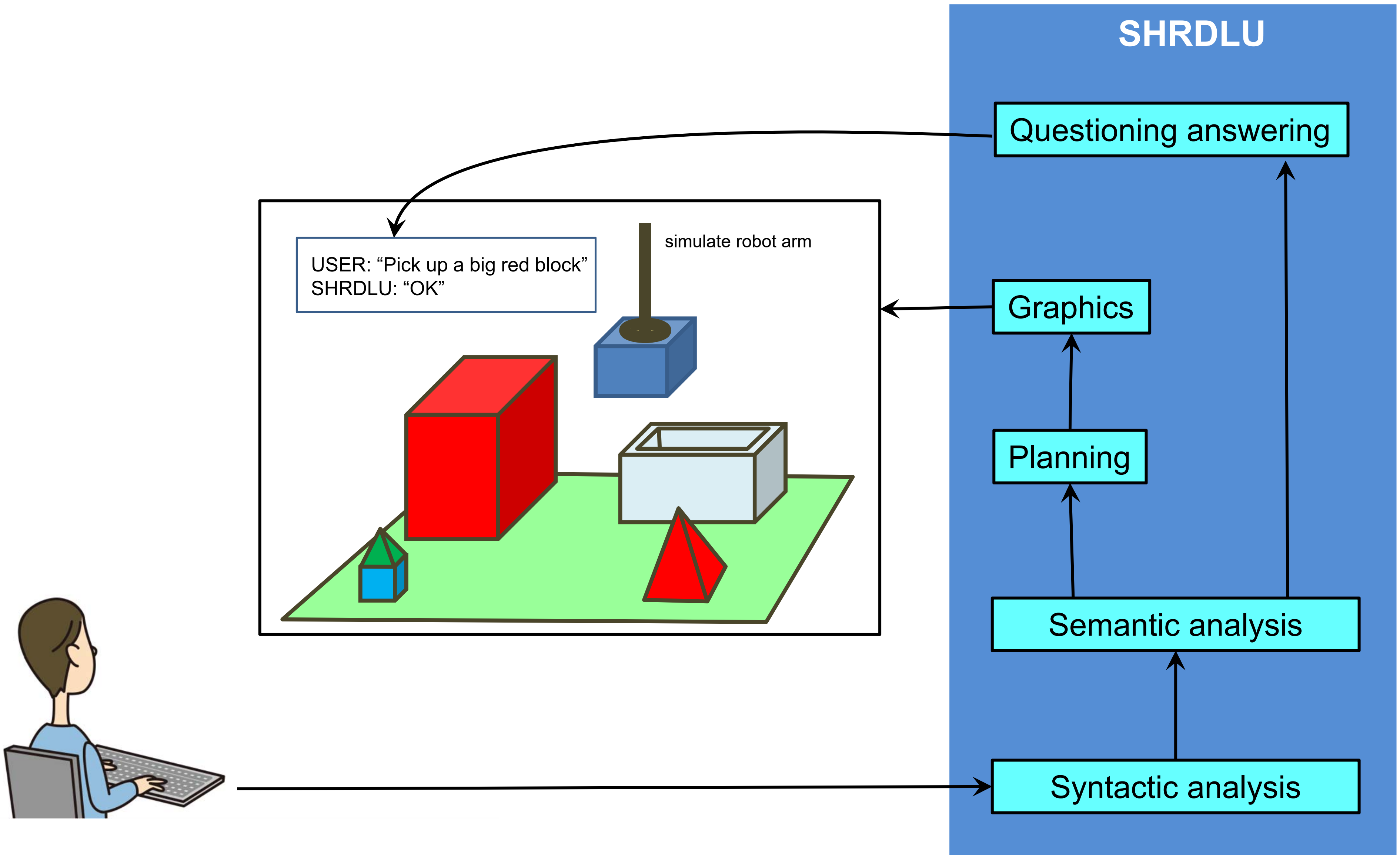
Early Natural Language Dialogue Systems

SHRDLU

- Natural language understanding system working on the "Blocks" world.
- Based on the belief that "a computer cannot deal with language unless it can understand the subject it is discussing"
- Answers questions, executes commands, and accepts information in normal English dialog.
- Uses semantic information and context to understand discourse and disambiguate sentences.
- Procedural knowledge representation <-> declarative KR

Winograd, Terry, *Understanding Natural Language,* New York: Academic Press, 1972. http://hci.stanford.edu/~winograd/shrdlu/

Early Natural Language Dialogue Systems



Winograd, Terry, *Understanding Natural Language,* New York: Academic Press, 1972. http://hci.stanford.edu/~winograd/shrdlu/

Early Natural Language Dialogue Systems

ELIZA: Natural language dialogue system without explicit knowledge about the discourse domain

Transformation rules: collection of key lists.

$$(K\ ((D_1)\ (R_{1,1})\ (R_{1,2})\cdots(R_{1,m_1}))$$
$$((D_2)\ (R_{2,1})\ (R_{2,2})\cdots(R_{2,m_2}))$$
$$\cdots$$
$$((D_n)\ (R_{n,1})\ (R_{n,2})\cdots(R_{n,m_n})))$$

where, K: keyword, D_i: decomposition template, $R_{i,j}$: Reassembly rule.

Example of a decomposition template and a reassembly rule:

D: (0 YOU 0 ME)

R: (WHAT MAKES YOU THINK I 3 YOU)

"It seems that you hate me"

-> "What makes you think I hate you"

Substitution rule. E.g.,

(YOURSELF = MYSELF)

(MY YOUR 5 (transformation rules))

Example:

"You are very helpful"

-> "I are very helpful"

by a simple substitution rule

-> "What makes you think I am very helpful"

by a decomposition template: (0 I are 0)

by a reassembly rule:

(What makes you think I am 4)

After $R_{i,j}$ was applied, an index in the transformation list is inserted to prevent the same reassembly rule from being applied in a row.

(How = (What)) allows the transformation rule for "how" to be equally applied to "what"

(MOTHER DLIST (/NOUN FAMILY))

allows the keyword "MOTHER" to be identified as a noun and as a member of the class "family". As a result, "MOTHER" will match "(/FAMILY)" in a decomposition rule.

[Weizenbaum 1966]

Story understanding and generation

"Mary socked John."
"Mary punched John."
"Mary hit John with her fist."
"John was socked by Mary."
"Marie a donne un coup de poing a Jean."
"Maria pego a Juan."

Margie ➡

Action: PROPEL
Actor: Mary
Object: Fist
From: Mary
To: John

Primitive	Meaning
ATRANS	Transfer of an abstract relationship (i.e. give)
PTRANS	Transfer of the physical location of an object (e.g., go)
PROPEL	Application of physical force to an object (e.g. push)
MOVE	Movement of a body part by its owner (e.g. kick)
GRASP	Grasping of an object by an action (e.g. throw)
INGEST	Ingesting of an object by an animal (e.g. eat)
EXPEL	Expulsion of something from the body of an animal (e.g. cry)
MTRANS	Transfer of mental information (e.g. tell)
MBUILD	Building new information out of old (e.g decide)
SPEAK	Producing of sounds (e.g. say)
ATTEND	Focusing of a sense organ toward a stimulus (e.g. listen)

[Schank 1975]

Story understanding and generation

Script-based understanding

SAM (Script Applier Mechanism)

[Cullingford 1981] Cullingford, Richard E: SAM and Micro SAM. In Roger C. Schank, & Christopher K. Riesbeck (Eds.), Inside computer understanding. Hillsdale, NJ: Erlbaum, 1981

FRUMP

[Dejong 1979] DeJong, Gerald F.: Skimming stories in real time: An experiment in integrated understanding (Technical Report YALE/DCS/tr158). New Haven, CT: Computer Science Department, Yale University, 1979

Story understanding and generation

Plan-based understanding

PAM

[Wilensky 1978] Wilensky, Robert: Understanding goal-based stories (Technical Report YALE/DCS/tr140). New Haven, CT: Computer Science Department, Yale University, 1978.

POLITICS

[Carbonell 1978] Carbonell, Jaime: Subjective understanding: Computer models of belief systems (Technical Report YALE/DCS/tr150). New Haven, CT: Computer Science Department, Yale University, 1978.

Dynamic Memory

IPP

[Lebowitz 1980] Lebowitz, Michael : \Generalization and memory in an integrated understanding system (Technical Report YALE/DCS/tr186). New Haven, CT: Computer Science Department, Yale University, 1980.

BORIS

[Lehnert 1983] Lehnert, Wendy G., Dyer, Michael G., Johnson, Peter N., Yang, C. J., Harley, Steve: BORIS -- An experiment in in-depth understanding of narratives. Artificial Intelligence, 20(1), 15-62., 1983.

CYRUS

[Kolodner 1984] Kolodner, Janet L.: Retrieval and organizational strategies in conceptual memory: A computer model. Hillsdale, NJ: Erlbaum, 1984.

Story telling

TALE-SPIN

[Meehan 1981] Meehan, James: TALE-SPIN and Micro TALE-SPIN. In Roger C. Schank, & Christopher K. Riesbeck (Eds.), Inside computer understanding. Hillsdale, NJ: Erlbaum, 1981.

Early Speech Dialogue Systems

The Hearsay-II Speech-Understanding System

Integrates multiple levels of information processing by knowledge sources coordinated by the blackboard model:

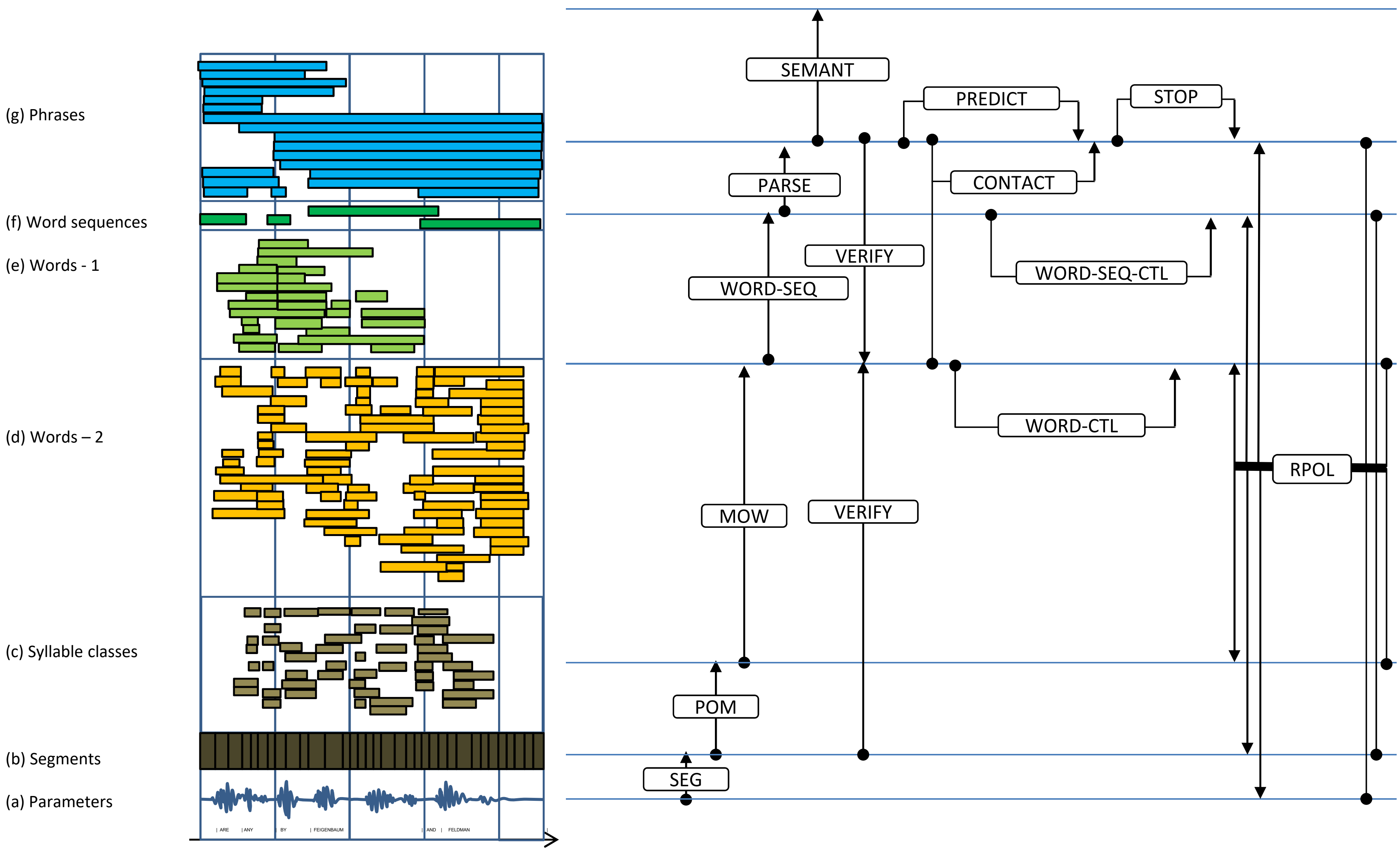
- Parameter
- Segment
- Syllable
- Word
- Word-sequence
- Phrase
- Data base interface

Combining top-down (hypothesis driven) and bottom-up (data-driven) processing.

Selective attention to allocate limited computing resources.

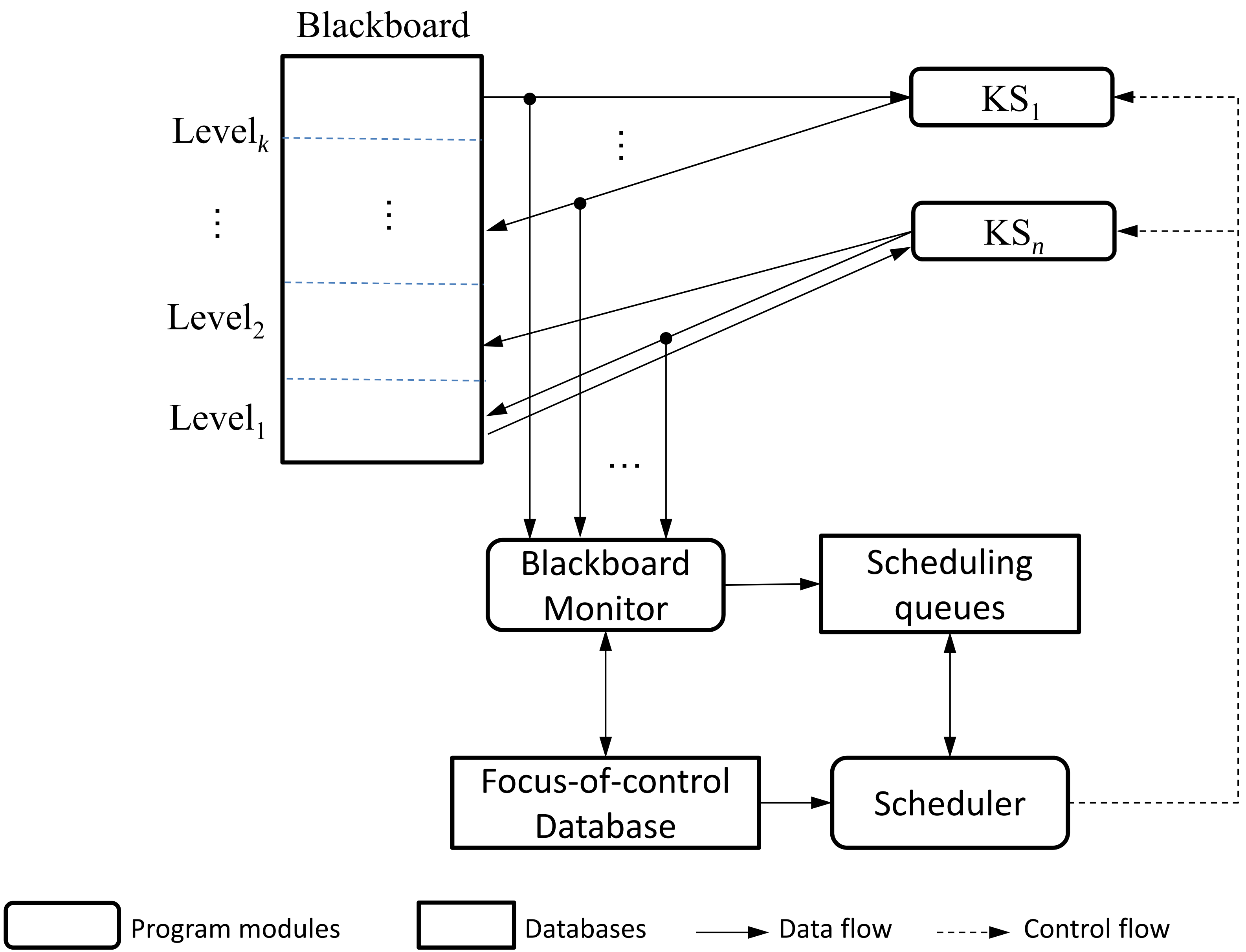
[Erman 1980]

The blackboard architecture



[Erman 1980]

The blackboard architecture



[Erman 1980]

Early Speech Dialogue Systems

PUT-THAT-THERE

The conjoint use of voice-input and gesture recognition to command events on a large format raster-scan graphic display in a "Media Room"

Commands

"Create a blue square there."
"Move the blue triangle to the right of the green square"
"Move <u>that</u> to the right of the green square." (with pointing)
"Put that <u>there</u>" (indicated by gesture)
"Make <u>that</u> smaller" (with pointing gesture)
"Make <u>that</u> (indicating some item) like <u>that</u> (indicating some other item)"
"Delete <u>that</u>" (pointing to some item)
"Call <u>that</u> … the calendar" (with pointing)

[Bolt 1980]

The Knowledge Navigator

"Phil": concrete image of a personal assistant.

[Apple Computer 1987]

CASA: Computers Are Social Actors

Social responses to computers

- not: conscious beliefs that computers are human or human-like.
- not: from user's ignorance or psychological or social dysfunctions
- not from: a belief that subjects are interacting with programmers
- rather: human-computer relationship is fundamentally social.

Nass 1994]

From Tool-based Computing to Assistive Interface

Requirements

- **Support interactive give and take**

  Assistants will need not only respond to questions but also ask questions

- **Recognize the costs of interaction and delay**

  It is inappropriate to require the user's confirmation of every decision made while carrying out a task.

- **Manage interruptions effectively**

  When it is necessary to initiate an interaction with the user, the assistant needs to do so carefully, recognizing the likelihood that the user is already occupied to some degree.

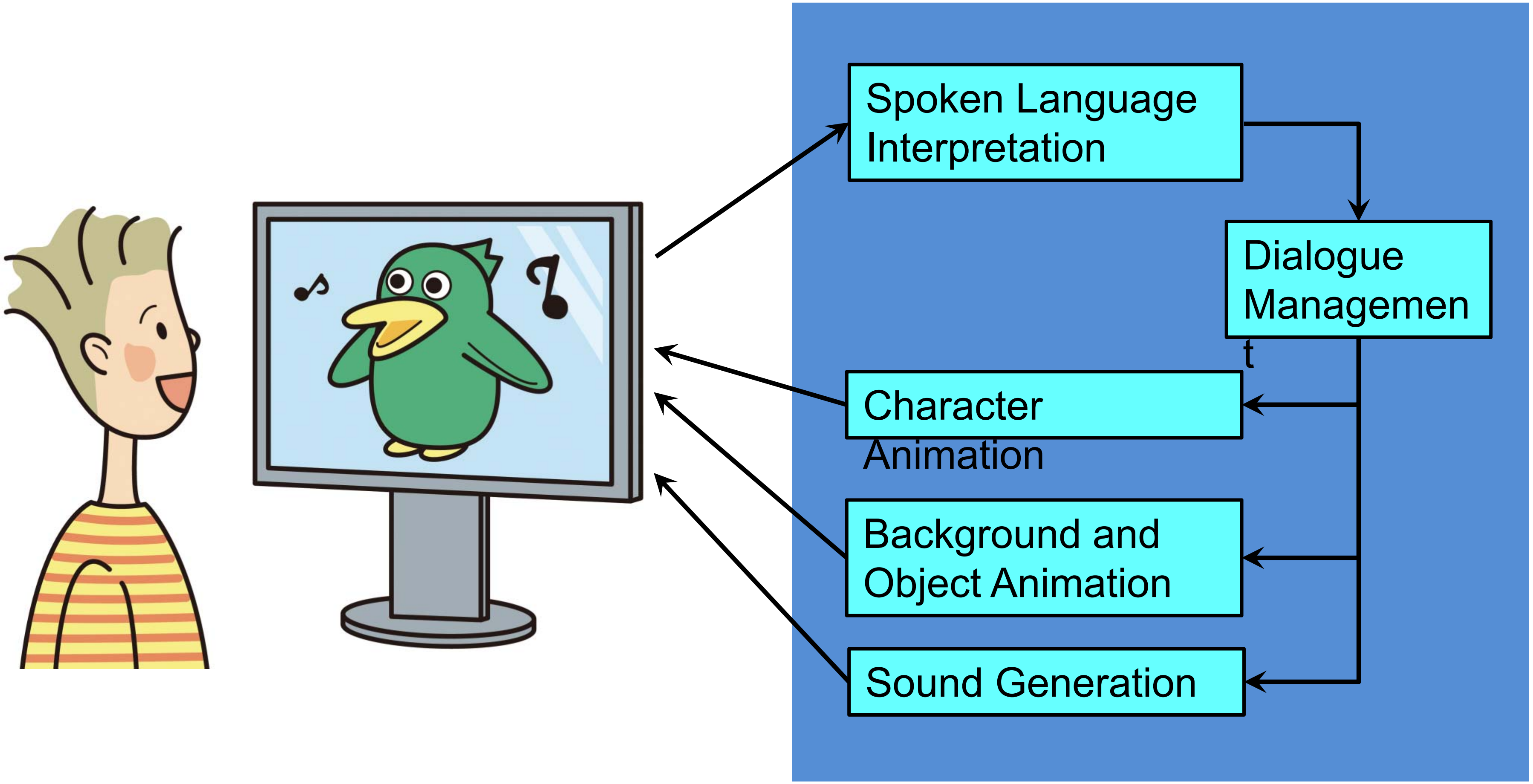
- **Acknowledge the social and emotional aspects of interaction**

  To become a comfortable working partner, a computer assistant will need to vary its behavior depending on such the task, the time of day, and the boss's mood.

=> The PERSONA project @ Microsoft Research

[Ball 1997]

The architecture of Peedy



[Ball 1997]

Believability

Believable agents

provides the illusion of life, and permits the audience's suspension of disbelief.
… critical in theater, film, animation, radio drama, …

Questions

i. What makes characters believable? What makes agents believable?
Are agents different from characters?

ii. What is the nature of current integrated architectures and situated agents?
How well do they support believability? Are there clear areas demanding study?

iii. How much breadth of capability is necessary to produce a believable agent? How much depth (competence) is necessary?

iv. What are the respective role of movement and language in achieving believability?

v. What is the role of context in establishing expectations in the user and thus is simplifying the task? For instance, how can the setting of an interface agent or the theme of a story help limit the technical requirements on agents?

vi. How can we measure believability and progress toward believability?
Why don't artists use "scientifically valid" techniques to evaluate the believability of their characters?

[Bates 1994]

Oz project

Interactive Drama :

Interactive drama concerns itself with building dramatically interesting virtual worlds inhabited by computer-controlled characters, within which the user (the player) experiences a story from a first person perspective. … (Bates 1992)

Drama = Character + Story + Presentation

[Mateas 1997]

Oz project

Personality

Rich personality should infuse everything that a character does. What makes characters interesting are their unique ways doing things.

Emotion

Characters exhibit their own emotions and respond to the emotions of others in personality-specific ways.

Self-motivation

Characters have their own internal drives and desires which they pursue whether or not others are interacting with them.

Change

Characters grow and change with time, in a manner consistent with their personality.

Social relationships

Characters engage in interactions with others in a manner consistent with their relationship. In turn, these relationships change as a result of the interaction.

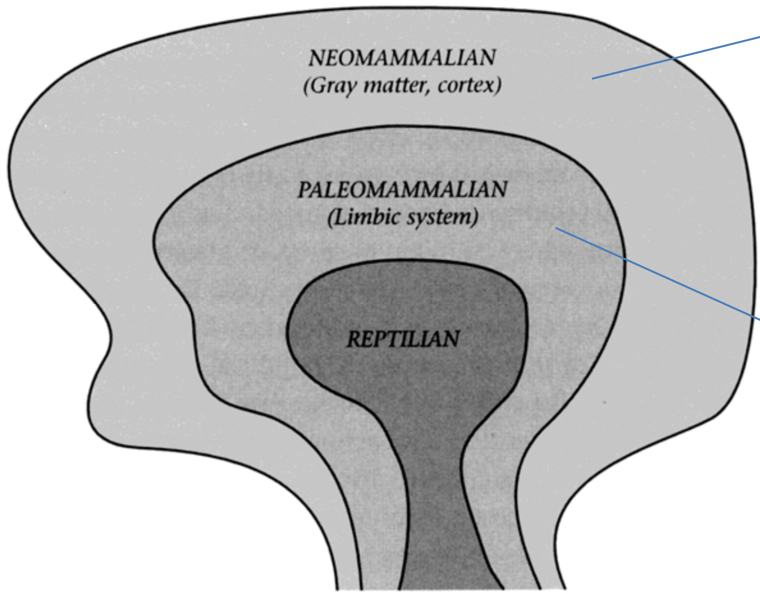
Illusion of life

Pursuing multiple, simultaneous goals and actions, having broad capabilities, and reacting quickly to stimuli in the environment.

[Mateas 1997]

Affective computing

– A principle to design emotional intelligence that can change the behavior according to the affective status of the partner.



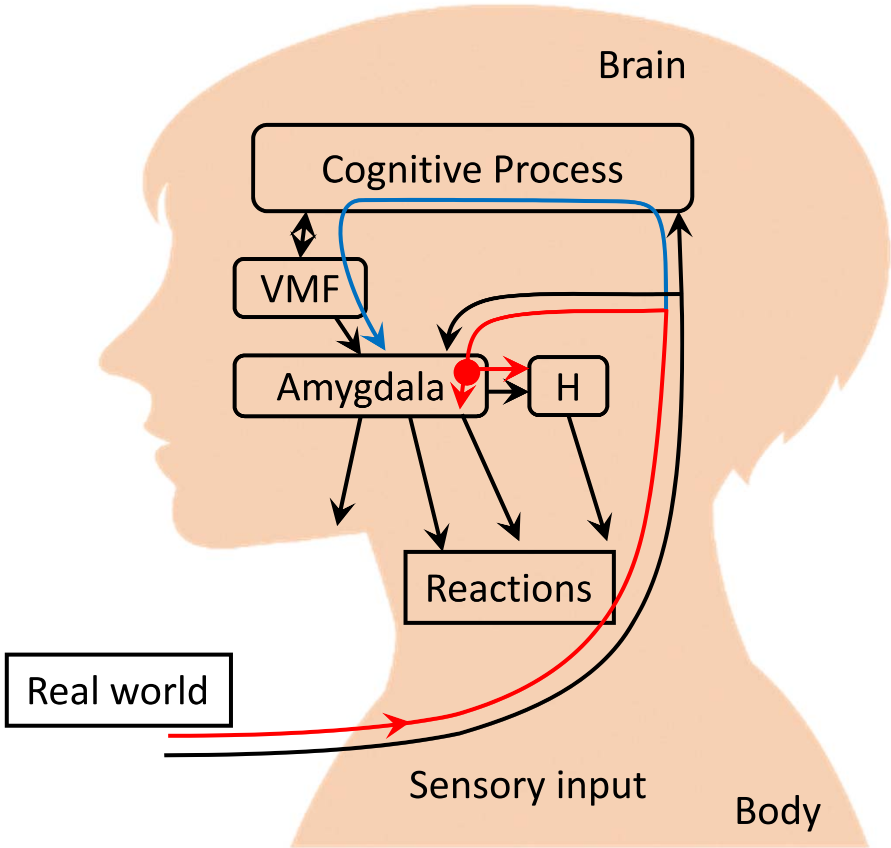
Majority of perceptual processing has been assumed to occur.

The limbic system is considered the primary seat of emotion, attention, and memory.

Paul MacLean's "triune brain"

[Picard 1997]

Antonio Damasio's "Descartes' Error – Emotion, Reason and the Human Brain"



H: Hypothalamus,
VMF: ventromedial prefrontal cortex

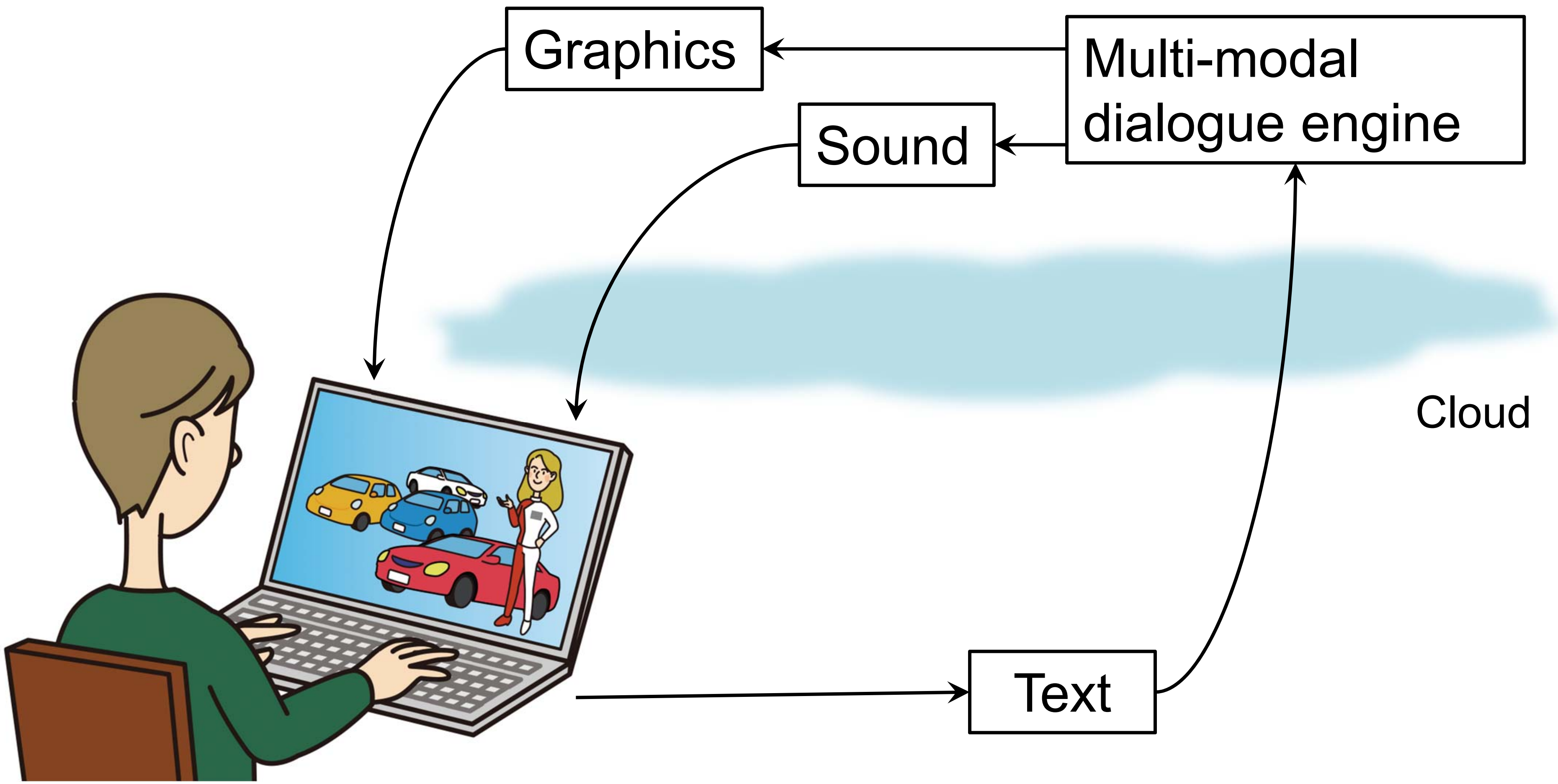
If you come to know that animal or object or situation *X* causes fear, you will have two ways of behaving toward *X*. *The first way* is innate; you do not control it. Moreover, it is not specific to *X*; a large number of creatures, objects, and circumstances can cause the response. *The second way* is based on your own experience and is specific to *X*. Knowing about *X* allows you to think ahead and predict the probability of its being present in a given environment, so that you can avoid *X*, preemptively, rather than just have to react to its presence in an emergency. …

Primary emotions depend on limbic system circuitry, the amygdala and anterior cingulate being the primary players. After an appropriate stimulus activates the amygdala, a number of responses ensue: internal responses, muscular responses, visceral responses, and responses to neurotransmitter nuclei and hypothalamus.

Secondary emotions utilize the machinery of primary emotions. The stimulus may still be processed directly via the amygdala but is now also analyzed in the thought process, and may activate frontal cortices (VM). VM acts via the amygdala. (p. 133-137)

[Damasio 1994]

Jennifer James



[Hayes-Roth 1998]

Oz project

Personality

Rich personality should infuse everything that a character does. What makes characters interesting are their unique ways doing things.

Emotion

Characters exhibit their own emotions and respond to the emotions of others in personality-specific ways.

Self-motivation

Characters have their own internal drives and desires which they pursue whether or not others are interacting with them.

Change

Characters grow and change with time, in a manner consistent with their personality.

Social relationships

Characters engage in interactions with others in a manner consistent with their relationship. In turn, these relationships change as a result of the interaction.

Illusion of life

Pursuing multiple, simultaneous goals and actions, having broad capabilities, and reacting quickly to stimuli in the environment.

[Mateas 1997]

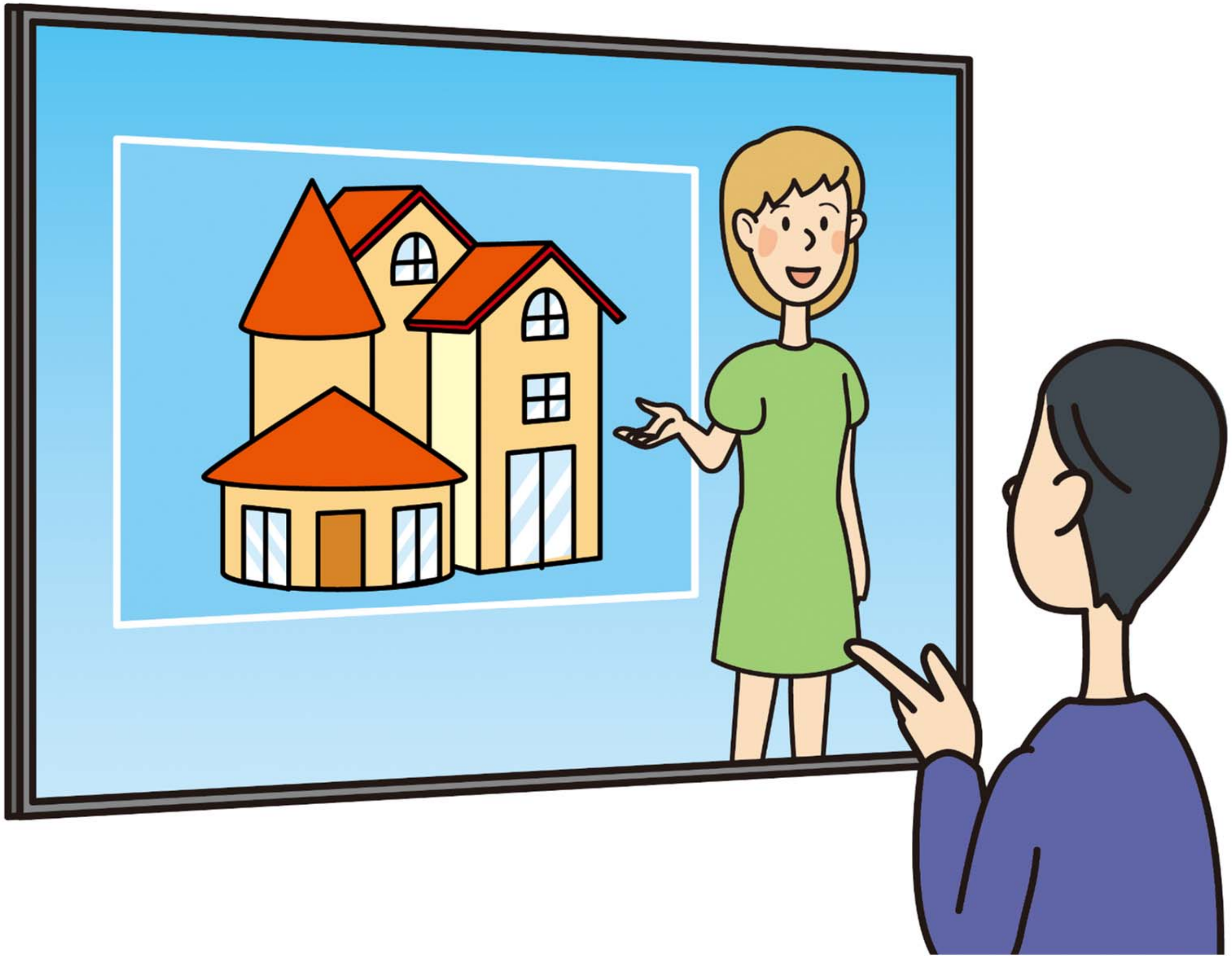
Rea

Implements the social, linguistic, and psychological conventions of conversation.

- Has a human-like body. Uses eye gaze, body posture, hand gestures, and facial displays to organize and regulate the conversation.
- The conversational model relies on the function of non-verbal behaviors as well as speech.
- A full symmetry between input and output modalities: not only respond to visual, audio and speech cues (such as speech, shifts in gaze, gesture, and non-speech audio but also generate these cues.

[Cassell 1999]

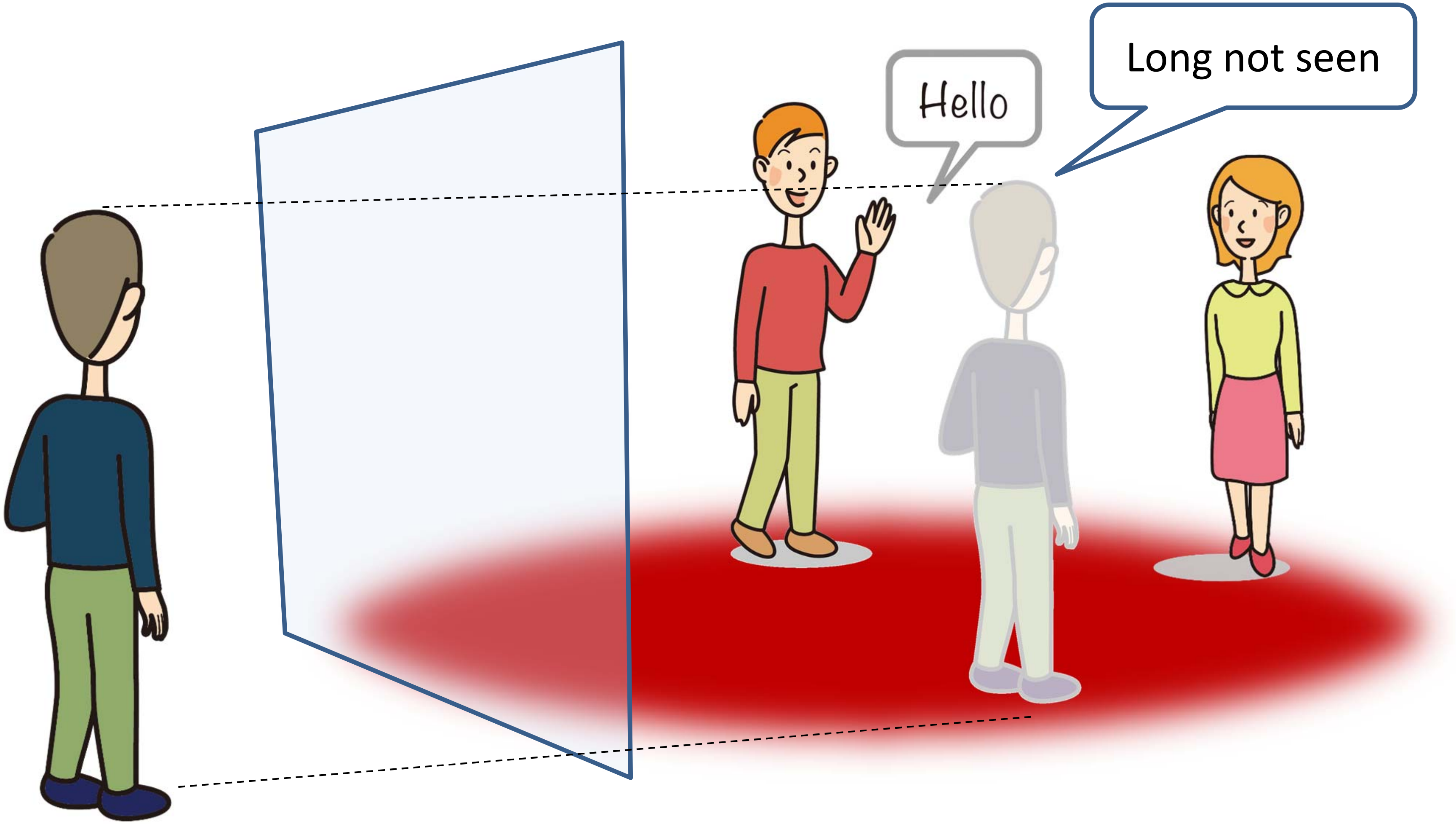
Rea



[Cassell 1999]

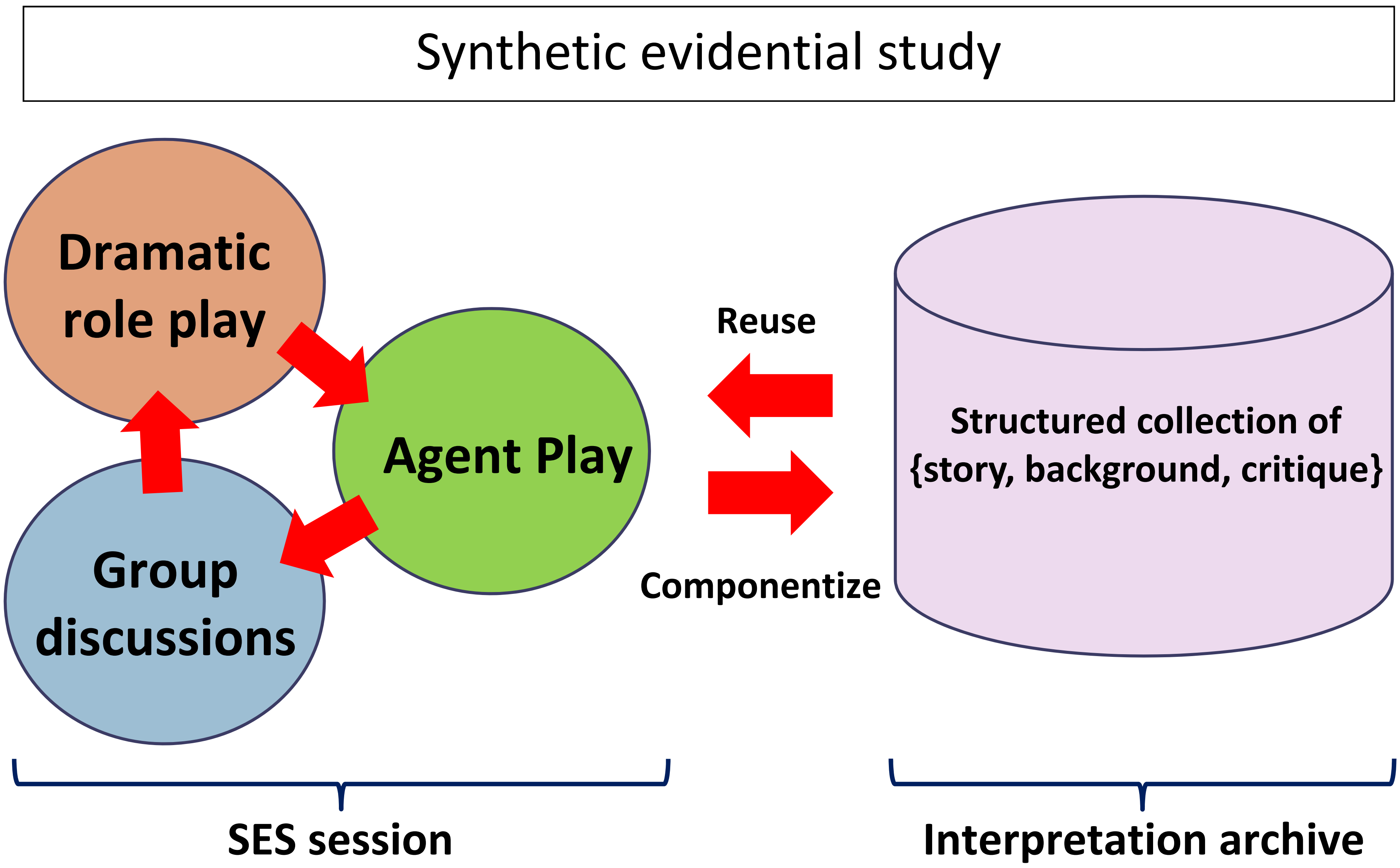
Façade

Interactive Drama



Dynamically reorganizing beats (story components) according to the input

[Mateas 2010]



Synthetic evidential study (SES) combines dramatic role play and group discussion to help people spin stories by bringing together partial thoughts and evidence.

Summary

1. Conversational artifacts have a long history of development.
2. The development was based on three lines of research: interactional systems, transactional systems, and cognitive systems.
3. Studies on interactional systems stated as researches on natural language dialogue system, followed by speech dialogue systems, multi-modal interaction systems, and embodied conversational agents (or intelligent virtual agents).
4. Studies on transactional systems focused on story generation and understanding systems.
5. Studies on cognitive systems center on affective computing and theory of mind systems.
6. Interactive drama integrates the three approaches.
7. Content production is critical.

References

[Apple Computer 1987] http://homepage.mac.com/ericestrada/Movies/iMovieTheater53.html

[Ball 1997] Gene Ball, Dan Ling, David Kurlander, John Miller, David Pugh, Tim Skelly, Andy Stankosky, David Thiel, Maarten Van Dantzich, and Trace Wax. Lifelike Computer Characters: The Persona Project at Microsoft Research. Software Agents. Jeffrey M. Bradshaw (ed.). AAAI/MIT Press. Menlo Park, CA. 1997.

[Bates 1994] Joseph Bates. The role of emotion in believable agents. Communications of ACM, Vol. 37, No. 7, pp. 122-125, 1994.

[Bolt 1980] Richard A. Bolt: Put-that-there": Voice and gesture at the graphics interface. In Proceedings of the 7th annual conference on Computer graphics and interactive techniques, Vol. 14, No. 3. (July 1980), pp. 262-270.

[Erman 1980] Lee D. Erman, Frederick Hayes-Roth, Victor R. Lesser, and D. Raj Reddy. The Hearsay-II Speech-Understanding System: Integrating Knowledge to Resolve Uncertainty, ACM Computing Surveys, Volume 12, Issue 2, Pages: 213 - 253, 1980.

[Cassell 1999] Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H. and Yan, H. (1999). "Embodiment in Conversational Interfaces: Rea." Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), pp. 520-527. Pittsburgh, PA.

[Green 1961] Bert F. Green, Jr., Alice K. Wolf, Carol Chomsky, and Kenneth Laughery: Baseball: An Automatic Question-Answerer Proceedings of the IRE-AIEE-ACM '61 (Western); western joint IRE-AIEE-ACM computer conference, 1961.

[Hayes-Roth 1998] Barbara Hayes-Roth: Jennifer James, celebrity auto spokesperson, International Conference on Computer Graphics and Interactive Techniques archive ACM SIGGRAPH 98 Conference abstracts and applications table of contents, P. 136, 1998.

[Mateas 2010] Michael Mateas and Andrew Stern. Writing Façade: A Case Study in Procedural Authorship. In Harrigan, P., Wardrip-Fruin, N.: Second Person: Role-Playing and Story in Games and Playable Media, MIT Press, pp. 183-207, 2010.

[Weizenbaum 1966] Joseph Weizenbaum. ELIZA -- A Computer Program for the study of natural language communication between man and machine, Communications of the ACM, Vol. 9, No. 1, pp. 36-45, 1966.

[Winograd 1972] Terry Winograd: Understanding Natural Language, Academic Press, 1972.

[Woods 1973] W. A. Woods: Progress in natural language understanding: an application to lunar geology, In Proceeding AFIPS '73 Proceedings of the June 4-8, 1973, national computer conference and exposition.