Speech recognition and synthesis

Introduction

Successful *Automatic Dialog Systems* must

- Handle numerous different users
- Incite effective user expectations
- Fail gracefully (eg, with human back-up)
- Allow multimodal interaction, if at all possible
- Allow user initiative
- *Automatic Dialog Systems* are as much an *ergonomic* as a *speech* technology problem

Many pictures (and their copyrights) are from
[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Introduction

Successful *Automatic Dialog Systems* must

- Handle numerous different users
- Incite effective user expectations
- Fail gracefully (eg, with human back-up)
- Allow multimodal interaction, if at all possible
- Allow user initiative
- *Automatic Dialog Systems* are as much an *ergonomic* as a *speech* technology problem

Many pictures (and their copyrights) are from
[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Introduction

Successful *Automatic Dialog Systems* must

- Handle numerous different users
- Incite effective user expectations
- Fail gracefully (eg, with human back-up)
- Allow multimodal interaction, if at all possible
- Allow user initiative
- *Automatic Dialog Systems* are as much an *ergonomic* as a *speech* technology problem

Many pictures (and their copyrights) are from
[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Introduction

Successful *Automatic Dialog Systems* must

- Handle numerous different users
- Incite effective user expectations
- Fail gracefully (eg, with human back-up)
- Allow multimodal interaction, if at all possible
- Allow user initiative
- *Automatic Dialog Systems* are as much an *ergonomic* as a *speech* technology problem

Many pictures (and their copyrights) are from
[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Introduction

Successful *Automatic Dialog Systems* must

- Handle numerous different users
- Incite effective user expectations
- Fail gracefully (eg, with human back-up)
- Allow multimodal interaction, if at all possible
- Allow user initiative
- *Automatic Dialog Systems* are as much an *ergonomic* as a *speech* technology problem

Many pictures (and their copyrights) are from
[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Introduction

Successful *Automatic Dialog Systems* must

- Handle numerous different users
- Incite effective user expectations
- Fail gracefully (eg, with human back-up)
- Allow multimodal interaction, if at all possible
- Allow user initiative
- *Automatic Dialog Systems* are as much an *ergonomic* as a *speech* technology problem

Many pictures (and their copyrights) are from
[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Introduction

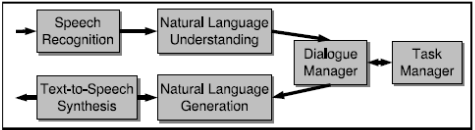


Figure 19.5 Simplified architecture of the components of a conversational agent.

Automatic Dialog Systems have the combined limitations of:

- ASR + NLP: The real bottleneck
- NLG + TTS: Normally not a problem
- Dialog management + database: A bottleneck in complex tasks

[Jurafsky and Martin(2000)]

Introduction

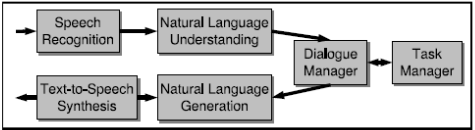


Figure 19.5 Simplified architecture of the components of a conversational agent.

Automatic Dialog Systems have the combined limitations of:

- **ASR + NLP**: The real bottleneck
- NLG + TTS: Normally not a problem
- Dialog management + database: A bottleneck in complex tasks

[Jurafsky and Martin(2000)]

Introduction

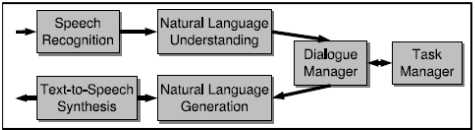


Figure 19.5 Simplified architecture of the components of a conversational agent.

Automatic Dialog Systems have the combined limitations of:

- **ASR + NLP**: The real bottleneck
- **NLG + TTS**: Normally not a problem
- Dialog management + database: A bottleneck in complex tasks

[Jurafsky and Martin(2000)]

Introduction

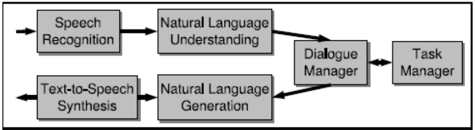


Figure 19.5 Simplified architecture of the components of a conversational agent.

Automatic Dialog Systems have the combined limitations of:

- **ASR + NLP**: The real bottleneck
- **NLG + TTS**: Normally not a problem
- **Dialog management + database**: A bottleneck in complex tasks

[Jurafsky and Martin(2000)]

Conversational Human-Computer Interaction: Practical dialogs

General conversations are much too complex. Limit *Automatic Dialog Systems* to practical dialogues

Dialogues that are focused on a concrete task, eg,

- Task-oriented
- Information seeking
- Advice and tutoring
- Command and control

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Conversational Human-Computer Interaction: Practical dialogs

General conversations are much too complex. Limit *Automatic Dialog Systems* to practical dialogues

Dialogues that are focused on a concrete task, eg,

- Task-oriented
- Information seeking
- Advice and tutoring
- Command and control

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Conversational Human-Computer Interaction: Practical dialogs

General conversations are much too complex. Limit *Automatic Dialog Systems* to practical dialogues

Dialogues that are focused on a concrete task, eg,

- Task-oriented
- Information seeking
- Advice and tutoring
- Command and control

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Conversational Human-Computer Interaction: Practical dialogs

General conversations are much too complex. Limit *Automatic Dialog Systems* to practical dialogues

Dialogues that are focused on a concrete task, eg,

- Task-oriented
- Information seeking
- Advice and tutoring
- Command and control

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Conversational Human-Computer Interaction

The Practical Dialogue Hypothesis

The conversational competence required for practical dialogues, while still complex, is significantly simpler to achieve than general human conversational competence

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Conversational Human-Computer Interaction

The Domain-Independence Hypothesis

Within the genre of practical dialogue, the bulk of the complexity in the language interpretation and dialogue management is independent of the task being performed

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Conversational Human-Computer Interaction

Technique Used	Example Task	Task Complexity	Dialogue Phenomena handled
Finite-state Script	Long-distance dialing	least complex	User answers questions
Frame-based	Getting train arrival and departure information		User asks questions, simple clarifications by system
Sets of Contexts	Travel booking agent		Shifts between predetermined topics
Plan-based Models	Kitchen design consultant		Dynamically generated topic structures, collaborative negotiation subdialogues
Agent-based Models	Disaster relief management	most complex	Different modalities (e.g., planned world and actual world)

Dialogue and task complexity

- Practical Dialogues
- Frame based (form-filling) is currently most used
- Set of frames complex due to switch (going back)
- Plan and Agent based require model-of-the-world

Conversational Human-Computer Interaction

Technique Used	Example Task	Task Complexity	Dialogue Phenomena handled
Finite-state Script	Long-distance dialing	least complex	User answers questions
Frame-based	Getting train arrival and departure information		User asks questions, simple clarifications by system
Sets of Contexts	Travel booking agent		Shifts between predetermined topics
Plan-based Models	Kitchen design consultant		Dynamically generated topic structures, collaborative negotiation subdialogues
Agent-based Models	Disaster relief management	most complex	Different modalities (e.g., planned world and actual world)

Dialogue and task complexity

- Practical Dialogues
- Frame based (form-filling) is currently most used
- Set of frames complex due to switch (going back)
- Plan and Agent based require model-of-the-world

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Conversational Human-Computer Interaction

Technique Used	Example Task	Task Complexity	Dialogue Phenomena handled
Finite-state Script	Long-distance dialing	least complex	User answers questions
Frame-based	Getting train arrival and departure information		User asks questions, simple clarifications by system
Sets of Contexts	Travel booking agent		Shifts between predetermined topics
Plan-based Models	Kitchen design consultant		Dynamically generated topic structures, collaborative negotiation subdialogues
Agent-based Models	Disaster relief management	most complex	Different modalities (e.g., planned world and actual world)

Dialogue and task complexity

- Practical Dialogues
- Frame based (form-filling) is currently most used
- Set of frames complex due to switch (going back)
- Plan and Agent based require model-of-the-world

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Conversational Human-Computer Interaction

Technique Used	Example Task	Task Complexity	Dialogue Phenomena handled
Finite-state Script	Long-distance dialing	least complex	User answers questions
Frame-based	Getting train arrival and departure information		User asks questions, simple clarifications by system
Sets of Contexts	Travel booking agent		Shifts between predetermined topics
Plan-based Models	Kitchen design consultant		Dynamically generated topic structures, collaborative negotiation subdialogues
Agent-based Models	Disaster relief management	most complex	Different modalities (e.g., planned world and actual world)

Dialogue and task complexity

- Practical Dialogues
- Frame based (form-filling) is currently most used
- Set of frames complex due to switch (going back)
- Plan and Agent based require model-of-the-world

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Conversational Human-Computer Interaction

Parameter	Possible Values
The train ID?	BN101, ...
The event?	Departure, arrival
The location?	Avon, Bath, Corning, ...
The date/time range?	Monday, Aug 3, afternoon, ...

Context for a train information task

- Frame based dialogue system
- Fill in forms, send query when ready
- Simple and robust
- Simplifies ASR+NLP tasks (pattern matching)

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Conversational Human-Computer Interaction

Parameter	Possible Values
The train ID?	BN101, ...
The event?	Departure, arrival
The location?	Avon, Bath, Corning, ...
The date/time range?	Monday, Aug 3, afternoon, ...

Context for a train information task

- Frame based dialogue system
- Fill in forms, send query when ready
- Simple and robust
- Simplifies ASR+NLP tasks (pattern matching)

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Conversational Human-Computer Interaction

Parameter	Possible Values
The train ID?	BN101, ...
The event?	Departure, arrival
The location?	Avon, Bath, Corning, ...
The date/time range?	Monday, Aug 3, afternoon, ...

Context for a train information task

- Frame based dialogue system
- Fill in forms, send query when ready
- Simple and robust
- Simplifies ASR+NLP tasks (pattern matching)

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Conversational Human-Computer Interaction

Parameter	Possible Values
The train ID?	BN101, ...
The event?	Departure, arrival
The location?	Avon, Bath, Corning, ...
The date/time range?	Monday, Aug 3, afternoon, ...

Context for a train information task

- Frame based dialogue system
- Fill in forms, send query when ready
- Simple and robust
- Simplifies ASR+NLP tasks (pattern matching)

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Challenges for Dialogue Systems

- Parsing Language in Practical Dialogues
- Integrating Dialogue and Task Performance
- Intention Recognition
- Mixed Initiative Dialogue

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems

Challenges for Dialogue Systems

- Parsing Language in Practical Dialogues
- Integrating Dialogue and Task Performance
- Intention Recognition
- Mixed Initiative Dialogue

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems

Challenges for Dialogue Systems

- Parsing Language in Practical Dialogues
- Integrating Dialogue and Task Performance
- Intention Recognition
- Mixed Initiative Dialogue

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems

Challenges for Dialogue Systems

- Parsing Language in Practical Dialogues
- Integrating Dialogue and Task Performance
- Intention Recognition
- Mixed Initiative Dialogue

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Parsing Language in Practical Dialogues

- Detailed semantic, "deep", representation
- Broad coverage NL grammars fail due to ambiguity
- Semantic restrictions could work
- Add domain-specific restrictions for tasks
- Apply *Grice's Maxims*
- Parsing based on *Speech Acts*

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Parsing Language in Practical Dialogues

- Detailed semantic, "deep", representation
- Broad coverage NL grammars fail due to ambiguity
- Semantic restrictions could work
- Add domain-specific restrictions for tasks
- Apply *Grice's Maxims*
- Parsing based on *Speech Acts*

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Parsing Language in Practical Dialogues

- Detailed semantic, "deep", representation
- Broad coverage NL grammars fail due to ambiguity
- Semantic restrictions could work
- Add domain-specific restrictions for tasks
- Apply *Grice's Maxims*
- Parsing based on *Speech Acts*

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Parsing Language in Practical Dialogues

- Detailed semantic, “deep”, representation
- Broad coverage NL grammars fail due to ambiguity
- Semantic restrictions could work
- Add domain-specific restrictions for tasks
- Apply *Grice's Maxims*
- Parsing based on *Speech Acts*

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Parsing Language in Practical Dialogues

- Detailed semantic, "deep", representation
- Broad coverage NL grammars fail due to ambiguity
- Semantic restrictions could work
- Add domain-specific restrictions for tasks
- Apply *Grice's Maxims*
- Parsing based on *Speech Acts*

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Parsing Language in Practical Dialogues

- Detailed semantic, "deep", representation
- Broad coverage NL grammars fail due to ambiguity
- Semantic restrictions could work
- Add domain-specific restrictions for tasks
- Apply *Grice's Maxims*
- Parsing based on *Speech Acts*

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Integrating Dialogue and Task Performance

- Complex tasks based on Agents
- Abstract **problem-solving** model:
- **Objectives**: The way we want the world to be
- **Solutions**: Courses of action to achieve objectives
- **Resources**: Objects and abstractions available
- **Situations**: The way the world currently might be

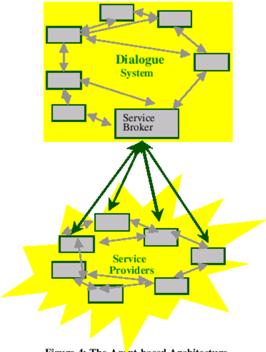


Figure 4: The Agent-based Architecture

Agent based architecture

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Integrating Dialogue and Task Performance

- Complex tasks based on Agents
- Abstract **problem-solving** model:
- **Objectives**: The way we want the world to be
- **Solutions**: Courses of action to achieve objectives
- **Resources**: Objects and abstractions available
- **Situations**: The way the world currently might be

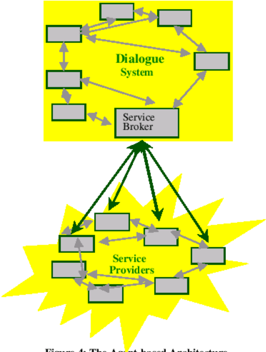


Figure 4: The Agent-based Architecture

Agent based architecture

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Integrating Dialogue and Task Performance

- Complex tasks based on Agents
- Abstract **problem-solving** model:
- **Objectives**: The way we want the world to be
- **Solutions**: Courses of action to achieve objectives
- **Resources**: Objects and abstractions available
- **Situations**: The way the world currently might be

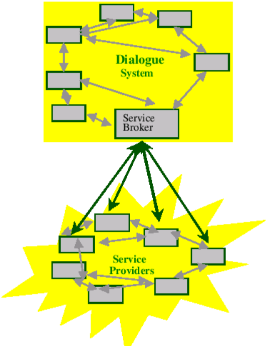


Figure 4: The Agent-based Architecture

Agent based architecture

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Integrating Dialogue and Task Performance

- Complex tasks based on Agents
- Abstract **problem-solving** model:
- **Objectives**: The way we want the world to be
- **Solutions**: Courses of action to achieve objectives
- **Resources**: Objects and abstractions available
- **Situations**: The way the world currently might be

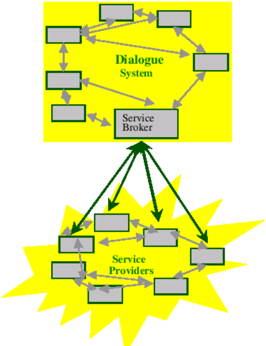


Figure 4: The Agent-based Architecture

Agent based architecture

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Integrating Dialogue and Task Performance

- Complex tasks based on Agents
- Abstract **problem-solving** model:
- **Objectives**: The way we want the world to be
- **Solutions**: Courses of action to achieve objectives
- **Resources**: Objects and abstractions available
- **Situations**: The way the world currently might be

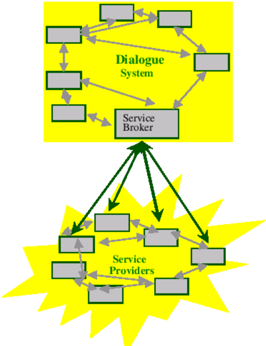


Figure 4: The Agent-based Architecture

Agent based architecture

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Integrating Dialogue and Task Performance

- Complex tasks based on Agents
- Abstract **problem-solving** model:
- **Objectives**: The way we want the world to be
- **Solutions**: Courses of action to achieve objectives
- **Resources**: Objects and abstractions available
- **Situations**: The way the world currently might be

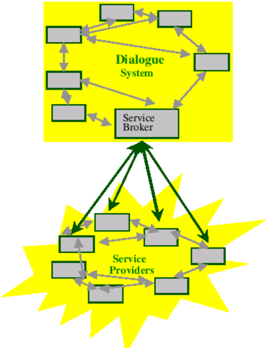


Figure 4: The Agent-based Architecture

Agent based architecture

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Intention Recognition

- Determine the goal of the user
- Can switch with every utterance
- Use implicatures
- Extrapolate from preceding actions
- Interpolate from "parent" (sub-)goals
- Is a probabilistic framework possible?

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Intention Recognition

- Determine the goal of the user
- Can switch with every utterance
- Use implicatures
- Extrapolate from preceding actions
- Interpolate from “parent” (sub-)goals
- Is a probabilistic framework possible?

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Intention Recognition

- Determine the goal of the user
- Can switch with every utterance
- Use implicatures
- Extrapolate from preceding actions
- Interpolate from "parent" (sub-)goals
- Is a probabilistic framework possible?

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Intention Recognition

- Determine the goal of the user
- Can switch with every utterance
- Use implicatures
- Extrapolate from preceding actions
- Interpolate from "parent" (sub-)goals
- Is a probabilistic framework possible?

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Intention Recognition

- Determine the goal of the user
- Can switch with every utterance
- Use implicatures
- Extrapolate from preceding actions
- Interpolate from "parent" (sub-)goals
- Is a probabilistic framework possible?

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Intention Recognition

- Determine the goal of the user
- Can switch with every utterance
- Use implicatures
- Extrapolate from preceding actions
- Interpolate from "parent" (sub-)goals
- Is a probabilistic framework possible?

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Mixed Initiative Dialogue

- Finite-state: typically fixed system-initiative
- Frame based: Fixed user/system-initiative (eg, Query-Answer)
- Fixed user-initiative: User does not know what information is needed
- Solution: Limited mixed initiative
- Full mixed-initiative: Both user and system can *barge-in*
- Mixed-initiative: Weight costs of interruption and system goals

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Mixed Initiative Dialogue

- Finite-state: typically fixed system-initiative
- Frame based: Fixed user/system-initiative (eg, Query-Answer)
- Fixed user-initiative: User does not know what information is needed
- Solution: Limited mixed initiative
- Full mixed-initiative: Both user and system can *barge-in*
- Mixed-initiative: Weight costs of interruption and system goals

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Mixed Initiative Dialogue

- Finite-state: typically fixed system-initiative
- Frame based: Fixed user/system-initiative (eg, Query-Answer)
- Fixed user-initiative: User does not know what information is needed
- Solution: Limited mixed initiative
- Full mixed-initiative: Both user and system can *barge-in*
- Mixed-initiative: Weight costs of interruption and system goals

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Mixed Initiative Dialogue

- Finite-state: typically fixed system-initiative
- Frame based: Fixed user/system-initiative (eg, Query-Answer)
- Fixed user-initiative: User does not know what information is needed
- Solution: Limited mixed initiative
- Full mixed-initiative: Both user and system can *barge-in*
- Mixed-initiative: Weight costs of interruption and system goals

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Mixed Initiative Dialogue

- Finite-state: typically fixed system-initiative
- Frame based: Fixed user/system-initiative (eg, Query-Answer)
- Fixed user-initiative: User does not know what information is needed
- Solution: Limited mixed initiative
- Full mixed-initiative: Both user and system can *barge-in*
- Mixed-initiative: Weight costs of interruption and system goals

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems: Challenges

Mixed Initiative Dialogue

- Finite-state: typically fixed system-initiative
- Frame based: Fixed user/system-initiative (eg, Query-Answer)
- Fixed user-initiative: User does not know what information is needed
- Solution: Limited mixed initiative
- Full mixed-initiative: Both user and system can *barge-in*
- Mixed-initiative: Weight costs of interruption and system goals

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

Spoken Dialogue Systems

1. *USR: We need to get the woman in Penfield to Strong*
2. *SYS: OK*
3. *USR: What vehicles are available?*
4. *SYS: There are ambulances in Pittsford and Webster*
5. *USR: OK. Use one from Pittsford*
6. *SYS: Do you know that Route 96 is blocked due to construction?*
7. *USR: Oh*
8. *USR: Let's use the interstate instead*
9. *SYS: OK. I'll dispatch the crew*

A short example of a practical dialog

- Both speakers use acknowledgements (*OK*, *Oh*)
- Semantic representation
- Intention recognition
- Goal driven
- System interrupts dialog (barge-in)

Spoken Dialogue Systems

1. *USR: We need to get the woman in Penfield to Strong*
2. *SYS: OK*
3. *USR: What vehicles are available?*
4. *SYS: There are ambulances in Pittsford and Webster*
5. *USR: OK. Use one from Pittsford*
6. *SYS: Do you know that Route 96 is blocked due to construction?*
7. *USR: Oh*
8. *USR: Let's use the interstate instead*
9. *SYS: OK. I'll dispatch the crew*

A short example of a practical dialog

- Both speakers use acknowledgements (*OK*, *Oh*)
- Semantic representation
- Intention recognition
- Goal driven
- System interrupts dialog (barge-in)

Spoken Dialogue Systems

1. *USR: We need to get the woman in Penfield to Strong*
2. *SYS: OK*
3. *USR: What vehicles are available?*
4. *SYS: There are ambulances in Pittsford and Webster*
5. *USR: OK. Use one from Pittsford*
6. *SYS: Do you know that Route 96 is blocked due to construction?*
7. *USR: Oh*
8. *USR: Let's use the interstate instead*
9. *SYS: OK. I'll dispatch the crew*

A short example of a practical dialog

- Both speakers use acknowledgements (*OK, Oh*)
- Semantic representation
- Intention recognition
- Goal driven
- System interrupts dialog (barge-in)

Spoken Dialogue Systems

1. *USR: We need to get the woman in Penfield to Strong*
2. *SYS: OK*
3. *USR: What vehicles are available?*
4. *SYS: There are ambulances in Pittsford and Webster*
5. *USR: OK. Use one from Pittsford*
6. *SYS: Do you know that Route 96 is blocked due to construction?*
7. *USR: Oh*
8. *USR: Let's use the interstate instead*
9. *SYS: OK. I'll dispatch the crew*

A short example of a practical dialog

- Both speakers use acknowledgements (*OK*, *Oh*)
- Semantic representation
- Intention recognition
- Goal driven
- System interrupts dialog (barge-in)

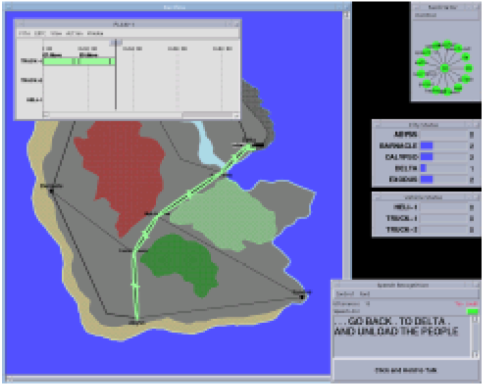
Spoken Dialogue Systems

1. *USR: We need to get the woman in Penfield to Strong*
2. *SYS: OK*
3. *USR: What vehicles are available?*
4. *SYS: There are ambulances in Pittsford and Webster*
5. *USR: OK. Use one from Pittsford*
6. *SYS: Do you know that Route 96 is blocked due to construction?*
7. *USR: Oh*
8. *USR: Let's use the interstate instead*
9. *SYS: OK. I'll dispatch the crew*

A short example of a practical dialog

- Both speakers use acknowledgements (*OK*, *Oh*)
- Semantic representation
- Intention recognition
- Goal driven
- System interrupts dialog (barge-in)

TRIPS

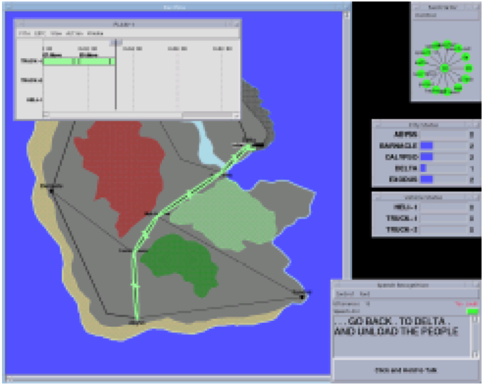


Interacting with TRIPS

- Multi modal interaction with current state shown
- Emergency Response System

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

TRIPS

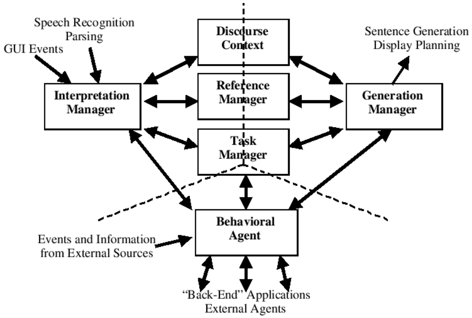


Interacting with TRIPS

- Multi modal interaction with current state shown
- Emergency Response System

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

TRIPS

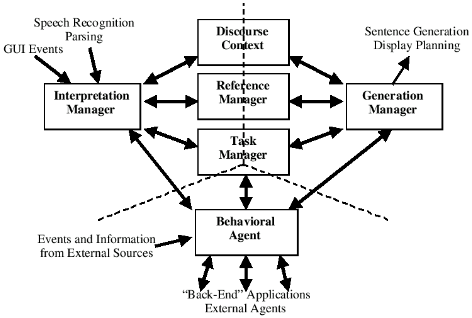


TRIPS system architecture

- Interpretation
- Generation
- Behavior

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

TRIPS

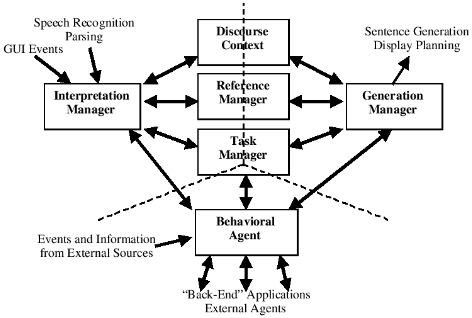


TRIPS system architecture

- Interpretation
- Generation
- Behavior

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

TRIPS



TRIPS system architecture

- Interpretation
- Generation
- Behavior

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

TRIPS

1. *USR: We need to get the woman in Penfield to Strong*

Reference resolution

- SS1: The set consisting of *USR* and *SYS* (general dialogue setting)
- WOM1: The Injured woman in Penfield previously discussed (discourse history)
- Strong Memorial Hospital (general world knowledge)

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

TRIPS

1. *USR: We need to get the woman in Penfield to Strong*

Reference resolution

- SS1: The set consisting of *USR* and *SYS* (general dialogue setting)
- WOM1: The Injured woman in Penfield previously discussed (discourse history)
- Strong Memorial Hospital (general world knowledge)

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

TRIPS

1. *USR: We need to get the woman in Penfield to Strong*

Reference resolution

- SS1: The set consisting of *USR* and *SYS* (general dialogue setting)
- WOM1: The Injured woman in Penfield previously discussed (discourse history)
- Strong Memorial Hospital (general world knowledge)

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

TRIPS

1. *USR: We need to get the woman in Penfield to Strong*

Reference resolution

- SS1: The set consisting of *USR* and *SYS* (general dialogue setting)
- WOM1: The Injured woman in Penfield previously discussed (discourse history)
- Strong Memorial Hospital (general world knowledge)

[Allen et al.(2001)Allen, Byron, Dzikovska, Ferguson, Galescu, and Stent]

OVIS

Public Transport Information System

- Deliver train travel information (station-to-station)
- Telephone based application
- Speech only
- Replaced existing human based service
- Based on an existing German system (Philips Aachen)
- Has been in active service (still is)
- Frame-based

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS

Public Transport Information System

- Deliver train travel information (station-to-station)
- Telephone based application
- Speech only
- Replaced existing human based service
- Based on an existing German system (Philips Aachen)
- Has been in active service (still is)
- Frame-based

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS

Public Transport Information System

- Deliver train travel information (station-to-station)
- Telephone based application
- Speech only
- Replaced existing human based service
- Based on an existing German system (Philips Aachen)
- Has been in active service (still is)
- Frame-based

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS

Public Transport Information System

- Deliver train travel information (station-to-station)
- Telephone based application
- Speech only
- Replaced existing human based service
- Based on an existing German system (Philips Aachen)
- Has been in active service (still is)
- Frame-based

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS

Public Transport Information System

- Deliver train travel information (station-to-station)
- Telephone based application
- Speech only
- Replaced existing human based service
- Based on an existing German system (Philips Aachen)
- Has been in active service (still is)
- Frame-based

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS

Public Transport Information System

- Deliver train travel information (station-to-station)
- Telephone based application
- Speech only
- Replaced existing human based service
- Based on an existing German system (Philips Aachen)
- Has been in active service (still is)
- Frame-based

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS

Public Transport Information System

- Deliver train travel information (station-to-station)
- Telephone based application
- Speech only
- Replaced existing human based service
- Based on an existing German system (Philips Aachen)
- Has been in active service (still is)
- Frame-based

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS

Spoken Dialogue System (SDS) components

1. Continuous HMM based Speech Recognition (CSR)
2. Natural Language Processing (NLP)
3. Dialogue Management (DM)
4. Text-To-Speech (TTS)

OVIS

Spoken Dialogue System (SDS) components

1. Continuous HMM based Speech Recognition (CSR)
2. Natural Language Processing (NLP)
3. Dialogue Management (DM)
4. Text-To-Speech (TTS)

OVIS

Spoken Dialogue System (SDS) components

1. Continuous HMM based Speech Recognition (CSR)
2. Natural Language Processing (NLP)
3. Dialogue Management (DM)
4. Text-To-Speech (TTS)

OVIS

Spoken Dialogue System (SDS) components

1. Continuous HMM based Speech Recognition (CSR)
2. Natural Language Processing (NLP)
3. Dialogue Management (DM)
4. Text-To-Speech (TTS)

OVIS

Skip *Wizard-of-Oz* or *Green-curtain* scenarios and build a working system from scratch.

Stages to build and train SDS

1. Make a first version of the SDS with available data (which need not be application-specific)
2. Ask a limited group of people to use this system, and store the dialogues
3. Use the recorded data (which are application-specific) to improve the SDS
4. Gradually increase the data and the number of users
5. Repeat steps [2], [3], and [4] until the system works satisfactorily

OVIS

Skip *Wizard-of-Oz* or *Green-curtain* scenarios and build a working system from scratch.

Stages to build and train SDS

1. Make a first version of the SDS with available data (which need not be application-specific)
2. Ask a limited group of people to use this system, and store the dialogues
3. Use the recorded data (which are application-specific) to improve the SDS
4. Gradually increase the data and the number of users
5. Repeat steps [2], [3], and [4] until the system works satisfactorily

OVIS

Skip *Wizard-of-Oz* or *Green-curtain* scenarios and build a working system from scratch.

Stages to build and train SDS

1. Make a first version of the SDS with available data (which need not be application-specific)
2. Ask a limited group of people to use this system, and store the dialogues
3. Use the recorded data (which are application-specific) to improve the SDS
4. Gradually increase the data and the number of users
5. Repeat steps [2], [3], and [4] until the system works satisfactorily

OVIS

Skip *Wizard-of-Oz* or *Green-curtain* scenarios and build a working system from scratch.

Stages to build and train SDS

1. Make a first version of the SDS with available data (which need not be application-specific)
2. Ask a limited group of people to use this system, and store the dialogues
3. Use the recorded data (which are application-specific) to improve the SDS
4. Gradually increase the data and the number of users
5. Repeat steps [2], [3], and [4] until the system works satisfactorily

OVIS

Skip *Wizard-of-Oz* or *Green-curtain* scenarios and build a working system from scratch.

Stages to build and train SDS

1. Make a first version of the SDS with available data (which need not be application-specific)
2. Ask a limited group of people to use this system, and store the dialogues
3. Use the recorded data (which are application-specific) to improve the SDS
4. Gradually increase the data and the number of users
5. Repeat steps [2], [3], and [4] until the system works satisfactorily

OVIS: Continuous Speech Recognition

Start training with the Polyphone multi-speaker corpus

- 2500 utterances
- Read speech
- Semi-spontaneous (read) speech
- Recorded over the phone
- For each speaker, 5 out 50 *Polyphone* sentences selected
- Phonetically rich sentences (*all* Dutch phonemes)
- 50 Dutch phone models (2 for each of /r/ and /l/)

OVIS: Continuous Speech Recognition

Start training with the Polyphone multi-speaker corpus

- 2500 utterances
- Read speech
- Semi-spontaneous (read) speech
- Recorded over the phone
- For each speaker, 5 out 50 *Polyphone* sentences selected
- Phonetically rich sentences (*all* Dutch phonemes)
- 50 Dutch phone models (2 for each of /r/ and /l/)

OVIS: Continuous Speech Recognition

Start training with the Polyphone multi-speaker corpus

- 2500 utterances
- Read speech
- Semi-spontaneous (read) speech
- Recorded over the phone
- For each speaker, 5 out 50 *Polyphone* sentences selected
- Phonetically rich sentences (*all* Dutch phonemes)
- 50 Dutch phone models (2 for each of /r/ and /l/)

OVIS: Continuous Speech Recognition

Start training with the Polyphone multi-speaker corpus

- 2500 utterances
- Read speech
- Semi-spontaneous (read) speech
- Recorded over the phone
- For each speaker, 5 out 50 *Polyphone* sentences selected
- Phonetically rich sentences (*all* Dutch phonemes)
- 50 Dutch phone models (2 for each of /r/ and /l/)

OVIS: Continuous Speech Recognition

Start training with the Polyphone multi-speaker corpus

- 2500 utterances
- Read speech
- Semi-spontaneous (read) speech
- Recorded over the phone
- For each speaker, 5 out 50 *Polyphone* sentences selected
- Phonetically rich sentences (*all* Dutch phonemes)
- 50 Dutch phone models (2 for each of /r/ and /l/)

OVIS: Continuous Speech Recognition

Start training with the Polyphone multi-speaker corpus

- 2500 utterances
- Read speech
- Semi-spontaneous (read) speech
- Recorded over the phone
- For each speaker, 5 out 50 *Polyphone* sentences selected
- Phonetically rich sentences (*all* Dutch phonemes)
- 50 Dutch phone models (2 for each of /r/ and /l/)

OVIS: Continuous Speech Recognition

Start training with the Polyphone multi-speaker corpus

- 2500 utterances
- Read speech
- Semi-spontaneous (read) speech
- Recorded over the phone
- For each speaker, 5 out 50 *Polyphone* sentences selected
- Phonetically rich sentences (*all* Dutch phonemes)
- 50 Dutch phone models (2 for each of /r/ and /l/)

OVIS: Pronunciation lexicon

Phoneme representations

- Names of stations from the ONOMASTICA database
- Lemma forms of other words from the CELEX database
- Remaining generated by a grapheme-to-phoneme converter
- Pronunciation variation initially not modelled

OVIS: Pronunciation lexicon

Phoneme representations

- Names of stations from the ONOMASTICA database
- Lemma forms of other words from the CELEX database
- Remaining generated by a grapheme-to-phoneme converter
- Pronunciation variation initially not modelled

OVIS: Pronunciation lexicon

Phoneme representations

- Names of stations from the ONOMASTICA database
- Lemma forms of other words from the CELEX database
- Remaining generated by a grapheme-to-phoneme converter
- Pronunciation variation initially not modelled

OVIS: Pronunciation lexicon

Phoneme representations

- Names of stations from the ONOMASTICA database
- Lemma forms of other words from the CELEX database
- Remaining generated by a grapheme-to-phoneme converter
- Pronunciation variation initially not modelled

OVIS: NLP and DM

NLP and DM taken from German original

- Date and time conventions adapted
- Interface with different train table format (eg, start of *tomorrow*)
- Adaptations for user preferences, eg, train numbers
- Collect volunteer queries from keyboard simulation
- Form based database query system with feed-back
- Allows user to correct the system

OVIS: NLP and DM

NLP and DM taken from German original

- Date and time conventions adapted
- Interface with different train table format (eg, start of *tomorrow*)
- Adaptations for user preferences, eg, train numbers
- Collect volunteer queries from keyboard simulation
- Form based database query system with feed-back
- Allows user to correct the system

OVIS: NLP and DM

NLP and DM taken from German original

- Date and time conventions adapted
- Interface with different train table format (eg, start of *tomorrow*)
- Adaptations for user preferences, eg, train numbers
- Collect volunteer queries from keyboard simulation
- Form based database query system with feed-back
- Allows user to correct the system

OVIS: NLP and DM

NLP and DM taken from German original

- Date and time conventions adapted
- Interface with different train table format (eg, start of *tomorrow*)
- Adaptations for user preferences, eg, train numbers
- Collect volunteer queries from keyboard simulation
- Form based database query system with feed-back
- Allows user to correct the system

OVIS: NLP and DM

NLP and DM taken from German original

- Date and time conventions adapted
- Interface with different train table format (eg, start of *tomorrow*)
- Adaptations for user preferences, eg, train numbers
- Collect volunteer queries from keyboard simulation
- Form based database query system with feed-back
- Allows user to correct the system

OVIS: NLP and DM

NLP and DM taken from German original

- Date and time conventions adapted
- Interface with different train table format (eg, start of *tomorrow*)
- Adaptations for user preferences, eg, train numbers
- Collect volunteer queries from keyboard simulation
- Form based database query system with feed-back
- Allows user to correct the system

OVIS: TTS

Speech generation (TTS)

- German original could not be used
- Concatenate utterance fragments
- Female voice

OVIS: TTS

Speech generation (TTS)

- German original could not be used
- Concatenate utterance fragments
- Female voice

OVIS: TTS

Speech generation (TTS)

- German original could not be used
- Concatenate utterance fragments
- Female voice

OVIS: Training

Database	utterances	source	duration (hours:min)
DB0	2500	Polyphone	4:42
DB1	1301	application	0:41
DB2	5496	application	3:47
DB3	6401	application	4:35
DB4	8000	application	5:55
DB5	10003	application	7:20

Databases used during development of the SDS

- Start with the *Polyphone* database (DB0)
- Collect volunteer responses from this system
- Retrain the system with the new speech and repeat
- DB1-5 are incremental, ie, DB5 contains all of DB4 etc.

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

Database	utterances	source	duration (hours:min)
DB0	2500	Polyphone	4:42
DB1	1301	application	0:41
DB2	5496	application	3:47
DB3	6401	application	4:35
DB4	8000	application	5:55
DB5	10003	application	7:20

Databases used during development of the SDS

- Start with the *Polyphone* database (DB0)
- Collect volunteer responses from this system
- Retrain the system with the new speech and repeat
- DB1-5 are incremental, ie, DB5 contains all of DB4 etc.

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

Database	utterances	source	duration (hours:min)
DB0	2500	Polyphone	4:42
DB1	1301	application	0:41
DB2	5496	application	3:47
DB3	6401	application	4:35
DB4	8000	application	5:55
DB5	10003	application	7:20

Databases used during development of the SDS

- Start with the *Polyphone* database (DB0)
- Collect volunteer responses from this system
- Retrain the system with the new speech and repeat
- DB1-5 are incremental, ie, DB5 contains all of DB4 etc.

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

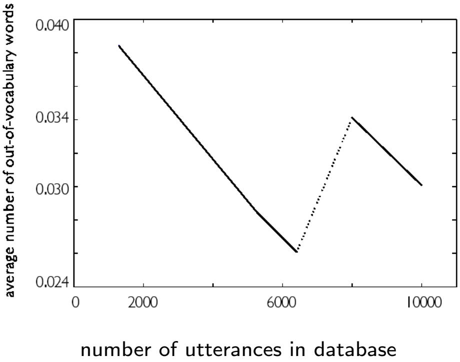
Database	utterances	source	duration (hours:min)
DB0	2500	Polyphone	4:42
DB1	1301	application	0:41
DB2	5496	application	3:47
DB3	6401	application	4:35
DB4	8000	application	5:55
DB5	10003	application	7:20

Databases used during development of the SDS

- Start with the *Polyphone* database (DB0)
- Collect volunteer responses from this system
- Retrain the system with the new speech and repeat
- DB1-5 are incremental, ie, DB5 contains all of DB4 etc.

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

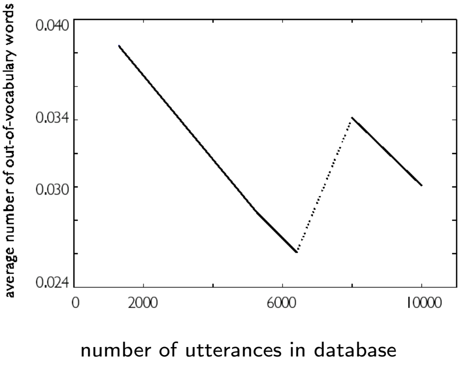


Out-of-vocabulary words per utterance vs. corpus size

- Number of OOV words is small
- DB0-DB3 small number of users
- After DB3 (6401 utterances) new users recruited

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

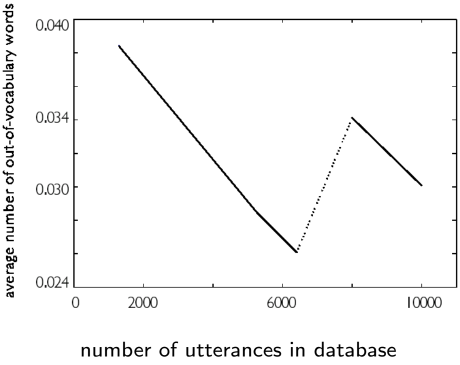


Out-of-vocabulary words per utterance vs. corpus size

- Number of OOV words is small
- DB0-DB3 small number of users
- After DB3 (6401 utterances) new users recruited

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training



Out-of-vocabulary words per utterance vs. corpus size

- Number of OOV words is small
- DB0-DB3 small number of users
- After DB3 (6401 utterances) new users recruited

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

System	P0 + L0	P02 + L0	P02 + L2
WG - WER	20.59	18.36	6.72
WG - SER	40.00	36.60	16.00
BS - WER	39.87	31.45	14.73
BS - SER	65.00	54.20	28.00

Performance level for different phoneme models (Pi) and language models (Lj). Evaluation is done with test database 1

- Training phoneme models on both DB0 (polyphone) and DB2 (application) reduced error rates
- Training language model on DB2 (application) reduced errors more
- Application specific data is more important for language modelling than phoneme modelling

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

System	P0 + L0	P02 + L0	P02 + L2
WG - WER	20.59	18.36	6.72
WG - SER	40.00	36.60	16.00
BS - WER	39.87	31.45	14.73
BS - SER	65.00	54.20	28.00

Performance level for different phoneme models (Pi) and language models (Lj). Evaluation is done with test database 1

- Training phoneme models on both DB0 (polyphone) and DB2 (application) reduced error rates
- Training language model on DB2 (application) reduced errors more
- Application specific data is more important for language modelling than phoneme modelling

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

System	P0 + L0	P02 + L0	P02 + L2
WG - WER	20.59	18.36	6.72
WG - SER	40.00	36.60	16.00
BS - WER	39.87	31.45	14.73
BS - SER	65.00	54.20	28.00

Performance level for different phoneme models (Pi) and language models (Lj). Evaluation is done with test database 1

- Training phoneme models on both DB0 (polyphone) and DB2 (application) reduced error rates
- Training language model on DB2 (application) reduced errors more
- Application specific data is more important for language modelling than phoneme modelling

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

System	P02 + L2	P03 + L2	P03 + L3	P3 + L2	P3 + L3
WG - WER	6.72	6.94	6.94	6.94	6.94
WG - SER	16.00	15.20	15.60	16.20	15.40
BS - WER	14.73	15.43	15.70	16.41	14.84
BS - SER	28.00	29.00	28.60	26.00	26.40

Performance level for different phoneme models (P02/3 vs P3) and language models (L2 vs L3). Evaluation is done with test database 1

- Increasing DB size from 5496 to 6401 utterances had little effect
- Leaving out Polyphone data (DB0) hardly had an effect
- Leaving out DB0 even decreased WER a little

WG: word-graph, BS: best sentence,

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

System	P02 + L2	P03 + L2	P03 + L3	P3 + L2	P3 + L3
WG - WER	6.72	6.94	6.94	6.94	6.94
WG - SER	16.00	15.20	15.60	16.20	15.40
BS - WER	14.73	15.43	15.70	16.41	14.84
BS - SER	28.00	29.00	28.60	26.00	26.40

Performance level for different phoneme models (P02/3 vs P3) and language models (L2 vs L3). Evaluation is done with test database 1

- Increasing DB size from 5496 to 6401 utterances had little effect
- Leaving out Polyphone data (DB0) hardly had an effect
- Leaving out DB0 even decreased WER a little

WG: word-graph, BS: best sentence,

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

System	P02 + L2	P03 + L2	P03 + L3	P3 + L2	P3 + L3
WG - WER	6.72	6.94	6.94	6.94	6.94
WG - SER	16.00	15.20	15.60	16.20	15.40
BS - WER	14.73	15.43	15.70	16.41	14.84
BS - SER	28.00	29.00	28.60	26.00	26.40

Performance level for different phoneme models (P02/3 vs P3) and language models (L2 vs L3). Evaluation is done with test database 1

- Increasing DB size from 5496 to 6401 utterances had little effect
- Leaving out Polyphone data (DB0) hardly had an effect
- Leaving out DB0 even decreased WER a little

WG: word-graph, BS: best sentence,

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

testDB	old	new		
System	P3 + L3	P3 + L3	P4 + L4	P5 + L5
WG - WER	6.94	8.87	6.81	6.69
WG - SER	15.40	17.80	14.40	13.80
BS - WER	14.84	15.27	12.93	14.02
BS - SER	26.40	25.40	24.20	24.60

Performance levels for different phoneme models (Pi) and language models (Lj). Evaluation is done with test database 1 (column 2: old) and 2 (columns 3-5: new)

- Test database 2 induced more errors
- DB4 (8,000 utterances) had lower WER again
- Increase to 10,000 utterances (DB5) had little effect

WG: word-graph, BS: best sentence,

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

testDB	old	new		
System	P3 + L3	P3 + L3	P4 + L4	P5 + L5
WG - WER	6.94	8.87	6.81	6.69
WG - SER	15.40	17.80	14.40	13.80
BS - WER	14.84	15.27	12.93	14.02
BS - SER	26.40	25.40	24.20	24.60

Performance levels for different phoneme models (Pi) and language models (Lj). Evaluation is done with test database 1 (column 2: old) and 2 (columns 3-5: new)

- Test database 2 induced more errors
- DB4 (8,000 utterances) had lower WER again
- Increase to 10,000 utterances (DB5) had little effect

WG: word-graph, BS: best sentence,

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS: Training

testDB	old	new		
System	P3 + L3	P3 + L3	P4 + L4	P5 + L5
WG - WER	6.94	8.87	6.81	6.69
WG - SER	15.40	17.80	14.40	13.80
BS - WER	14.84	15.27	12.93	14.02
BS - SER	26.40	25.40	24.20	24.60

Performance levels for different phoneme models (Pi) and language models (Lj). Evaluation is done with test database 1 (column 2: old) and 2 (columns 3-5: new)

- Test database 2 induced more errors
- DB4 (8,000 utterances) had lower WER again
- Increase to 10,000 utterances (DB5) had little effect

WG: word-graph, BS: best sentence,

[Strik et al.(1997)Strik, Russel, van den Heuvel, Cucchiarini, and Boves]

OVIS

Pronunciation variation and non-speech sounds

- A single pronunciation per word gives problems
- Eg, /ɣɛldərɔp/ vs. /ɣɛldrɔp/ and /ɑmsədɑm/ vs. /ɑmstərdɑm/
- Different sources causes inconsistencies
- People use several different variants
- Variant in lexicon not the "best" one

OVIS

Pronunciation variation and non-speech sounds

- A single pronunciation per word gives problems
- Eg, /ɣɛldərɔp/ vs. /ɣɛldrɔp/ and /ɑmsədɑm/ vs. /ɑmstərdɑm/
- Different sources causes inconsistencies
- People use several different variants
- Variant in lexicon not the "best" one

OVIS

Pronunciation variation and non-speech sounds

- A single pronunciation per word gives problems
- Eg, /ɣɛldərɔp/ vs. /ɣɛldrɔp/ and /ɑmsədɑm/ vs. /ɑmstərdɑm/
- Different sources causes inconsistencies
- People use several different variants
- Variant in lexicon not the "best" one

OVIS

Pronunciation variation and non-speech sounds

- A single pronunciation per word gives problems
- Eg, /ɣɛldərɔp/ vs. /ɣɛldrɔp/ and /ɑmsədɑm/ vs. /ɑmstərdɑm/
- Different sources causes inconsistencies
- People use several different variants
- Variant in lexicon not the "best" one

OVIS

Pronunciation variation and non-speech sounds

- A single pronunciation per word gives problems
- Eg, /ɣɛldərɔp/ vs. /ɣɛldrɔp/ and /ɑmsədɑm/ vs. /ɑmstərdɑm/
- Different sources causes inconsistencies
- People use several different variants
- Variant in lexicon not the “best” one

OVIS: Conclusions

It actually worked!

- Adapt an existing frame-based system
- Bootstrap on actual usage
- Collect and train more
- Use robust DM
- Use human fall-back

OVIS: Conclusions

It actually worked!

- Adapt an existing frame-based system
- Bootstrap on actual usage
- Collect and train more
- Use robust DM
- Use human fall-back

OVIS: Conclusions

It actually worked!

- Adapt an existing frame-based system
- Bootstrap on actual usage
- Collect and train more
- Use robust DM
- Use human fall-back

OVIS: Conclusions

It actually worked!

- Adapt an existing frame-based system
- Bootstrap on actual usage
- Collect and train more
- Use robust DM
- Use human fall-back

OVIS: Conclusions

It actually worked!

- Adapt an existing frame-based system
- Bootstrap on actual usage
- Collect and train more
- Use robust DM
- Use human fall-back

Further Reading I

James F. Allen, Donna K. Byron, Myroslava Dzikovska, George Ferguson, Lucian Galescu, and Amanda Stent.
Toward conversational human-computer interaction.
AI Magazine, Winter, 2001.
URL
http://www.cs.rochester.edu/research/cisd/pubs/2001/allen-et-al-aimag2001.pdf.

P. Boersma.
Praat, a system for doing phonetics by computer.
Glot International, 5:341–345, 2001.
URL http://www.Praat.org/.

P. Boersma and D. Weenink.
Praat 4.2: doing phonetics by computer.
Computer program: http://www.Praat.org/, 2004.
URL http://www.Praat.org/.

FSF.
GNU General Public License.
Web, June 1991.
URL http://www.gnu.org/licenses/gpl.html.

Daniel Jurafsky and James H. Martin.
Speech and Language Processing.
Prentice-Hall, 2000.
ISBN 0-13-095069-6.
URL http://www.cs.colorado.edu/~martin/slp.html.
Updates at http://www.cs.colorado.edu/

Further Reading II

Helmer Strik, Albert Russel, Henk van den Heuvel, Catia Cucchiarini, and Lou Boves.
A spoken dialog system for the dutch public transport information service.
Int. Journal of Speech Technology, 2:121–131, 1997.
URL http://lands.let.ru.nl/literature/strik.1996.4.ps.
Link is to an older version.

W. Wesseling and R. J. J. H. van Son.
Timing of experimentally elicited minimal responses as quantitative evidence for the use of intonation in projecting TRPs.
In *Proceedings of Interspeech2005*, Lisbon, 2005.

Wieneke Wesseling and R.J.J.H. Van Son.
Early Preparation of Experimentally Elicited Minimal Responses.
In *Proceedings of SIGdial 2005*, September 2005.
URL http://www.fon.hum.uva.nl/rob/Publications/ArtikelSIGdial2005.pdf.

Appendix A: Implicatures

Conversations contain rules of inference

Conversational Maxims of Grice

- Quantity: Be exactly as informative as required
  - Not less informative
  - Not more informative
- Quality: Speak the truth
  - Do not say what you believe is false
  - Do not say that for which you lack evidence
- Relevance: Be relevant
- Manner: Be perspicuous
  - Avoid obscurity
  - Avoid ambiguity
  - Be brief
  - Be orderly

Back to Challenges

Appendix A: Implicatures

Conversations contain rules of inference

Conversational Maxims of Grice

- **Quantity**: Be *exactly* as informative as required
  - Not less informative
  - Not more informative
- Quality: Speak the truth
  - Do not say what you believe is false
  - Do not say that for which you lack evidence
- Relevance: Be relevant
- Manner: Be perspicuous
  - Avoid obscurity
  - Avoid ambiguity
  - Be brief
  - Be orderly

Back to Challenges

Appendix A: Implicatures

Conversations contain rules of inference

Conversational Maxims of Grice

- **Quantity**: Be *exactly* as informative as required
  - Not *less* informative
  - Not more informative
- Quality: Speak the truth
  - Do not say what you believe is false
  - Do not say that for which you lack evidence
- Relevance: Be relevant
- Manner: Be perspicuous
  - Avoid obscurity
  - Avoid ambiguity
  - Be brief
  - Be orderly

Back to Challenges

Appendix A: Implicatures

Conversations contain rules of inference

Conversational Maxims of Grice

- **Quantity**: Be *exactly* as informative as required
  - Not *less* informative
  - Not *more* informative
- Quality: Speak the truth
  - Do not say what you believe is false
  - Do not say that for which you lack evidence
- Relevance: Be relevant
- Manner: Be perspicuous
  - Avoid obscurity
  - Avoid ambiguity
  - Be brief
  - Be orderly

Back to Challenges

Appendix A: Implicatures

Conversations contain rules of inference

Conversational Maxims of Grice

- **Quantity**: Be *exactly* as informative as required
  - Not *less* informative
  - Not *more* informative
- **Quality**: Speak the *truth*
  - Do not say what you believe is false
  - Do not say that for which you lack evidence
- Relevance: Be relevant
- Manner: Be perspicuous
  - Avoid obscurity
  - Avoid ambiguity
  - Be brief
  - Be orderly

Back to Challenges

Appendix A: Implicatures

Conversations contain rules of inference

Conversational Maxims of Grice

- **Quantity**: Be *exactly* as informative as required
  - Not *less* informative
  - Not *more* informative
- **Quality**: Speak the *truth*
  - Do not say what you believe is *false*
  - Do not say that for which you lack evidence
- Relevance: Be relevant
- Manner: Be perspicuous
  - Avoid obscurity
  - Avoid ambiguity
  - Be brief
  - Be orderly

Back to Challenges

Appendix A: Implicatures

Conversations contain rules of inference

Conversational Maxims of Grice

- **Quantity**: Be *exactly* as informative as required
  - Not *less* informative
  - Not *more* informative
- **Quality**: Speak the *truth*
  - Do not say what you believe is *false*
  - Do not say that for which you lack *evidence*
- Relevance: Be relevant
- Manner: Be perspicuous
  - Avoid obscurity
  - Avoid ambiguity
  - Be brief
  - Be orderly

Back to Challenges

Appendix A: Implicatures

Conversations contain rules of inference

Conversational Maxims of Grice

- **Quantity**: Be *exactly* as informative as required
  - Not *less* informative
  - Not *more* informative
- **Quality**: Speak the *truth*
  - Do not say what you believe is *false*
  - Do not say that for which you lack *evidence*
- **Relevance**: Be relevant
- Manner: Be perspicuous
  - Avoid obscurity
  - Avoid ambiguity
  - Be brief
  - Be orderly

Back to Challenges

Appendix A: Implicatures

Conversations contain rules of inference

Conversational Maxims of Grice

- **Quantity**: Be *exactly* as informative as required
  - Not *less* informative
  - Not *more* informative
- **Quality**: Speak the *truth*
  - Do not say what you believe is *false*
  - Do not say that for which you lack *evidence*
- **Relevance**: Be relevant
- **Manner**: Be *perspicuous*
  - Avoid obscurity
  - Avoid ambiguity
  - Be brief
  - Be orderly

Back to Challenges

Appendix A: Implicatures

Conversations contain rules of inference

Conversational Maxims of Grice

- **Quantity**: Be *exactly* as informative as required
  - Not *less* informative
  - Not *more* informative
- **Quality**: Speak the *truth*
  - Do not say what you believe is *false*
  - Do not say that for which you lack *evidence*
- **Relevance**: Be relevant
- **Manner**: Be *perspicuous*
  - Avoid *obscurity*
  - Avoid ambiguity
  - Be brief
  - Be orderly

Back to Challenges

Appendix A: Implicatures

Conversations contain rules of inference

Conversational Maxims of Grice

- **Quantity**: Be *exactly* as informative as required
  - Not *less* informative
  - Not *more* informative
- **Quality**: Speak the *truth*
  - Do not say what you believe is *false*
  - Do not say that for which you lack *evidence*
- **Relevance**: Be relevant
- **Manner**: Be *perspicuous*
  - Avoid *obscurity*
  - Avoid *ambiguity*
  - Be brief
  - Be orderly

Back to Challenges

Appendix A: Implicatures

Conversations contain rules of inference

Conversational Maxims of Grice

- **Quantity**: Be *exactly* as informative as required
  - Not *less* informative
  - Not *more* informative
- **Quality**: Speak the *truth*
  - Do not say what you believe is *false*
  - Do not say that for which you lack *evidence*
- **Relevance**: Be relevant
- **Manner**: Be *perspicuous*
  - Avoid *obscurity*
  - Avoid *ambiguity*
  - Be *brief*
  - Be orderly

Back to Challenges

Appendix A: Implicatures

Conversations contain rules of inference

Conversational Maxims of Grice

- **Quantity**: Be *exactly* as informative as required
  - Not *less* informative
  - Not *more* informative
- **Quality**: Speak the *truth*
  - Do not say what you believe is *false*
  - Do not say that for which you lack *evidence*
- **Relevance**: Be relevant
- **Manner**: Be *perspicuous*
  - Avoid *obscurity*
  - Avoid *ambiguity*
  - Be *brief*
  - Be *orderly*

Back to Challenges

Copyright License

The GNU General Public License I

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software—to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

The GNU General Public License II

The GNU General Public License III

1. You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.
   You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.
2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:
   1. You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.
   2. You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.
   3. If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

The GNU General Public License IV

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.
Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.
In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:
    1. Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,
    2. Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,
    3. Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b above.)

The GNU General Public License V

The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.
If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.
5. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.
6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

The GNU General Public License VI

7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.
   If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.
   It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.
   This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.
8. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

The GNU General Public License VII

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.
   Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.
10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

No Warranty

11. Because the program is licensed free of charge, there is no warranty for the program, to the extent permitted by applicable law. Except when otherwise stated in writing the copyright holders and/or other parties provide the program "as is" without warranty of any kind, either expressed or implied, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The entire risk as to the quality and performance of the program is with you. Should the program prove defective, you assume the cost of all necessary servicing, repair or correction.

The GNU General Public License VIII

The GNU General Public License IX

Appendix: How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

> *one line to give the program's name and a brief idea of what it does.*
> *Copyright (C) yyyy name of author*
> *This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.*
>
> *This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.*
>
> *You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301, USA.*

Also add information on how to contact you by electronic and paper mail.

If the program is interactive, make it output a short notice like this when it starts in an interactive mode:

> *Gnomovision version 69, Copyright (C) yyyy name of author*
> *Gnomovision comes with ABSOLUTELY NO WARRANTY; for details type 'show w'.*
> *This is free software, and you are welcome to redistribute it under certain conditions;*
> *type 'show c' for details.*

The GNU General Public License X

The hypothetical commands `show w` and `show c` should show the appropriate parts of the General Public License. Of course, the commands you use may be called something other than `show w` and `show c`; they could even be mouse-clicks or menu items—whatever suits your program.
You should also get your employer (if you work as a programmer) or your school, if any, to sign a "copyright disclaimer" for the program, if necessary. Here is a sample; alter the names:

> *Yoyodyne, Inc., hereby disclaims all copyright interest in the program*
> *'Gnomovision' (which makes passes at compilers) written by James Hacker.*
> *signature of Ty Coon, 1 April 1989*
> *Ty Coon, President of Vice*

This General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Library General Public License instead of this License.