Machine Translation (MT)

History, Theory, Problems and Usage

Introduction

Introduction

- Please translate the following sentence into English:
  „In den letzten Jahren wurde die Forschung auf dem Gebiet der maschinellen Übersetzung vorangetrieben, da die Anzahl zu übersetzender Dokumente ständig steigt."

Introduction

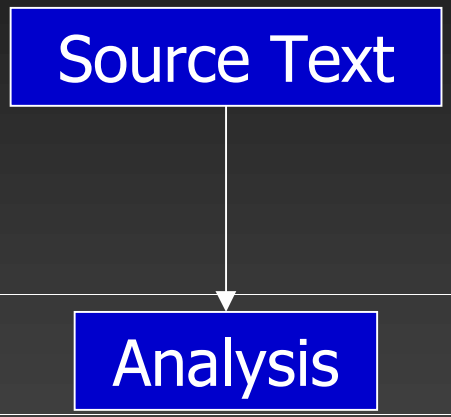
- Please translate the following sentence into English:
  „In den letzten Jahren wurde die Forschung auf dem Gebiet der maschinellen Übersetzung vorangetrieben, da die Anzahl zu übersetzender Dokumente ständig steigt."
- Now try to imagine how a computer would do this translation without knowing its meaning.

Translation Process

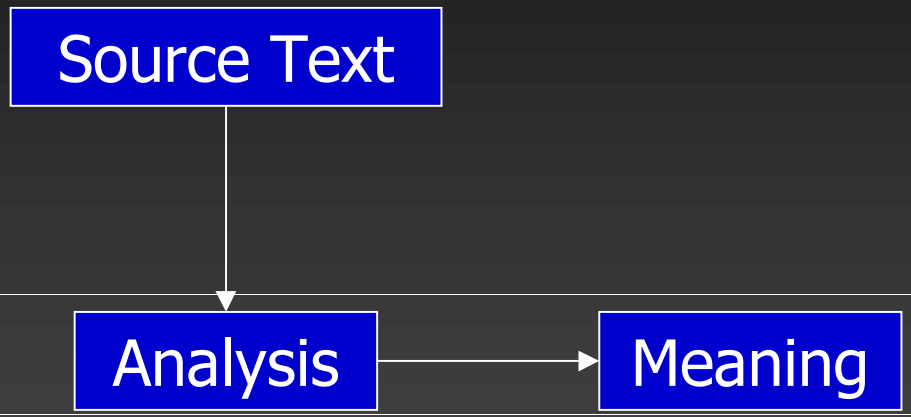
Translation Process

Source Text

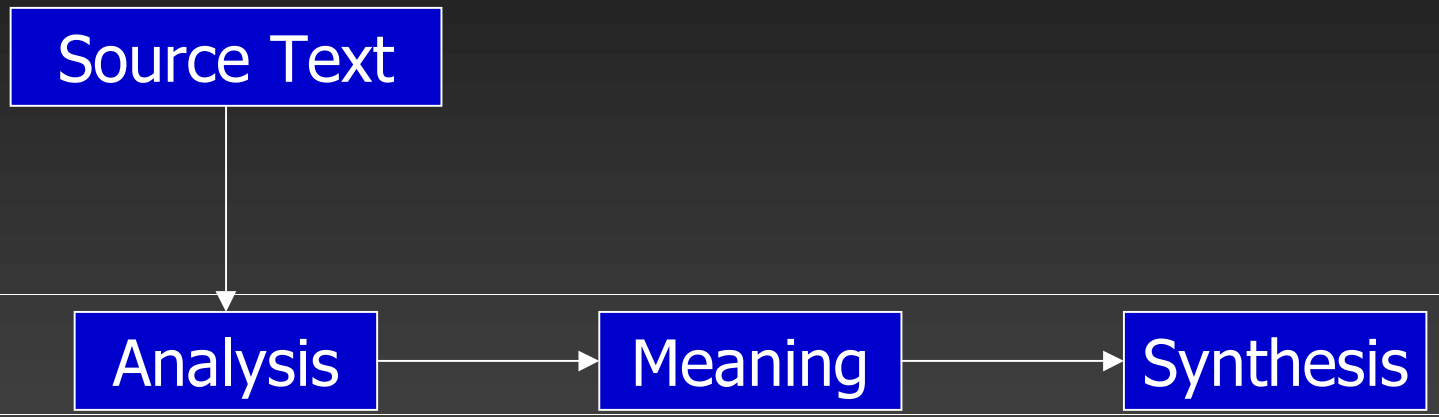
Translation Process



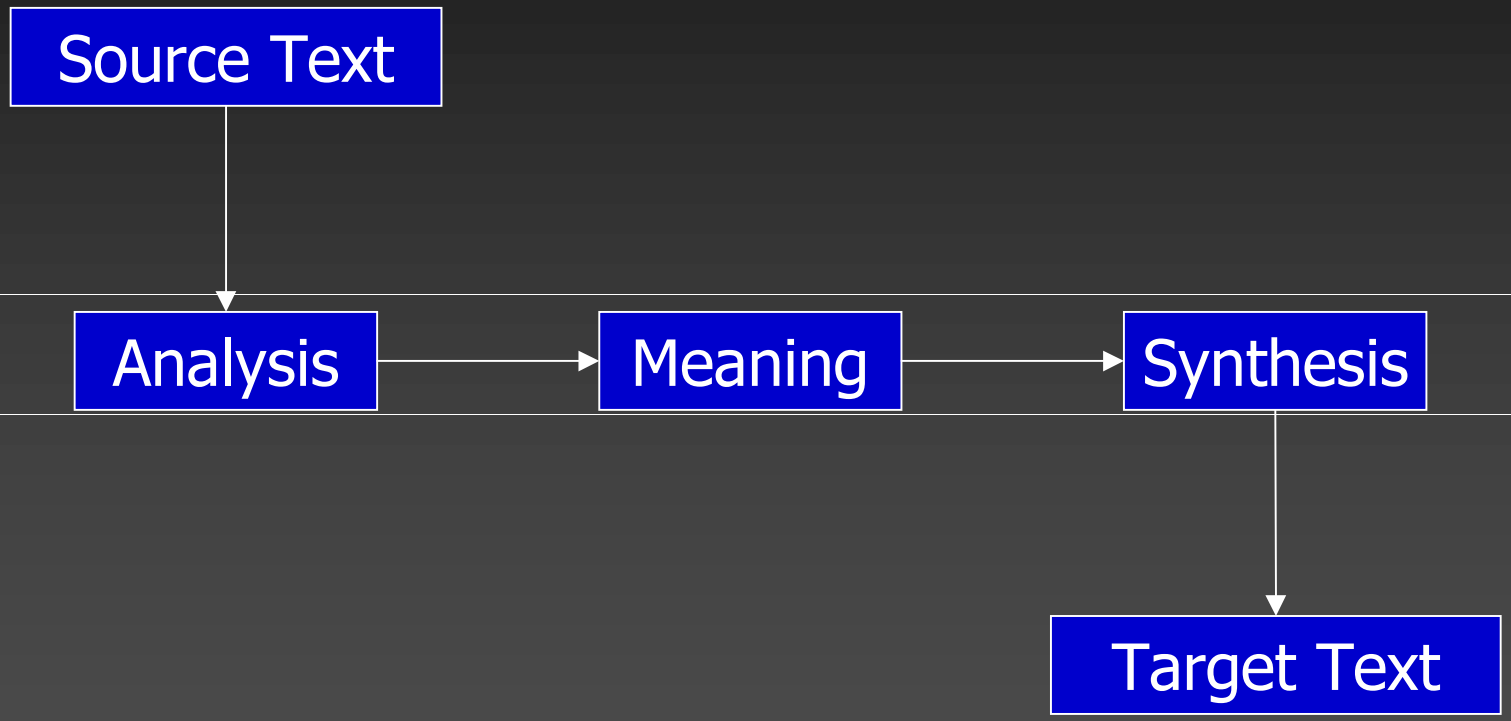
Translation Process



Translation Process



Translation Process



Historical Overview

Historical Overview

- four periods:
  - optimistic beginnings
  - disillusion
  - 70ies: partial successes
  - commercial application

Historical Overview

- four periods:
  - optimistic beginnings
  - disillusion
  - 70ies: partial successes
  - commercial application
- generations of translation techniques

History
- Optimistic Beginnings -

History
- Optimistic Beginnings -

- 1933 first machine supported translation systems (France, Russia)

History
- Optimistic Beginnings -

- 1933 first machine supported translation systems (France, Russia)
- 1942 first computer -> condition for development of machine translation created

History
- Optimistic Beginnings -

- 1933 first machine supported translation systems (France, Russia)
- 1942 first computer -> condition for development of machine translation created
- very optimistic attitude towards the problems, high expectations

History
- Optimistic Beginnings -

- 1933 first machine supported translation systems (France, Russia)
- 1942 first computer -> condition for development of machine translation created
- very optimistic attitude towards the problems, high expectations
- 1952 first MT conference

History
- Optimistic Beginnings -

- 1933 first machine supported translation systems (France, Russia)
- 1942 first computer -> condition for development of machine translation created
- very optimistic attitude towards the problems, high expectations
- 1952 first MT conference
- 1954 Georgetown Experiment -> enormous success, research intensified

History
- Disillusion -

History
- Disillusion -

- problems got more clearly

History
- Disillusion -

- problems got more clearly
- expectations not fulfilled

History
- Disillusion -

- problems got more clearly
- expectations not fulfilled
- 1964 founding of a committee to examine the future of MT

History
- Disillusion -

- problems got more clearly
- expectations not fulfilled
- 1964 founding of a committee to examine the future of MT
- 1966 ALPAC Report: very negative

History
- Disillusion -

- problems got more clearly
- expectations not fulfilled
- 1964 founding of a committee to examine the future of MT
- 1966 ALPAC Report: very negative
- much less research funds

History
- Disillusion -

- problems got more clearly
- expectations not fulfilled
- 1964 founding of a committee to examine the future of MT
- 1966 ALPAC Report: very negative
- much less research funds
- interest in MT decreased

History
- 70ies -

History
- 70ies -

- revival of MT research

History
- 70ies -

- revival of MT research
- much more realism

History
- 70ies -

- revival of MT research
- much more realism
- new research field: Computer Linguistics

History
- 70ies -

- revival of MT research
- much more realism
- new research field: Computer Linguistics
- development of grammatical formalism

History
- Commercial Application -

History
- Commercial Application -

- amount of documents, that have to be translated, steadily increasing

History
- Commercial Application -

- amount of documents, that have to be translated, steadily increasing
- MT now taken seriously and used commercially

History
- Commercial Application -

- amount of documents, that have to be translated, steadily increasing
- MT now taken seriously and used commercially
- successful research especially in Japan

History
- Commercial Application -

- amount of documents, that have to be translated, steadily increasing
- MT now taken seriously and used commercially
- successful research especially in Japan
- today: MT systems widely spread

History
- Commercial Application -

- amount of documents, that have to be translated, steadily increasing
- MT now taken seriously and used commercially
- successful research especially in Japan
- today: MT systems widely spread
- limits of application are marked clearly

Generations of Translation Techniques

Generations of Translation Techniques

- word-by-word-translation
  - simple comparison of dictionaries

Generations of Translation Techniques

- word-by-word-translation
  - simple comparison of dictionaries
- partial inclusion of syntax
  - syntactical rules about the structure of sentences extracted from the dictionaries

Generations of Translation Techniques

- word-by-word-translation
  - simple comparison of dictionaries
- partial inclusion of syntax
  - syntactical rules about the structure of sentences extracted from the dictionaries
- total inclusion of syntax
  - separate rules for syntactical analysis and synthesis

Generations of Translation Techniques

- word-by-word-translation
  - simple comparison of dictionaries
- partial inclusion of syntax
  - syntactical rules about the structure of sentences extracted from the dictionaries
- total inclusion of syntax
  - separate rules for syntactical analysis and synthesis
- partial inclusion of semantics
  - extra semantic rules to deal with ambiguities

Generations of Translation Techniques

- word-by-word-translation
  - simple comparison of dictionaries
- partial inclusion of syntax
  - syntactical rules about the structure of sentences extracted from the dictionaries
- total inclusion of syntax
  - separate rules for syntactical analysis and synthesis
- partial inclusion of semantics
  - extra semantic rules to deal with ambiguities

Theoretical Background

Theoretical Background

- What is an adequate translation?

Theoretical Background

- What is an adequate translation?
- Where does the meaning come from?

Theoretical Background

- What is an adequate translation?
- Where does the meaning come from?
- now: selection of the many translation problems

Theoretical Background

- What is an adequate translation?
- Where does the meaning come from?
- now: selection of the many translation problems
- practical exercises

Theoretical Background

- What is an adequate translation?
- Where does the meaning come from?
- now: selection of the many translation problems
- practical exercises

Problems of Machine Translation

- polysemy, homonymy
- syntactial ambiguities
- referential ambiguity
- Fuzzy Hedges
- synonyms
- metaphors and symbols
- new vocabulary developments

Problems of MT
- Polysemy -

Problems of MT
- Polysemy -

- **Polysem:** A word whose meanings are diverged or radiated.

Problems of MT
- Polysemy -

- **Polysem:** A word whose meanings are diverged or radiated.
- the proper translation is difficult to find even for a human translator

Problems of MT
- Polysemy -

- **Polysem:** A word whose meanings are diverged or radiated.
- the proper translation is difficult to find even for a human translator
- difficult to distinguish from Homonymy

Problems of MT
- Polysemy -

- **Polysem:** A word whose meanings are diverged or radiated.
- the proper translation is difficult to find even for a human translator
- difficult to distinguish from Homonymy
- examples:

Problems of MT
- Polysemy -

- **Polysem:** A word whose meanings are diverged or radiated.
- the proper translation is difficult to find even for a human translator
- difficult to distinguish from Homonymy
- examples:
  - "walk" might mean "Spaziergang" or "Spazieren gehen"

Problems of MT
- Polysemy -

- **Polysem:** A word whose meanings are diverged or radiated.
- the proper translation is difficult to find even for a human translator
- difficult to distinguish from Homonymy
- examples:
  - "walk" might mean "Spaziergang" or "Spazieren gehen"
  - "free" might mean "frei", "unabhängig" etc.

Problems of MT
- Homonymy -

Problems of MT
- Homonymy -

- **Homonym:** Several independent words which "share" the same linguistic body.

Problems of MT
- Homonymy -

- **Homonym:** Several independent words which "share" the same linguistic body.
- difficult to translate, often depends on context and semantics

Problems of MT
- Homonymy -

- **Homonym:** Several independent words which "share" the same linguistic body.
- difficult to translate, often depends on context and semantics
- examples:

Problems of MT
- Homonymy -

- **Homonym:** Several independent words which "share" the same linguistic body.
- difficult to translate, often depends on context and semantics
- examples:
  - "Reif" might mean "ring", "bracelet" or "white frost"

Problems of MT
- Homonymy -

- **Homonym:** Several independent words which "share" the same linguistic body.
- difficult to translate, often depends on context and semantics
- examples:
  - "Reif" might mean "ring", "bracelet" or "white frost"
  - "screen" might mean "Schirm", "Leinwand", "Tarnung", "Raster", "Abschirmung"

Problems of MT
- Homonymy -

- **Homonym:** Several independent words which "share" the same linguistic body.
- difficult to translate, often depends on context and semantics
- examples:
  - "Reif" might mean "ring", "bracelet" or "white frost"
  - "screen" might mean "Schirm", "Leinwand", "Tarnung", "Raster", "Abschirmung"

Problems of MT
- Syntactical Ambiguities -

Problems of MT
- Syntactical Ambiguities -

- structure of sentence not only depends on type of words but often also on semantics

Problems of MT
- Syntactical Ambiguities -

- structure of sentence not only depends on type of words but often also on semantics
- *"Flying planes can be dangerous."* is ambigious, words can be grouped in two ways

Problems of MT
- Syntactical Ambiguities -

- structure of sentence not only depends on type of words but often also on semantics
- *"Flying planes can be dangerous."* is ambigious, words can be grouped in two ways
  - "(Flying planes) can be dangerous." or

Problems of MT
- Syntactical Ambiguities -

- structure of sentence not only depends on type of words but often also on semantics
- *"Flying planes can be dangerous."* is ambigious, words can be grouped in two ways
  - "(Flying planes) can be dangerous." or
  - "(Flying) (planes) can be dangerous."

Problems of MT
- Syntactical Ambiguities -

- structure of sentence not only depends on type of words but often also on semantics
- *"Flying planes can be dangerous."* is ambigious, words can be grouped in two ways
  - "(Flying planes) can be dangerous." or
  - "(Flying) (planes) can be dangerous."

Problems of MT
- Referential Ambiguity -

Problems of MT
- Referential Ambiguity -

- pronouns refer to certain words but it is often not obvious to which

Problems of MT
- Referential Ambiguity -

- pronouns refer to certain words but it is often not obvious to which
- references might even cross sentence boundaries

Problems of MT
- Referential Ambiguity -

Problems of MT
- Referential Ambiguity -

- examples:

Problems of MT
- Referential Ambiguity -

- examples:
  - *"I missed my cat. It seemed to have disappeared."* The computer needs to figure out which word is referenced by "it".

Problems of MT
- Referential Ambiguity -

- examples:
  - *"I missed my cat. It seemed to have disappeared."* The computer needs to figure out which word is referenced by "it".
  - *"My cat was chasing a mouse. It played with it."*
    Everybody knows that the cat is playing with the mouse. But how is a computer supposed to know?

Problems of MT
- Referential Ambiguity -

- examples:
  - *"I missed my cat. It seemed to have disappeared."* The computer needs to figure out which word is referenced by "it".
  - *"My cat was chasing a mouse. It played with it."*
    Everybody knows that the cat is playing with the mouse. But how is a computer supposed to know?

Problems of MT
- Fuzzy Hedges -

Problems of MT
- Fuzzy Hedges -

- vague words, terms and expressions

Problems of MT
- Fuzzy Hedges -

- vague words, terms and expressions
- very language dependent

Problems of MT
- Fuzzy Hedges -

- vague words, terms and expressions
- very language dependent
- difficult to translate

Problems of MT
- Fuzzy Hedges -

- vague words, terms and expressions
- very language dependent
- difficult to translate
- examples:

Problems of MT
- Fuzzy Hedges -

- vague words, terms and expressions
- very language dependent
- difficult to translate
- examples:
  - "in a sense", "irgendwie", "very"

Problems of MT
- Fuzzy Hedges -

- vague words, terms and expressions
- very language dependent
- difficult to translate
- examples:
  - "in a sense", "irgendwie", "very"
  - *"High quality fully automatic machine translation is considered to be virtually impossible."*

Problems of MT
- Fuzzy Hedges -

- vague words, terms and expressions
- very language dependent
- difficult to translate
- examples:
  - "in a sense", "irgendwie", "very"
  - *"High quality fully automatic machine translation is considered to be virtually impossible."*

Problems of MT
- Metaphors and Symbols -

Problems of MT
- Metaphors and Symbols -

- metaphors and symbols depend on the underlying culture and history

Problems of MT
- Metaphors and Symbols -

- metaphors and symbols depend on the underlying culture and history
- often cannot be translated (Chinese sayings sometimes just do not make sense to non-Chinese people)

Problems of MT
- Metaphors and Symbols -

- metaphors and symbols depend on the underlying culture and history
- often cannot be translated (Chinese sayings sometimes just do not make sense to non-Chinese people)
- idiomatic dictionaries may be used to ease translation

Problems of MT
- Metaphors and Symbols -

Problems of MT
- Metaphors and Symbols -

- example:

Problems of MT
- Metaphors and Symbols -

- example:
  - "Mit eiserner Miene feuerte er seinen treuesten Mitarbeiter."

Problems of MT
- Metaphors and Symbols -

- example:
  - "Mit eiserner Miene feuerte er seinen treuesten Mitarbeiter."
  - corresponding English idiom: "with a stony expression"

Problems of MT
- New Developments -

Problems of MT
- New Developments -

- languages are dynamic

Problems of MT
- New Developments -

- languages are dynamic
- new words created

Problems of MT
- New Developments -

- languages are dynamic
- new words created
- proper names of new technologies

Problems of MT
- New Developments -

- languages are dynamic
- new words created
- proper names of new technologies
- example:
  - Secure Shell
  - Telnet

Problems of MT
- New Developments -

- languages are dynamic
- new words created
- proper names of new technologies
- example:
  - Secure Shell
  - Telnet

Problems of MT
- Synonyms -

Problems of MT
- Synonyms -

- there are often several words with almost the same meaning

Problems of MT
- Synonyms -

- there are often several words with almost the same meaning
- it is difficult to choose the right one

Problems of MT
- Synonyms -

- there are often several words with almost the same meaning
- it is difficult to choose the right one
- it depends on context, style and semantics

Problems of MT
- Synonyms -

- there are often several words with almost the same meaning
- it is difficult to choose the right one
- it depends on context, style and semantics

Real World Usage

Real World Usage

- three categories

Real World Usage

- three categories
  - human translation with machine support

Real World Usage

- three categories
  - human translation with machine support
  - machine translation with human support

Real World Usage

- three categories
  - human translation with machine support
  - machine translation with human support
  - fully automated translation

Human Translation with Machine Support

Human Translation with Machine Support

- fast & easy access to to dicitionaries, thesauri etc.

Human Translation with Machine Support

- fast & easy access to to dicitionaries, thesauri etc.
- system provides suggestions and alternatives for translation

Human Translation with Machine Support

- fast & easy access to to dicitionaries, thesauri etc.
- system provides suggestions and alternatives for translation
- human translator decides

Human Translation with Machine Support

- fast & easy access to to dicitionaries, thesauri etc.
- system provides suggestions and alternatives for translation
- human translator decides
- extendable dictionaries

Machine Translation with Human Support

Machine Translation with Human Support

- computer controls the translation process

Machine Translation with Human Support

- computer controls the translation process
- computer asks the human if it comes across problems like ambiguities

Machine Translation with Human Support

- computer controls the translation process
- computer asks the human if it comes across problems like ambiguities
- interaction takes place when the computer requests it

Machine Translation with Human Support

- computer controls the translation process
- computer asks the human if it comes across problems like ambiguities
- interaction takes place when the computer requests it

Fully Automated Translation

Fully Automated Translation

- speed over quality

Fully Automated Translation

- speed over quality
- useful to get an overall idea of text contents

Fully Automated Translation

- speed over quality
- useful to get an overall idea of text contents
- pre- and postprocessing usually neccessary

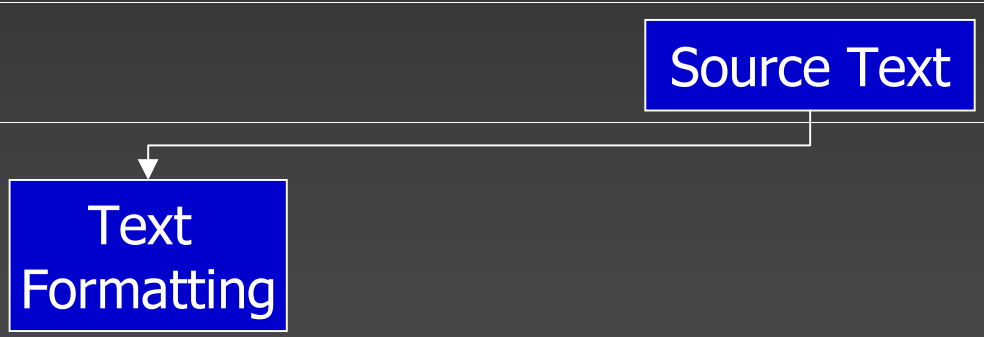
Fully Automated Translation

- speed over quality
- useful to get an overall idea of text contents
- pre- and postprocessing usually neccessary

Source Text

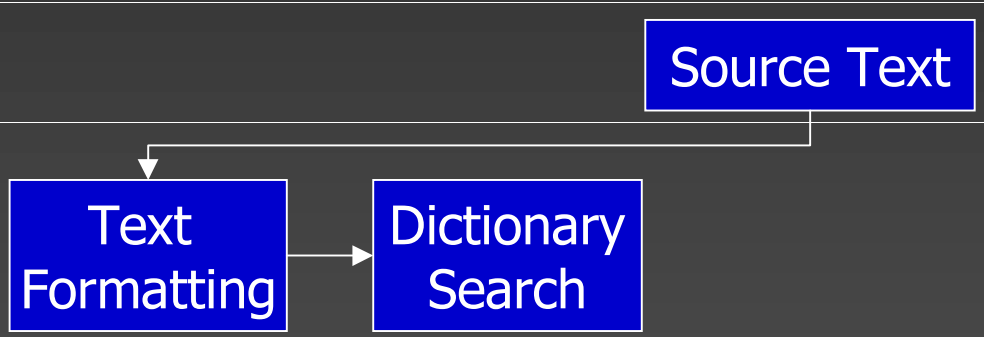
Fully Automated Translation

- speed over quality
- useful to get an overall idea of text contents
- pre- and postprocessing usually neccessary



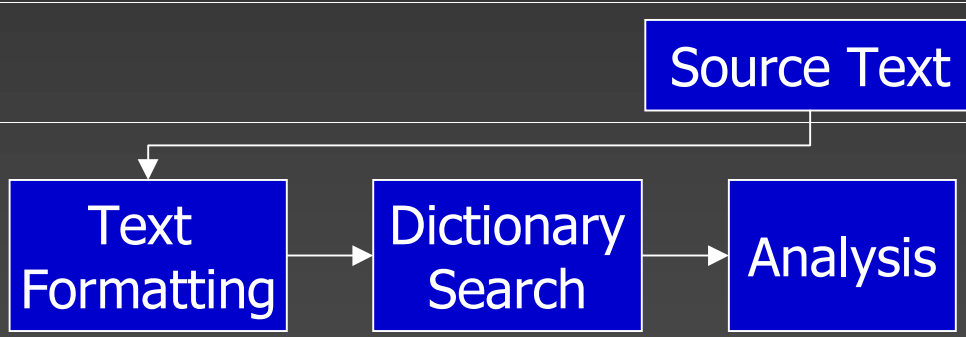
Fully Automated Translation

- speed over quality
- useful to get an overall idea of text contents
- pre- and postprocessing usually neccessary



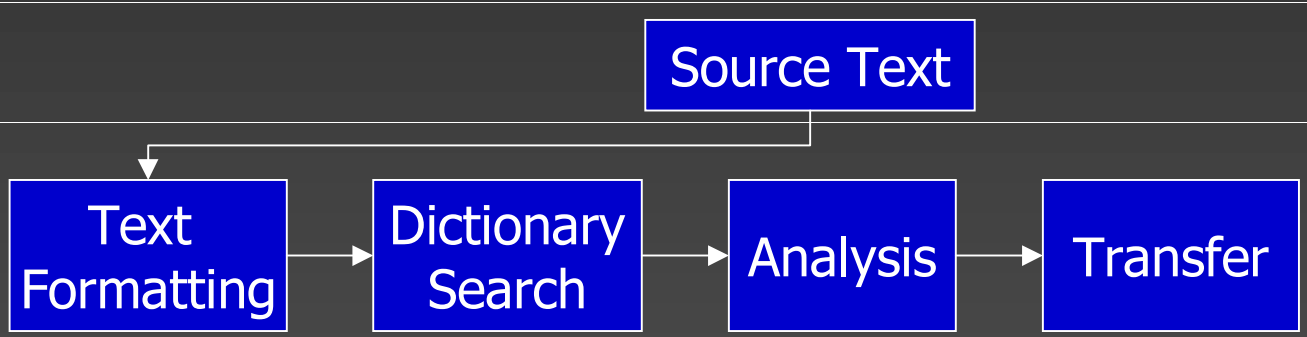
Fully Automated Translation

- speed over quality
- useful to get an overall idea of text contents
- pre- and postprocessing usually neccessary



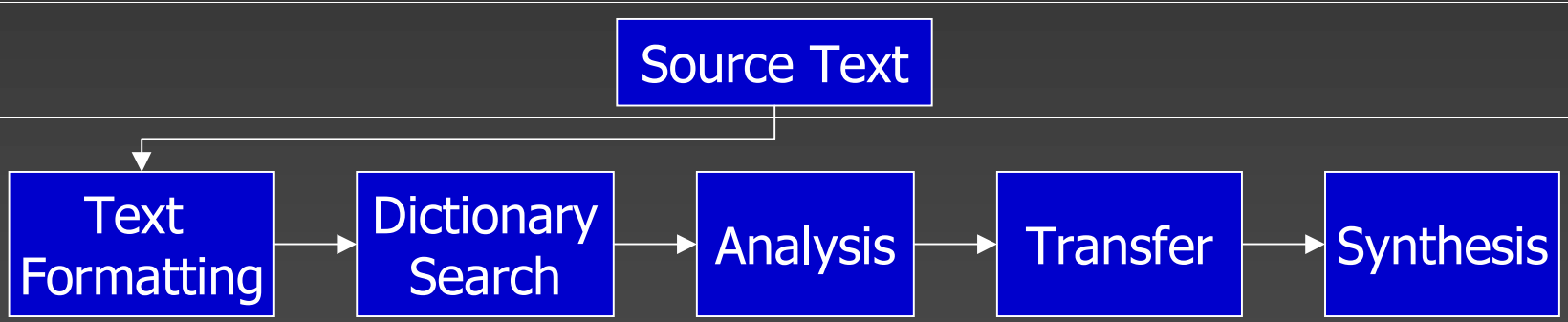
Fully Automated Translation

- speed over quality
- useful to get an overall idea of text contents
- pre- and postprocessing usually neccessary



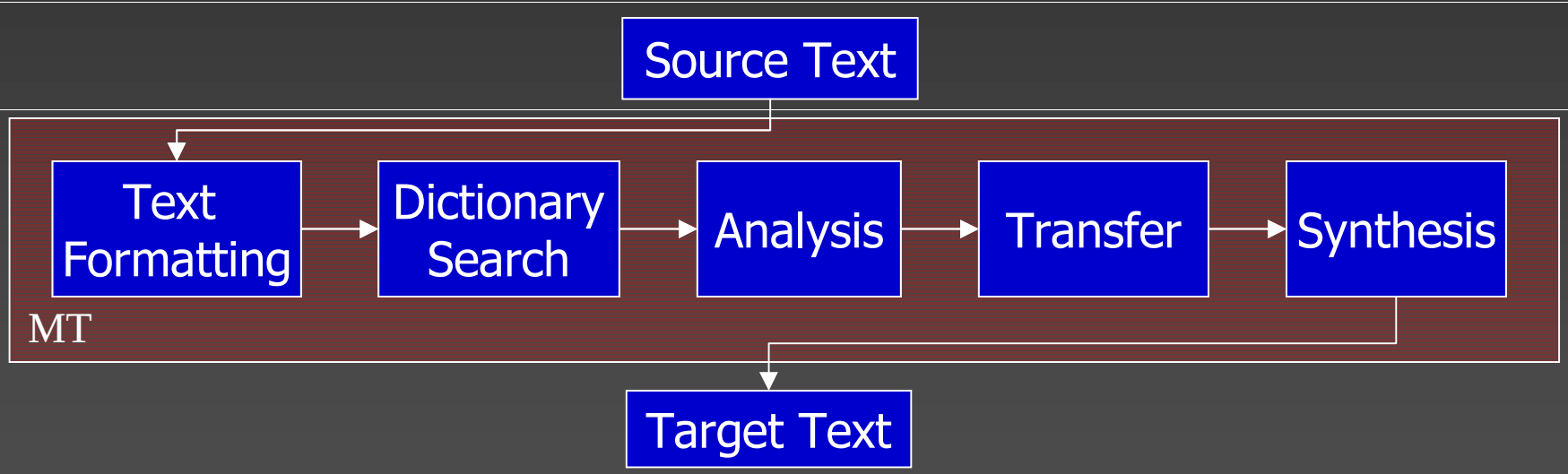
Fully Automated Translation

- speed over quality
- useful to get an overall idea of text contents
- pre- and postprocessing usually neccessary



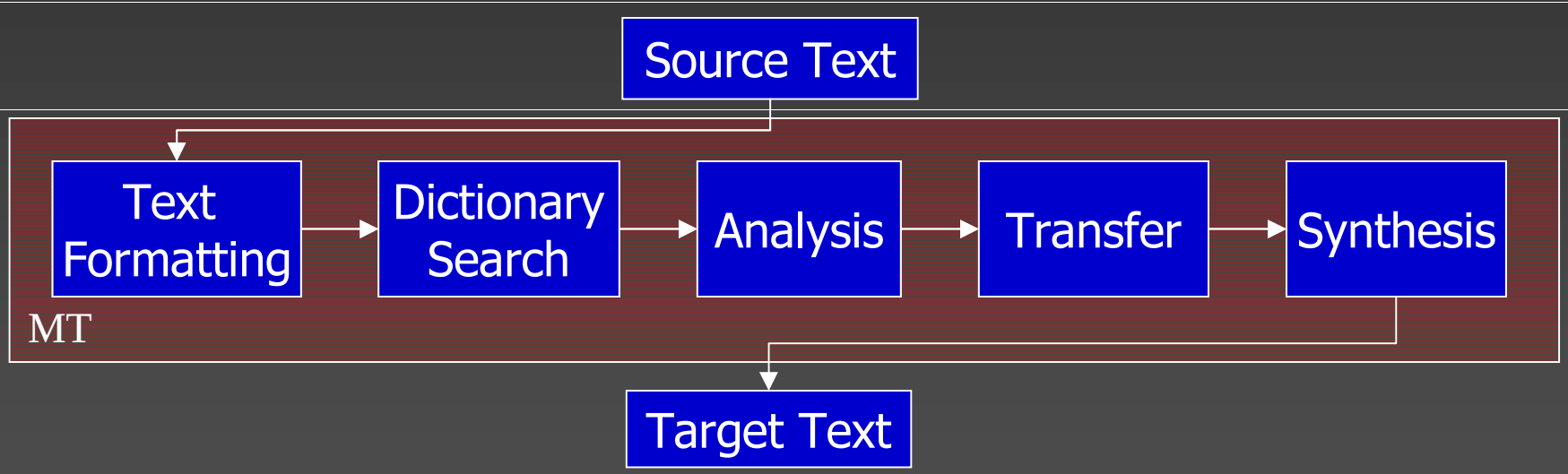
Fully Automated Translation

- speed over quality
- useful to get an overall idea of text contents
- pre- and postprocessing usually neccessary



Fully Automated Translation

- speed over quality
- useful to get an overall idea of text contents
- pre- and postprocessing usually neccessary



Current Research

Current Research

- architectures of translation systems
  - rule based paradigm
  - data oriented paradigm

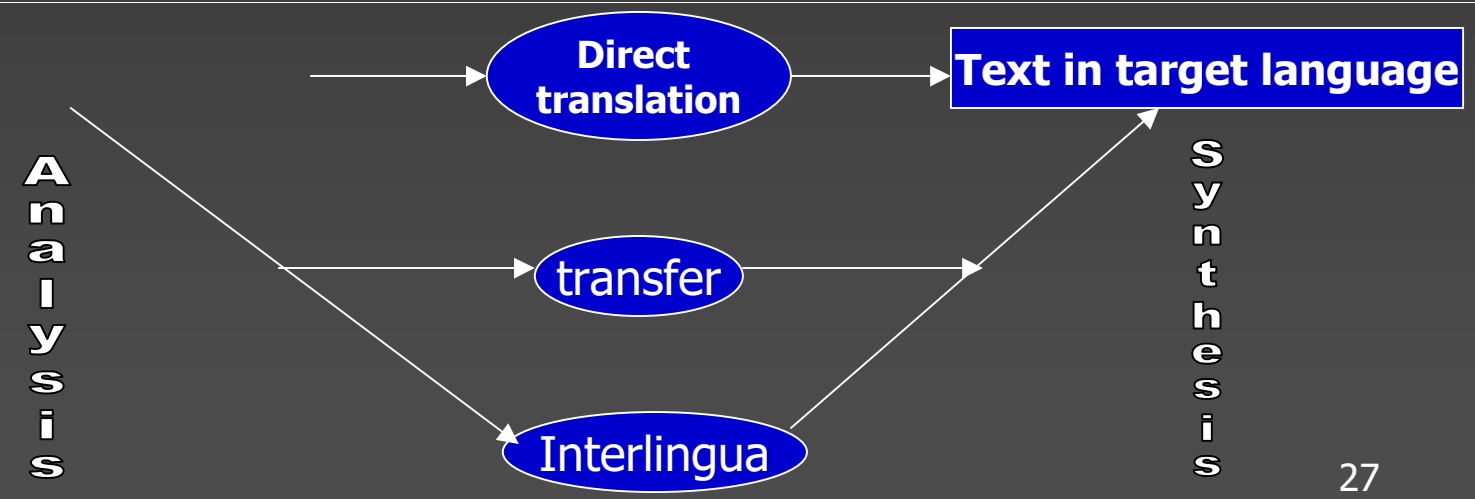
Current Research

- architectures of translation systems
  - rule based paradigm
  - data oriented paradigm
- machine interpreting

Current Research

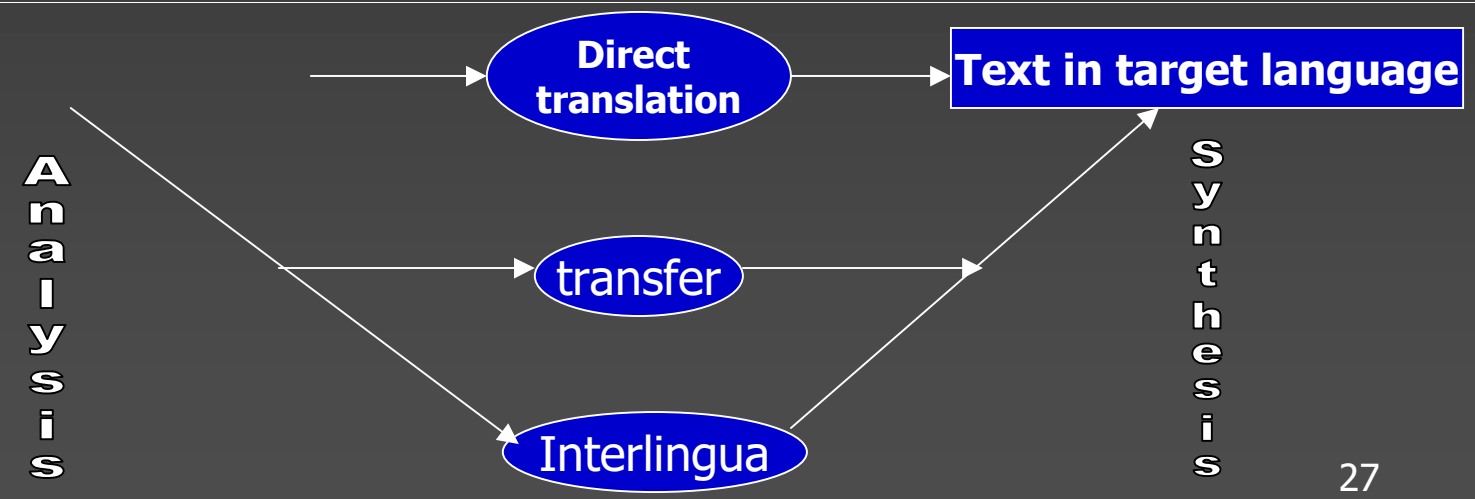
- architectures of translation systems
  - rule based paradigm
  - data oriented paradigm
- machine interpreting
- Artificial Intelligence

Architectures: Rule Based Paradigm



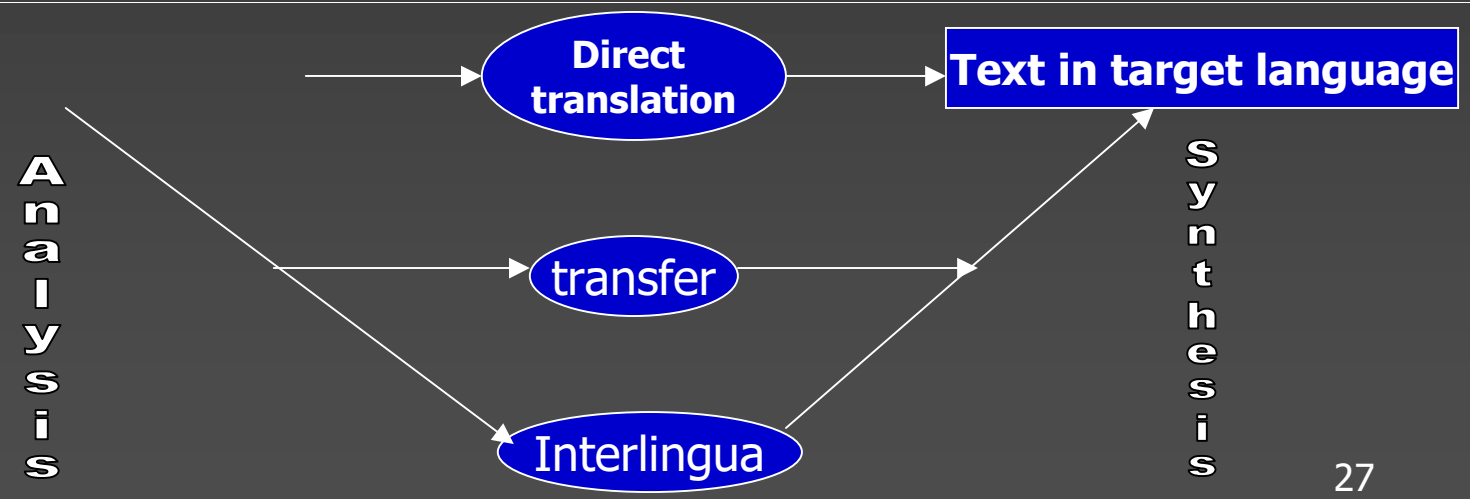
Architectures: Rule Based Paradigm

- basic functionality represented by rules



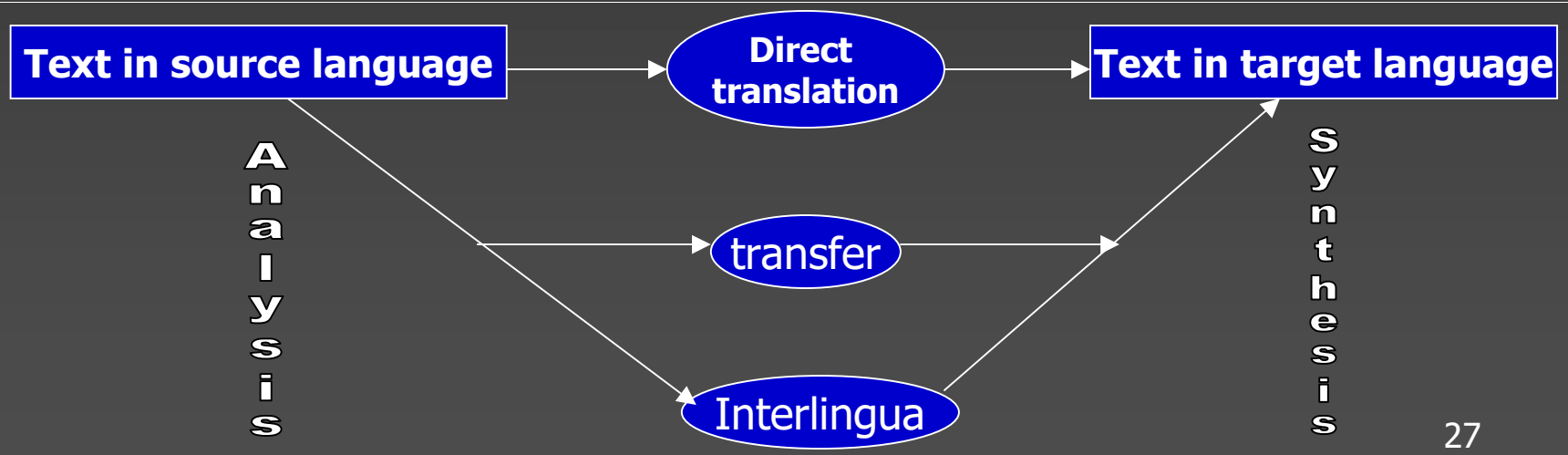
Architectures: Rule Based Paradigm

- basic functionality represented by rules
- translation strategies:
  - direct translation
  - transfer approach
  - interlingua approach



Architectures: Rule Based Paradigm

- basic functionality represented by rules
- translation strategies:
  - direct translation
  - transfer approach
  - interlingua approach



Architectures:
Data Oriented Paradigm

Architectures: Data Oriented Paradigm

- statistical MT

Architectures: Data Oriented Paradigm

- statistical MT
  - starting point: each sentence of a language is a possible translation of a sentence in another language

Architectures: Data Oriented Paradigm

- statistical MT
  - starting point: each sentence of a language is a possible translation of a sentence in another language
  - works with assigned probabilities

Architectures: Data Oriented Paradigm

- statistical MT
  - starting point: each sentence of a language is a possible translation of a sentence in another language
  - works with assigned probabilities
  - no linguistic knowledge necessary

Architectures: Data Oriented Paradigm

- statistical MT
  - starting point: each sentence of a language is a possible translation of a sentence in another language
  - works with assigned probabilities
  - no linguistic knowledge necessary
- example based MT

Architectures: Data Oriented Paradigm

- statistical MT
  - starting point: each sentence of a language is a possible translation of a sentence in another language
  - works with assigned probabilities
  - no linguistic knowledge necessary
- example based MT
  - basis: former produced translations

Architectures: Data Oriented Paradigm

- statistical MT
  - starting point: each sentence of a language is a possible translation of a sentence in another language
  - works with assigned probabilities
  - no linguistic knowledge necessary
- example based MT
  - basis: former produced translations
  - analogue forming of new translations

Machine Interpreting

Machine Interpreting

- very young research field

Machine Interpreting

- very young research field
- approaches from MT and Speech Technology combined

Machine Interpreting

- very young research field
- approaches from MT and Speech Technology combined
- need to deal with phenomena of spontaneous language ("ähm")

Machine Interpreting

- very young research field
- approaches from MT and Speech Technology combined
- need to deal with phenomena of spontaneous language ("ähm")
- very high level of difficulty

Artificial Intelligence

Artificial Intelligence

- field of computer science

Artificial Intelligence

- field of computer science
- try to develop computer systems that can think and learn by themselves

Artificial Intelligence

- field of computer science
- try to develop computer systems that can think and learn by themselves
- Neural Networks

Artificial Intelligence

- field of computer science
- try to develop computer systems that can think and learn by themselves
- Neural Networks
- future: will help to achieve better results in MT

Artificial Intelligence

- field of computer science
- try to develop computer systems that can think and learn by themselves
- Neural Networks
- future: will help to achieve better results in MT

Artificial Intelligence

- field of computer science
- try to develop computer systems that can think and learn by themselves
- Neural Networks
- future: will help to achieve better results in MT

Conclusion

Conclusion

Current machine translation systems are already very helpful, but not perfect. There are linguistic problems that cannot be satisfyingly solved by computers unable to think like human beings. Maybe in the future, further progress in Artificial Intelligence will help to solve the remaining problems.

Conclusion

Current machine translation systems are already very helpful, but not perfect. There are linguistic problems that cannot be satisfyingly solved by computers unable to think like human beings. Maybe in the future, further progress in Artificial Intelligence will help to solve the remaining problems.