Linguistics for Language Technologies, Language Technologies for Linguistics

Adam Przepiórkowski

INSTITUTE OF COMPUTER SCIENCE
POLISH ACADEMY OF SCIENCES
ul. Jana Kazimierza 5, 01-248 Warsaw

HLT Days 2012

27 September 2012

Outline

1 Lx4LT

2 LT4Lx

3 Lx4LT again?

4 Conclusion

Lx4LT 1

Karen Sparck Jones, *Natural Language Processing: a historical review*, 2001 – dominating approaches in Natural Language Processing (NLP):

- late 1940s – late 1960s: early Machine Translation (focus on syntax),
- late 1960s – late 1970s: AI-flavoured (knowledge, semantics, inference),
- **late 1970s – early 1990s: grammatico-logical tradition (unification grammar formalisms, logical representations; syntax, semantics, discourse)**,
- from early 1990s: revival of finite state methods, shallow and statistical approaches, applications (MT, Automatic Summarisation, Question Answering, Information Extraction, etc.).

Lx4LT 1

Karen Sparck Jones, *Natural Language Processing: a historical review*, 2001 – dominating approaches in Natural Language Processing (NLP):

- late 1940s – late 1960s: early Machine Translation (focus on syntax),
- late 1960s – late 1970s: AI-flavoured (knowledge, semantics, inference),
- late 1970s – early 1990s: grammatico-logical tradition (unification grammar formalisms, logical representations; syntax, semantics, discourse),
- from early 1990s: revival of finite state methods, shallow and statistical approaches, applications (MT, Automatic Summarisation, Question Answering, Information Extraction, etc.).

Lx4LT 1

Karen Sparck Jones, *Natural Language Processing: a historical review*, 2001 – dominating approaches in Natural Language Processing (NLP):

- late 1940s – late 1960s: early Machine Translation (focus on syntax),
- late 1960s – late 1970s: AI-flavoured (knowledge, semantics, inference),
- late 1970s – early 1990s: grammatico-logical tradition (unification grammar formalisms, logical representations; syntax, semantics, discourse),
- from early 1990s: revival of finite state methods, shallow and statistical approaches, applications (MT, Automatic Summarisation, Question Answering, Information Extraction, etc.).

Lx4LT 1

Karen Sparck Jones, *Natural Language Processing: a historical review*, 2001 – dominating approaches in Natural Language Processing (NLP):

- late 1940s – late 1960s: early Machine Translation (focus on syntax),
- late 1960s – late 1970s: AI-flavoured (knowledge, semantics, inference),
- **late 1970s – early 1990s: grammatico-logical tradition (unification grammar formalisms, logical representations; syntax, semantics, discourse)**,
- from early 1990s: revival of finite state methods, shallow and statistical approaches, applications (MT, Automatic Summarisation, Question Answering, Information Extraction, etc.).

Lx4LT 1

Karen Sparck Jones, *Natural Language Processing: a historical review*, 2001 – dominating approaches in Natural Language Processing (NLP):

- late 1940s – late 1960s: early Machine Translation (focus on syntax),
- late 1960s – late 1970s: AI-flavoured (knowledge, semantics, inference),
- **late 1970s – early 1990s: grammatico-logical tradition (unification grammar formalisms, logical representations; syntax, semantics, discourse)**,
- from early 1990s: revival of finite state methods, shallow and statistical approaches, applications (MT, Automatic Summarisation, Question Answering, Information Extraction, etc.).

Lx4LT 1

Karen Sparck Jones, *Natural Language Processing: a historical review*, 2001 – dominating approaches in Natural Language Processing (NLP):

- late 1940s – late 1960s: early Machine Translation (focus on syntax),
- late 1960s – late 1970s: AI-flavoured (knowledge, semantics, inference),
- **late 1970s – early 1990s: grammatico-logical tradition (unification grammar formalisms, logical representations; syntax, semantics, discourse)**,
- from early 1990s: revival of finite state methods, shallow and statistical approaches, applications (MT, Automatic Summarisation, Question Answering, Information Extraction, etc.).

Every time we fire a linguist, the performance of our system goes up – Frederick Jelinek, IBM, around 1985–1988.

Lx4LT 2

The grammatico-logical tradition (late 1970s, 1980s):

- Lexical-Functional Grammar (LFG): Joan Bresnan and Ronald Kaplan,
- Generalized Phrase Structure Grammar (GPSG): Gerald Gazdar et al.,
- Head-driven Phrase Structure Grammar (HPSG): Carl Pollard and Ivan Sag,
- Discourse Representation Theory (DRT): Hans Kamp.

LFG and HPSG parsers still actively developed, and even used commercially. DRT occasionally used as the semantic representation language.

Lx4LT 2

The grammatico-logical tradition (late 1970s, 1980s):

- Lexical-Functional Grammar (LFG): Joan Bresnan and Ronald Kaplan,
- Generalized Phrase Structure Grammar (GPSG): Gerald Gazdar et al.,
- Head-driven Phrase Structure Grammar (HPSG): Carl Pollard and Ivan Sag,
- Discourse Representation Theory (DRT): Hans Kamp.

LFG and HPSG parsers still actively developed, and even used commercially. DRT occasionally used as the semantic representation language.

Lx4LT 2

The grammatico-logical tradition (late 1970s, 1980s):

- Lexical-Functional Grammar (LFG): Joan Bresnan and Ronald Kaplan,
- Generalized Phrase Structure Grammar (GPSG): Gerald Gazdar et al.,
- Head-driven Phrase Structure Grammar (HPSG): Carl Pollard and Ivan Sag,
- Discourse Representation Theory (DRT): Hans Kamp.

LFG and HPSG parsers still actively developed, and even used commercially. DRT occasionally used as the semantic representation language.

Lx4LT 3

Yorick Wilks, *The History of Natural Language Processing and Machine Translation*, 2005 – interesting discussion of other influences of linguistics on NLP, e.g.:

- Charles Fillmore, *The Case for Case*, 1968 – basis for Semantic Role Labelling, FrameNet, etc.,
- M.A.K. Halliday's Systemic-Functional Grammar (since early 1960s) has influenced, e.g.:
  - Winograd's MIT Block World application (late 1960s, early 1970s),
  - the design of Functional Unification Grammar (early 1980s),
  - and Sentiment Analysis (via the Appraisal Theory of mid-1990s),
- Steven Abney's 1990s work on chunking (and shallow parsing in general).

On the other hand, hardly any direct influence of Chomsky's transformational approaches.

Yorick Wilks, *The History of Natural Language Processing and Machine Translation*, 2005 – interesting discussion of other influences of linguistics on NLP, e.g.:

- Charles Fillmore, *The Case for Case*, 1968 – basis for Semantic Role Labelling, FrameNet, etc.,
- M.A.K. Halliday's Systemic-Functional Grammar (since early 1960s) has influenced, e.g.:
  - Winograd's MIT Block World application (late 1960s, early 1970s),
  - the design of Functional Unification Grammar (early 1980s),
  - and Sentiment Analysis (via the Appraisal Theory of mid-1990s),
- Steven Abney's 1990s work on chunking (and shallow parsing in general).

On the other hand, hardly any direct influence of Chomsky's transformational approaches.

Yorick Wilks, *The History of Natural Language Processing and Machine Translation*, 2005 – interesting discussion of other influences of linguistics on NLP, e.g.:

- Charles Fillmore, *The Case for Case*, 1968 – basis for Semantic Role Labelling, FrameNet, etc.,
- M.A.K. Halliday's Systemic-Functional Grammar (since early 1960s) has influenced, e.g.:
  - Winograd's MIT Block World application (late 1960s, early 1970s),
  - the design of Functional Unification Grammar (early 1980s),
  - and Sentiment Analysis (via the Appraisal Theory of mid-1990s),
- Steven Abney's 1990s work on chunking (and shallow parsing in general).

On the other hand, hardly any direct influence of Chomsky's transformational approaches.

Yorick Wilks, *The History of Natural Language Processing and Machine Translation*, 2005 – interesting discussion of other influences of linguistics on NLP, e.g.:

- Charles Fillmore, *The Case for Case*, 1968 – basis for Semantic Role Labelling, FrameNet, etc.,
- M.A.K. Halliday's Systemic-Functional Grammar (since early 1960s) has influenced, e.g.:
  - Winograd's MIT Block World application (late 1960s, early 1970s),
  - the design of Functional Unification Grammar (early 1980s),
  - and Sentiment Analysis (via the Appraisal Theory of mid-1990s),
- Steven Abney's 1990s work on chunking (and shallow parsing in general).

On the other hand, hardly any direct influence of Chomsky's transformational approaches.

Yorick Wilks, *The History of Natural Language Processing and Machine Translation*, 2005 – interesting discussion of other influences of linguistics on NLP, e.g.:

- Charles Fillmore, *The Case for Case*, 1968 – basis for Semantic Role Labelling, FrameNet, etc.,
- M.A.K. Halliday's Systemic-Functional Grammar (since early 1960s) has influenced, e.g.:
  - Winograd's MIT Block World application (late 1960s, early 1970s),
  - the design of Functional Unification Grammar (early 1980s),
  - and Sentiment Analysis (via the Appraisal Theory of mid-1990s),
- Steven Abney's 1990s work on chunking (and shallow parsing in general).

On the other hand, hardly any direct influence of Chomsky's transformational approaches.

A lot of hidden linguistics in **annotated** corpora:

- **morphosyntactically** – for training taggers,
- **syntactically** – for training parsers,
- **semantically** – for training Word Sense Disambiguation systems, Semantic Labelling systems, etc.

Applying supervised machine learning techniques to such resources requires no linguistic knowledge, but **writing annotation guidelines – and then supervising the annotation – requires considerable linguistic expertise.**

A lot of hidden linguistics in **annotated** corpora:

- **morphosyntactically** – for training taggers,
- **syntactically** – for training parsers,
- **semantically** – for training Word Sense Disambiguation systems, Semantic Labelling systems, etc.

Applying supervised machine learning techniques to such resources requires no linguistic knowledge, but **writing annotation guidelines – and then supervising the annotation – requires considerable linguistic expertise**.

Outline

1 Lx4LT

2 LT4Lx

3 Lx4LT again?

4 Conclusion

Increasing flow of knowledge in the other other direction: from Language Technologies to Linguistics.

Most obvious, but also rather trivial: the use of language corpora in **lexicography**.

Increasing flow of knowledge in the other other direction: from Language Technologies to Linguistics.

Most obvious, but also rather trivial: the use of language corpora in **lexicography**.

“Peter Sokolowski, editor at large for Merriam-Webster Inc., thumbs through the index card files at the dictionary publisher’s headquarters in Springfield, Mass.” (Associated Press, 2011)

More interesting: using data mining techniques to discover linguistic facts – **annotation mining**.

Tibor Kiss et al., 2010, *A Logistic Regression Model of Determiner Omission in PPs*:

- why *to the doctor*, *from the cinema*, but *from school*, *in jail*, *by train*, etc.?
- research on German, on a corpus annotated with parts of speech, preposition senses, etc.,
- result – a multicausal phenomenon:
  - the sense of the preposition,
  - the sense of the noun,
  - syntactic complexity of the embedded noun phrase.

More interesting: using data mining techniques to discover linguistic facts – **annotation mining**.

Tibor Kiss et al., 2010, *A Logistic Regression Model of Determiner Omission in PPs*:

- why *to the doctor*, *from the cinema*, but *from school*, *in jail*, *by train*, etc.?
- research on German, on a corpus annotated with parts of speech, preposition senses, etc.,
- result – a multicausal phenomenon:
  - the sense of the preposition,
  - the sense of the noun,
  - syntactic complexity of the embedded noun phrase.

More interesting: using data mining techniques to discover linguistic facts – **annotation mining**.

Tibor Kiss et al., 2010, *A Logistic Regression Model of Determiner Omission in PPs*:

- why *to the doctor*, *from the cinema*, but *from school*, *in jail*, *by train*, etc.?
- research on German, on a corpus annotated with parts of speech, preposition senses, etc.,
- result – a multicausal phenomenon:
  - the sense of the preposition,
  - the sense of the noun,
  - syntactic complexity of the embedded noun phrase.

More interesting: using data mining techniques to discover linguistic facts – **annotation mining**.

Tibor Kiss et al., 2010, *A Logistic Regression Model of Determiner Omission in PPs*:

- why *to the doctor*, *from the cinema*, but *from school*, *in jail*, *by train*, etc.?
- research on German, on a corpus annotated with parts of speech, preposition senses, etc.,
- result – a multicausal phenomenon:
  - the sense of the preposition,
  - the sense of the noun,
  - syntactic complexity of the embedded noun phrase.

CESAR
CEntral And South-east europeAn Resources

Another example – valence acquisition:

I went to Warsaw. HLT Days 2012 was great.

1. Partition the data into sentential clauses.
2. Identify verbs and argument phrases in each clause.
3. Record each verb-argument(s) combination as a valence frame for the verb.

META-NET

Another example – valence acquisition:

I went to Warsaw. HLT Days 2012 was great.

1. Partition the data into sentential clauses.
2. Identify verbs and argument phrases in each clause.
3. Record each verb-argument(s) combination as a valence frame for the verb.

Another example – valence acquisition:

$$\text{I}_{NP_{nom}}\ \underline{\text{went}}\ [\text{to Warsaw}]_{PP_{to}}$$

1. Partition the data into sentential clauses.
2. Identify verbs and argument phrases in each clause.
3. Record each verb-argument(s) combination as a valence frame for the verb.

Another example – valence acquisition:

go: 1. NP_{nom}, PP_{to}
2. ...

1. Partition the data into sentential clauses.
2. Identify verbs and argument phrases in each clause.
3. Record each verb-argument(s) combination as a valence frame for the verb.

CESAR
CEntral And South-east europeAn Resources

1. Partition the data into sentential clauses.
2. Identify verbs and argument phrases in each clause.
3. Record each verb-argument(s) combination as a valence frame for the verb.

Noise produced by:

- morphosyntactic annotation,
- automatic identification of arguments.

Practically always **two stages**:

- **linguistic**: collection of evidence,
- **statistical**: deciding which evidence is reliable.

CESAR
CEntral And South-east europeAn Resources

1. Partition the data into sentential clauses.
2. Identify verbs and argument phrases in each clause.
3. Record each verb-argument(s) combination as a valence frame for the verb.

Noise produced by:

- morphosyntactic annotation,
- automatic identification of arguments.

Practically always **two stages**:

- **linguistic**: collection of evidence,
- **statistical**: deciding which evidence is reliable.

CESAR
CEntral And South-east europeAn Resources

1. Partition the data into sentential clauses.
2. Identify verbs and argument phrases in each clause.
3. Record each verb-argument(s) combination as a valence frame for the verb.

Noise produced by:

- morphosyntactic annotation,
- automatic identification of arguments.

Practically always **two stages**:

- **linguistic**: collection of evidence,
- **statistical**: deciding which evidence is reliable.

CESAR
CEntral And South-east europeAn Resources

1. Partition the data into sentential clauses.
2. Identify verbs and argument phrases in each clause.
3. Filter attested verb-argument(s) combinations based on their statistical distribution.
4. Record each remaining verb-argument(s) combination as a valence frame for the verb.

Noise produced by:

- morphosyntactic annotation,
- automatic identification of arguments.

Practically always **two stages**:

- **linguistic**: collection of evidence,
- **statistical**: deciding which evidence is reliable.

New fascinating area: **unsupervised learning of linguistic information**.

Survey paper: Harald Hammarström, Lars Borin, *Unsupervised Learning of Morphology*, CL 2011:

- **input**: raw (unannotated) natural language text data,
- **output**: a description of the morphological structure of the language of the input text,
- **with**: as little supervision (parameters, thresholds, model selection during development, etc.), as possible.

Results so far: not practically usable or theoretically exciting, but promising.

There is also ongoing research on unsupervised learning of higher linguistic levels, e.g.:

Chris Biemann, 2011, *Structure Discovery in Natural Language*.

New fascinating area: **unsupervised learning of linguistic information**.

Survey paper: Harald Hammarström, Lars Borin, *Unsupervised Learning of Morphology*, CL 2011:

- **input**: raw (unannotated) natural language text data,
- **output**: a description of the morphological structure of the language of the input text,
- **with**: as little supervision (parameters, thresholds, model selection during development, etc.), as possible.

Results so far: not practically usable or theoretically exciting, but promising.

There is also ongoing research on unsupervised learning of higher linguistic levels, e.g.:

Chris Biemann, 2011, *Structure Discovery in Natural Language*.

New fascinating area: **unsupervised learning of linguistic information**.

Survey paper: Harald Hammarström, Lars Borin, *Unsupervised Learning of Morphology*, CL 2011:

- **input**: raw (unannotated) natural language text data,
- **output**: a description of the morphological structure of the language of the input text,
- **with**: as little supervision (parameters, thresholds, model selection during development, etc.), as possible.

Results so far: not practically usable or theoretically exciting, but promising.

There is also ongoing research on unsupervised learning of higher linguistic levels, e.g.:

Chris Biemann, 2011, *Structure Discovery in Natural Language*.

New fascinating area: **unsupervised learning of linguistic information**.

Survey paper: Harald Hammarström, Lars Borin, *Unsupervised Learning of Morphology*, CL 2011:

- **input**: raw (unannotated) natural language text data,
- **output**: a description of the morphological structure of the language of the input text,
- **with**: as little supervision (parameters, thresholds, model selection during development, etc.), as possible.

Results so far: not practically usable or theoretically exciting, but promising.

There is also ongoing research on unsupervised learning of higher linguistic levels, e.g.:

Chris Biemann, 2011, *Structure Discovery in Natural Language*.

Outline

1 Lx4LT

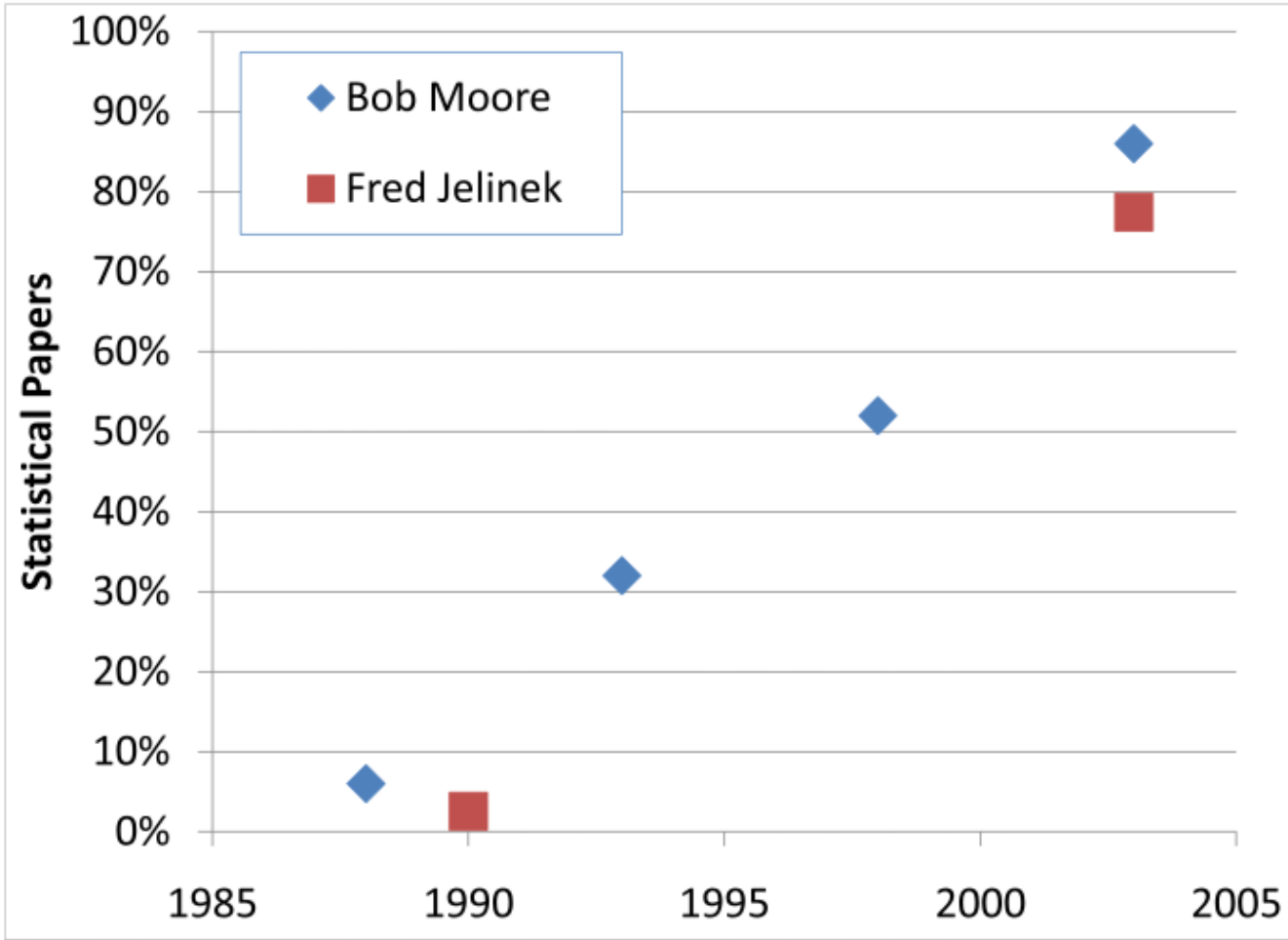
2 LT4Lx

3 Lx4LT again?

4 Conclusion

Lx4LT again? 1

Kenneth Church, *A Pendulum Swung Too Far*, LiLT 2011:



Statistical papers at ACL meetings.

Lx4LT again? 2

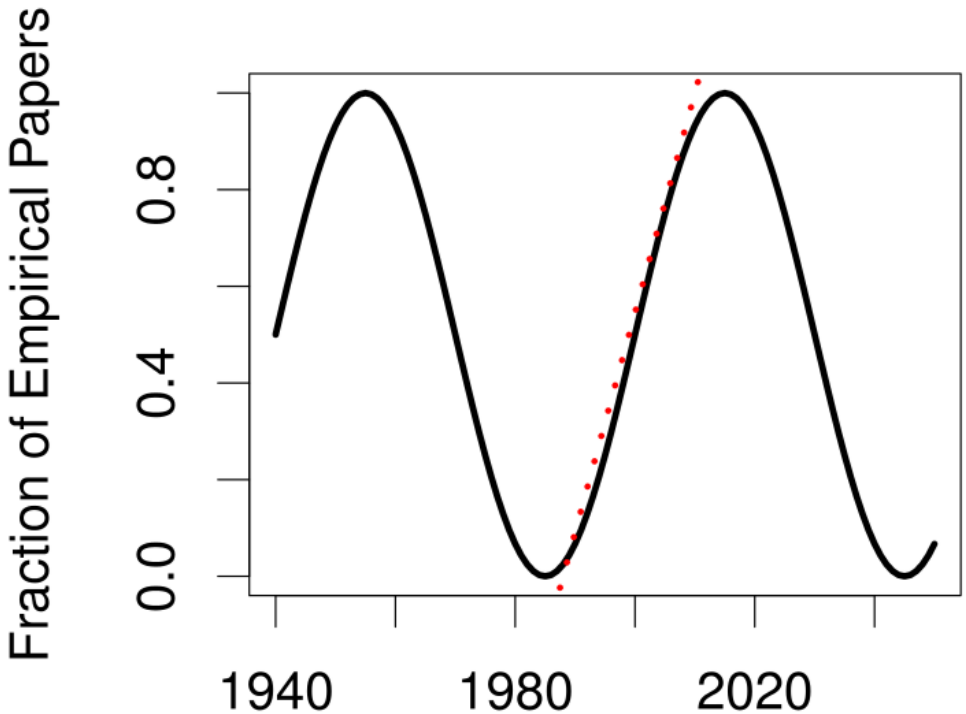
Kenneth Church about late 1980s and now:

> *In addition to everything else that was going on at the time, we wanted to make room for a little work of a different kind... It is nice to see the field come together as it has, but we may have been too successful. Not only have we succeeded in making room for what we were interested in, but now there is no longer much room for anything else.*

Lx4LT again? 3

Kenneth Church's view on the history of NLP:

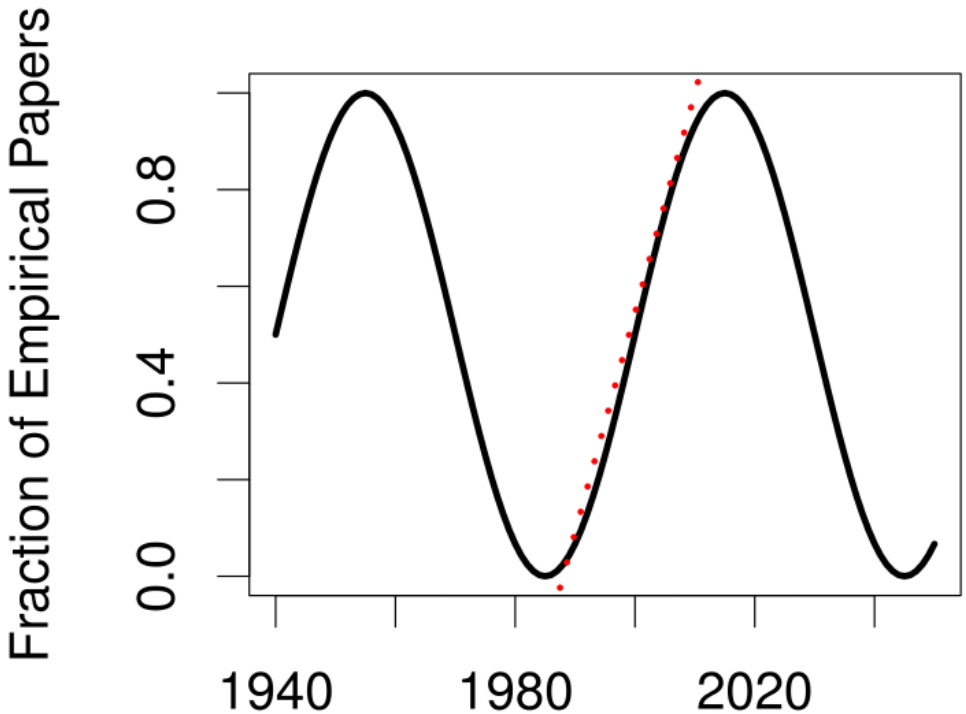
- 1950s: Empiricism (Shannon, Skinner, Firth, Harris),
- 1970s: Rationalism (Chomsky, Minsky),
- 1990s: Empiricism (IBM Speech Group, AT&T Bell Labs),
- 2010s: A Return to Rationalism?



Lx4LT again? 3

Kenneth Church's view on the history of NLP:

- 1950s: Empiricism (Shannon, Skinner, Firth, Harris),
- 1970s: Rationalism (Chomsky, Minsky),
- 1990s: Empiricism (IBM Speech Group, AT&T Bell Labs),
- 2010s: A Return to Rationalism?



Lx4LT again? 4

Some indicators of this trend:

- increasing linguistic knowledge injected into the best Machine Translation systems,
- recognised need for more prosody in speech systems – but they require the understing of the information structure,
- Watson uses a large coverage grammar of English (and still *Requires Deeper NLP*) – Włodek Zadrożny today,
- also emphasis of COLING 2012 on linguistic content.

Lx4LT again? 4

Some indicators of this trend:

- increasing linguistic knowledge injected into the best Machine Translation systems,
- recognised need for more prosody in speech systems – but they require the understing of the information structure,
- Watson uses a large coverage grammar of English (and still *Requires Deeper NLP*) – Włodek Zadrożny today,
- also emphasis of COLING 2012 on linguistic content.

Lx4LT again? 4

Some indicators of this trend:

- increasing linguistic knowledge injected into the best Machine Translation systems,
- recognised need for more prosody in speech systems – but they require the understing of the information structure,
- Watson uses a large coverage grammar of English (and still *Requires Deeper NLP*) – Włodek Zadrożny today,
- also emphasis of COLING 2012 on linguistic content.

Lx4LT again? 4

Some indicators of this trend:

- increasing linguistic knowledge injected into the best Machine Translation systems,
- recognised need for more prosody in speech systems – but they require the understing of the information structure,
- Watson uses a large coverage grammar of English (and still *Requires Deeper NLP*) – Włodek Zadrożny today,
- also emphasis of COLING 2012 on linguistic content.

JLM

(With apologies for this shameless plug.)

Outline

1 Lx4LT

2 LT4Lx

3 Lx4LT again?

4 Conclusion

Conclusions

Main points:

- 1980s: very close cooperation between Lx and LT,
- mainly in the area of parsing,
- later: perceived abandonment of Lx by LT,
- but actually Lx continues to influence LT,
- currently: more and more usefulness of LT for Lx,
- beginning of another rationalist turn.

Exciting times ahead for more fruitful and bi-directional cooperation between Language Technologies and Linguistics in sight!

Conclusions

Main points:

- 1980s: very close cooperation between Lx and LT,
- mainly in the area of parsing,
- later: perceived abandonment of Lx by LT,
- but actually Lx continues to influence LT,
- currently: more and more usefulness of LT for Lx,
- beginning of another rationalist turn.

Exciting times ahead for more fruitful and bi-directional cooperation between Language Technologies and Linguistics in sight!

Conclusions

Main points:

- 1980s: very close cooperation between Lx and LT,
- mainly in the area of parsing,
- later: perceived abandonment of Lx by LT,
- but actually Lx continues to influence LT,
- currently: more and more usefulness of LT for Lx,
- beginning of another rationalist turn.

Exciting times ahead for more fruitful and bi-directional cooperation between Language Technologies and Linguistics in sight!

Conclusions

Main points:

- 1980s: very close cooperation between Lx and LT,
- mainly in the area of parsing,
- later: perceived abandonment of Lx by LT,
- but actually Lx continues to influence LT,
- currently: more and more usefulness of LT for Lx,
- beginning of another rationalist turn.

Exciting times ahead for more fruitful and bi-directional cooperation between Language Technologies and Linguistics in sight!

Conclusions

Main points:

- 1980s: very close cooperation between Lx and LT,
- mainly in the area of parsing,
- later: perceived abandonment of Lx by LT,
- but actually Lx continues to influence LT,
- currently: more and more usefulness of LT for Lx,
- beginning of another rationalist turn.

Exciting times ahead for more fruitful and bi-directional cooperation between Language Technologies and Linguistics in sight!

Thank you for your attention!