


Historical Models and Serial Sources

Prof. Dr.-Ing. Michael Piotrowski

<michael.piotrowski@unil.ch>

 @true_mxp

Digital Approaches Towards Serial Publications · Brussels, September 11, 2017

Digital Humanities

Let's first look at computational linguistics

Linguistics has a “vantage point” for observing the digital humanities, because it has essentially completed the transformation from “armchair linguistics” to an empirical science using formal models.

Theoretical CL

[“mathematical linguistics”]

Formal language theory

Grammar theory

Computational learning theory

⋮

Applied CL

[“corpus linguistics”, NLP]

Application of theories and methods
to *concrete* languages

Let's first look at computational linguistics

Linguistics has a “vantage point” for observing the digital humanities, because it has essentially completed the transformation from “armchair linguistics” to an empirical science using formal models.

Theoretical CL

[“mathematical linguistics”]

Formal language theory

Grammar theory

Computational learning theory

⋮

Applied CL

[“corpus linguistics”, NLP]

Application of theories and methods
to *concrete* languages

Let's first look at computational linguistics

Linguistics has a “vantage point” for observing the digital humanities, because it has essentially completed the transformation from “armchair linguistics” to an empirical science using formal models.

Theoretical CL

[“mathematical linguistics”]

Formal language theory

Grammar theory

Computational learning theory

⋮

Theory of formal modeling of NL

Applied CL

[“corpus linguistics”, NLP]

Application of theories and methods
to *concrete* languages

Formal modeling of NL

Let's first look at computational linguistics

Linguistics has a “vantage point” for observing the digital humanities, because it has essentially completed the transformation from “armchair linguistics” to an empirical science using formal models.

Theoretical CL

[“mathematical linguistics”]

Formal language theory

Grammar theory

Computational learning theory

⋮

Theory of formal modeling of NL

→ “Actually” math or CS

Applied CL

[“corpus linguistics”, NLP]

Application of theories and methods
to *concrete* languages

Formal modeling of NL

→ “Actually” linguistics

Let's look at computational linguistics

- ▶ Theoretical CL studies the means and methods of constructing formal models in linguistics.
- ▶ Applied CL is concerned with the construction of formal models of natural languages and with the methodology of constructing such models.
- ▶ Can we apply the same reasoning to digital humanities?

Let's look at computational linguistics

- ▶ Theoretical CL studies the means and methods of constructing formal models in linguistics.
→ **Metascience**
- ▶ Applied CL is concerned with the construction of formal models of natural languages and with the methodology of constructing such models.
- ▶ Can we apply the same reasoning to digital humanities?

Let's look at computational linguistics

- ▶ Theoretical CL studies the means and methods of constructing formal models in linguistics.
→ **Metascience**
- ▶ Applied CL is concerned with the construction of formal models of natural languages and with the methodology of constructing such models.
→ **Science**
- ▶ Can we apply the same reasoning to digital humanities?

Let's look at computational linguistics

- ▶ Theoretical CL studies the means and methods of constructing formal models in linguistics.
→ **Metascience**
- ▶ Applied CL is concerned with the construction of formal models of natural languages and with the methodology of constructing such models.
→ **Science**
- ▶ Can we apply the same reasoning to digital humanities?

Definition (Digital humanities)

The digital humanities study the means and methods of constructing formal models in the humanities.

→ concerned with “construction materials” for formal models: **metascience**.

Definition (Digital history, DH applied to history)

Digital history is concerned with the construction of formal models of historical circumstances and with the methodology of constructing such models.

→ digital history creates *concrete* formal models of its research objects. Correspondingly: Digital philology, digital musicology, etc.—“applied DH”, *subfields of their disciplines*.

Definition (Digital humanities)

The digital humanities study the means and methods of constructing formal models in the humanities.

→ concerned with “construction materials” for formal models: **metascience**.

Definition (Digital history, DH applied to history)

Digital history is concerned with the construction of formal models of historical circumstances and with the methodology of constructing such models.

→ digital history creates *concrete* formal models of its research objects. Correspondingly: Digital philology, digital musicology, etc.—“applied DH”, *subfields of their disciplines*.

Definition (Digital humanities)

The digital humanities study the means and methods of constructing formal models in the humanities.

→ concerned with “construction materials” for formal models: **metascience**.

Definition (Digital history, DH applied to history)

Digital history is concerned with the construction of formal models of historical circumstances and with the methodology of constructing such models.

→ digital history creates *concrete* formal models of its research objects. Correspondingly: Digital philology, digital musicology, etc.—“applied DH”, *subfields of their disciplines*.

Formal Models

What do we mean by “model”?

We use the term in the sense of Stachowiak’s *Allgemeiner Modelltheorie* ‘General Model Theory’ [Sta73].

Three fundamental properties of models:

Mapping Models are always models *of something*, viz., mappings from, representations of natural or artificial originals, which can themselves be models.

Reduction Models generally *do not capture all attributes* of the original they represent, but only those that seem relevant to the model creators and/or model users.

Pragmatism Models are *not per se uniquely assigned* to their originals. They fulfill their replacement function

- a for particular–cognitive and/or acting–subjects that use the model,
- b within particular time intervals, and
- c restricted to particular mental or actual operations.

What do we mean by “model”?

We use the term in the sense of Stachowiak’s *Allgemeiner Modelltheorie* ‘General Model Theory’ [Sta73].

Three fundamental properties of models:

Mapping Models are always models *of something*, viz., mappings from, representations of natural or artificial originals, which can themselves be models.

Reduction Models generally *do not capture all attributes* of the original they represent, but only those that seem relevant to the model creators and/or model users.

Pragmatism Models are *not per se uniquely assigned* to their originals. They fulfill their replacement function

- a) for particular–cognitive and/or acting–subjects that use the model,
- b) within particular time intervals, and
- c) restricted to particular mental or actual operations.

What do we mean by “model”?

We use the term in the sense of Stachowiak’s *Allgemeiner Modelltheorie* ‘General Model Theory’ [Sta73].

Three fundamental properties of models:

Mapping Models are always models *of something*, viz., mappings from, representations of natural or artificial originals, which can themselves be models.

Reduction Models generally *do not capture all attributes* of the original they represent, but only those that seem relevant to the model creators and/or model users.

Pragmatism Models are *not per se uniquely assigned* to their originals. They fulfill their replacement function

- a for particular–cognitive and/or acting–subjects that use the model,
- b within particular time intervals, and
- c restricted to particular mental or actual operations.

What do we mean by “model”?

We use the term in the sense of Stachowiak’s *Allgemeiner Modelltheorie* ‘General Model Theory’ [Sta73].

Three fundamental properties of models:

Mapping Models are always models *of something*, viz., mappings from, representations of natural or artificial originals, which can themselves be models.

Reduction Models generally *do not capture all attributes* of the original they represent, but only those that seem relevant to the model creators and/or model users.

Pragmatism Models are *not per se uniquely assigned* to their originals. They fulfill their replacement function

- a for particular–cognitive and/or acting–subjects that use the model,
- b within particular time intervals, and
- c restricted to particular mental or actual operations.

What do we mean by “model”?

We use the term in the sense of Stachowiak’s *Allgemeiner Modelltheorie* ‘General Model Theory’ [Sta73].

Three fundamental properties of models:

Mapping Models are always models *of something*, viz., mappings from, representations of natural or artificial originals, which can themselves be models.

Reduction Models generally *do not capture all attributes* of the original they represent, but only those that seem relevant to the model creators and/or model users.

Pragmatism Models are *not per se uniquely assigned* to their originals. They fulfill their replacement function

- a for particular–cognitive and/or acting–subjects that use the model,
- b within particular time intervals, and
- c restricted to particular mental or actual operations.

What do we mean by “model”?

We use the term in the sense of Stachowiak’s *Allgemeiner Modelltheorie* ‘General Model Theory’ [Sta73].

Three fundamental properties of models:

Mapping Models are always models *of something*, viz., mappings from, representations of natural or artificial originals, which can themselves be models.

Reduction Models generally *do not capture all attributes* of the original they represent, but only those that seem relevant to the model creators and/or model users.

Pragmatism Models are *not per se uniquely assigned* to their originals. They fulfill their replacement function

- a for particular–cognitive and/or acting–subjects that use the model,
- b within particular time intervals, and
- c restricted to particular mental or actual operations.

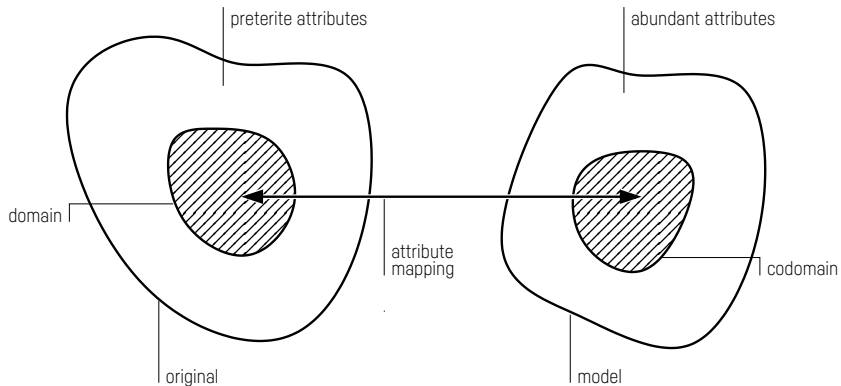
Models

- ▶ Arbitrary objects can be described as individuals characterized by a finite number of attributes
- ▶ Attributes can be characteristics and properties of individuals, relations between individuals, properties of properties, properties of relations, etc.

Models

- ▶ Arbitrary objects can be described as individuals characterized by a finite number of attributes
- ▶ Attributes can be characteristics and properties of individuals, relations between individuals, properties of properties, properties of relations, etc.

Original-model-mapping



Modeling as an everyday task

- ▶ The construction of models is an everyday task; we construct models all the time.
- ▶ In his influential 1971 paper “Counterintuitive behavior of social systems,” Forrester points out:

Each of us uses models constantly. Every person in private life and in business instinctively uses models for decision making. The mental images in one's head about one's surroundings are models. One's head does not contain real families, businesses, cities, governments, or countries. One uses selected concepts and relationships to represent real systems. A mental image is a model. All decisions are taken on the basis of models. All laws are passed on the basis of models. All executive actions are taken on the basis of models. [For95, p. 4]

→ « Nous ne raisonnons que sur des modèles. » [Paul Valéry]

Modeling as an everyday task

- ▶ The construction of models is an everyday task; we construct models all the time.
- ▶ In his influential 1971 paper “Counterintuitive behavior of social systems,” Forrester points out:

Each of us uses models constantly. Every person in private life and in business instinctively uses models for decision making. The mental images in one's head about one's surroundings are models. One's head does not contain real families, businesses, cities, governments, or countries. One uses selected concepts and relationships to represent real systems. A mental image is a model. All decisions are taken on the basis of models. All laws are passed on the basis of models. All executive actions are taken on the basis of models. [For95, p. 4]

→ « Nous ne raisonnons que sur des modèles. » [Paul Valéry]

Modeling as an everyday task

- ▶ The construction of models is an everyday task; we construct models all the time.
- ▶ In his influential 1971 paper “Counterintuitive behavior of social systems,” Forrester points out:

Each of us uses models constantly. Every person in private life and in business instinctively uses models for decision making. The mental images in one's head about one's surroundings are models. One's head does not contain real families, businesses, cities, governments, or countries. One uses selected concepts and relationships to represent real systems. A mental image is a model. All decisions are taken on the basis of models. All laws are passed on the basis of models. All executive actions are taken on the basis of models. [For95, p. 4]

→ « Nous ne raisonnons que sur des modèles. » [Paul Valéry]

Modeling as an everyday task

- ▶ The construction of models is an everyday task; we construct models all the time.
- ▶ In his influential 1971 paper “Counterintuitive behavior of social systems,” Forrester points out:

Each of us uses models constantly. Every person in private life and in business instinctively uses models for decision making. The mental images in one's head about one's surroundings are models. One's head does not contain real families, businesses, cities, governments, or countries. One uses selected concepts and relationships to represent real systems. A mental image is a model. All decisions are taken on the basis of models. All laws are passed on the basis of models. All executive actions are taken on the basis of models. [For95, p. 4]

→ « Nous ne raisonnons que sur des modèles. » [Paul Valéry]

Models in science

When it comes to science, Epstein thus notes:

The choice, then, is not whether to build models; it's whether to build explicit ones. In explicit models, assumptions are laid out in detail, so we can study exactly what they entail. On these assumptions, this sort of thing happens. When you alter the assumptions that is what happens. By writing explicit models, you let others replicate your results. [Eps08]

Models in the humanities?

The most common misunderstanding about science is that scientists seek and find truth. They don't—they make and test models. [...] Making sense of anything means making models that can predict outcomes and accommodate observations. Truth is a model. [Neil Gershenfeld]

- ▶ All scientific and scholarly research constructs models of their objects of research.
- ▶ In order to understand a complex object (phenomenon, situation, ...), you need to understand its parts and how they interrelate with each other.
- This is exactly what a model describes.
- ⦿ In contrast to the natural sciences, models in the humanities are traditionally *not formal* and only partially explicit;
- ▶ narratives are *informal* models [of scholars' mental models].

Models in the humanities?

The most common misunderstanding about science is that scientists seek and find truth. They don't—they make and test models. [...] Making sense of anything means making models that can predict outcomes and accommodate observations. Truth is a model. (Neil Gershenfeld)

- ▶ **All** scientific and scholarly research constructs models of their objects of research.
- ▶ In order to understand a complex object (phenomenon, situation, ...), you need to understand its parts and how they interrelate with each other.
- This is exactly what a model describes.
- ◎ In contrast to the natural sciences, models in the humanities are traditionally *not formal* and only partially explicit;
- ▶ narratives are *informal* models (of scholars' mental models).

Models in the humanities?

The most common misunderstanding about science is that scientists seek and find truth. They don't—they make and test models. [...] Making sense of anything means making models that can predict outcomes and accommodate observations. Truth is a model. [Neil Gershenfeld]

- ▶ **All** scientific and scholarly research constructs models of their objects of research.
- ▶ In order to understand a complex object (phenomenon, situation, ...), you need to understand its parts and how they interrelate with each other.
- This is exactly what a model describes.
- ⦿ In contrast to the natural sciences, models in the humanities are traditionally *not formal* and only partially explicit;
 - ▶ narratives are *informal* models [of scholars' mental models].

Models in the humanities?

The most common misunderstanding about science is that scientists seek and find truth. They don't—they make and test models. [...] Making sense of anything means making models that can predict outcomes and accommodate observations. Truth is a model. [Neil Gershenfeld]

- ▶ **All** scientific and scholarly research constructs models of their objects of research.
- ▶ In order to understand a complex object (phenomenon, situation, ...), you need to understand its parts and how they interrelate with each other.
- This is exactly what a model describes.
- ⦿ In contrast to the natural sciences, models in the humanities are traditionally *not formal* and only partially explicit;
 - ▶ narratives are *informal* models [of scholars' mental models].

Models in the humanities?

The most common misunderstanding about science is that scientists seek and find truth. They don't—they make and test models. [...] Making sense of anything means making models that can predict outcomes and accommodate observations. Truth is a model. [Neil Gershenfeld]

- ▶ **All** scientific and scholarly research constructs models of their objects of research.
- ▶ In order to understand a complex object (phenomenon, situation, ...), you need to understand its parts and how they interrelate with each other.
- This is exactly what a model describes.
- ◎ In contrast to the natural sciences, models in the humanities are traditionally *not formal* and only partially explicit;
 - ▶ narratives are *informal* models [of scholars' mental models].

Models in the humanities?

The most common misunderstanding about science is that scientists seek and find truth. They don't—they make and test models. [...] Making sense of anything means making models that can predict outcomes and accommodate observations. Truth is a model. [Neil Gershenfeld]

- ▶ **All** scientific and scholarly research constructs models of their objects of research.
- ▶ In order to understand a complex object (phenomenon, situation, ...), you need to understand its parts and how they interrelate with each other.
- This is exactly what a model describes.
- ⦿ In contrast to the natural sciences, models in the humanities are traditionally *not formal* and only partially explicit;
- ▶ narratives are *informal* models [of scholars' mental models].

What do we mean by “formal”?

Слово «формальный» не означает ничего, кроме как «логически последовательный + однозначный + абсолютно явный». [GM69, p. 9]

The word “formal” means nothing more than “logically coherent + unambiguous + explicit.”

- ▶ A formal model is thus more than explicit.
- ▶ There are different degrees of formalization.
- ▶ We are primarily interested in a degree of formalization that allows models to be **processed and manipulated by computers**.

What do we mean by “formal”?

Слово «формальный» не означает ничего, кроме как «логически последовательный + однозначный + абсолютно явный». [GM69, p. 9]

The word “formal” means nothing more than “logically coherent + unambiguous + explicit.”

- ▶ A formal model is thus more than explicit.
- ▶ There are different degrees of formalization.
- ▶ We are primarily interested in a degree of formalization that allows models to be processed and manipulated by computers.

What do we mean by “formal”?

Слово «формальный» не означает ничего, кроме как «логически последовательный + однозначный + абсолютно явный». [GM69, p. 9]

The word “formal” means nothing more than “logically coherent + unambiguous + explicit.”

- ▶ A formal model is thus more than explicit.
- ▶ There are different degrees of formalization.
- ▶ We are primarily interested in a degree of formalization that allows models to be **processed and manipulated by computers**.

What do we mean by “formal”?

Слово «формальный» не означает ничего, кроме как «логически последовательный + однозначный + абсолютно явный». [GM69, p. 9]

The word “formal” means nothing more than “logically coherent + unambiguous + explicit.”

- ▶ A formal model is thus more than explicit.
- ▶ There are different degrees of formalization.
- ▶ We are primarily interested in a degree of formalization that allows models to be **processed and manipulated by computers**.

What do we mean by “formal”?

Слово «формальный» не означает ничего, кроме как «логически последовательный + однозначный + абсолютно явный». [GM69, p. 9]

The word “formal” means nothing more than “logically coherent + unambiguous + explicit.”

- ▶ A formal model is thus more than explicit.
- ▶ There are different degrees of formalization.
- ▶ We are primarily interested in a degree of formalization that allows models to be **processed and manipulated by computers**.

Corpora as Models

Corpora *are* models

- ▶ Corpora are clearly models in the sense of Stachowiak:
- ▶ in linguistics, they are models of language constructed by sampling utterances in order to study linguistic phenomena.
- ▶ Since *all* of language is inaccessible, one of the main goals in corpus design [= modeling] is to ensure *representativeness* in order to minimize distortion in the mapping from the original [the language] to the model [the corpus].
- ▶ This includes balancing the corpus for features such as genre, topic, authors, diachronic and diatopic distribution, etc. Subcorpora, e.g., different genres or eras, are designed to be comparable by using the same number and size of samples.

Corpora *are* models

- ▶ Corpora are clearly models in the sense of Stachowiak:
- ▶ in linguistics, they are models of language constructed by sampling utterances in order to study linguistic phenomena.
- ▶ Since *all* of language is inaccessible, one of the main goals in corpus design [= modeling] is to ensure *representativeness* in order to minimize distortion in the mapping from the original [the language] to the model [the corpus].
- ▶ This includes balancing the corpus for features such as genre, topic, authors, diachronic and diatopic distribution, etc. Subcorpora, e.g., different genres or eras, are designed to be comparable by using the same number and size of samples.

Corpora *are* models

- ▶ Corpora are clearly models in the sense of Stachowiak:
- ▶ in linguistics, they are models of language constructed by sampling utterances in order to study linguistic phenomena.
- ▶ Since *all* of language is inaccessible, one of the main goals in corpus design (= modeling) is to ensure *representativeness* in order to minimize distortion in the mapping from the original (the language) to the model (the corpus).
- ▶ This includes balancing the corpus for features such as genre, topic, authors, diachronic and diatopic distribution, etc. Subcorpora, e.g., different genres or eras, are designed to be comparable by using the same number and size of samples.

Corpora *are* models

- ▶ Corpora are clearly models in the sense of Stachowiak:
- ▶ in linguistics, they are models of language constructed by sampling utterances in order to study linguistic phenomena.
- ▶ Since *all* of language is inaccessible, one of the main goals in corpus design (= modeling) is to ensure *representativeness* in order to minimize distortion in the mapping from the original (the language) to the model (the corpus).
- ▶ This includes balancing the corpus for features such as genre, topic, authors, diachronic and diatopic distribution, etc. Subcorpora, e.g., different genres or eras, are designed to be comparable by using the same number and size of samples.

Corpus design

- ▶ Whatever is missing from the corpus will also be missing from all subsequent analyses; overrepresentation and underrepresentation are hard to adjust for afterward.
- Corpus design is thus of utmost importance for the validity of any results derived from the corpus.
- ▶ In corpus linguistics, Douglas Biber [Bib90, Bib93] has laid the foundation for theoretical reflection on corpus design:

The use of computer-based corpora provides a solid empirical foundation for general purpose language tools and descriptions, and enables analyses of a scope not otherwise possible. However, a corpus must be “representative” in order to be appropriately used as the basis for generalizations concerning a language as a whole [...] [Bib93, p. 243]

Corpora in linguistics

- ▶ The next step is to model the language phenomenon under research.
- ▶ This involves the construction of formal models, e.g., in the form of *queries* to the corpus.
- ▶ It is sometimes possible to formulate exact queries, when the phenomenon is observable on the *surface level*, e.g., when one is interested in a specific word form.
- ▶ In other cases, additional *annotation layers* are needed, such as part-of-speech tags, lemmas, or morphological analyses.
- ▶ When dealing with historical texts (whether in a diachronic or synchronic fashion), it may be necessary to consider older spellings or syntactic constructions.

Corpora in linguistics

- ▶ The next step is to model the language phenomenon under research.
- ▶ This involves the construction of formal models, e.g., in the form of *queries* to the corpus.
- ▶ It is sometimes possible to formulate exact queries, when the phenomenon is observable on the *surface level*, e.g., when one is interested in a specific word form.
- ▶ In other cases, additional *annotation layers* are needed, such as part-of-speech tags, lemmas, or morphological analyses.
- ▶ When dealing with historical texts (whether in a diachronic or synchronic fashion), it may be necessary to consider older spellings or syntactic constructions.

Corpora in linguistics

- ▶ The next step is to model the language phenomenon under research.
- ▶ This involves the construction of formal models, e.g., in the form of *queries* to the corpus.
- ▶ It is sometimes possible to formulate exact queries, when the phenomenon is observable on the *surface level*, e.g., when one is interested in a specific word form.
- ▶ In other cases, additional *annotation layers* are needed, such as part-of-speech tags, lemmas, or morphological analyses.
- ▶ When dealing with historical texts (whether in a diachronic or synchronic fashion), it may be necessary to consider older spellings or syntactic constructions.

Corpora in linguistics

- ▶ The next step is to model the language phenomenon under research.
- ▶ This involves the construction of formal models, e.g., in the form of *queries* to the corpus.
- ▶ It is sometimes possible to formulate exact queries, when the phenomenon is observable on the *surface level*, e.g., when one is interested in a specific word form.
- ▶ In other cases, additional *annotation layers* are needed, such as part-of-speech tags, lemmas, or morphological analyses.
- ▶ When dealing with historical texts (whether in a diachronic or synchronic fashion), it may be necessary to consider older spellings or syntactic constructions.

Corpora in linguistics

- ▶ The next step is to model the language phenomenon under research.
- ▶ This involves the construction of formal models, e.g., in the form of *queries* to the corpus.
- ▶ It is sometimes possible to formulate exact queries, when the phenomenon is observable on the *surface level*, e.g., when one is interested in a specific word form.
- ▶ In other cases, additional *annotation layers* are needed, such as part-of-speech tags, lemmas, or morphological analyses.
- ▶ When dealing with historical texts (whether in a diachronic or synchronic fashion), it may be necessary to consider older spellings or syntactic constructions.

Sampled vs. complete corpora

- ▶ Most corpora in linguistics aim to model relatively large subsets of language (↗ representativeness by sampling).
- ▶ Linguistics is, for the most part, not concerned with the meaning of texts.
- One newspaper article is as good as another.
- ▶ More narrowly focused corpora may in fact be *complete*, e.g., a corpus of all diaries of Thomas Mann is complete in the sense that it contains *all* of the diaries.
- ▶ In such cases, no *linguistic* modeling takes place; also usually are not intended for linguistic research, but, e.g., for literary studies.

Sampled vs. complete corpora

- ▶ Most corpora in linguistics aim to model relatively large subsets of language (↗ representativeness by sampling).
- ▶ Linguistics is, for the most part, not concerned with the meaning of texts.
- One newspaper article is as good as another.
- ▶ More narrowly focused corpora may in fact be *complete*, e.g., a corpus of all diaries of Thomas Mann is complete in the sense that it contains *all* of the diaries.
- ▶ In such cases, no *linguistic* modeling takes place; also usually are not intended for linguistic research, but, e.g., for literary studies.

Sampled vs. complete corpora

- ▶ Most corpora in linguistics aim to model relatively large subsets of language (↗ representativeness by sampling).
- ▶ Linguistics is, for the most part, not concerned with the meaning of texts.
- ➔ One newspaper article is as good as another.
- ▶ More narrowly focused corpora may in fact be *complete*, e.g., a corpus of all diaries of Thomas Mann is complete in the sense that it contains *all* of the diaries.
- ▶ In such cases, no *linguistic* modeling takes place; also usually are not intended for linguistic research, but, e.g., for literary studies.

Sampled vs. complete corpora

- ▶ Most corpora in linguistics aim to model relatively large subsets of language (↗ representativeness by sampling).
- ▶ Linguistics is, for the most part, not concerned with the meaning of texts.
- One newspaper article is as good as another.
- ▶ More narrowly focused corpora may in fact be *complete*, e.g., a corpus of all diaries of Thomas Mann is complete in the sense that it contains *all* of the diaries.
- ▶ In such cases, no *linguistic* modeling takes place; also usually are not intended for linguistic research, but, e.g., for literary studies.

Sampled vs. complete corpora

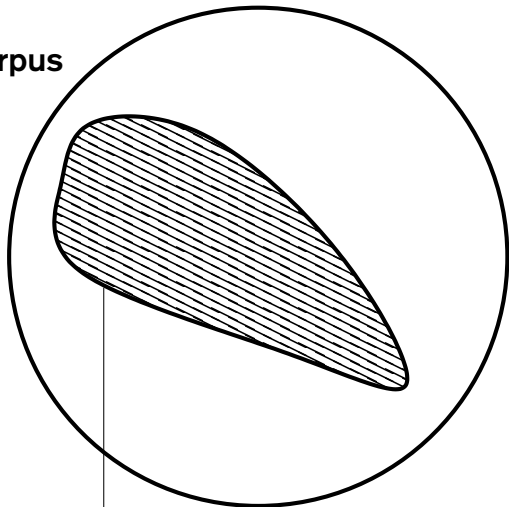
- ▶ Most corpora in linguistics aim to model relatively large subsets of language (↗ representativeness by sampling).
- ▶ Linguistics is, for the most part, not concerned with the meaning of texts.
- One newspaper article is as good as another.
- ▶ More narrowly focused corpora may in fact be *complete*, e.g., a corpus of all diaries of Thomas Mann is complete in the sense that it contains *all* of the diaries.
- ▶ In such cases, no *linguistic* modeling takes place; also usually are not intended for linguistic research, but, e.g., for literary studies.

Sampled vs. complete corpora

- ▶ In such cases, *most* of the research is concerned with the content of the corpus, but phenomena outside the corpus may also be of concern, such as authors (who have a biography and an *œuvre* not modeled by the corpus) or external events.
- ▶ A question like “how did Thomas Mann *describe* the outbreak of World War II?” is concerned with the *textual description inside* the corpus, but in relation to an event *outside* the corpus;
- ▶ the question “how did Thomas Mann *react* to the outbreak of World War II?” on the basis of his diaries moves further parts of the research outside the corpus; it becomes a model of Mann’s historical emotional state.
- ➔ A linguistic corpus is incomplete, but the phenomenon under study is completely contained within.
- ➔ A literary corpus may be complete, but the phenomenon under study is not fully contained within it.

Linguistics

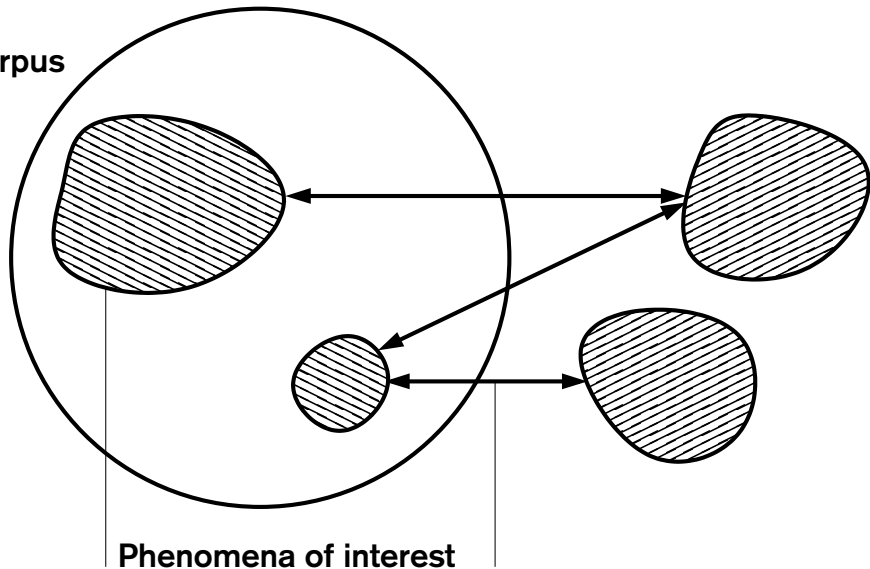
Corpus



Phenomenon of interest

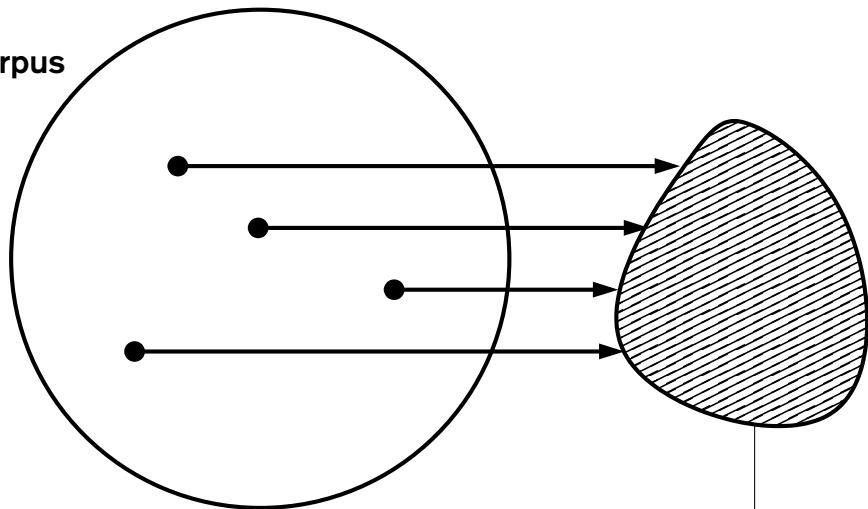
Literary studies

Corpus



History

Corpus



Phenomenon of interest

Conclusion

Conclusion

- ▶ Digital humanities is all about formal modeling.
- ▶ A corpus is a model in the sense of Stachowiak: it is a reduced mapping of some original for a particular purpose.
- ▶ Corpus-based research means constructing a formal model to test against the corpus
- ▶ There is much experience in corpus linguistics, but the research object of history is very different from that of linguistics.
- ▶ Both models (corpus and query) **must be documented**, otherwise the advantages of formal models are lost.
- ▶ The main problem may be that many scholars are not aware of their modeling.
- Many black boxes and much hand-waving.

Conclusion

- ▶ Digital humanities is all about formal modeling.
- ▶ A corpus is a model in the sense of Stachowiak: it is a reduced mapping of some original for a particular purpose.
- ▶ Corpus-based research means constructing a formal model to test against the corpus
- ▶ There is much experience in corpus linguistics, but the research object of history is very different from that of linguistics.
- ▶ Both models [corpus and query] **must be documented**, otherwise the advantages of formal models are lost.
- ▶ The main problem may be that many scholars are not aware of their modeling.
- Many black boxes and much hand-waving.

Conclusion

- ▶ Digital humanities is all about formal modeling.
- ▶ A corpus is a model in the sense of Stachowiak: it is a reduced mapping of some original for a particular purpose.
- ▶ Corpus-based research means constructing a formal model to test against the corpus
- ▶ There is much experience in corpus linguistics, but the research object of history is very different from that of linguistics.
- ▶ Both models [corpus and query] **must be documented**, otherwise the advantages of formal models are lost.
- ▶ The main problem may be that many scholars are not aware of their modeling.
- Many black boxes and much hand-waving.

Conclusion

- ▶ Digital humanities is all about formal modeling.
- ▶ A corpus is a model in the sense of Stachowiak: it is a reduced mapping of some original for a particular purpose.
- ▶ Corpus-based research means constructing a formal model to test against the corpus
 - ↔ Querying a corpus means constructing a formal model of the phenomenon under research.
- ▶ There is much experience in corpus linguistics, but the research object of history is very different from that of linguistics.
- ▶ Both models (corpus and query) **must be documented**, otherwise the advantages of formal models are lost.
- ▶ The main problem may be that many scholars are not aware of their modeling.
- Many black boxes and much hand-waving.

Conclusion

- ▶ Digital humanities is all about formal modeling.
- ▶ A corpus is a model in the sense of Stachowiak: it is a reduced mapping of some original for a particular purpose.
- ▶ Corpus-based research means constructing a formal model to test against the corpus
 - ↔ Querying a corpus means constructing a formal model of the phenomenon under research.
- ▶ There is much experience in corpus linguistics, but the research object of history is very different from that of linguistics.
 - ▶ Both models (corpus and query) **must be documented**, otherwise the advantages of formal models are lost.
 - ▶ The main problem may be that many scholars are not aware of their modeling.
- Many black boxes and much hand-waving.

Conclusion

- ▶ Digital humanities is all about formal modeling.
- ▶ A corpus is a model in the sense of Stachowiak: it is a reduced mapping of some original for a particular purpose.
- ▶ Corpus-based research means constructing a formal model to test against the corpus
 - ↔ Querying a corpus means constructing a formal model of the phenomenon under research.
- ▶ There is much experience in corpus linguistics, but the research object of history is very different from that of linguistics.
- ▶ Both models (corpus and query) **must be documented**, otherwise the advantages of formal models are lost.
- ▶ The main problem may be that many scholars are not aware of their modeling.
- Many black boxes and much hand-waving.

Conclusion

- ▶ Digital humanities is all about formal modeling.
- ▶ A corpus is a model in the sense of Stachowiak: it is a reduced mapping of some original for a particular purpose.
- ▶ Corpus-based research means constructing a formal model to test against the corpus
 - ↔ Querying a corpus means constructing a formal model of the phenomenon under research.
- ▶ There is much experience in corpus linguistics, but the research object of history is very different from that of linguistics.
- ▶ Both models (corpus and query) **must be documented**, otherwise the advantages of formal models are lost.
- ▶ The main problem may be that many scholars are not aware of their modeling.
- Many black boxes and much hand-waving.


Conclusion

- ▶ Digital humanities is all about formal modeling.
 - ▶ A corpus is a model in the sense of Stachowiak: it is a reduced mapping of some original for a particular purpose.
 - ▶ Corpus-based research means constructing a formal model to test against the corpus
 - ↔ Querying a corpus means constructing a formal model of the phenomenon under research.
 - ▶ There is much experience in corpus linguistics, but the research object of history is very different from that of linguistics.
 - ▶ Both models (corpus and query) **must be documented**, otherwise the advantages of formal models are lost.
 - ▶ The main problem may be that many scholars are not aware of their modeling.
- **Many black boxes and much hand-waving.**

Historical Models and Serial Sources

Prof. Dr.-Ing. Michael Piotrowski

<michael.piotrowski@unil.ch>

 @true_mxp

Digital Approaches Towards Serial Publications · Brussels, September 11, 2017

References I

-  Douglas Biber, *Methodological issues regarding corpus-based analyses of linguistic variation*, *Literary and Linguistic Computing* **5** [1990], no. 4, 257–269 [english].
-  _____, *Representativeness in corpus design*, *Literary and Linguistic Computing* **8** [1993], no. 4, 243–257 [english].
-  Joshua M. Epstein, *Why model?*, *Journal of Artificial Societies and Social Simulation* **11** [2008], no. 4, 12+ [english].
-  Jay W. Forrester, *Counterintuitive behavior of social systems*, 1995, This is a revised version of the 1971 article published in *Technology Review*.
-  Aleksej Vsevolodovič Gladkij and Igor Aleksandrovič Mel'čuk, *Elementy matematičeskoj lingvistiki*, Nauka, Moskva, 1969 [russian].
-  Herbert Stachowiak, *Allgemeine Modelltheorie*, Springer, Wien, New York, 1973 [german].