

Assessing the Dimensions of Factuality in Biomedical Text

Katarina Ragna Krüger

13. September 2017 (GSCL2017)



“Assessing the Dimensions of Factuality in Biomedical Text”

vorgelegt bei:
Prof. Dr. Manfred Stede
Angewandte Computerlinguistik
Department Linguistik
Universität Potsdam

Anwendungsfall: NLP-Pipeline z.B. für *Text Mining*

Anwendungsfall: NLP-Pipeline z.B. für *Text Mining*

Beispiel: Arztbrief

Die o.g. Allgemeinsymptome in Kombination mit Fieber über 38,0°C, Aortenklappeninsuffizienz, Proteinurie, Hepatosplenomegalie (vorbefundlich durch den Hausarzt), Thrombozytopenie, erhöhtem CRP und dem Nachweis von Strep. oralis in der Blutkultur lassen am ehesten den Verdacht einer Endokarditis lenta zu.

Anwendungsfall: NLP-Pipeline z.B. für *Text Mining*

Beispiel: Arztbrief

Die o.g. Allgemeinsymptome in Kombination mit **Fieber** über 38,0°C, **Aortenklappeninsuffizienz**, **Proteinurie**, **Hepatosplenomegalie** (vorbefundlich durch den Hausarzt), **Thrombozytopenie**, **erhöhtem CRP** und dem Nachweis von **Strep. oralis** in der Blutkultur lassen am ehesten den Verdacht einer Endokarditis lenta zu.

⇒ System extrahiert z.B. die **Symptome**

Anwendungsfall: NLP-Pipeline z.B. für *Text Mining*

Beispiel: Fachartikel

Moreover, simian virus 40 enhancer activity was blocked by the MnlI-AluI fragment in HeLa cells but not in B cells. Thus, the novel enhancer element identified in this study is probably a target site for both positive and negative factors.

Anwendungsfall: NLP-Pipeline z.B. für *Text Mining*

Beispiel: Fachartikel

Moreover, **simian virus 40 enhancer activity was blocked by the MnlI-AluI fragment in HeLa cells** but not in B cells. Thus, the novel enhancer element identified in this study is probably a target site for both positive and negative factors.

⇒ System extrahiert z.B. die **Ergebnisse**

Zentrale Probleme: Kontexte

Zentrale Probleme: Kontexte

1 Negationen

Zentrale Probleme: Kontexte

1 Negationen

Beispiel: Arztbrief

Leber normal groß, echoreiches, homogenes Parenchym ohne distale Schallabschwächung, sonographischer Aspekt einer Fettleber, Pankreas normal groß, Milz nicht vergrößert, Nieren normal groß, kein Harnstau, unauffällige Gefäße, keine freie Flüssigkeit, Harnblase ohne Wandverdickung, keine pathologischen intraluminalen Echostrukturen.

Zentrale Probleme: Kontexte

1 Negationen

Beispiel: Arztbrief

Leber normal groß, echoreiches, homogenes Parenchym ohne **distale Schallabschwächung**, sonographischer Aspekt einer **Fettleber**, Pankreas normal groß, Milz nicht **vergrößert**, Nieren normal groß, kein **Harnstau**, unauffällige Gefäße, keine **freie Flüssigkeit**, Harnblase ohne **Wandverdickung**, keine **pathologischen intraluminalen Echostrukturen**.

Zentrale Probleme: Kontexte

1 Negationen

Beispiel: Arztbrief

Leber normal groß, echoreiches, homogenes Parenchym **ohne distale Schallabschwächung**, sonographischer Aspekt einer **Fettleber**, Pankreas normal groß, Milz **nicht vergrößert**, Nieren normal groß, **kein Harnstau**, unauffällige Gefäße, **keine freie Flüssigkeit**, Harnblase **ohne Wandverdickung**, **keine pathologischen intraluminalen Echostrukturen**.

Zentrale Probleme: Kontexte

- 1 Negationen
- 2 Hedging – Vagheitsausdrücke

Zentrale Probleme: Kontexte

- 1 Negationen
- 2 Hedging – Vagheitsausdrücke

Beispiel: Arztbrief

Elektrophysiologisch konnte als Ursache der Symptomatik eine distal betonte Polyneuropathie nachgewiesen werden, möglicherweise im Rahmen einer schon länger vorbestehenden Störung mit sekundärer Medikamenteninduktion (Antimykotika?). Wobei auch durch einen vorbestehenden entzündlichen Prozess bei anamnestisch geschildertem Erysipel eine Mitinduktion denkbar wäre.

Zentrale Probleme: Kontexte

- 1 Negationen
- 2 Hedging – Vagheitsausdrücke

Beispiel: Arztbrief

Elektrophysiologisch konnte als Ursache der Symptomatik eine **distal betonte Polyneuropathie** nachgewiesen werden, möglicherweise im Rahmen einer schon länger **vorbestehenden Störung mit sekundärer Medikamenteninduktion (Antimykotika?)**. Wobei auch durch einen **vorbestehenden entzündlichen Prozess** bei anamnestisch geschildertem **Erysipel** eine **Mitinduktion** denkbar wäre.

Zentrale Probleme: Kontexte

- 1 Negationen
- 2 Hedging – Vagheitsausdrücke

Beispiel: Arztbrief

Elektrophysiologisch konnte als Ursache der Symptomatik eine **distal betonte Polyneuropathie** nachgewiesen werden, **möglicherweise** im Rahmen einer schon länger **vorbestehenden Störung mit sekundärer Medikamenteninduktion (Antimykotika?)**. Wobei auch durch einen **vorbestehenden entzündlichen Prozess** bei anamnestisch geschildertem **Erysipel** eine **Mitinduktion denkbar wäre**.

Zentrale Probleme: Kontexte

- 1 Negationen
- 2 Hedging – Vagheitsausdrücke

Beispiel: Fachartikel

These results indicate that in monocytic cell lineage, HIV-1 could mimic some differentiation/activation stimuli allowing nuclear NF-KB expression.

Zentrale Probleme: Kontexte

- 1 Negationen
- 2 Hedging – Vagheitsausdrücke

Beispiel: Fachartikel

These results indicate that **in monocytic cell lineage, HIV-1 could mimic some differentiation/activation stimuli allowing nuclear NF-KB expression.**

Zentrale Probleme: Kontexte

- 1 Negationen
- 2 Hedging – Vagheitsausdrücke

Beispiel: Fachartikel

These results **indicate that** in monocytic cell lineage, HIV-1 **could** mimic some differentiation/activation stimuli allowing nuclear NF-KB expression.

Zentrale Probleme: Kontexte

- 1 Negationen
- 2 Hedging – Vagheitsausdrücke

Ignorieren der Kontexte führt zu **falsch-positiven** Extrakten!

Zentrale Probleme: Kontexte

- 1 Negationen
- 2 Hedging – Vagheitsausdrücke

Ignorieren der Kontexte führt zu **falsch-positiven** Extrakten!

⇒ Lösung: automatische Erkenner für Negationen und Hedging

Vielfältige Formen von Hedging

- Substantive
- Verben
- Modalverben
- adjektivische/adverbiale Ausdrücke
- Konjunktionen
- komplexe Ausdrücke

Vielfältige Formen von Hedging

- Substantive
- Verben
- Modalverben
- adjektivische/adverbiale Ausdrücke
- Konjunktionen
- komplexe Ausdrücke

Beispiel: Arztbrief

Als möglicher Ursprung des mutmaßlich mehrzeitig verlaufenen thrombembolischen Ereignisses fand sich eine Oberschenkelvenenthrombose rechts.

Vielfältige Formen von Hedging

- Substantive
- Verben
- Modalverben
- adjektivische/adverbiale Ausdrücke
- Konjunktionen
- komplexe Ausdrücke

Beispiel: Arztbrief

Als **möglicher** Ursprung des **mutmaßlich** mehrzeitig verlaufenen thrombembolischen Ereignisses fand sich eine Oberschenkelvenenthrombose rechts.

Vielfältige Formen von Hedging

- Substantive
- Verben
- Modalverben
- adjektivische/adverbiale Ausdrücke
- Konjunktionen
- komplexe Ausdrücke

Beispiel: Arztbrief

Als **möglicher** (Ursprung des **mutmaßlich** (mehrzeitig verlaufenen) thrombembolischen Ereignisses) fand sich eine Oberschenkelvenenthrombose rechts.

Vielfältige Formen von Hedging

- Substantive
- Verben
- Modalverben
- adjektivische/adverbiale Ausdrücke
- Konjunktionen
- komplexe Ausdrücke

⇒ Wichtig: Bestimmung des jeweiligen **Skopus** der Hedges!

Leistungen bisheriger Hedging-Erkenner

Leistungen bisheriger Hedging-Erkenner

Ansätze:

- 1 Regel-basierte Verfahren (lokale **Grammatiken**)
- 2 Klassifikation mit **Machine Learning**-Verfahren (z.B. *SVM*, *CRF*)

Leistungen bisheriger Hedging-Erkenner

Ansätze:

- 1 Regel-basierte Verfahren (lokale **Grammatiken**)
- 2 Klassifikation mit **Machine Learning**-Verfahren (z.B. *SVM*, *CRF*)

Ergebnisse:

- z.B. CoNLL-2010 Shared Task:
Hedge-Identifikation: F-Score ≈ 85
Skopus-Bestimmung: F-Score ≈ 60

Was ist dann noch das Problem?

Was ist dann noch das Problem?

Faktualität ist kein binäres Konzept!

Was ist dann noch das Problem?

Faktualität ist kein binäres Konzept!

Beispiel: Arztbrief 1

This **may** represent a viral illness.

Was ist dann noch das Problem?

Faktualität ist kein binäres Konzept!

Beispiel: Arztbrief 1

This **may** represent a viral illness.

Beispiel: Arztbrief 2

Findings **most likely** represent reactive airways disease **or** viral infection with left lower lobe subsegmental atelectasis.

Was ist dann noch das Problem?

Faktualität ist kein binäres Konzept!

Beispiel: Arztbrief 1

This **may** represent a viral illness.

Beispiel: Arztbrief 2

Findings **most likely** represent reactive airways disease **or** viral infection with left lower lobe subsegmental atelectasis.

Verschiedene Grade von Faktualität

⇒ **verschiedene Grade von Hedging**

Was ist dann noch das Problem?

Faktualität ist kein binäres Konzept!

Beispiel: Arztbrief 1

This **may** represent a viral illness.

Beispiel: Arztbrief 2

Findings **most likely** represent reactive airways disease **or** viral infection with left lower lobe subsegmental atelectasis.

Verschiedene Grade von Faktualität

⇒ **verschiedene Grade von Hedging**

Ignorieren der Faktualitätsgrade führt zu **falsch-negativen** Ergebnissen!

Projekt

Fernziel: automatisches Erkennen der Grade von Faktualität in Texten

.....

Projekt

Fernziel: automatisches Erkennen der Grade von Faktualität in Texten

.....

Vorhaben: Erstellen eines Korpus mit **annotierten Faktualitätsgraden**
(Gebiet: Biologie/Medizin) als Trainingsgrundlage

Projekt

Fernziel: automatisches Erkennen der Grade von Faktualität in Texten

.....

Vorhaben: Erstellen eines Korpus mit **annotierten Faktualitätsgraden**
(Gebiet: Biologie/Medizin) als Trainingsgrundlage

Grundlage: *BioScope*-Korpus

Projekt

Fernziel: automatisches Erkennen der Grade von Faktualität in Texten

.....

Vorhaben: Erstellen eines Korpus mit **annotierten Faktualitätsgraden**
(Gebiet: Biologie/Medizin) als Trainingsgrundlage

Grundlage: *BioScope*-Korpus

Methode: *Crowdsourcing*

BioScope-Korpus (Vincze et al. 2008)

BioScope-Korpus (Vincze et al. 2008)

- u.a. CoNLL-2010 Shared Task-Datensatz
- Textsorten:
 - 1 1.273 biomed. Abstracts
 - 2 9 biomed. Fachartikel
 - 3 1.954 radiol. Arztbriefe
- vorannotiert: Negationen, Hedges (mit Skopi)

Vorbereitung der Daten

BioScope-Korpus

Vorbereitung der Daten

BioScope-Korpus



Randomisierte Auswahl der Testsätze



Extraktion der Sätze aus XML



Prüfung der Sätze

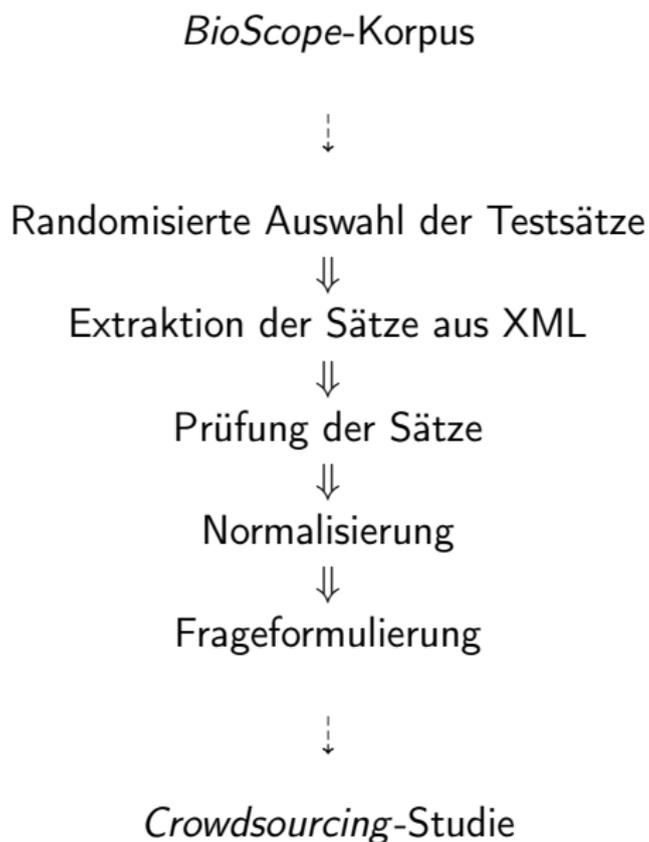


Normalisierung



Frageformulierung

Vorbereitung der Daten



Normalisierung der Testsätze (nach Saurí 2008)

```
<sentence id="S1.7">These results <xcope id="X1.7.2"><cue type="speculation" ref="X1.7.2">indicate that</cue> in monocytic cell lineage, HIV-1 <xcope id="X1.7.1"><cue type="speculation" ref="X1.7.1">could</cue> mimic some differentiation/activation stimuli allowing nuclear NF-KB expression</xcope></xcope>.</sentence>
```

Normalisierung der Testsätze (nach Saurí 2008)

```
<sentence id="S1.7">These results <xcope id="X1.7.2"><cue type="speculation" ref="X1.7.2">indicate that</cue> in monocytic cell lineage, HIV-1 <xcope id="X1.7.1"><cue type="speculation" ref="X1.7.1">could</cue> mimic some differentiation/activation stimuli allowing nuclear NF-KB expression</xcope></xcope>.</sentence>
```



These results **indicate that** in monocytic cell lineage, HIV-1 **could** mimic some differentiation/activation stimuli allowing nuclear NF-KB expression.

Normalisierung der Testsätze (nach Saurí 2008)

```
<sentence id="S1.7">These results <xcope id="X1.7.2"><cue type="speculation" ref="X1.7.2">indicate that</cue> in monocytic cell lineage, HIV-1 <xcope id="X1.7.1"><cue type="speculation" ref="X1.7.1">could</cue> mimic some differentiation/activation stimuli allowing nuclear NF-KB expression</xcope></xcope>.</sentence>
```



These results **indicate that** in monocytic cell lineage, HIV-1 **could** mimic some differentiation/activation stimuli allowing nuclear NF-KB expression.



Based on your reading, do you think that...

in monocytic cell lineage, HIV-1 mimics some differentiation/activation stimuli allowing nuclear NF-KB expression?

Crowdsourcing-Studie

45%

These results indicate that in monocytic cell lineage, HIV-1 could mimic some differentiation/activation stimuli allowing nuclear NF-KB expression.

Based on your reading, do you think that

In monocytic cell lineage, HIV-1 mimics some differentiation/activation stimuli allowing nuclear NF-KB expression?

	certainly	probably	possibly	possibly not	probably not	certainly not	unknown (no claims made)
happened/is happening/will happen	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

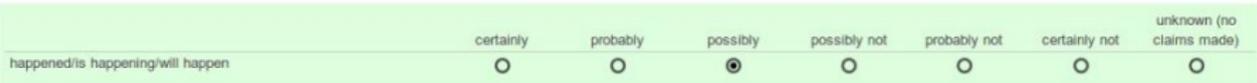
Crowdsourcing-Studie

45%

These results indicate that in monocytic cell lineage, HIV-1 could mimic some differentiation/activation stimuli allowing nuclear NF-KB expression.

Based on your reading, do you think that

In monocytic cell lineage, HIV-1 mimics some differentiation/activation stimuli allowing nuclear NF-KB expression?



- 7-Punkte-Skala für Faktualität (nach de Marneffe, Manning und Potts 2012):

CT+

PR+

PS+

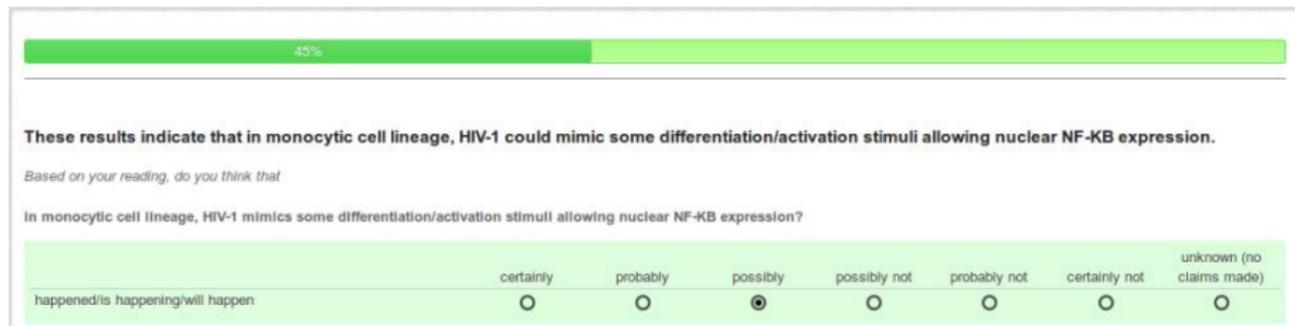
PS-

PR-

CT-

Uu

Crowdsourcing-Studie



- 7-Punkte-Skala für Faktualität (nach de Marneffe, Manning und Potts 2012):

CT+ PR+ PS+ PS- PR- CT- Uu

- Plattform: Prolific Academic Ltd

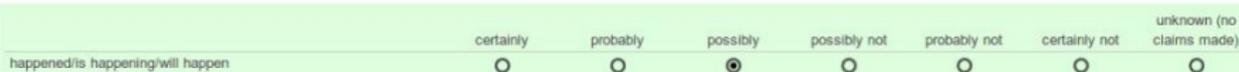
Crowdsourcing-Studie

45%

These results indicate that in monocytic cell lineage, HIV-1 could mimic some differentiation/activation stimuli allowing nuclear NF-KB expression.

Based on your reading, do you think that

In monocytic cell lineage, HIV-1 mimics some differentiation/activation stimuli allowing nuclear NF-KB expression?

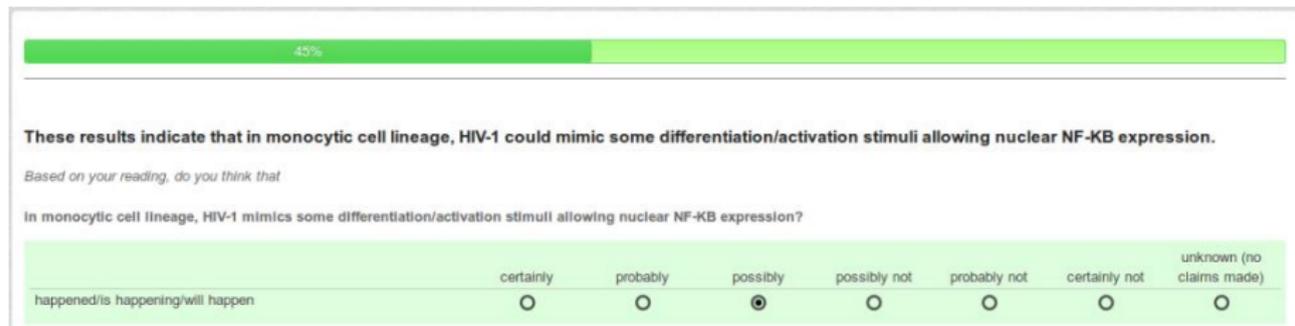


- 7-Punkte-Skala für Faktualität (nach de Marneffe, Manning und Potts 2012):

CT+ PR+ PS+ PS- PR- CT- Uu

- Plattform: Prolific Academic Ltd
- 293 Probanden

Crowdsourcing-Studie



- 7-Punkte-Skala für Faktualität (nach de Marneffe, Manning und Potts 2012):

CT+ PR+ PS+ PS- PR- CT- Uu

- Plattform: Prolific Academic Ltd
- 293 Probanden
- 690 Testsätze (ca. 60 – 70% Hedges), je 10-fach bewertet

Ergebnisse

Ergebnisse

- Erreichte Mehrheitsvoten für 73,62% der Testsätze

Ergebnisse

- Erreichte Mehrheitsvoten für 73,62% der Testsätze
- *Inter-Annotator-Agreement* (Fleiss' Kappa) moderat: gesamt 0,49

Ergebnisse

- Erreichte Mehrheitsvoten für 73,62% der Testsätze
- *Inter-Annotator-Agreement* (Fleiss' Kappa) moderat: gesamt 0,49
- Robuste Hedges:
 - Modalverben (*can, could, may, might*)
 - bestimmte Adverbien/Adjektive (*likely, possibly, potential*)
 - bestimmte Vollverben (*appear, suggest*)

Ergebnisse

- Erreichte Mehrheitsvoten für 73,62% der Testsätze
- *Inter-Annotator-Agreement* (Fleiss' Kappa) moderat: gesamt 0,49
- Robuste Hedges:
 - Modalverben (*can, could, may, might*)
 - bestimmte Adverbien/Adjektive (*likely, possibly, potential*)
 - bestimmte Vollverben (*appear, suggest*)
- systematische Wechselwirkungen von Hedges
 - mit anderen Hedges (z.B. *suggest + may*)
 - mit Negationen (z.B. *not + appear*)
 - mit anderen Ausdrücken (z.B. *strongly + indicate*)

Ergebnisse

- Erreichte Mehrheitsvoten für 73,62% der Testsätze
- *Inter-Annotator-Agreement* (Fleiss' Kappa) moderat: gesamt 0,49
- Robuste Hedges:
 - Modalverben (*can, could, may, might*)
 - bestimmte Adverbien/Adjektive (*likely, possibly, potential*)
 - bestimmte Vollverben (*appear, suggest*)
- systematische Wechselwirkungen von Hedges
 - mit anderen Hedges (z.B. *suggest + may*)
 - mit Negationen (z.B. *not + appear*)
 - mit anderen Ausdrücken (z.B. *strongly + indicate*)
- Unterschiede bei Arztbriefen:
 - deutlich mehr Hedging durch Adj/Adv, weniger durch Vollverben
 - tw. abweichende Hedge-Ausdrücke

Schlussfolgerungen

- 1 Faktualität wird **weitgehend einheitlich** bewertet.

Schlussfolgerungen

- ① Faktualität wird **weitgehend einheitlich** bewertet.
- ② Faktualität lässt sich anhand der **7-Punkte-Skala** graduieren.

Schlussfolgerungen

- 1 Faktualität wird **weitgehend einheitlich** bewertet.
- 2 Faktualität lässt sich anhand der **7-Punkte-Skala** graduieren.
- 3 Einige Hedges sind wesentlich **verlässlicher** als andere.

Schlussfolgerungen

- 1 Faktualität wird **weitgehend einheitlich** bewertet.
- 2 Faktualität lässt sich anhand der **7-Punkte-Skala** graduieren.
- 3 Einige Hedges sind wesentlich **verlässlicher** als andere.
- 4 Es gibt Unterschiede zwischen den **Textsorten**.

Schlussfolgerungen

- 1 Faktualität wird **weitgehend einheitlich** bewertet.
- 2 Faktualität lässt sich anhand der **7-Punkte-Skala** graduieren.
- 3 Einige Hedges sind wesentlich **verlässlicher** als andere.
- 4 Es gibt Unterschiede zwischen den **Textsorten**.
- 5 **Kontexte** müssen berücksichtigt werden.

Ausblick

Forschungsbeitrag des Projekts:

- erstes größeres, mit Faktualitätsgraden annotiertes biomed. Korpus
⇒ erste verlässliche Trainingsgrundlage für Klassifikatoren
(verfügbar für weitere Forschung)
- *Proof of Concept*: Graduierung von Faktualität in Fachtexten
- Erprobung von *Crowdsourcing* als neue ling. Methode
- Detailergebnisse zur Verwendung von Hedging in drei Textsorten

Forschungsbeitrag des Projekts:

- erstes größeres, mit Faktualitätsgraden annotiertes biomed. Korpus
⇒ erste verlässliche Trainingsgrundlage für Klassifikatoren
(verfügbar für weitere Forschung)
- *Proof of Concept*: Graduierung von Faktualität in Fachtexten
- Erprobung von *Crowdsourcing* als neue ling. Methode
- Detailergebnisse zur Verwendung von Hedging in drei Textsorten

Folgeprojekt bei Prof. Dr. Stede, Angewandte Computerlinguistik, UP:

- Konstruktion eines Klassifikators zur automatischen Extraktion von Informationen und ihren Faktualitätsgraden aus biomed. Texten
- probabilistisches Modell (vgl. de Marneffe, Manning und Potts 2012)

Literatur

- de Marneffe, Marie-Catherine, Christopher D. Manning und Christopher Potts (2012). “Did it happen? The pragmatic complexity of veridicality assessment”. In: *Computational Linguistics* 38.2, S. 301–333.
- Saurí, Roser (2008). “A Factuality Profiler for Eventualities in Text”. Diss. Waltham, MA.
- Vincze, Veronika, György Szarvas, Richárd Farkas, György Móra und János Csirik (2008). “The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes”. In: *BMC Bioinformatics* Suppl 11.9, S9. DOI: 10.1186/1471-2105-9-S11-S9.