

Basic evaluation idea

Ground truth + pred (kraken) evaluation

- Calculates page-wise error score using both the ground truth (gt) and the pred (kraken)
- Better because we have more info

Pred evaluation

- Calculates page-wise error score **only** using the pred (kraken)

If we have the Ground truth evaluator we can see if our individual predictors match.

Overlay evaluation approach

105

"... kam!" Diese Billsschrift machte tiefen Eindruck auf den Kaiser und bestärkte ihn in dem Vorjaze, die Leibeigenchaft sobald als möglich aufzuheben. Schon unter Maria Theresia wurden mit Zustimmung Josefs II. einige Gesetze zum Wohle der Bauern erlassen; es wurden J. B. Beamte eingesetzt, die Kreishauptleute, welche den Auftrag erhalten, die Bauern gegen ungerechte Behandlung von Seite ihrer Gutsbesitzer in Schutz zu nehmen, es wurde den Bauern gestattet, sich von der Leibeigenchaft durch eine Summe Geldes loszukaufen u. a. Aber ganz half dies doch nicht. Als Josef II. Alleinherrsherr wurde, war daher eine seiner ersten Regierungshandlungen, daß er die Leibeigenhaft aufhob (1781). Damit wurde der Zustand des Bauernstandes im wesentlichen derselbe, wie er noch heute ist; nur die Frontdienste blieben zum Teil, wurden aber vermindert und genau geregelt, damit kein Gutsbesitzer von den Bauern mehr verlangen könne, als ihm zufam. (Unter Kaiser Ferdinand I. dem Güttigen und dem gegenwärtig regierenden Monarchen, Sr. Majestät dem Kaiser Franz Joseph I., wurden die Frontdienste ebenfalls nach und nach aufgehoben.)

Mit Recht wird daher das Andenken Josefs II. besonders von dem Bauernstande noch heute hochgeehrt; einen Beweis dafür geben unter anderem die zahllosen Kaiser Joseph-Denkäbler, welche zum Andenken an den großen Volkskaiser in allen Teilen der österreichischen Monarchie errichtet worden sind.

Bur Wiederholung:

Napoleon I. Die große französische Revolution 1789. Herkunft Napoleons. Napoleon als Kaiser von Frankreich. Sein Streben nach der Weltherrschaft. Untergang des Deutschen Reiches durch Napoleon. Sein Verfahren gegen die deutschen Fürsten. Niederklegung des kaiserlichen Kaiserthums durch Kaiser Franz II. 1806. Entstehung des Titels "Kaiser von Österreich" (1804).

Österreichs Freiheitkampf 1809. Vorbereitungen zum Kriege. Erzherzog Karl als Kriegsminister. Errichtung der Landwehr.

Österreichs Kriegserklärung. Begeisterung für den Krieg. Rückszug der Österreicher. Der Heldenkampf bei Ebelsberg. (Radeckly.)

Eroberung Wiens durch Napoleon I. Napoleons Übergang über die Donau. Erster Tag der Schlacht bei Aspern (21. Mai). Zweiter Schlachttag.

23. Andreas Hofer.

Als Napoleon Kaiser der Franzosen war und in seinem Übermuth nach belieben die Länder Europas zerstörte, um die losgetrennten Gebiete entweder für sich selbst zu behalten oder an seine Verwandten und Anhänger zu verschenken, riss er nach einem blutigen, für ihn sieg-

106

Wohltätige Maßregeln der Kaiserin Maria Theresia

Aufhebung der Leibeigenhaft (1781)

Bur Wiederholung:

Napoleon I. Die große französische Revolution 1789. Herkunft Napoleons. Napoleon als Kaiser von Frankreich. Sein Streben nach der Weltherrschaft. Untergang des Deutschen Reiches durch Napoleon. Sein Verfahren gegen die deutschen Fürsten. Niederklegung des deutschen Kaiserthums durch Kaiser Franz II. 1806. Entstehung des Titels "Kaiser von Österreich" (1804).

Österreichs Freiheitkampf 1809. Vorbereitungen zum Kriege. Erzherzog Karl als Kriegsminister. Errichtung der Landwehr.

Österreichs Kriegserklärung. Begeisterung für den Krieg. Rückszug der Österreicher. Der Heldenkampf bei Ebelsberg. (Radeckly.)

Eroberung Wiens durch Napoleon I. Napoleons Übergang über die Donau. Erster Tag der Schlacht bei Aspern (21. Mai). Zweiter Schlachttag.

23. Andreas Hofer.

Als Napoleon Kaiser der Franzosen war und in seinem Übermuth nach belieben die Länder Europas zerstörte, um die losgetrennten Gebiete entweder für sich selbst zu behalten oder an seine Verwandten und Anhänger zu verschenken, riss er nach einem blutigen, für ihn sieg-

106

Wohltätige Maßregeln der Kaiserin Maria Theresia

Aufhebung der Leibeigenhaft (1781)

Bur Wiederholung:

Napoleon I. Die große französische Revolution 1789. Herkunft Napoleons. Napoleon als Kaiser von Frankreich. Sein Streben nach der Weltherrschaft. Untergang des Deutschen Reiches durch Napoleon. Sein Verfahren gegen die deutschen Fürsten. Niederklegung des deutschen Kaiserthums durch Kaiser Franz II. 1806. Entstehung des Titels "Kaiser von Österreich" (1804).

Österreichs Freiheitkampf 1809. Vorbereitungen zum Kriege. Erzherzog Karl als Kriegsminister. Errichtung der Landwehr.

Österreichs Kriegserklärung. Begeisterung für den Krieg. Rückszug der Österreicher. Der Heldenkampf bei Ebelsberg. (Radeckly.)

Eroberung Wiens durch Napoleon I. Napoleons Übergang über die Donau. Erster Tag der Schlacht bei Aspern (21. Mai). Zweiter Schlachttag.

23. Andreas Hofer.

Zerstörung Tirols von Österreich.

Als Napoleon Kaiser der Franzosen war und in seinem Übermuth nach belieben die Länder Europas zerstörte, um die losgetrennten Gebiete entweder für sich selbst zu behalten oder an seine Verwandten und Anhänger zu verschenken, riss er nach einem blutigen, für ihn sieg-

Raw kranken pred → Overlay with gt → Segment errors

Current rules

- **Good** if a pred *paragraph* matches with gt *paragraph*
- **Good** if a any pred box matches a gt *non-paragraph*.
So basically we say „If it has its own box to go with marginalia it's good no matter the .xml RegionType“
- **Bad** if there are boxes in pred that are NOT also in gt
- **Bad** if any marginalia in gt does not also have a corresponding box in pred
- **Count** the errors (we can weigh them)
- Box overlay has a threshhold with boxes match

In our example

106
„... kam!“ Diese Billsschrift machte tiefen Eindruck auf den Kaiser und bestärkte ihn in dem Vorjahr, die Leibeigenchaft sobald als möglich aufzugeben. Schon unter Maria Theresia wurden mit Zustimmung Josephs II. einige Gesetze zum Wohle der Bauern erlassen; es wurden J. B. Beamte eingesetzt, die Kreishauptleute, welche den Auftrag erhielten, die Bauern gegen ungerechte Behandlung von Seite ihrer Untertanen in Schutz zu nehmen, es wurde den Bauern gestattet, sich von der Leibeigenchaft durch eine Summe Geldes loszukaufen u. a. Aber ganz half dies doch nicht. Als Josef II. Alleinherrsherr wurde, war daher eine seiner ersten Regierungshandlungen, daß er die Leibeigenchaft aufhob (1781). Damit wurde der Zustand des Bauernstandes im wesentlichen derselbe, wie er noch heute ist; nur die Frontdienste blieben zum Teil, wurden aber vermindert und genau geregelt, damit kein Gutsbesitzer von den Bauern mehr verlangen könne, als ihm zufam. (Unter Kaiser Ferdinand I. dem Gültigen und dem gegenwärtig regierenden Monarchen, Sr. Majestät dem Kaiser Franz Joseph I., wurden die Frontdienste ebenfalls nach und nach aufgehoben.) Mit Recht wird daher das Andenken Josephs II. besonders von dem Bauernstande noch heute hochgeehrt; einen Beweis dafür geben unter anderem die zahllosen Kaiser Joseph-Denkämler, welche zum Andenken an den großen Volkskaiser in allen Teilen der österreichischen Monarchie errichtet worden sind.

Bur Wiederholung:

Napoleon I. Die große französische Revolution 1789. Herkunft Napoleon. Napoleon als Kaiser von Frankreich. Sein Streben nach der Weltherrschaft. Untergang des Deutschen Reiches durch Napoleon. Sein Verfahren gegen die deutschen Fürsten. Niederkunft des deutschen Kaiserthums durch Kaiser Franz II. 1806. Entstehung des Titels „Kaiser von Österreich“ (1804).

Österreichs Freiheitshumpf 1809. Vorbereitungen zum Kriege. Erzherzog Karl als Kriegsminister. Errichtung der Landwehr.

Österreichs Kriegserklärung. Begeisterung für den Krieg. Rüfung der Österreicher. Der Heldenkampf bei Ebelsberg. (Blodetly.)

Erobierung Wiens durch Napoleon I. Napoleons Übergang über die Donau. Erster Tag der Schlacht bei Aspern (21. Mai). Zweiter Schlachttag.

23. Andreas Hofer.

Als Napoleon Kaiser der Franzosen war und in seinem Übermuth nach Belieben die Länder Europas zerstückelte, um die losgetrennten Gebiete entweder für sich selbst zu behalten oder an seine Verwandten und Anhänger zu verschenken, rüß er nach einem blutigen, für ihn sieg-

Has it's own box so good

Not it's own box so BAD

Not marked as paragraph in gt so should have it's own box (i guess)

106
„... kam!“ Diese Billsschrift machte tiefen Eindruck auf den Kaiser und bestärkte ihn in dem Vorjahr, die Leibeigenchaft sobald als möglich aufzugeben. Schon unter Maria Theresia wurden mit Zustimmung Josephs II. einige Gesetze zum Wohle der Bauern erlassen; es wurden J. B. Beamte eingesetzt, die Kreishauptleute, welche den Auftrag erhielten, die Bauern gegen ungerechte Behandlung von Seite ihrer Untertanen in Schutz zu nehmen, es wurde den Bauern gestattet, sich von der Leibeigenchaft durch eine Summe Geldes loszukaufen u. a. Aber ganz half dies doch nicht. Als Josef II. Alleinherrsherr wurde, war daher eine seiner ersten Regierungshandlungen, daß er die Leibeigenchaft aufhob (1781). Damit wurde der Zustand des Bauernstandes im wesentlichen derselbe, wie er noch heute ist; nur die Frontdienste blieben zum Teil, wurden aber vermindert und genau geregelt, damit kein Gutsbesitzer von den Bauern mehr verlangen könne, als ihm zufam. (Unter Kaiser Ferdinand I. dem Gültigen und dem gegenwärtig regierenden Monarchen, Sr. Majestät dem Kaiser Franz Joseph I., wurden die Frontdienste ebenfalls nach und nach aufgehoben.) Mit Recht wird daher das Andenken Josephs II. besonders von dem Bauernstande noch heute hochgeehrt; einen Beweis dafür geben unter anderem die zahllosen Kaiser Joseph-Denkämler, welche zum Andenken an den großen Volkskaiser in allen Teilen der österreichischen Monarchie errichtet worden sind.

Aufhebung der Leibeigenchaft (1781).

106
„... kam!“ Diese Billsschrift machte tiefen Eindruck auf den Kaiser und bestärkte ihn in dem Vorjahr, die Leibeigenchaft sobald als möglich aufzugeben. Schon unter Maria Theresia wurden mit Zustimmung Josephs II. einige Gesetze zum Wohle der Bauern erlassen; es wurden J. B. Beamte eingesetzt, die Kreishauptleute, welche den Auftrag erhielten, die Bauern gegen ungerechte Behandlung von Seite ihrer Untertanen in Schutz zu nehmen, es wurde den Bauern gestattet, sich von der Leibeigenchaft durch eine Summe Geldes loszukaufen u. a. Aber ganz half dies doch nicht. Als Josef II. Alleinherrsherr wurde, war daher eine seiner ersten Regierungshandlungen, daß er die Leibeigenchaft aufhob (1781). Damit wurde der Zustand des Bauernstandes im wesentlichen derselbe, wie er noch heute ist; nur die Frontdienste blieben zum Teil, wurden aber vermindert und genau geregelt, damit kein Gutsbesitzer von den Bauern mehr verlangen könne, als ihm zufam. (Unter Kaiser Ferdinand I. dem Gültigen und dem gegenwärtig regierenden Monarchen, Sr. Majestät dem Kaiser Franz Joseph I., wurden die Frontdienste ebenfalls nach und nach aufgehoben.) Mit Recht wird daher das Andenken Josephs II. besonders von dem Bauernstande noch heute hochgeehrt; einen Beweis dafür geben unter anderem die zahllosen Kaiser Joseph-Denkämler, welche zum Andenken an den großen Volkskaiser in allen Teilen der österreichischen Monarchie errichtet worden sind.

Bur Wiederholung:

Napoleon I. Die große französische Revolution 1789. Herkunft Napoleon. Napoleon als Kaiser von Frankreich. Sein Streben nach der Weltherrschaft. Untergang des Deutschen Reiches durch Napoleon. Sein Verfahren gegen die deutschen Fürsten. Niederkunft des deutschen Kaiserthums durch Kaiser Franz II. 1806. Entstehung des Titels „Kaiser von Österreich“ (1804).

Österreichs Freiheitshumpf 1809. Vorbereitungen zum Kriege. Erzherzog Karl als Kriegsminister. Errichtung der Landwehr.

Österreichs Kriegserklärung. Begeisterung für den Krieg. Rüfung der Österreicher. Der Heldenkampf bei Ebelsberg. (Blodetly.)

Erobierung Wiens durch Napoleon I. Napoleons Übergang über die Donau. Erster Tag der Schlacht bei Aspern (21. Mai). Zweiter Schlachttag.

23. Andreas Hofer.

Losreisung Tirols von Österreich.

Als Napoleon Kaiser der Franzosen war und in seinem Übermuth nach Belieben die Länder Europas zerstückelte, um die losgetrennten Gebiete entweder für sich selbst zu behalten oder an seine Verwandten und Anhänger zu verschenken, rüß er nach einem blutigen, für ihn sieg-

Raw kranken pred → Overlay with gt → Segment errors

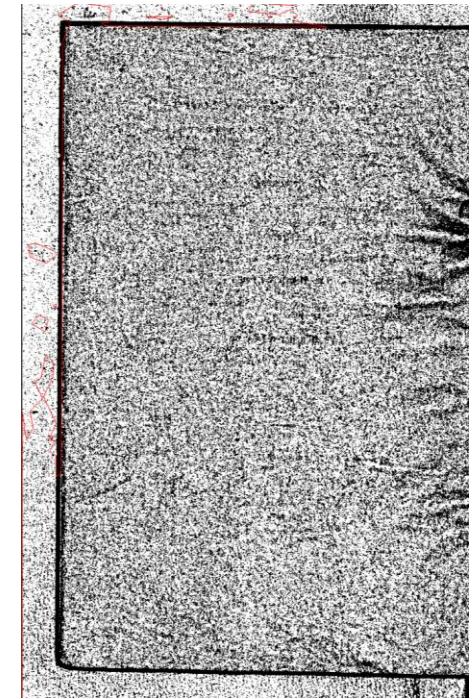
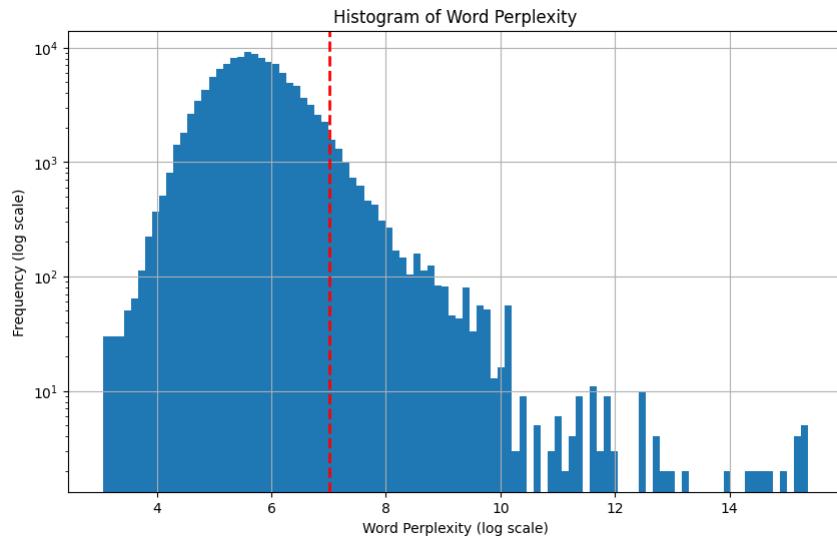
Chunk-based approach

NLP - Praktikum

Alzbeta Hrabosova

Threshold for error detection

- Define global threshold rather than page based
- Removed extremely faulty pages (for threshold detection only)
- IQR-based upper bound – **threshold** above which word perplexity is unusually high based on typical (cleaner) pages.



gab es in den Häusern
reicher Bürger mitunter
sehr geräumige Höfe
(vgl. Fig. 16), weil man
einen Raum brauchte,
wo Waren auf- und ab-
geladen werden konnten.

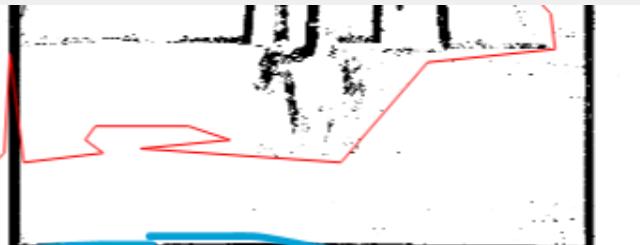
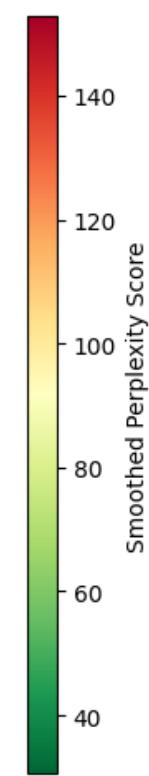
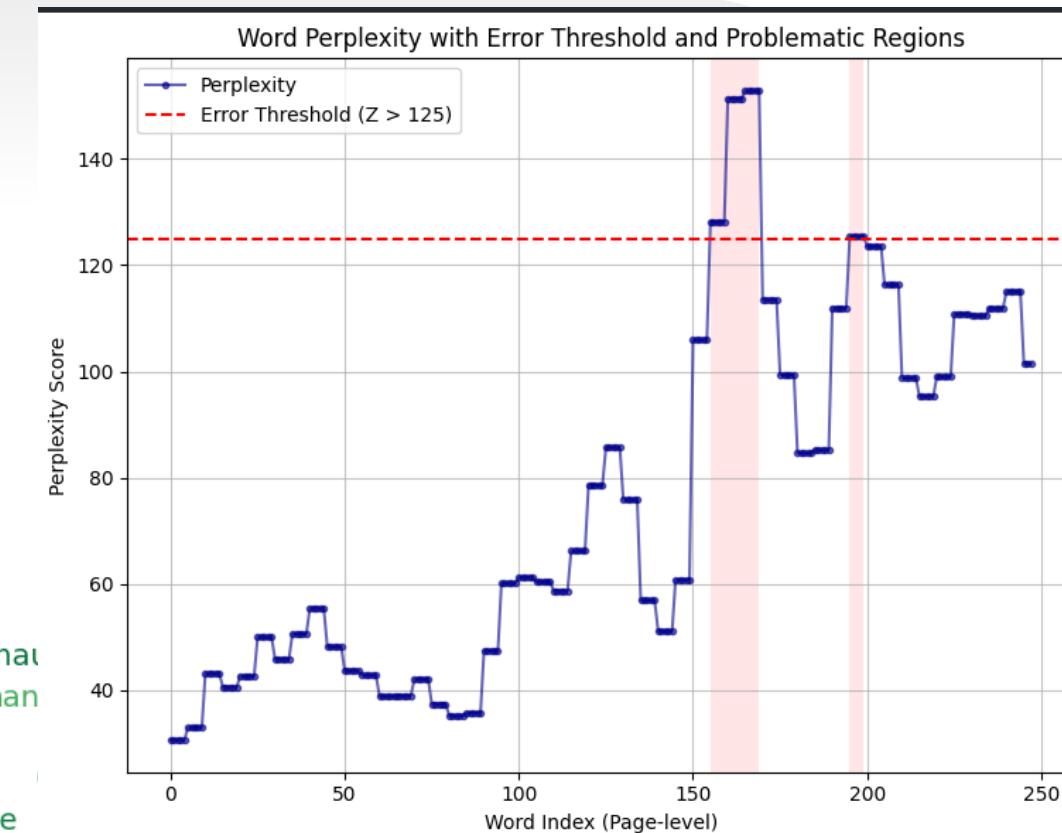


Fig. 14. Ein mittelalterliches Stabtor.

war in den Städten
des Mittelalters fast durchweg unbekannt; wer zur Nachtzeit das Haus verließ, mußte, wenn nicht gerade Vollmond war, eine Fackel oder sonst ein Licht mitnehmen. Es war daher auch ziemlich gefährlich, sich zur Nachtzeit auf die Gasse zu wagen; nächtliche Überfälle, Raub- und Mordtaten waren häufiger als heutzutage. Auch waren die

*) In Fig. 15 sind fast alle abgebildeten Häuser solche Riegelbauten. Typus, Bild aus der Geschichte für Knabenbürgerschulen. II. 2. Aufl.

mußte, so waren die Plätze klein, die Gassen eng, wie man von Wien, Prag und anderen Städten sehen kann (Fig. 15). gewöhnlich so, daß sie der Gasse ihre schmale Seite (die möglichst viel Licht zu gewinnen, baute man auch wohl die Gasse vorsprangen, wodurch die Gassen winkelig (zickzackförmig) haben sich hie und da noch erhalten. Die Häuser waren auch häufiger waren sogenannte Fach oder Riegelbauten.) Bei allen mitunter sehr geräumige Höfe (vgl. Fig. 16), weil man einen werden konnten. Straßenbeleuchtung war in den Städten 15 mittelalterliches Stadttor. Fehlen der des Mittelalters fast Hausstrahlenverließ, mußte, wenn nicht gerade Vollmond war, mitnehmen. Es war daher auch ziemlich gefährlich, tung. sich nächtliche Überfälle, RaubUnsiccerund Mordtaten waren häufiger 15 sind fast alle abgebildeten Häuser solche Riegelbauten. Knabenbürgerschulen. II. 2. Aufl. heit.



Difficulties

- No errors detected because of no context errors

3. Die Reformation in Ungarn und Siebenbürgen.

Litteratur: Haner, historia ecclesiarum Transsilvanicarum. Frankof. et Lips. 1694. — Lampe, historia ecclesiae reformatae in Hungaria et Transsilvania. Traj. 1725. — Fabó, Gemälde aus der Geschichte des ungarischen Protestantismus. Pest, 1868. — Geschichte der ev. Kirche in Ungarn. Herausg. von Merle d'Aubigné. Berlin, 1854. — Monumenta Evangelicorum Aug. conf. historica, ed. Fabó. 3 Bde. Pest, 1861—65. — Urkundenbuch der ev. Landeskirche A. B. in Siebenbürgen. Herausg. von G. D. Deutsch. 2 Bde. Hermannstadt, 1862—83.

Bald nach dem Beginne der grossen Kirchenreformation in Deutschland fanden die reformatorischen Ideen, zunächst durch die Schriften Luthers, auch in Ungarn Eingang und verbreiteten sich in wenigen Jahren in solchem Masse, dass Regierung und Reichstag in den Jahren 1523—25 sich veranlasst sahen, mit strengen Verordnungen und Gesetzen (Todesstrafe und Güterkonfiskation) dagegen aufzutreten. Sie kamen nicht zur Ausführung, um so weniger, als selbst in den Kreisen der Vornehmsten die neuen Lehren Freunde fanden (Briefwechsel Luthers mit Königin Maria). Nach der Katastrophe von Mohács gestalteten sich die Verhältnisse ungemein günstig für die weitere Ausbreitung der Reformation. Die durch den Tod mehrerer

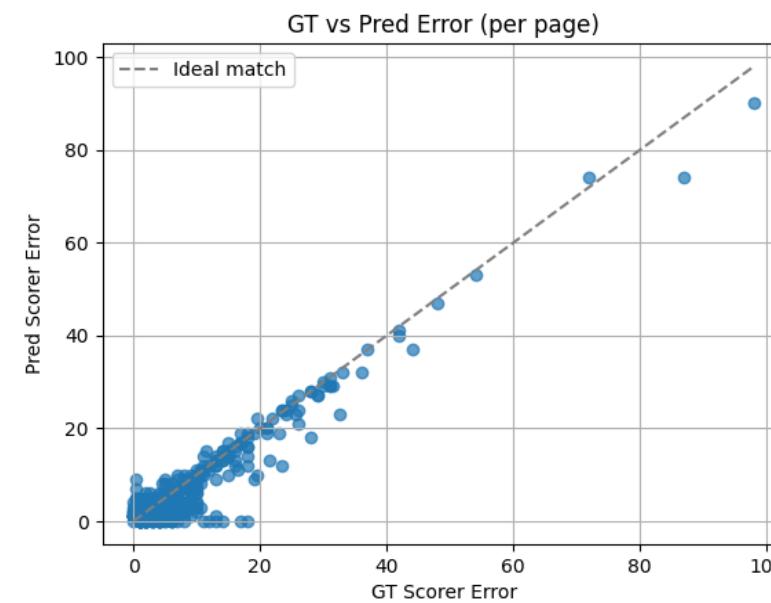
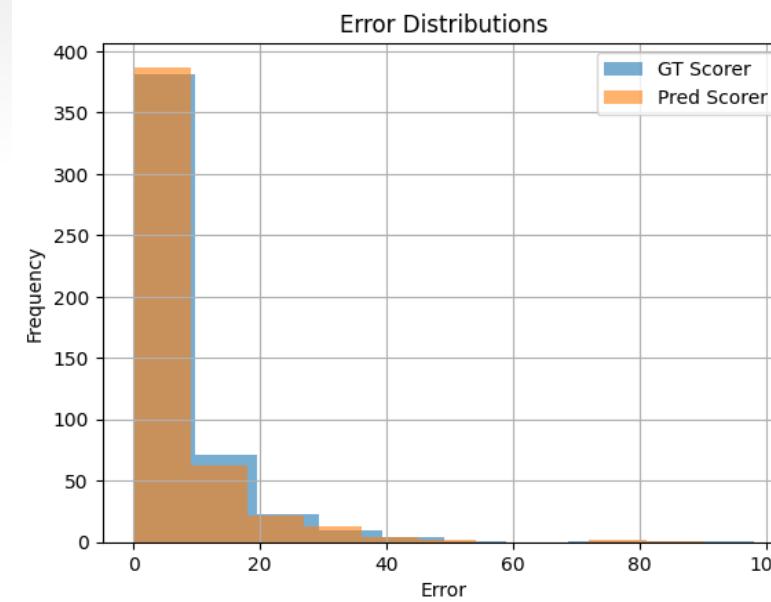
3. Die Reformation in Ungarn und Siebenbürgen.

Litteratur: Haner, historia ecclesiarum Transsilvanicarum. Frankof. et Lips. 1694. — Lampe, historia ecclesiae reformatae in Hungaria et Transsilvania. Traj. 1725. — Fabó, Gemälde aus der Geschichte des ungarischen Protestantismus. Pest, 1868. — Geschichte der ev. Kirche in Ungarn. Herausg. von Merle d'Aubigné. Berlin, 1854. — Monumenta Evangelicorum Aug. conf. historica, ed. Fabó. 3 Bde. Pest, 1861—65. — Urkundenbuch der ev. Landeskirche A. B. in Siebenbürgen. Herausg. von G. D. Deutsch. 2 Bde. Hermannstadt, 1862—83.

Bald nach dem Beginne der grossen Kirchenreformation in Deutschland fanden die reformatorischen Ideen, zunächst durch die Schriften Luthers, auch in Ungarn Eingang und verbreiteten sich in wenigen Jahren in solchem Masse, dass Regierung und Reichstag in den Jahren 1523—25 sich veranlasst sahen, mit strengen Verordnungen und Gesetzen (Todesstrafe und Güterkonfiskation) dagegen aufzutreten. Sie kamen nicht zur Ausführung, um so weniger, als selbst in den Kreisen der Vornehmsten die neuen Lehren Freunde fanden (Briefwechsel Luthers mit Königin Maria). Nach der Katastrophe von Mohács gestalteten sich die Verhältnisse ungemein günstig für die weitere Ausbreitung der Reformation. Die durch den Tod mehrerer

Evaluation

Metric	GT Scorer	Pred Scorer
Count	493	493.0000
Mean	7.6288	6.4503
Median	4.5	3.0000
Stdev	10.1078	9.5719
Min	0	0.0000
Max	98	90.0000
MAE	7.6288	6.4503
Correlation	0.9552	-
GT vs Pred RMSE	3.2135	-



Questions

- What is an error?

— 19 —

bis zum Jahre 330 v. Ch., in welchem ihm der macedonische König Alexander der Große ein Ende machte.

3. Culturverhältnisse der Iranier.

Unter den iranischen Stämmen war der persische der bedeutendste. Aus Persern bestand der Kern der Kriegsmacht, ihren Rat hörte der König an und sie allein waren von Steuern frei. Der König war der oberste Richter und Gesetzgeber; der Masse des Volkes unnahbar, thronte er in seiner Residenz zu Susa, von den 10.000 Unsterblichen bewacht. (Sie wurden so genannt, weil ihre Zahl immer die gleiche blieb, da jede Lücke gleich ergänzt wurde). Das Land war in Satrapien eingeteilt, deren Zahl je nach der Zeit verschieden und deren Verwaltung sehr einfach war.

Die Machtfülle der Satrapen war eine große, sie wurden jedoch von Spähern bewacht und schon der Verdacht eines Abfalls wurde grausam bestraft. In Armenien und Pontus blieben die Könige als Vasallen des Großkönigs bestehen; auch in den anderen Ländern ließ man die religiösen und nationalen Eigentümlichkeiten unangetastet.

— 19 —

bis zum Jahre 330 v. Ch., in welchem ihm der macedonische König Alexander der Große ein Ende machte.

3. Culturverhältnisse der Iranier.

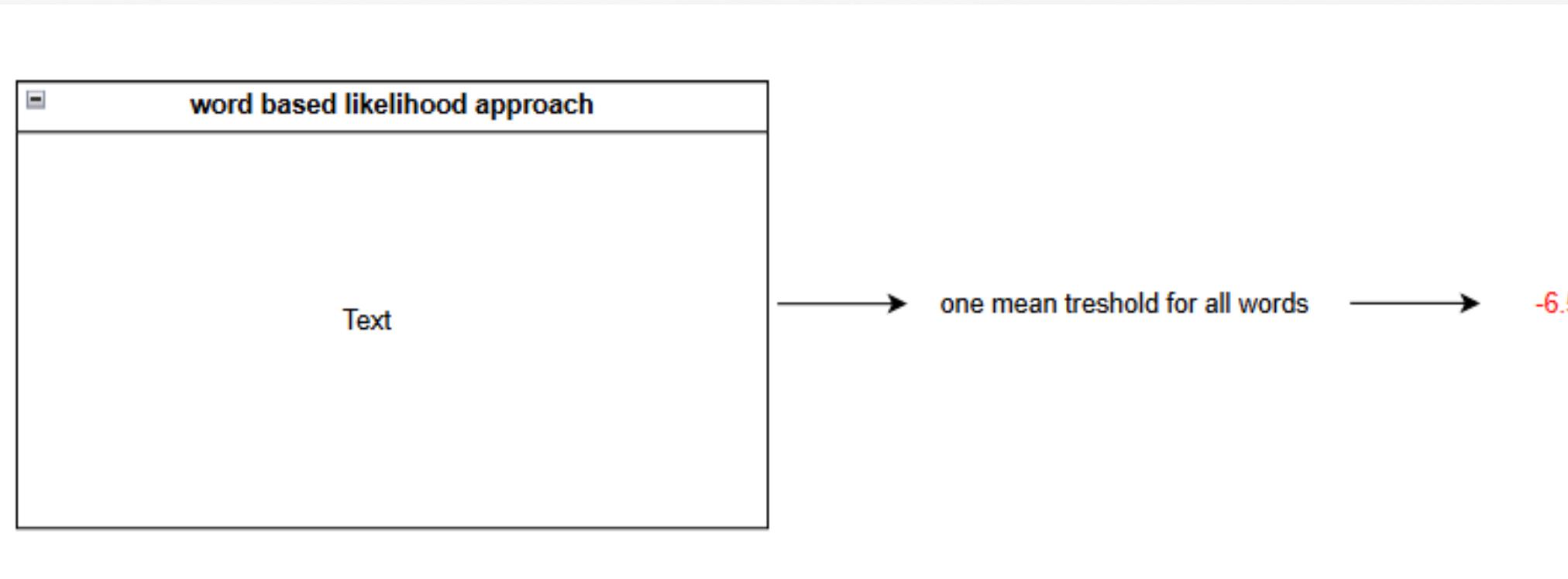
Unter den iranischen Stämmen war der persische der bedeutendste. Aus Persern bestand der Kern der Kriegsmacht, ihren Rat hörte der König an und sie allein waren von Steuern frei. Der König war der oberste Richter und Gesetzgeber; der Masse des Volkes unnahbar, thronte er in seiner Residenz zu Susa, von den 10.000 Unsterblichen bewacht. (Sie wurden so genannt, weil ihre Zahl immer die gleiche blieb, da jede Lücke gleich ergänzt wurde). Das Land war in Satrapien eingeteilt, deren Zahl je nach der Zeit verschieden und deren Verwaltung sehr einfach war.

Die Machtfülle der Satrapen war eine große, sie wurden jedoch von Spähern bewacht und schon der Verdacht eines Abfalls wurde grausam bestraft. In Armenien und Pontus blieben die Könige als Vasallen des Großkönigs bestehen; auch in den anderen Ländern ließ man die religiösen und nationalen Eigentümlichkeiten unangetastet.

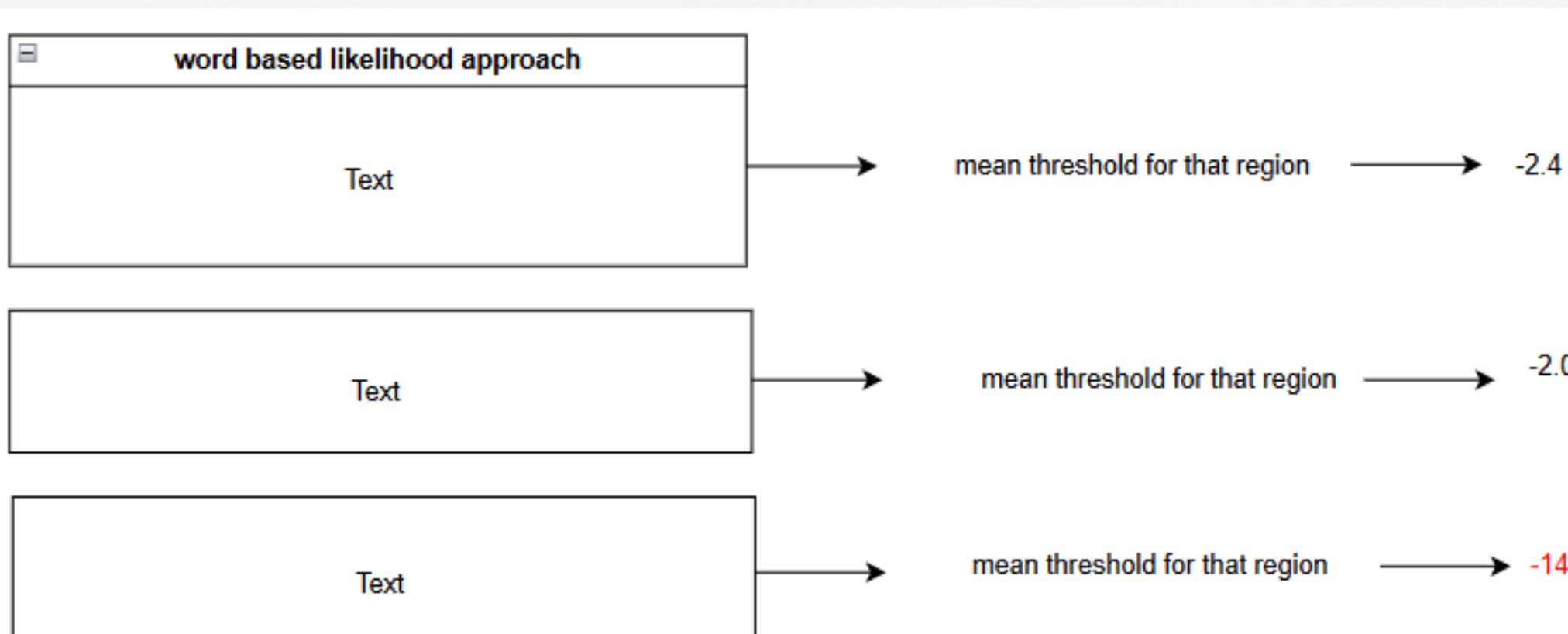
Likelihood Analysis

Presented by: Kerem Keptig

Word-Based Likelihood Approach & Splitting into Paragraphs



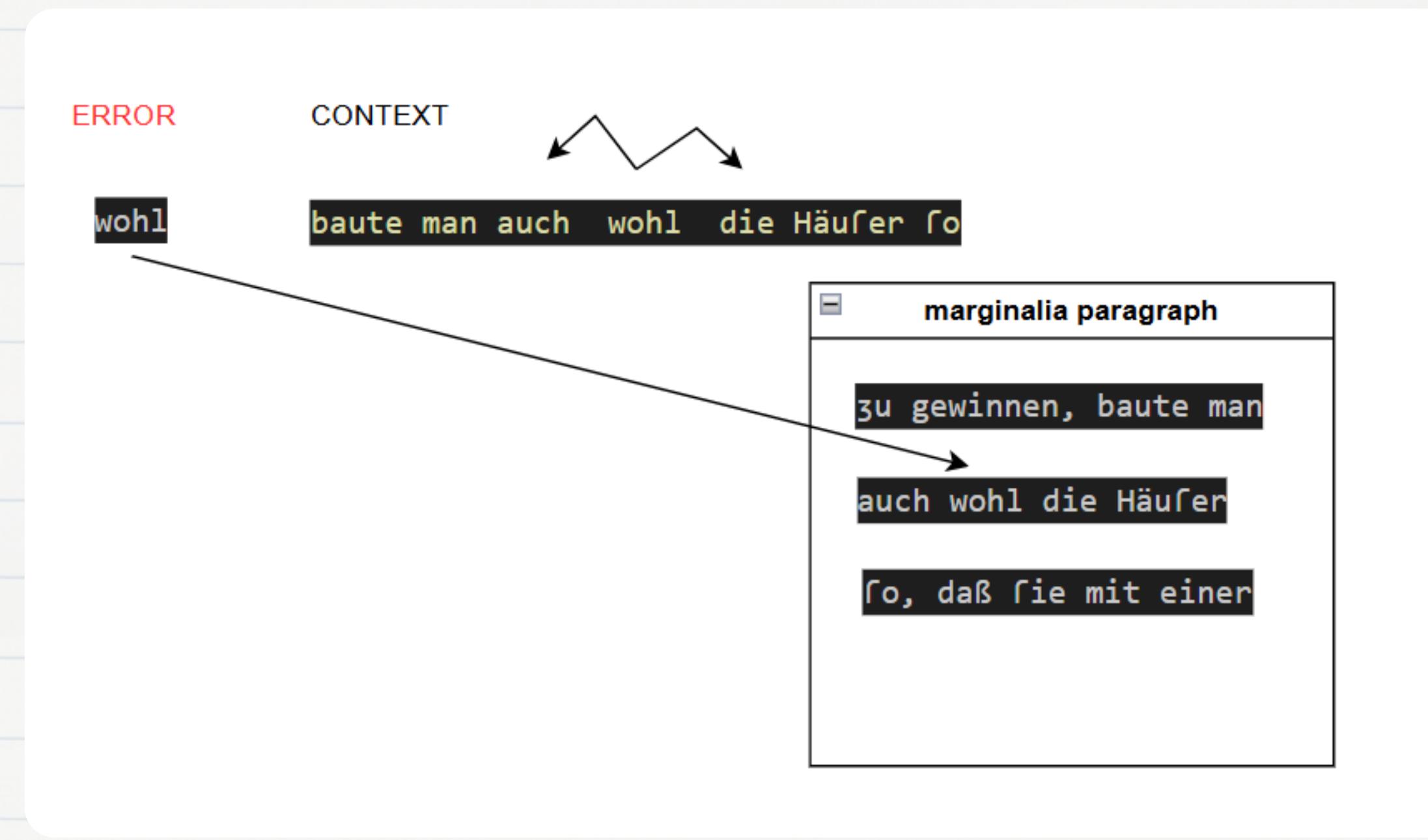
- I have one global threshold for whole page



- For each paragraph I have new thresholds based on mean likelihood

currently it does not calculate for each paragraph

Error Label Checking



- First, I check each paragraph in the Marginalia dataset for the target word.
- If the word is found, I check its surrounding words (previous and next).
- If one of the surrounding words matches the expected context, I consider this a correct placement and label it as a false positive (FP).

Error Label Checking

ERROR

wohl



"marginalia", "footnote", "page-number", "header", "decoration", "other"

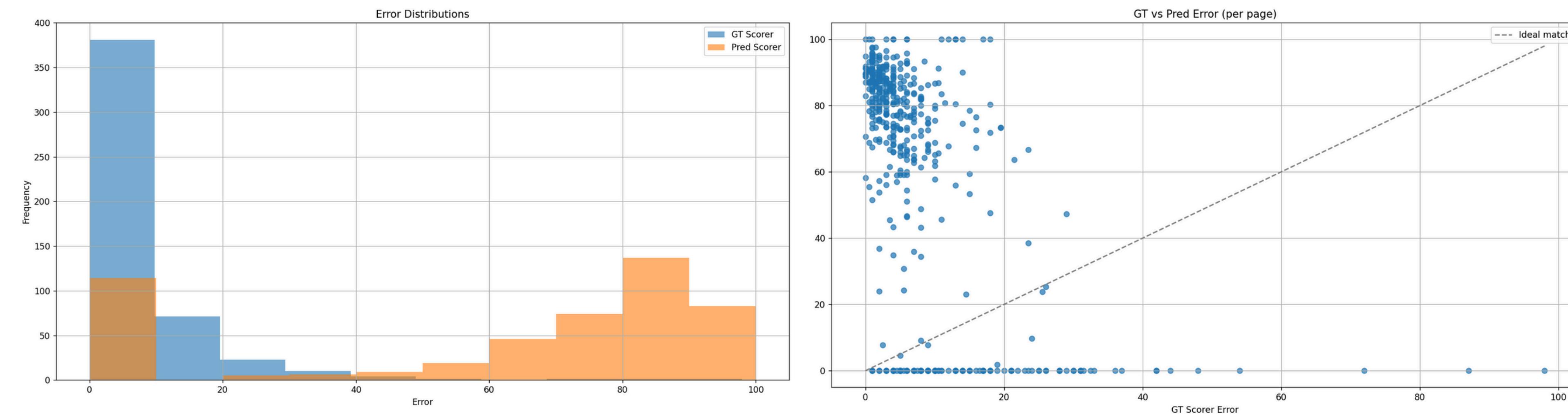
- If the surrounding words do not match, I then check if the word exists in the error labels.
- If the word appears in the error labels, I add it without checking the surrounding words because in some cases, the context may vary due to the error appearing in a mixed format.

GT Scorer vs Prediction Scorer

- Each page is assigned a single error score. This score is calculated by aggregating the quality of word-level matches between the predicted errors and the ground-truth marginalia.

Metric	GT Scorer	Pred Scorer
Count	493	493.0000
Mean	7.6288	60.7703
Median	4.5	76.9231
Stdev	10.1078	35.5901
Min	0	0.0000
Max	98	100.0000
MAE	7.6288	60.7703
Correlation	-0.5316	-
GT vs Pred RMSE	67.6146	-

Evaluation



Syntactic OCR Error Detection

Presented by: Arslan Ahmed

Which Unit of Analysis Should We Use?

OCR Error Detection can be:

- Line-based
- Sentence-based
- Overlay-window based

For syntactic error detection, sentence-based is more appropriate because syntax rules depend on full sentence context (e.g., verb structure)

Next, how to split the paragraphs into sentences?

Model-Based Splitting

Used spaCy's German model (de_core_news_sm)

Issues:

- Breaks on abbreviations like "Fig.", "z.B."
- Breaks incorrectly on semicolon (;)
- Struggles with historical character forms and OCR artifacts

Common abbreviations: "z.B.", "u.a.", "u.U.", "d.h.", "ggf.", "bzw.", "etc.", "sog.", "ca.", "Nr.", "Dr.", "Prof.", "bspw.", "vgl.", "Fig.", "Abb.", "Hr.", "Fr.", "bspw.", "evtl.", "ggf.", "bsp.", "i.d.R.", "Jan.", "Feb.", "Mrz.", "Apr.", "Mai.", "Jun.", "Jul.", "Aug.", "Sep.", "Okt.", "Nov.", "Dez.", "Jh.", "St.", "Abs.", "S.", "Tab.", "Bd.", "Hrsg.", "Kap.", "Lit.", "s.", "Anm.", "bspw.", "dgl.", "u.Ä.", "usw.", "z.T.", "z.Z.", "zzgl.".

Rule-Based Sentence Splitting – Our Approach

Abbreviation protection: "Fig. 8. Darius mit..." stays together

Historical char normalization: 'f' → 's', 'ʒ' → 'z'

Plätʒe klein, die Gaffʃen... → Plätze klein, die Gassen...

Hyphenation fix at line breaks:

gez wöhnlich → gewöhnlich

gekom- men → gekommen

Rule-Based Sentence Splitting – Example

Page: '1724666584_00000024'

Text: Histiäus wurde belohnt, Miltiades floh in seine griechische Heimat. Fig. 8. Darius mit Dienern (über ihm Ormuzd). Wandbild im Palast zu Persepolis. 16 Rückzug der Skythen. Rückzug der Perser. Die Griechen bei der Brücke.

Sentences extracted

'Histiäus wurde belohnt, Miltiades floh in seine griechische Heimat.'

'Fig. 8. Darius mit Dienern (über ihm Ormuzd).'

'Wandbild im Palast zu Persepolis. 16 Rückzug der Skythen.'

'Rückzug der Perser.'

'Die Griechen bei der Brücke.'

Syntactic Error Detection

Two Signals Used for Detecting Syntactic Errors

1. Finite Verb Presence

- We expect every valid German sentence to contain at least one finite verb

Example:

- "Die Reste der großen Armee." → No finite verb → ✗ Page: 173736378X_00000120.xml

2. NER Density (Named Entity Density)

- Calculated as: *number of named entities / total tokens*
- High density → possible sentence segmentation error
- NER density > 0.3

Example:

- "Flucht Napoleons." → NER density too high → ✗

Page: 173736378X_00000120.xml

Evaluation

Each page assigned an error score based on sentence quality

Ground truth scorer + prediction scorer comparison

Metric	GT Scorer	Pred Scorer
Count	493	493.0000
Mean	7.6288	63.3898
Median	4.5	61.5385
Stdev	10.1078	26.1683
Min	0	0.0000
Max	98	100.0000
MAE	7.6288	63.3898
Correlation	0.3595	-
GT vs Pred RMSE	60.8676	-

Evaluation

