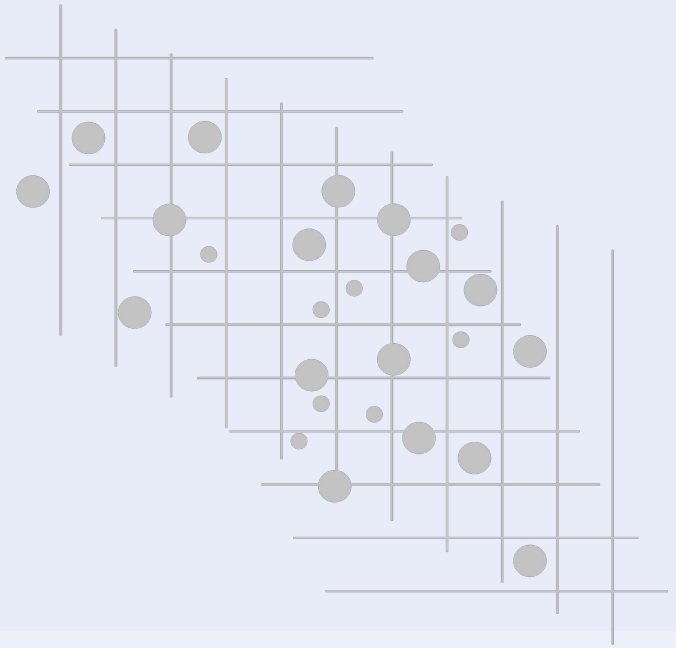
Bauhaus Universität Weimar
Faculty of Media
Degree Programme: M.Sc. Digital Engineering

COHERENCE BASED TEXT QUALITY IN SEARCH SUPPORTED WRITING

MASTER'S THESIS DEFENCE



1 Referee: Prof. Dr. Benno Stein

2 Referee: Prof. Dr. Ing. Volker Rodehorst

Supervised By:

Dr. Michael Völske

Dr. Magdalena Wolska

Presenter:

Bibek khadayat

Date:

OUTLINE

INTRODUCTION

- Definition and different factors
- Area of Impacts
- Search support writing
- Research Questions

CONCLUSION

- Conclusions
- Future Works

DATA AND DATASETS

- About Datasets
- Data Acquisition Setup
- Data Collected
- Prior Analysis

RESULTS AND ANALYSIS

- Results for research questions 1
- Results for research questions 2
- Results for research questions 3

be more precise: better

DATA PREPROCESSING

- Data Extraction
- Data cleaning
- Identification of major changes
- Data Analysis

EDITING TYPES

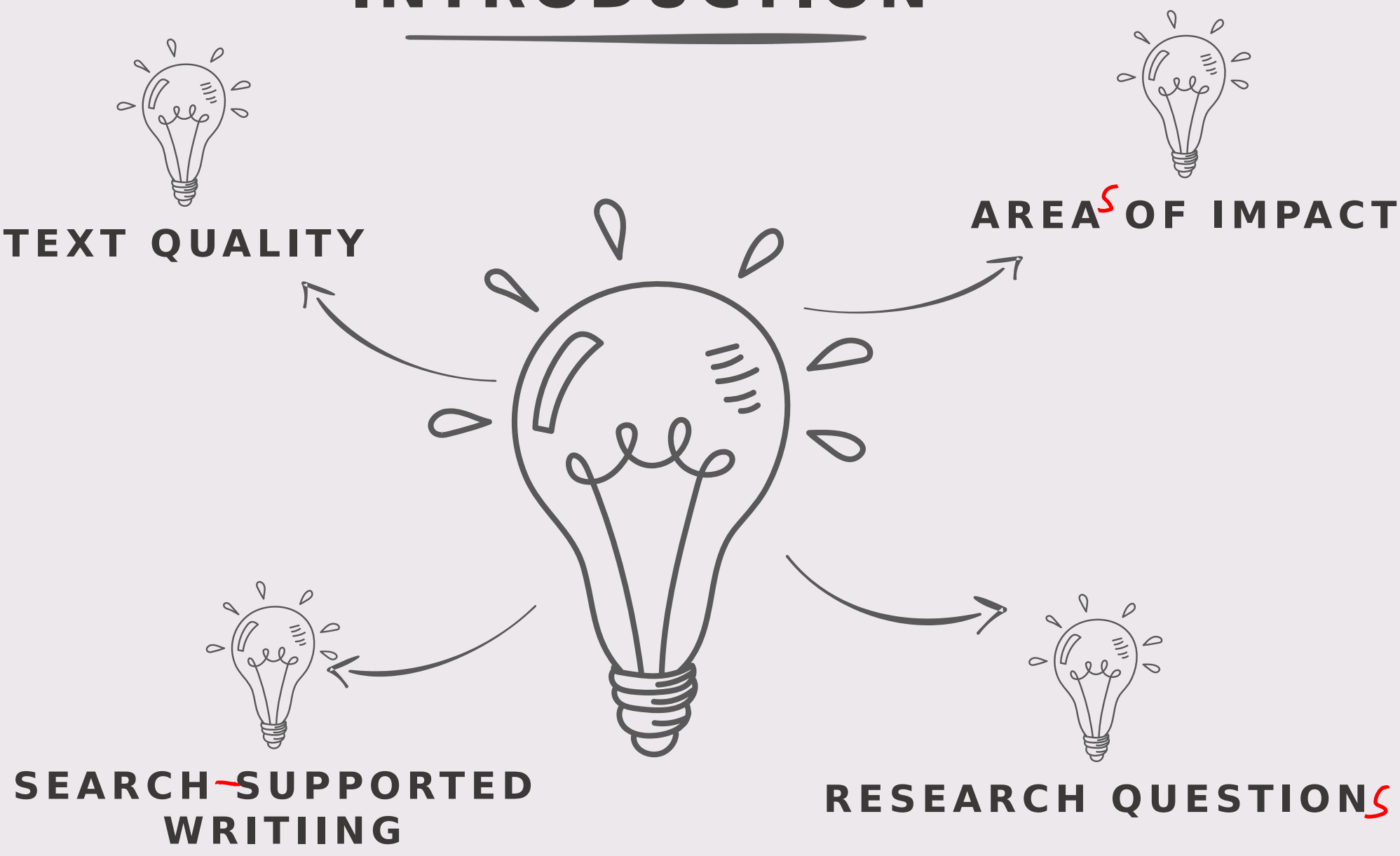
- About Editing Types
- Example of editing types.
- Automation of editing type

Better: automated detection

DIFFERENT TEXT QUALITY MEASURES

- Methodology for Coherence
- Methodology for Type-Token Ratio
- Methodology for Readability

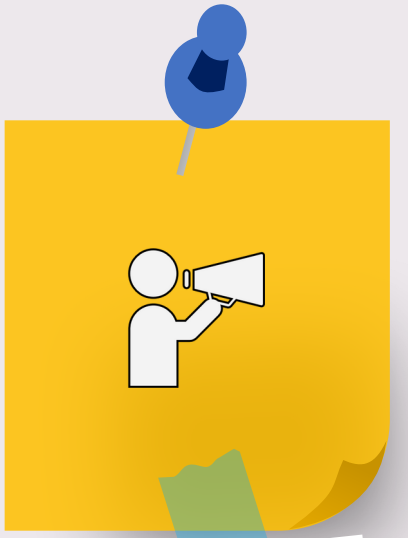
INTRODUCTION



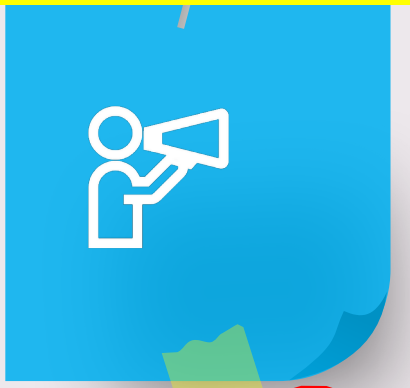
INTRODUCTION

TEXT QUALITY

I'd suggest to remove these text fields -- to much to read during the talk.You don't need to define what text is.Y



Text is used in linguistics to refer to any passage, spoken, or written, of whatever length, that does form a unified whole.



Quality text is the text which is easily understandable by everyone.

Text quality is judge on the basis of spelling, vocabulary, grammar, structure, organization, readability, and so on.

INTRODUCTION

TEXT QUALITY

- COHERENCE
- COHESION
- READABILITY

INTRODUCTION

TEXT QUALITY

Coherence	Cohesion
Nepal is a small and beautiful country. Many tourists visit Nepal because of its beauty and nature. White mountains, green forests, wildlife reserves etc., lure people from different countries.	**Nepal** is a small country. **Nepal** is located in between India and China. **Its** area is 147516 sq. kilometers.

INTRODUCTION

AREAS OF IMPACT

01 Web search recommendation

02 Automatic Summarization

03 Word Predictio n

04 Writing Assessment

INTRODUCTION

SEARCH SUPPORTED WRITING

(01) and (02) are the same thing -- I'd suggest to mention this only once.(m

01 Use of Internet for research purpose.

?

02 Gather information and data with the help of search engine.

03 Use these information in writing

INTRODUCTION

RESEARCH QUESTIONs

RQ1 What are the different types of editing techniques in search-supported writing ?

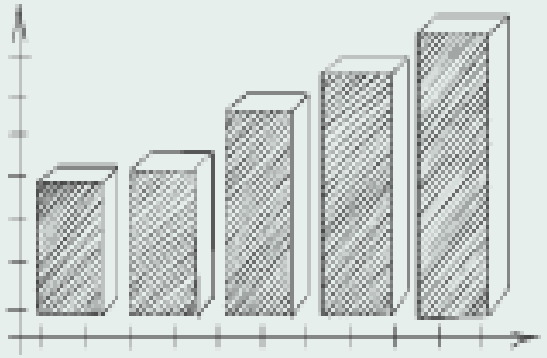
RQ2 How do different types of editing affect Coherence, Type-Token ratio, and Readability?

RQ3 How do different essay writing strategies affect Coherence, Type-Token ratio, and Readability?

DATA AND DATASETS

DESCRIPTION OF DATASETS

DATA ACQUISITION SETUP



PRIOR ANALYSIS

COLLECTED DATA

DATA AND DATASETS

DESCRIPTION OF DATASETS

Text Reuse Corpus 2012 (Webis-TRC-12)

150 long essays written by ~~the~~ professional writer~~s~~

Information ~~are used~~ from ClueWeb09 corpus

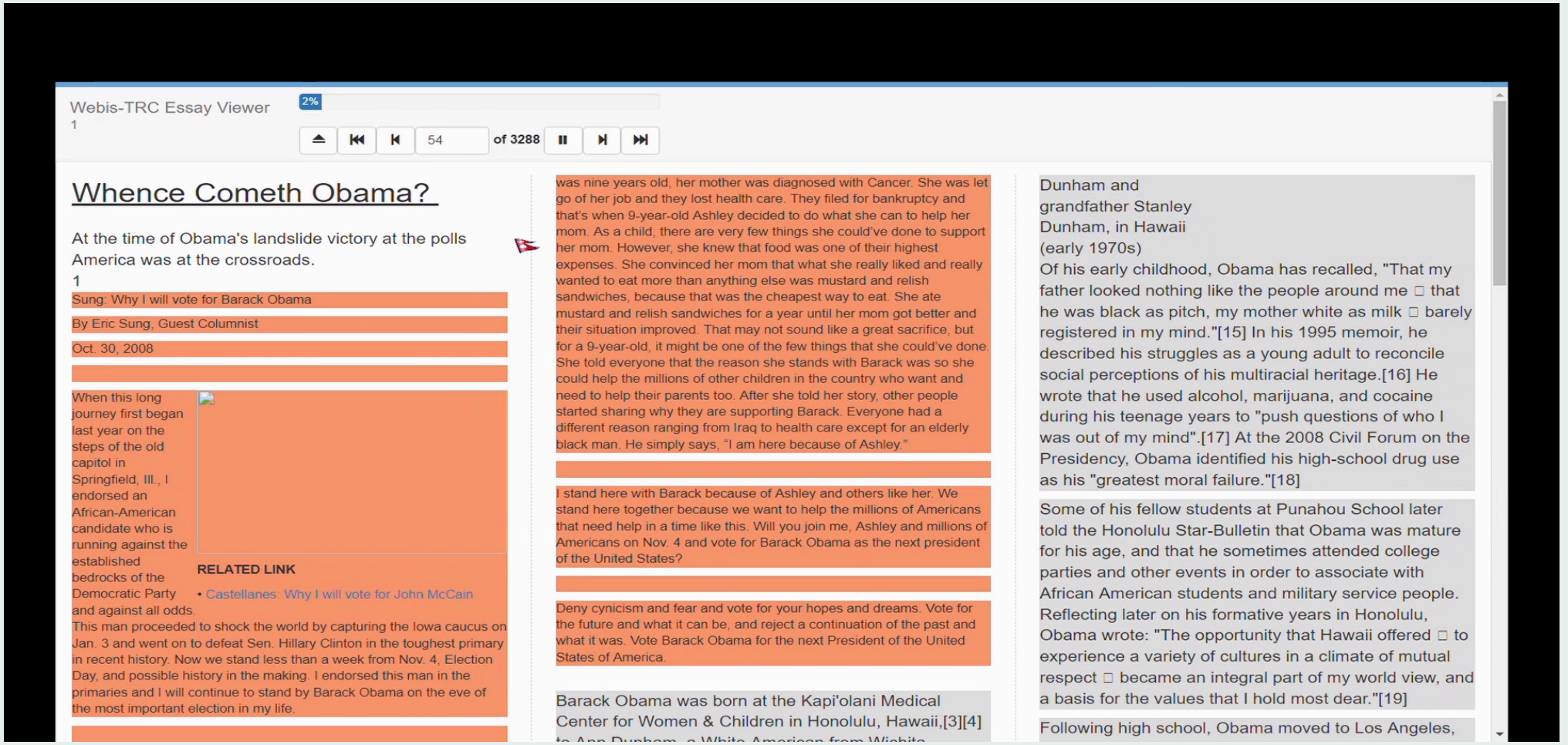
~~Contain~~ detailed interaction logs ~~file~~

+ Revision history!

~~Dataset is~~ available ~~in~~ on webis website ~~and can be explore how it is created in web visualization tool~~.

DATA AND DATASETS

DESCRIPTION OF DATASETS



DATA AND DATASETS

DATA ACQUISITION SETUP

Editor, Search Engine, and Datasets were provided to writer.

Dataset used: a set of topics and a set of web pages to search.

ChatNoir is used as search engine

DATA AND DATASETS

COLLECTED DATA

Writer or Author create different revisions during writing process in each essays or topics.

Essay

Final revisions of most of the essays are around 5000 words.

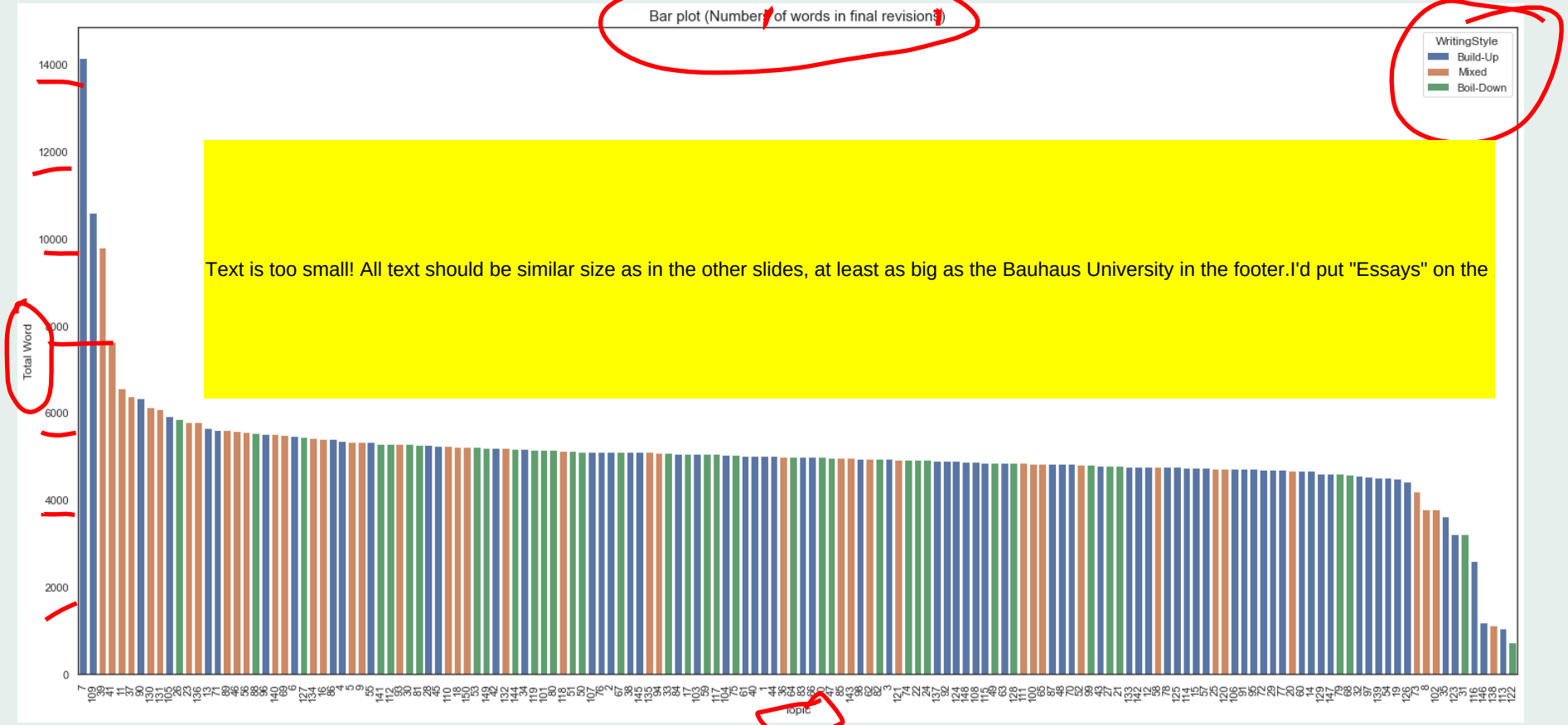
150 long essays were created with numbers of revisions in each essays by 12 professional writers.

you already said this on a previous sli

Instead, you could put here a recap/summary table likeEssays: 15

DATA AND DATASETS

COLLECTED DATA



DATA AND DATASETS

LEGEND

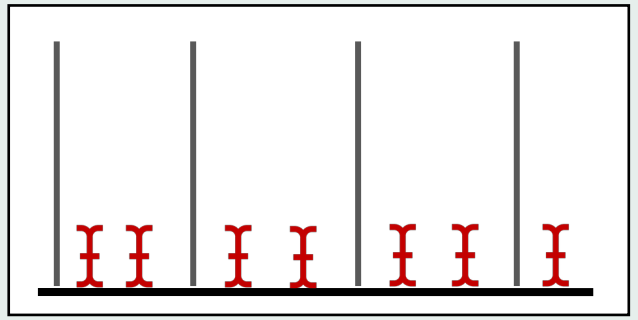
Research and copy paste

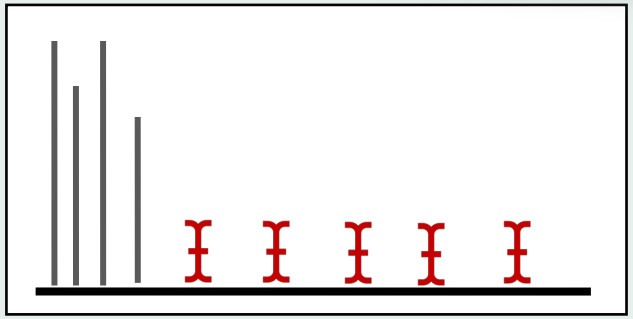
Edit

PRIOR ANALYSIS

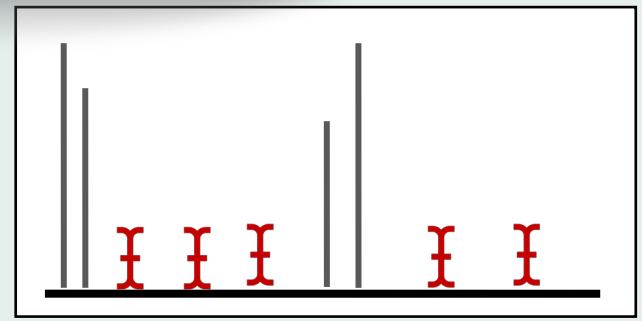
Writing Style

Build-up is a fashion, in which continuous lengthening of the essay in a regular period of time over the whole period of writing is done.



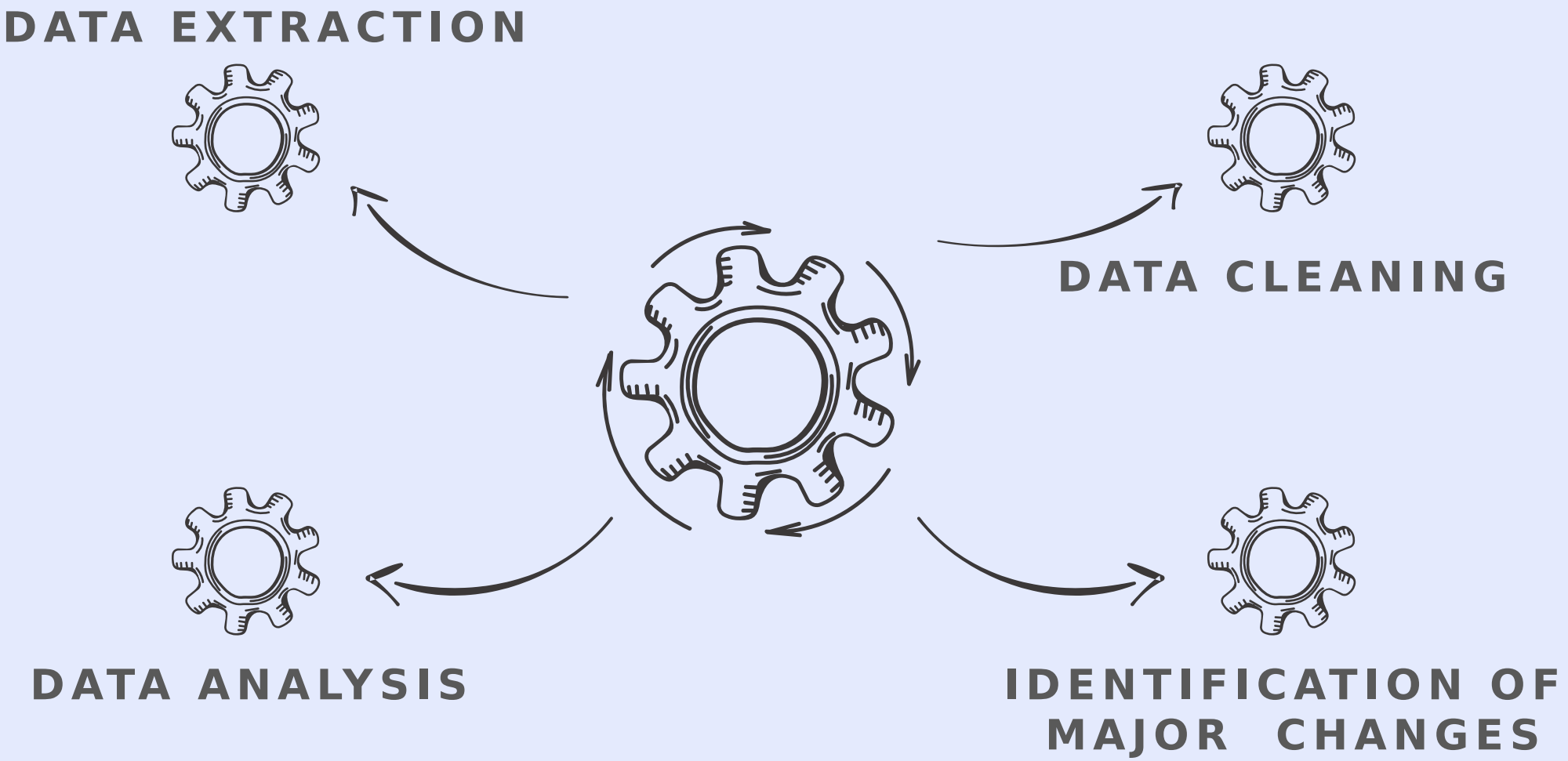


Boil-down is a fashion in which first quick length growth and then shorting happen.



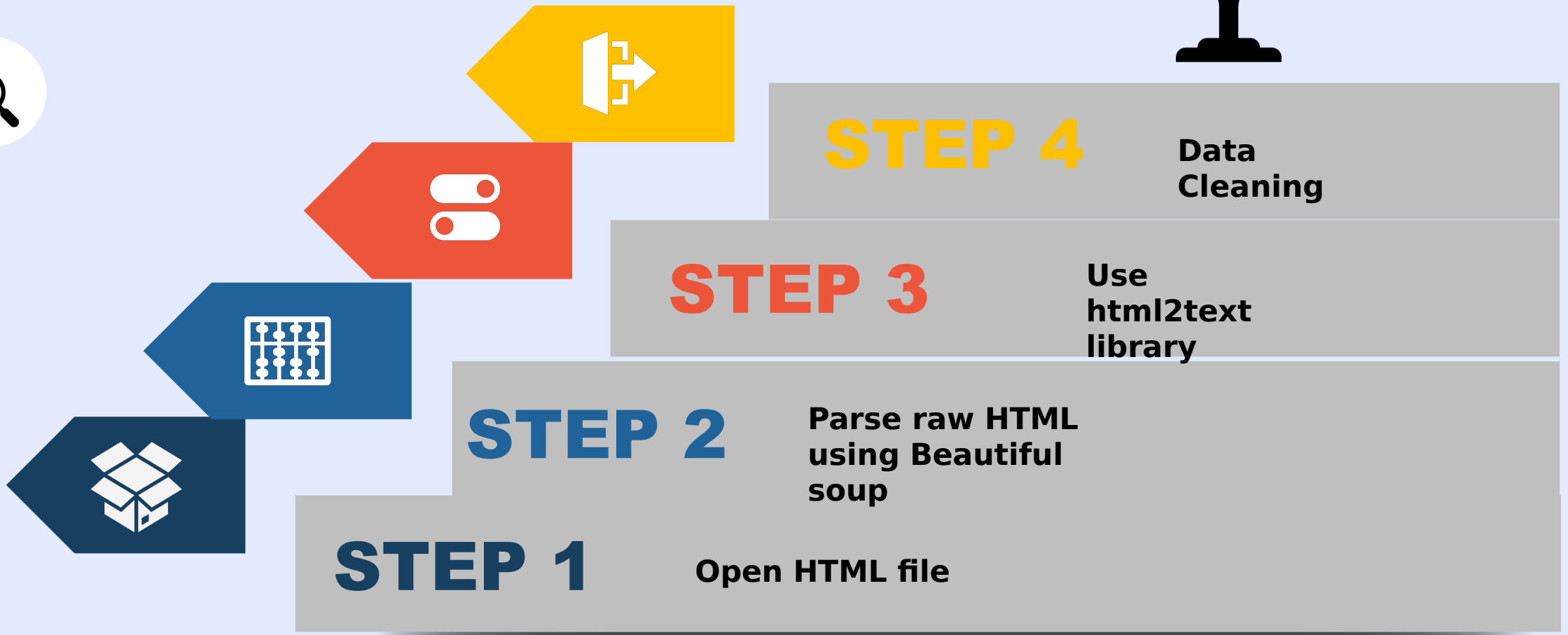
Mixed writing style have included both Build-Up and Boil-down.

DATA PREPROCESSING



DATA PREPROCESSING

DATA EXTRACTION AND DATA CLEANING



DATA PREPROCESSING

IDENTIFICATION OF MAJOR CHANGES

01 Too much of data to visualize

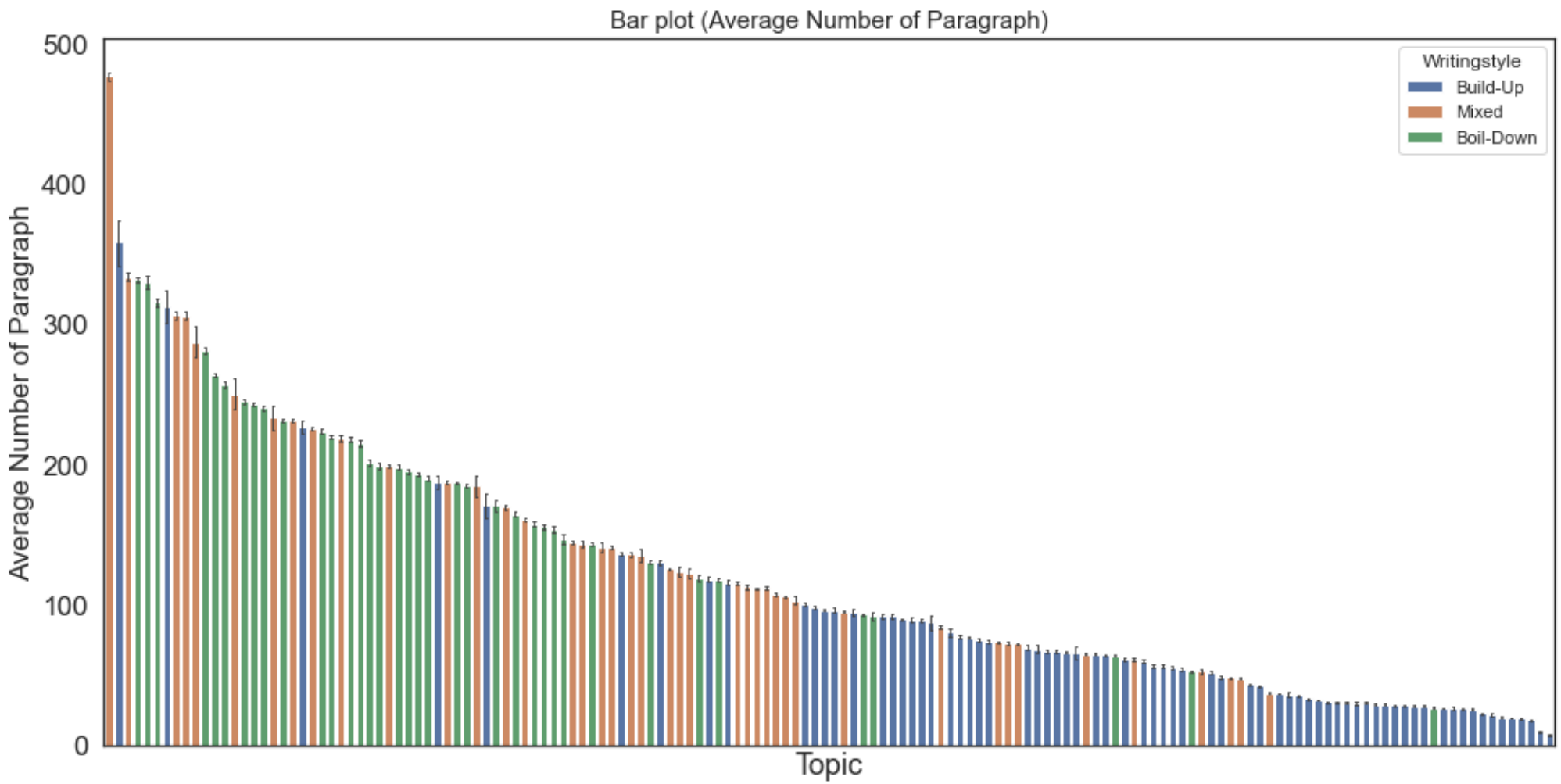
02 Check coherence score across adjacent revision

03 Select only major change

DATA PREPROCESSING

DATA ANALYSIS

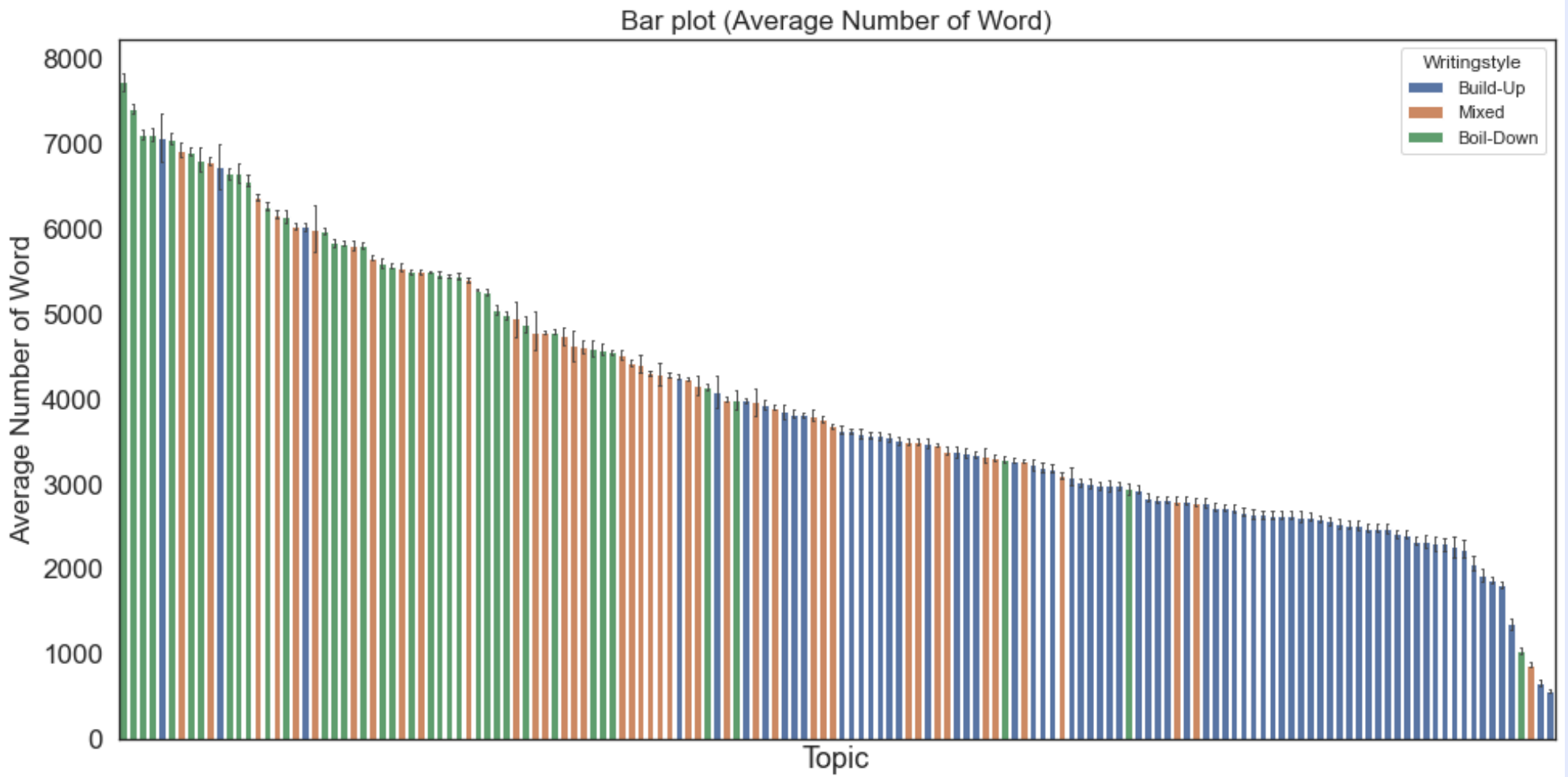
Average number of paragraphs.



DATA PREPROCESSING

DATA ANALYSIS

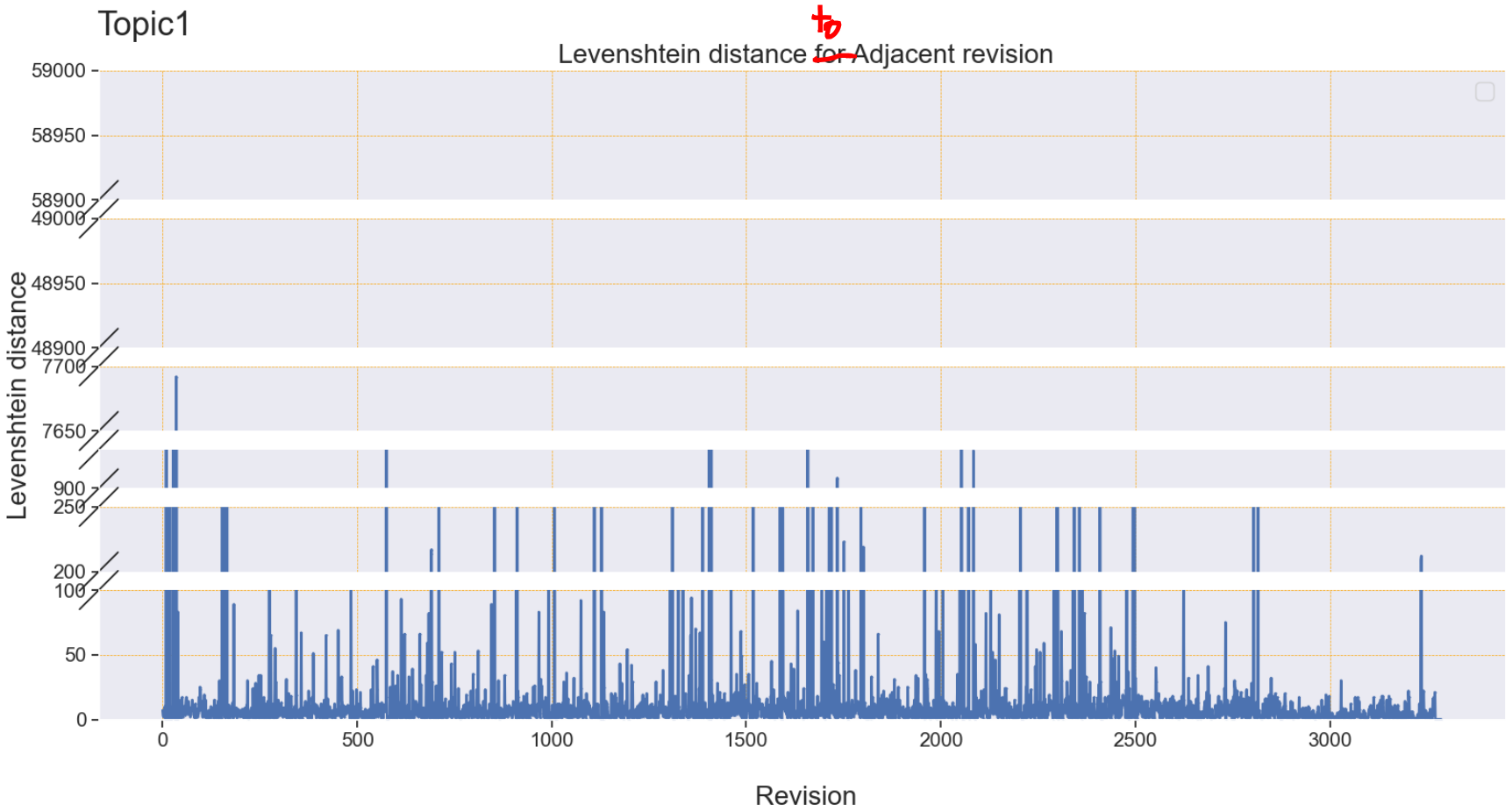
Average number of words.



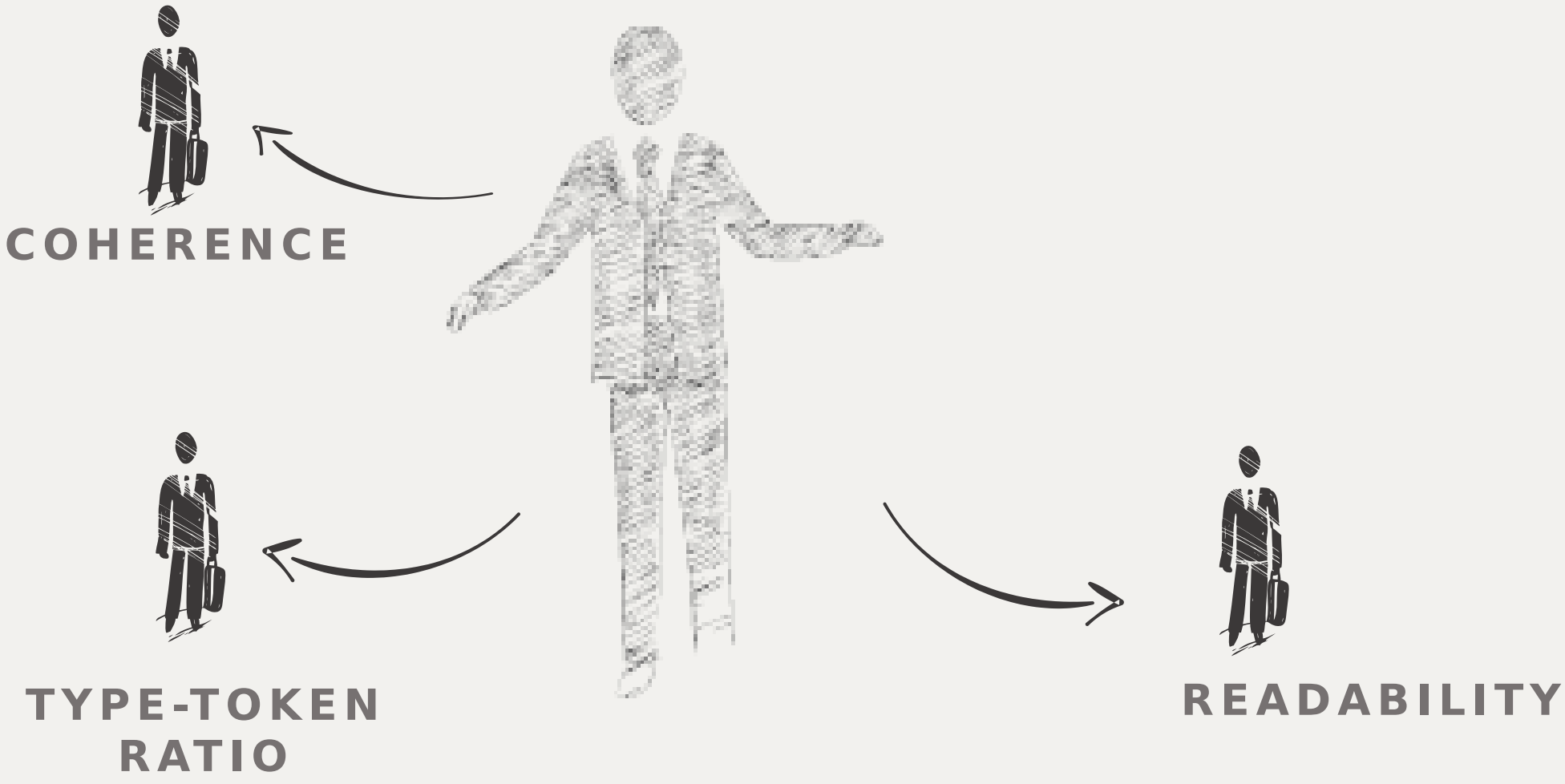
DATA PREPROCESSING

DATA ANALYSIS

The Levenshtein distance.



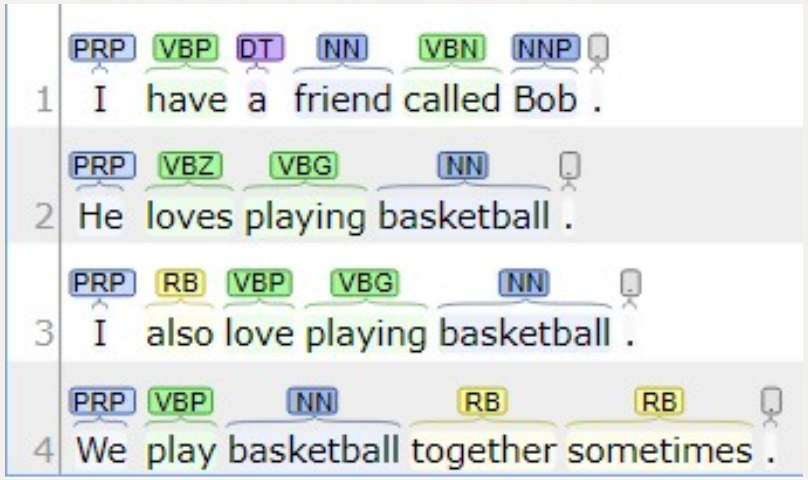
TEXT QUALITY MEASURES



TEXT QUALITY MEASURES

COHERENCE (Modeling Local Coherence: An Entity-based Approach; Regina Barzilay, Mirella Lapata)

For example we take "I have a friend called Bob. He loves playing basketball. I also love playing basketball. We play basketball together sometimes."



Index	I	friend	Bob	basketball	we
0	s	s	o	–	–
1	–	s	–	o	–
2	s	–	–	o	–
3	–	–	–	o	s

TEXT QUALITY MEASURES

COHERENCE (ENTITY GRID)

Index	I	friend	Bob	basketball	we
0	('S', '-')	('S', 'S')	('O', '-')	('-', 'O')	('-', '-')
1	('-', 'S')	('S', '-')	('-', '-')	('O', 'O')	('-', '-')
2	('S', '-')	('-', '-')	('-', '-')	('O', 'O')	('-', 'S')

Index	SS	SO	SX	S-	OS	OO	OX	O-	XS	XO	XX	X-	-S	-O	-X	--
0	0.0666667	0	0	0.2	0	0.133333	0	0.0666667	0	0	0	0	0.133333	0.0666667	0	0.333333

TEXT QUALITY MEASURES

READABILITY

- Flesch reading Ease Formula.
  206.835 − (1.015 × ASL) − (84.6 × ASW)
- Textstat python library.

Data: List of clean Text
Result: Readability_Score

```
Text ← [cleaned_text];        /* [cleaned_text] is lists of
 preprocessed paragraphs [p1,p2,p3,..] where P1,P2,p3
 are paragraph */
for i = 0 to len(Text) do
    text = Text[i];
    /* Check empty element to avoid zero exception     */
    if len(text) > 0 then
        Readability_Score ← flesch_reading_ease(text) ;
    else
        Readability_Score ← 0 ;
    end
end
```

TEXT QUALITY MEASURES

TYPE-TOKEN RATIO

```
at the time of obamas landslide victory at the poll america was at the
crossroad


Counter({'at': 3, 'the': 3, 'time': 1, 'of': 1, 'obamas': 1, 'landslide':
1, 'victory': 1, 'poll': 1, 'america': 1, 'was': 1, 'crossroad': 1})
```

I'd remove the Counter line, it's redundant with the table below

	Word	Frequency
0	at	3
1	the	3
2	time	1
3	of	1
4	obamas	1
5	landslide	1
6	victory	1
7	poll	1
8	america	1
9	was	1
10	crossroad	1

```
Type=11---Total Token=15


TTR Score=0.7333333333333333
```

so many significant digits aren't necessary

RESEARCH QUESTION RQ1 SETUP

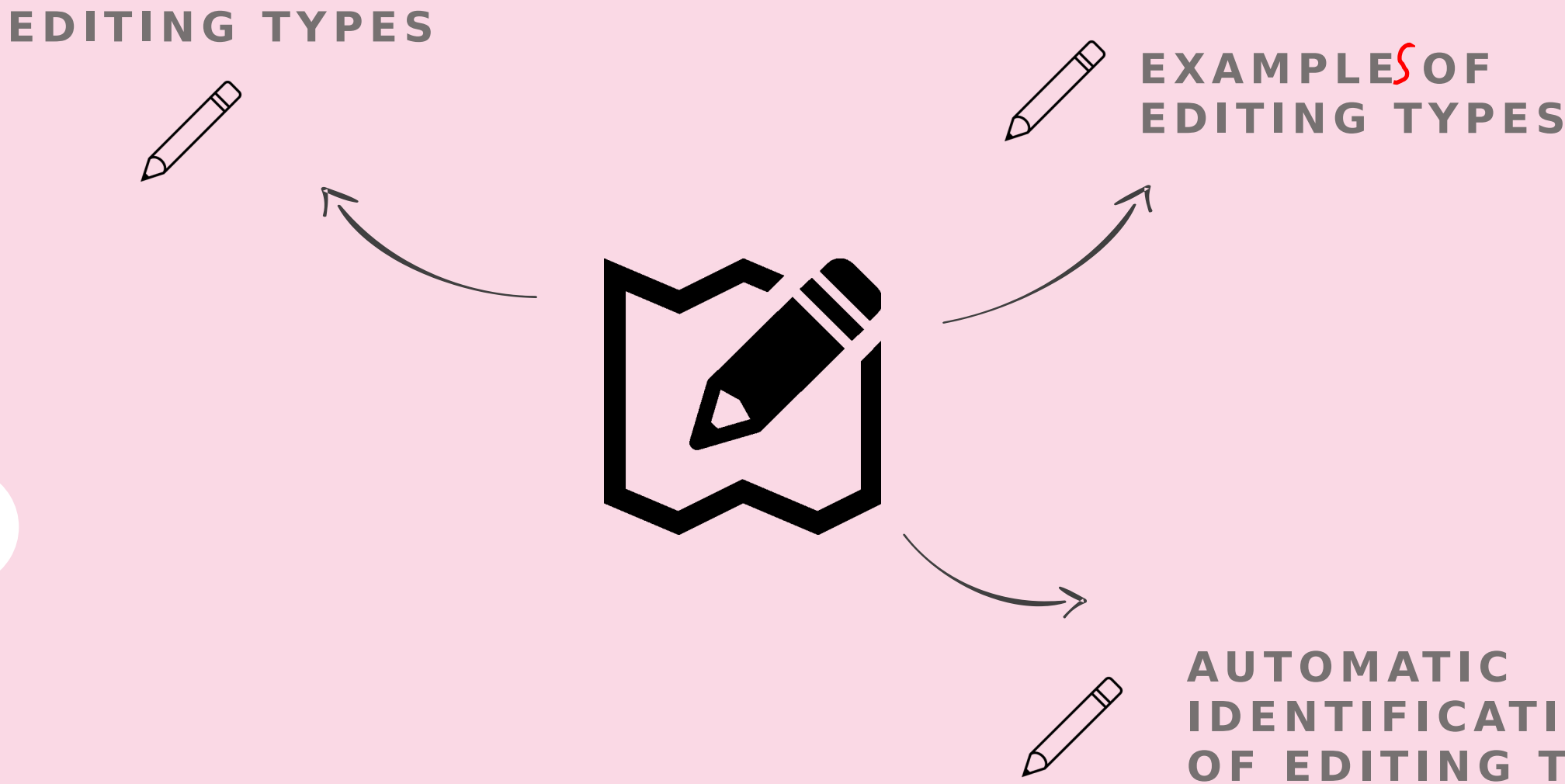
RQ1 What are the different types of editing techniques in search-supported writing ?

RQ2 How do different types of editing affect Coherence, Type-Token ratio, and Readability?

On these slides, can you make the circle around the non-active RQs gray instead of colored? (I.e. here, only RQ1 would be green, RQ2 and RQ3 would be gray)

RQ3 How do different essay writing strategies affect Coherence, Type-Token ratio, and Readability?

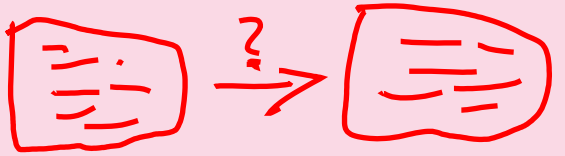
EDITING TYPES



EDITING TYPES

Change the ordering:- analyse pairs of adjacent revisions

EDITING TYPES



maybe add figure with two actual adjacent revisions

01 ~~Editing types mean~~ how the second revision ~~is~~ is edited compared to ~~first~~ the revision

give the number instead of "different"

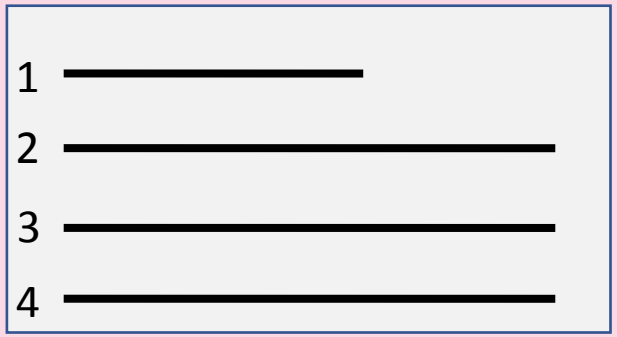
02 ~~During analysing dataset~~ we identify ~~different~~ editing types

03 We analyse across pairs of adjacent revisions

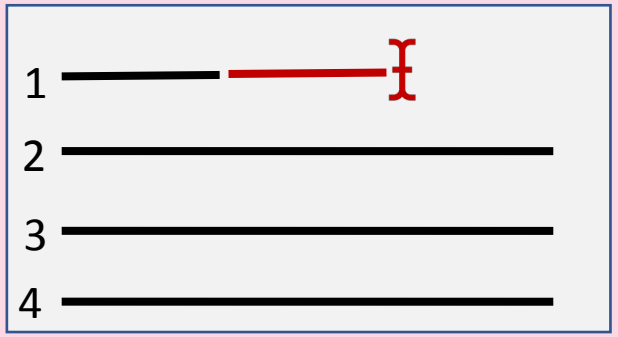
EDITING TYPES

EXAMPLE OF EDITING TYPES

Block Edit



Before



After

French Lick Resort and Casino

Oh Wow! Did you catch the latest news

French Lick Resort has embarked on an incredible $500 million historic restoration and expansion. In addition to the restoration of the French Lick Springs Hotel, including 443 fully renovated guest rooms, a new event center, exciting retail shops, and fully restored public areas, we are proud to announce the addition of the French Lick Casino, returning gaming to the Springs Valley for the first time since 1949. And, with the reopening of the luxurious West Baden Springs Hotel, French Lick has truly created the Midwest's premiere resort and casino destination.

French Lick Resort and Casino

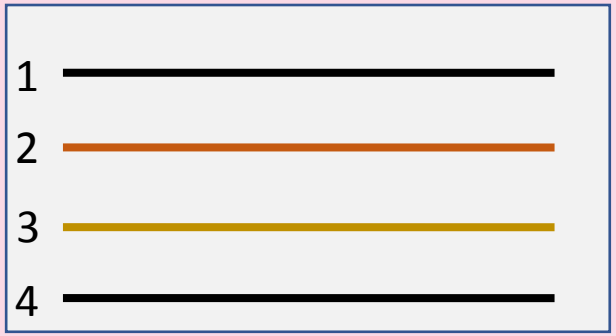
Oh Wow! Did you catch the latest news from Indiana? The State with the motto "The Crossroads of America" is not just a great place to watch motor races like the Indy 500, it's also a great place

French Lick Resort has embarked on an incredible $500 million historic restoration and expansion. In addition to the restoration of the French Lick Springs Hotel, including 443 fully renovated guest rooms, a new event center, exciting retail shops, and fully restored public areas, we are proud to announce the addition of the French Lick Casino, returning gaming to the Springs Valley for the first time since 1949. And, with the reopening of the luxurious West Baden Springs Hotel, French Lick has truly created the Midwest's premiere resort and casino destination.

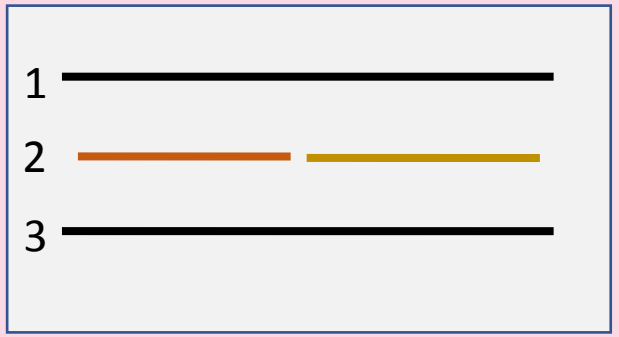
EDITING TYPES

EXAMPLE OF EDITING TYPES

Block Merged



Before



After

Right now (drum roll please) ... Places like Circle City Escorts have every variety of independent escort you could dream of in Indiana. Indiana Escort Referrals recommends Naughtynightlife.com - your free guide to independent escorts, escort agencies and erotic madame and monsieur masseurs. Fancy a blond escort for the night? Escort Service in Circle City can provide the companion of your dreams. Feel like a taste of your own favorite fetish? Heaven 'n Heels everywhere in Indiana has a directory of the most elegant, beautiful and erotic Indiana independent escorts that belong in paradise.

What a fantastic place to locate a Casino. No wonder the designers chose to add a touch of wonderland to the magic state when they built French Lick resort and Casino on

8670 W. State Road 56
French Lick, IN 47432 (Map it)

French Lick Resort has embarked on an incredible $500 million historic restoration and expansion. In addition to the restoration of the French Lick Springs Hotel, including 443 fully renovated guest rooms, a new event center, exciting retail shops, and fully restored public areas,

Right now (drum roll please) ... Places like Circle City Escorts have every variety of independent escort you could dream of in Indiana. Indiana Escort Referrals recommends Naughtynightlife.com - your free guide to independent escorts, escort agencies and erotic madame and monsieur masseurs. Fancy a blond escort for the night? Escort Service in Circle City can provide the companion of your dreams. Feel like a taste of your own favorite fetish? Heaven 'n Heels everywhere in Indiana has a directory of the most elegant, beautiful and erotic Indiana independent escorts that belong in paradise.

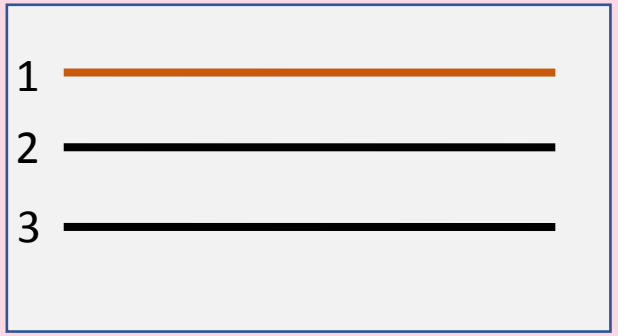
What a fantastic place to locate a Casino. No wonder the designers chose to add a touch of wonderland to the magic state when they built French Lick resort and Casino on 8670 W. State Road 56
French Lick, IN 47432 (Map it)

French Lick Resort has embarked on an incredible $500 million historic restoration and expansion. In addition to the restoration of the French Lick Springs Hotel, including 443 fully renovated guest rooms, a new event center, exciting retail shops, and fully restored public areas,

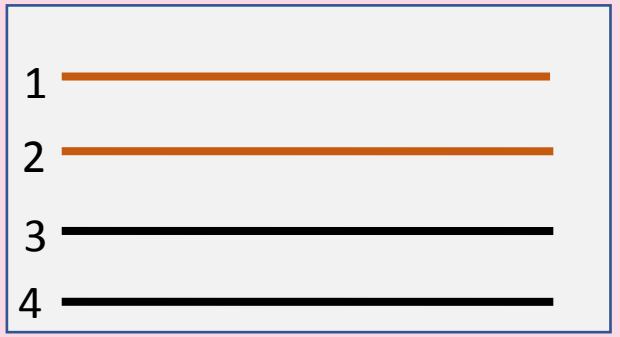
EDITING TYPES

EXAMPLE OF EDITING TYPES

Block Split



Before



After

French Lick Resort and Casino

Oh Wow! Did you catch the latest news French Lick Resort has embarked on an incredible $500 million historic restoration and expansion. In addition to the restoration of the French Lick Springs Hotel, including 443 fully renovated guest rooms, a new event center, exciting retail shops, and fully restored public areas, we are proud to announce the addition of the French Lick Casino, returning gaming to the Springs Valley for the first time since 1949. And, with the reopening of the luxurious West Baden Springs Hotel, French Lick has truly created the Midwest's premiere resort and casino destination.

For generations, our beautiful retreat has offered a scenic environment in which to relax and enjoy nature. Guests can stroll shaded walkways and visit the famous gazebo housing the Pluto mineral springs, nestled amidst lush gardens of colorful flowers and carefully trimmed greenery. The shaded walkways provide quiet solitude at mid-day or for an evening stroll. Our manicured grounds also provide an impeccable backdrop for all kinds of events, from weddings and corporate picnics, to family cookouts.

French Lick Resort and Casino

Oh Wow! Did you catch the latest news

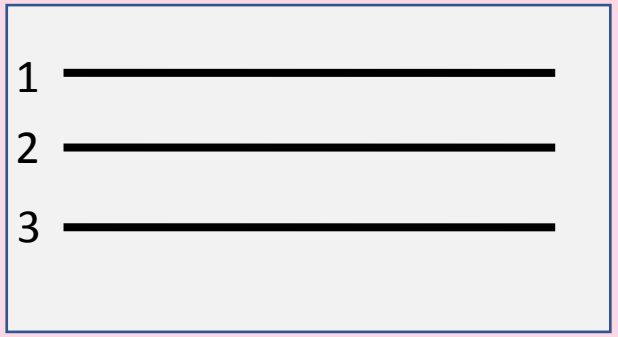
French Lick Resort has embarked on an incredible $500 million historic restoration and expansion. In addition to the restoration of the French Lick Springs Hotel, including 443 fully renovated guest rooms, a new event center, exciting retail shops, and fully restored public areas, we are proud to announce the addition of the French Lick Casino, returning gaming to the Springs Valley for the first time since 1949. And, with the reopening of the luxurious West Baden Springs Hotel, French Lick has truly created the Midwest's premiere resort and casino destination.

For generations, our beautiful retreat has offered a scenic environment in which to relax and enjoy nature. Guests can stroll shaded walkways and visit the famous gazebo housing the Pluto mineral springs, nestled amidst lush gardens of colorful flowers and carefully trimmed greenery. The shaded walkways provide quiet solitude at mid-day or for an evening stroll. Our manicured grounds also provide an impeccable backdrop for all kinds of events, from weddings and corporate picnics, to family cookouts.

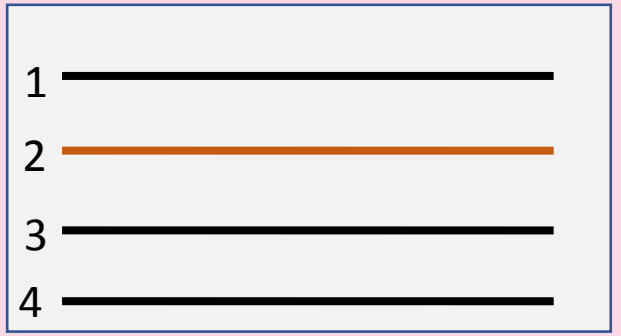
EDITING TYPES

EXAMPLE OF EDITING TYPES

Block Insertion



Before



After

French Lick Indiana got its name from early French settlers and the ineral licksin the area. French traders came to the area and discovered the mineral springs bubbling from the ground in the vicinity of what is now French Lick. Wildlife came to lick the mineral deposits left on the ground and rocks.In the early 1800's settlers began to bottle and sell the "Pluto Water" from the springs. In the early 1800's Doc Bowles built the first hotel, a three story frame building. The community thrived and there was an influx of tourist traffic coming to drink and soak in the mineral waters. In the 1850's French Lick was a key station in the "underground railway".The French Lick Springs Resort and Spa was built in the late 1800's. Tom Taggert purchased the property in 1901 and, with the help of the Monon Railroad, the former Indianapolis mayor turned the sleepy little resort into an international attraction. Many hoosiers traveled to French Lick by train. The old train depot remains in downtown French Lick.

Today French Lick and West Baden remain a favorite hoosier vacation destination along with Brown County Indiana.

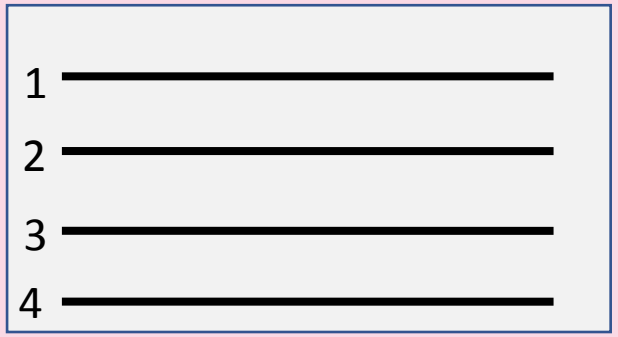
French Lick Resort and Casino.Write an advertising brochure for the French Lick Resort and Casino in Indiana. Interesting things could be the history of the casino, discounted packages for staying at the resort, are there close by other casinos, what could be job opportunities, etc.

French Lick Resort and Casino.Write an advertising brochure for the French Lick Resort and Casino in Indiana. Interesting things could be the history of the casino, discounted packages for staying at the resort, are there close by other casinos, what could be job opportunities, etc.

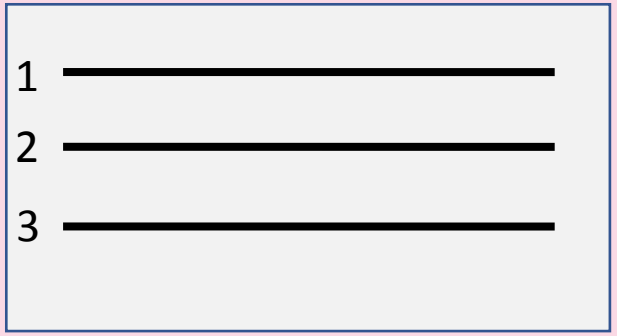
EDITING TYPES

EXAMPLE OF EDITING TYPES

Block Deletion



Before



After

On the other hand (if you prefer)

Historical Sites	Lincoln Boyhood National Memorial, George Rogers Clark National Historical Park, Amish Acres, Conner Prairie Pioneer Settlement, Historic Fort Wayne
Points of Interest	Wyandotte Cave, Indianapolis Motor Speedway, Indiana Dunes, Holiday World, Brown County craft shops

What a fantastic place to locate a Casino. No wonder the designers chose to add a touch of wonderland to the magic state when they added French Lick Resort and Casino at 8670 West State Road, 56 French Lick, Indianapolis, 47432.

On the other hand (if you prefer)

What a fantastic place to locate a Casino. No wonder the designers chose to add a touch of wonderland to the magic state when they added French Lick Resort and Casino at 8670 West State Road, 56 French Lick, Indianapolis, 47432.

Incredibly, French Lick Resort has now embarked on an absolutely amazing $500 million

EDITING TY

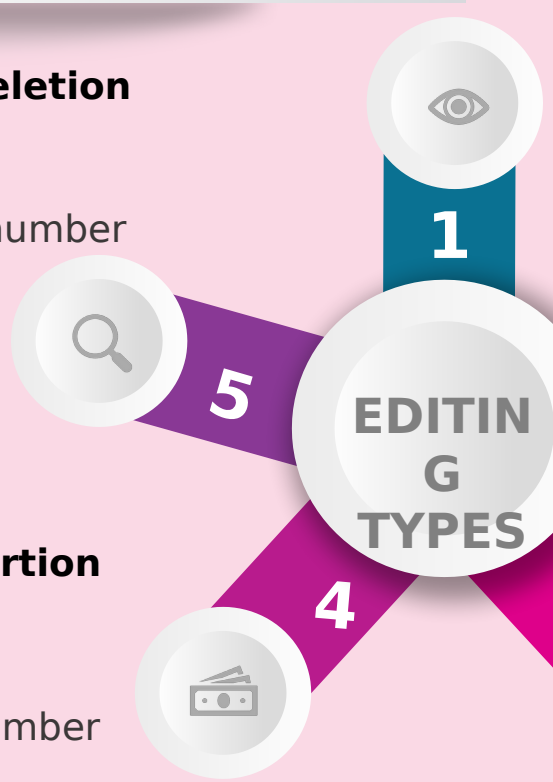
these operators comparing the same quant

AUTOMATIC IDENTIFICATION OF EDITING TYPES

01 Number of characters and the number of paragraphs plays a vital role

02 We check the number of characters and number of paragraphs in adjacent revisions.

I wouldn't make this a new, separate point since you already said this. A



Block deletion

number of paragraphs >number of paragraphs
number of characters >number of characters

Block edit

number of paragraphs = number of paragraphs
number of characters =! number of characters

Block merged

number of paragraphs > number of paragraphs
number of characters = number of characters

Block insertion

number of paragraphs < number of paragraphs
number of characters <number of characters

Block split

number of paragraphs < number of paragraphs
number of characters = number of characters

RESULT AND ANALYSIS

EDITING TYPES ?

EFFECTS OF EDITING TYPES?

EFFECTS OF WRITING STYLES?

RESULT AND ANALYSIS

RESEARCH QUESTION RQ1

RQ1 What are the different types of editing techniques in search-supported writing ?

RQ2 How do different types of editing affect Coherence, Type-Token ratio, and Readability?

RQ3 How do different essay writing strategies affect Coherence, Type-Token ratio, and Readability?

RESULT AND ANALYSIS

RQ1: DIFFERENT EDITING TYPES

Editing Types	Description
Block edit	Blocks are manually edited
Block merged	Two different blocks are merged
Block split	One block split into two
Block insertion	New block is inserted
Block deletion	Block is deleted

RESULT AND ANALYSIS

RESEARCH QUESTION RQ2

RQ1 What are the different types of editing techniques in search-supported writing ?

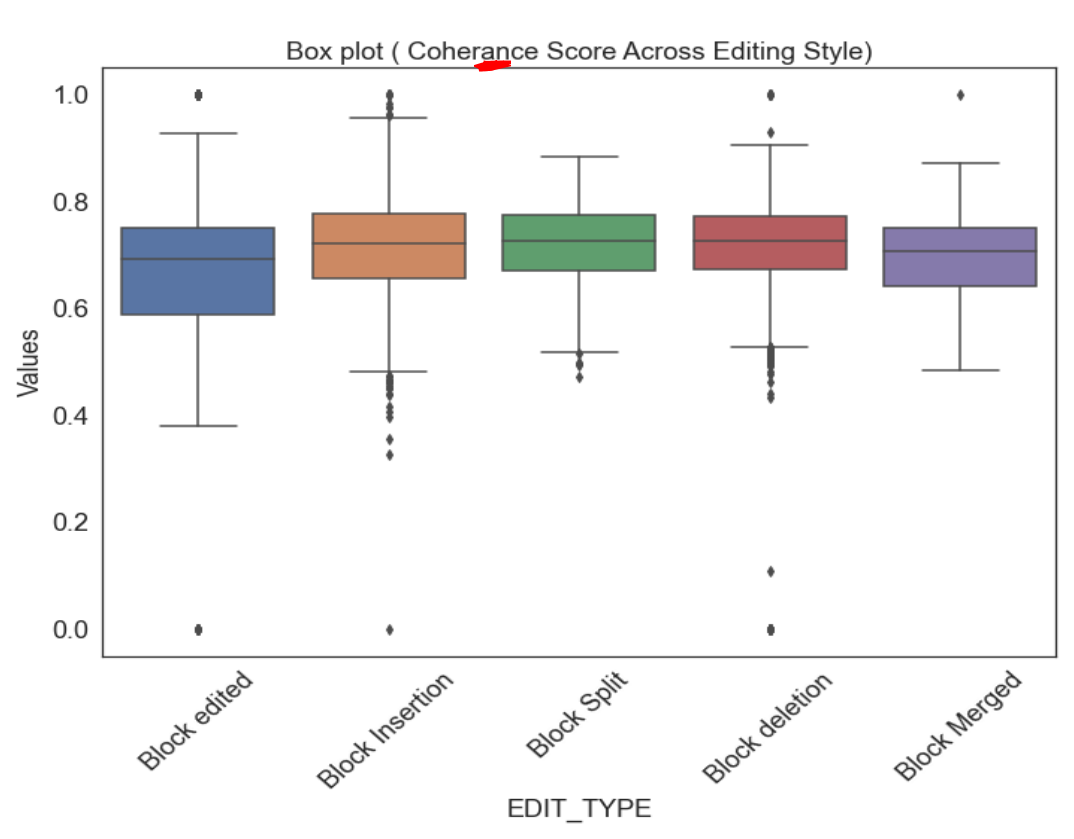
RQ2 How do different types of editing affect Coherence, Type-Token ratio, and Readability?

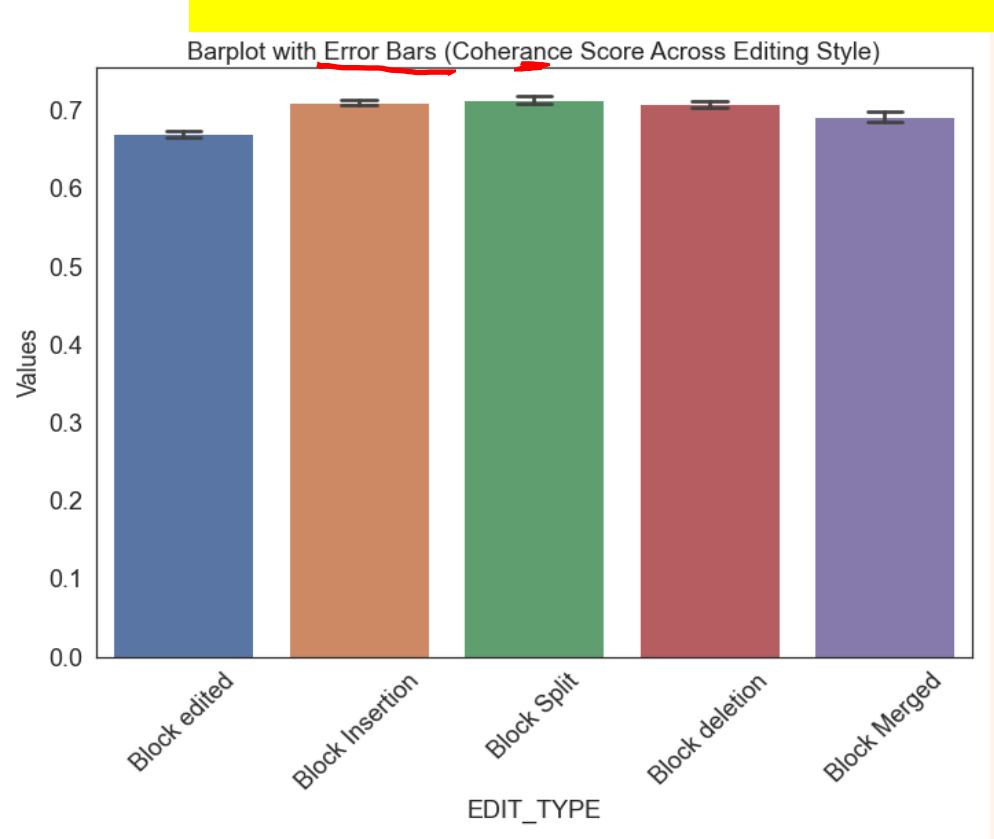
RQ3 How do different essay writing strategies affect Coherence, Type-Token ratio, and Readability?

RESULT AND ANALYSIS

RQ2: EFFECTS OF EDITING TYPES ON TEXT QUALITY

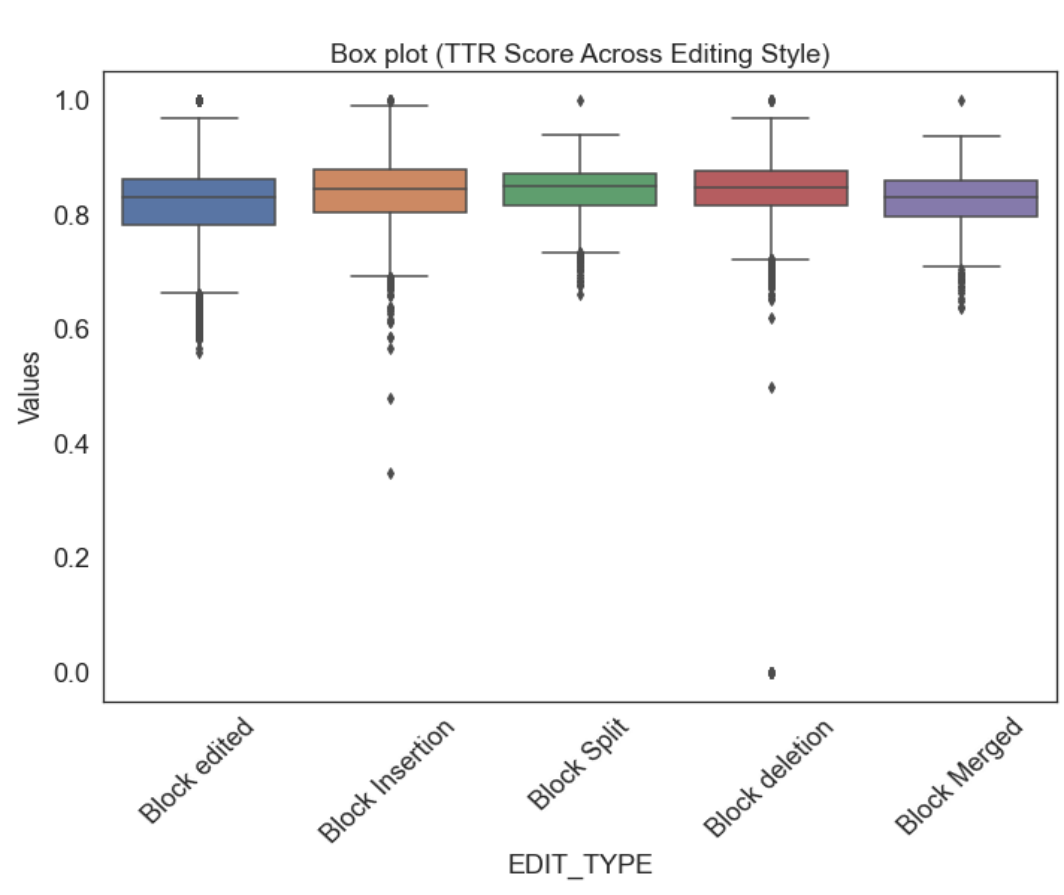
If these represent 95% confidence intervals, write that on the slide!

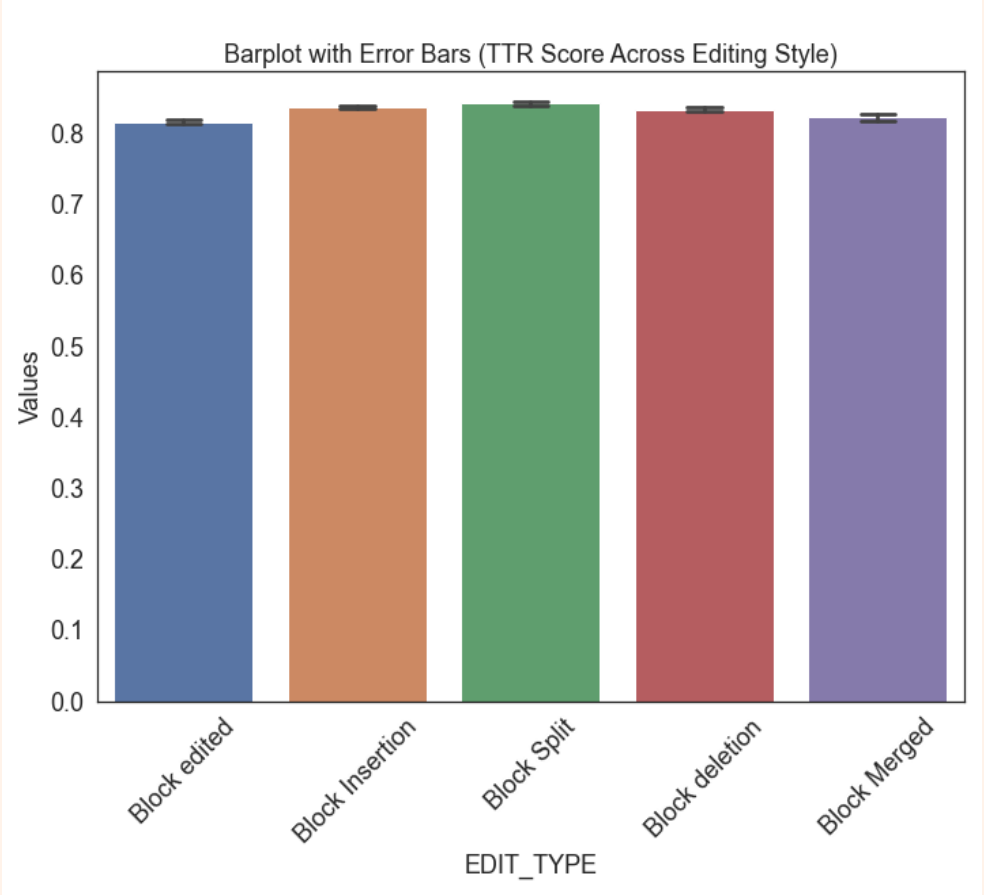




RESULT AND ANALYSIS

RQ2: EFFECTS OF EDITING TYPES ON TEXT QUALITY

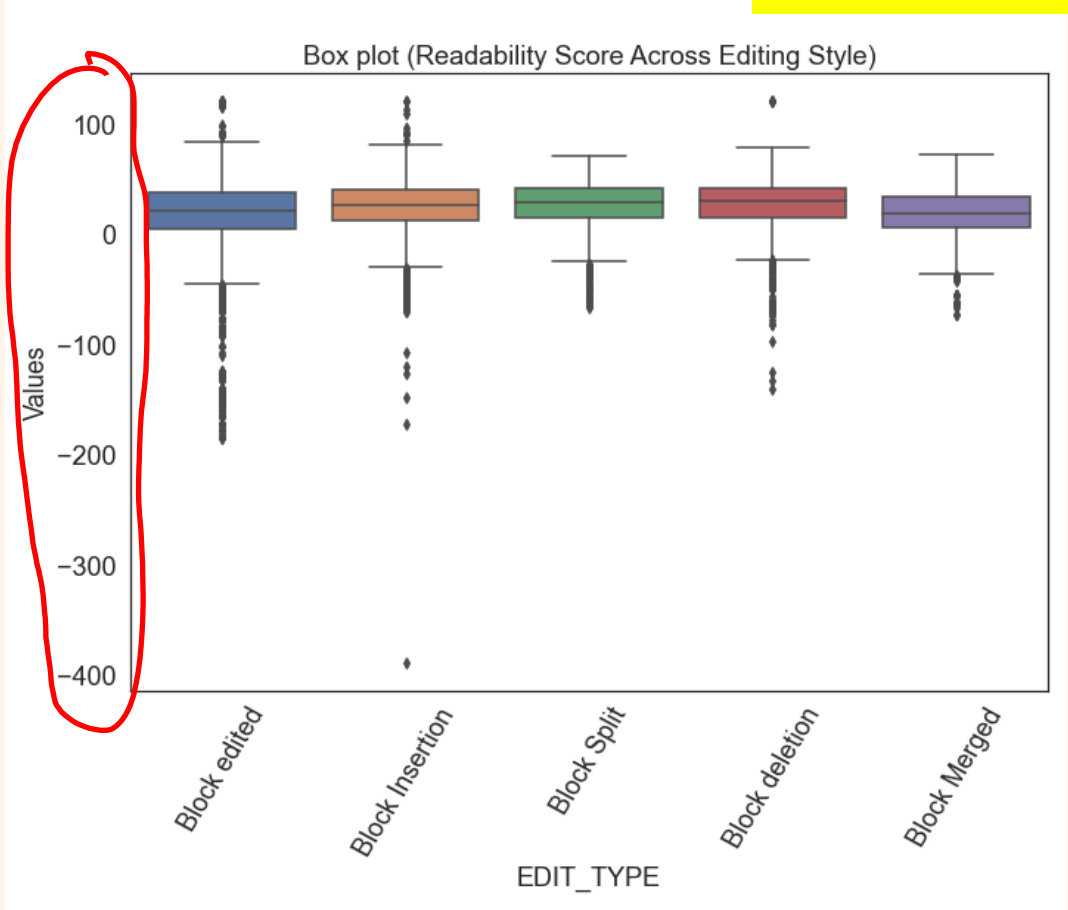


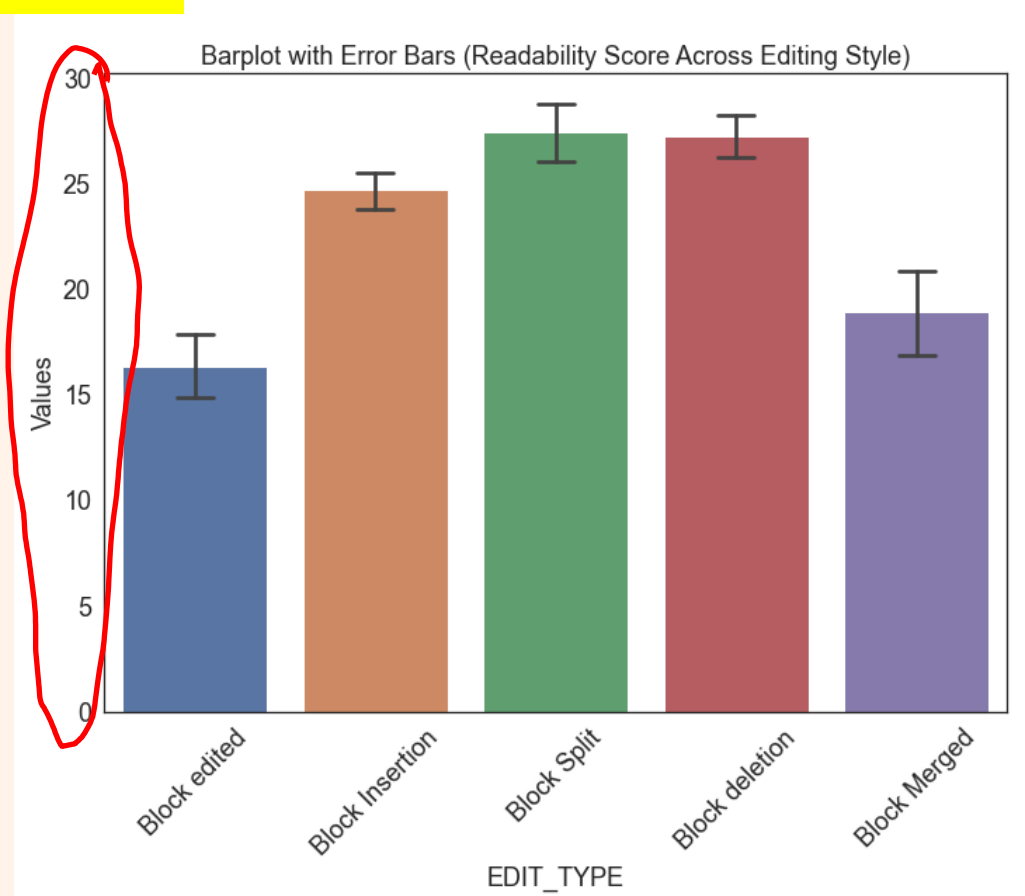


RESULT AND ANALYSIS

RQ2: EFFECTS OF EDITING TYPES ON TEXT QUALITY

why are the scales so different here?





RESULT AND ANALYSIS

RESEARCH QUESTION RQ3

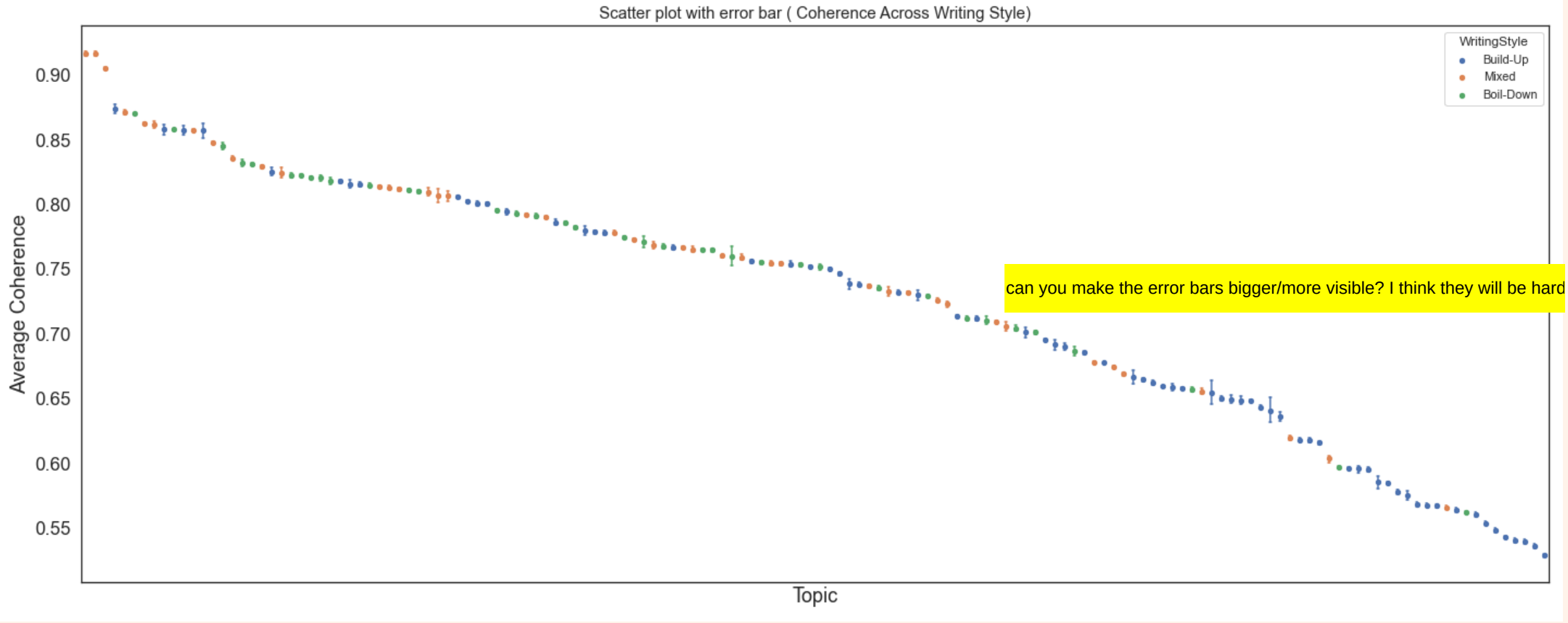
RQ1 What are the different types of editing techniques in search-supported writing ?

RQ2 How do different types of editing affect Coherence, Type-Token ratio, and Readability?

RQ3 How do different essay writing strategies affect Coherence, Type-Token ratio, and Readability?

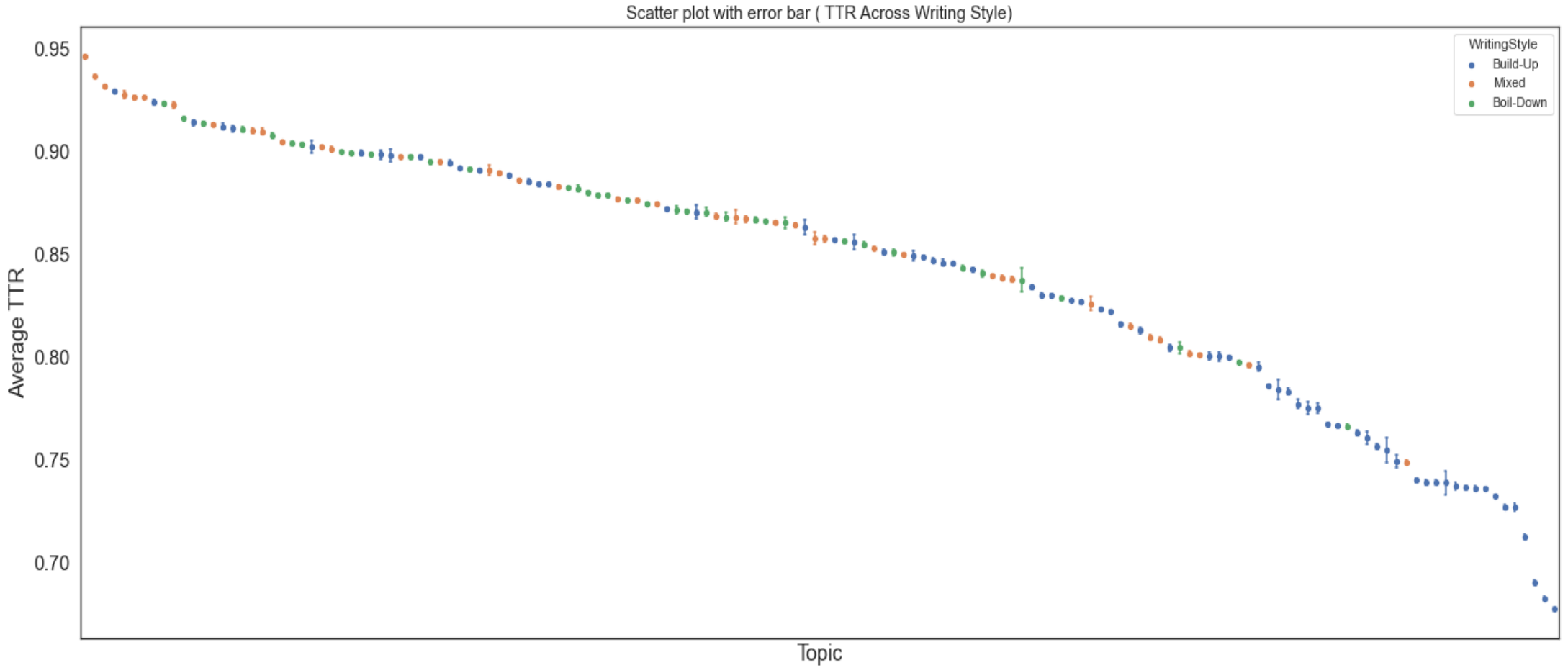
RESULT AND ANALYSIS

RQ3: EFFECTS OF WRITING STYLES ON TEXT QUALITY



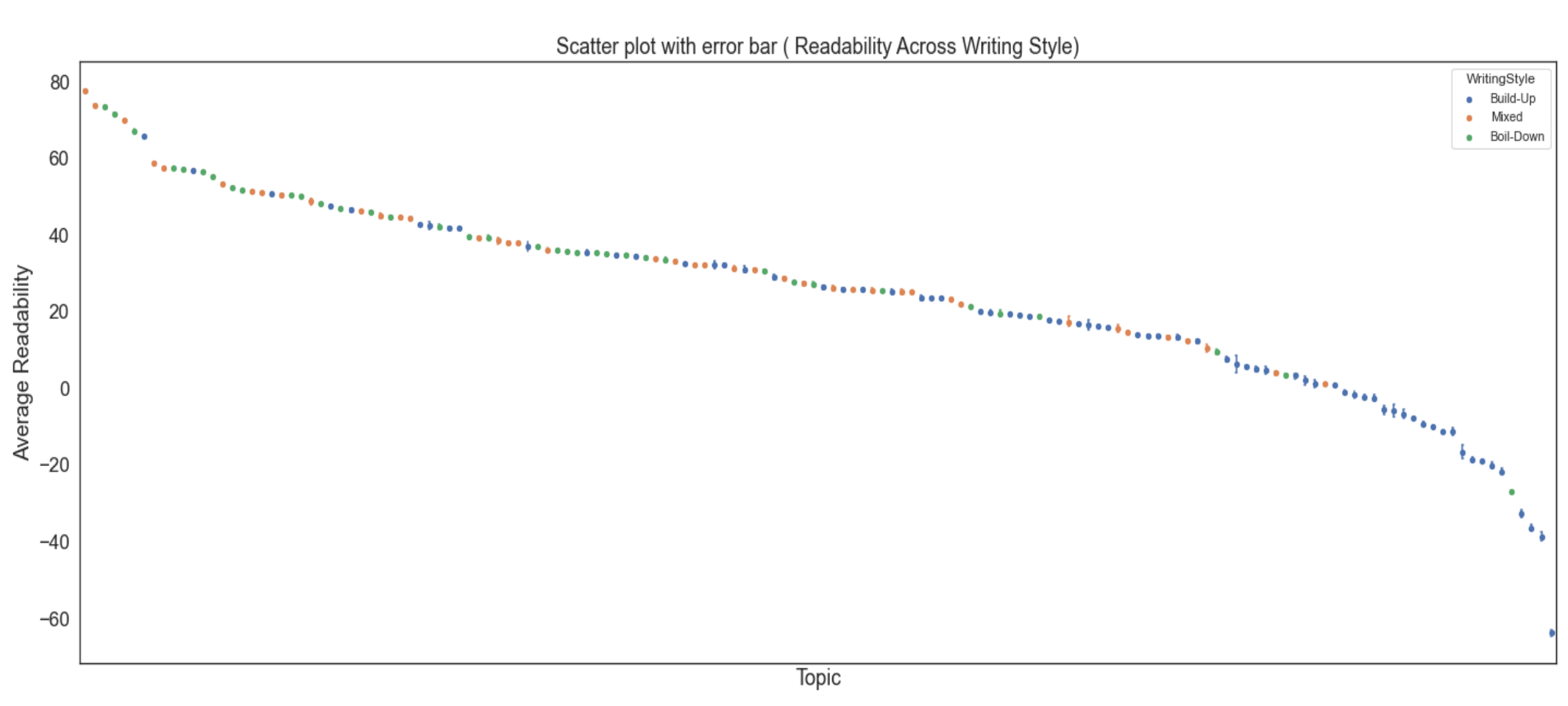
RESULT AND ANALYSIS

RQ3: EFFECTS OF WRITING STYLES ON TEXT QUALITY



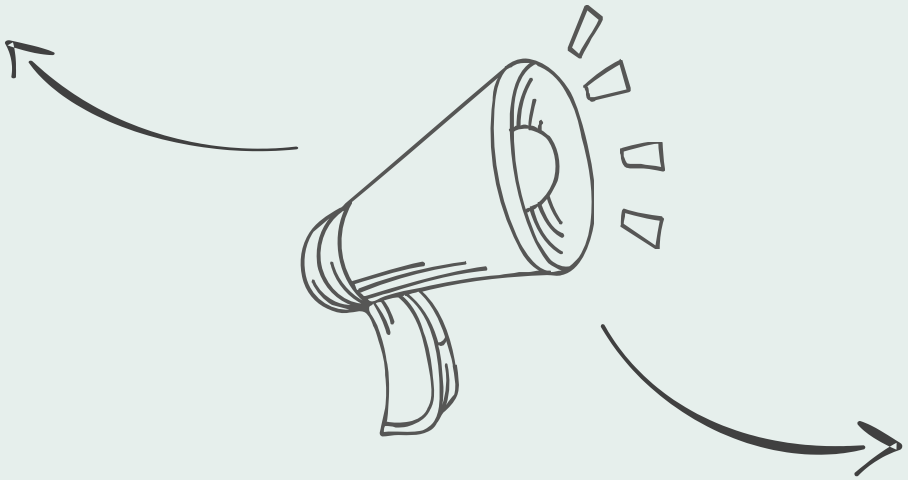
RESULT AND ANALYSIS

RQ3: EFFECTS OF WRITING STYLES ON TEXT QUALITY



CONCLUSION

MAIN FINDINGS



FUTURE WORK

CONCLUSION

MAIN FINDINGS

RQ1 ~~Different~~ editing types ? name them

Editing types affect readability, but Coherence and Type token ratio have same pattern of scores in every editing types. **RQ 2**

RQ 3 Most of the essays with build-up writing are having lower coherence, TTR, and readability scores

Coherence, TTR, and readability are higher in mixed writing **RQ 3**

CONCLUSION

FUTURE WORK

01 ~~Investigate and explain the~~ lower coherence in Build-up writing.

These are so similar that they should be the same point? Also, I'd again avoid full ser

~~Investigate the reason and explain the reason for~~ lower TTR and Readability in Build-up writing. 02

03 ~~Investigate~~ how ~~edits to the essay that introduce~~ new entities ~~in particular~~ affect ~~the~~ coherence score.

How coherence is affected by cohesion. 04

Thank you!