TurkLang 2021

THE STUDY OF "OATH, WAR AND LOOT" CONCEPTS IN SEYITBEK EPOS REGARDING COLLOCATION WITH COMPUTATIONAL LINGUISTICS METHODS

21-23 SEPTEMBER 2021, KYZYL, REPUBLIC OF TUVA, RUSSIA

PREPARED BY;

Assoc. Prof. Meriç GÜVEN

- Uşak University
- Uşak, TURKEY

İbrahim Atakan KUBİLAY, PhD

- Dokuz Eylül University
- İzmir, TURKEY

Yasemin (KARADEMİRLİDAĞ) KUBİLAY, PhD Student

- Uşak University
- Uşak, TURKEY

OUTLINE

- Introduction
  - Collocations
  - Seyitbek Epos
  - Concepts
- Processing
  - Preprocessing
  - Metrics
- Results

INTRODUCTION

WHAT IS A COLLOCATION?

- Collocations are word groups consisting of two or more words. They have a certain (higher than usual) usage frequency and limitations on how they are used.
- We can't replace a word in a collocation with any other word from a relatively large number of possibilities.
- Example: `wash the car` is NOT a collocation, because we can replace `car` with many options like `clothes`, `dishes`, `bicycle`,…etc.
- Example: `fast food` is a collocation, because we can`t replace `fast` with `quick`.

COLLOCATIONS AND LINGUISTIC CONTEXT

- Collocations also require linguistic context to detect. Therefore, they give hints about the cultural and historical development of a language. There may not be any reason for the preference of one word for the other. This makes it difficult for non-native speakers to understand collocations.
- Example: There is nothing semantically wrong with `quick train` but instead `fast train` is used as the *correct* collocation.
- Example: `Işığı yak.` (Turkish) Translation: `Turn on the light`. Literal translation: `Burn the light`.

SEYITBEK EPOS

- Seyitbek Epos is part of the three-part epos, telling the battles of Kyrghyz People with Kalmuks. It`s a continuation of Kurmanbek(Father of Seyitbek) Epos.
- Seyitbek is one of the important eposes of the Turkish World and is a masterpiece depicting the struggle for freedom that relates the facts of Kyrghyz History in an epic narrative.
- The epos has a very special and important place in Kyrghyz language, literature, history, and culture.
- It was first published in 1971 by famous folklorist Orozbey Urmambetov. Another, shorter version was published by Nurdin Adiyev.

HOW WERE THE TERMS CHOSEN?

PROCESSING

PRE-PROCESSING

- We have used the Modern Turkish translation of the Epos by Yasemin (KARADEMİRLİDAĞ) KUBİLAY.
- The text was preprocessed using the NLTK (Natural Language Processing Toolkit) package in Python programming language.
- Punctuation marks, numerals and common words in Turkish (stopwords) were removed.

METRICS

MUTUAL INFORMATION

- Mutual Information is a measure of mutual dependence between two random variables(here words).

- $MI = \log \frac{P(X,Y)}{P(X)P(Y)}$

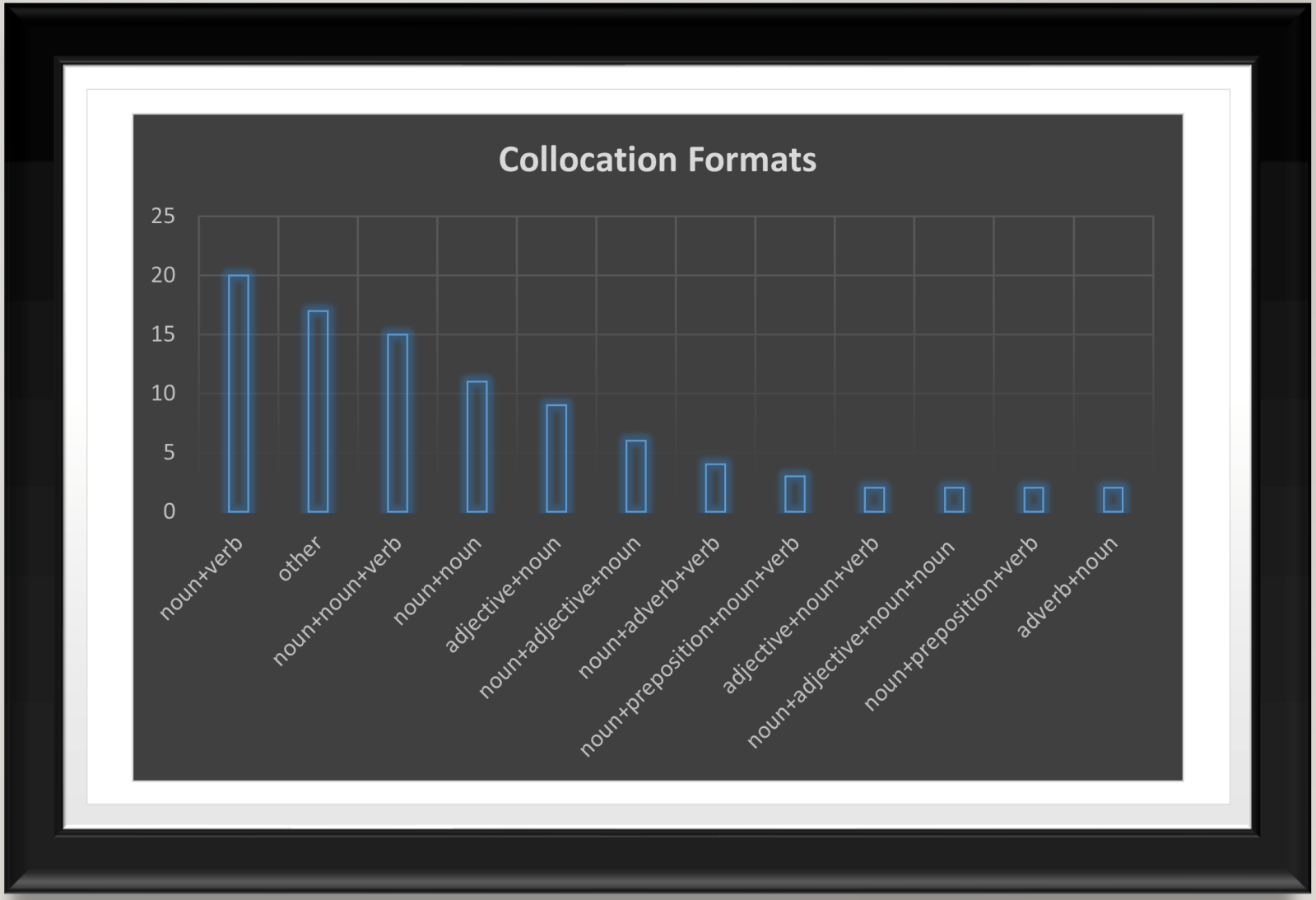
T-SCORE

- T-score is a statistical method that uses mean and variance (measures of central tendency) to test a null hypothesis. The null hypothesis in this case is that `the two words are independent, i.e., NOT collocations`.

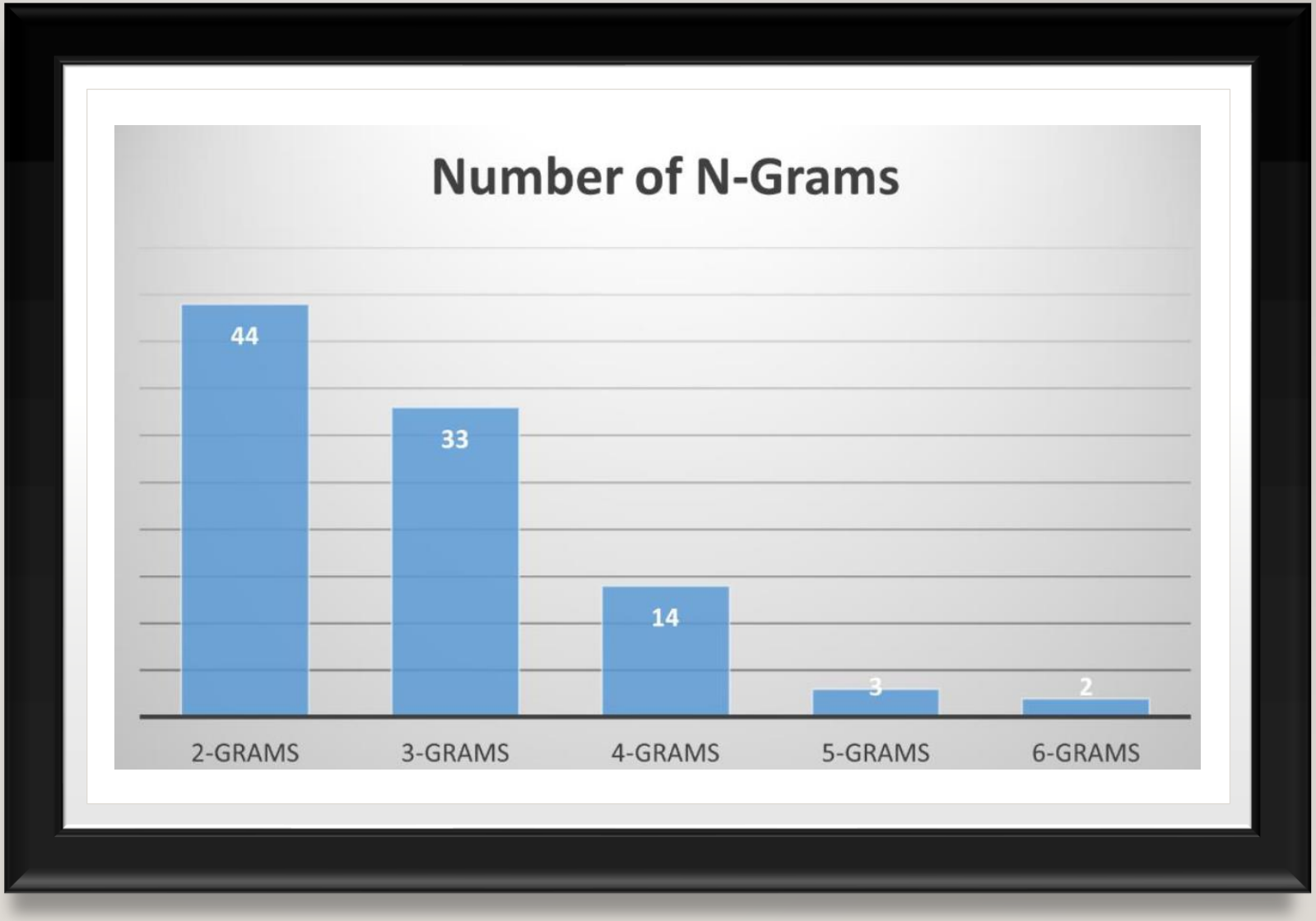
- $t = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{N}}}$

RESULTS

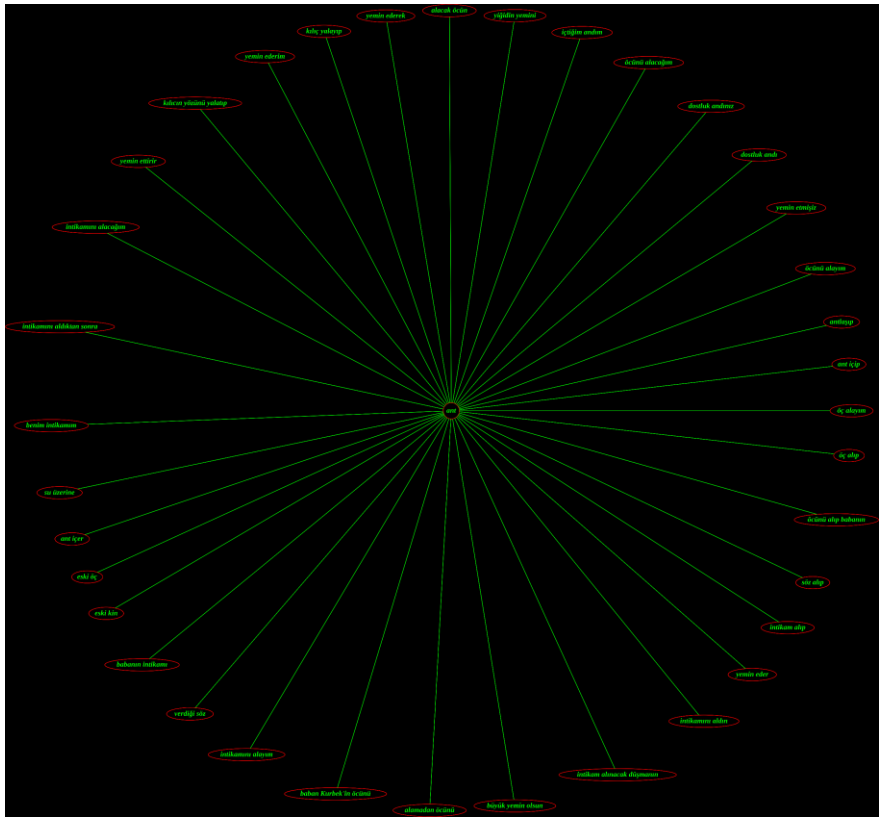
COLLOCATION FORMATS



N-GRAMS

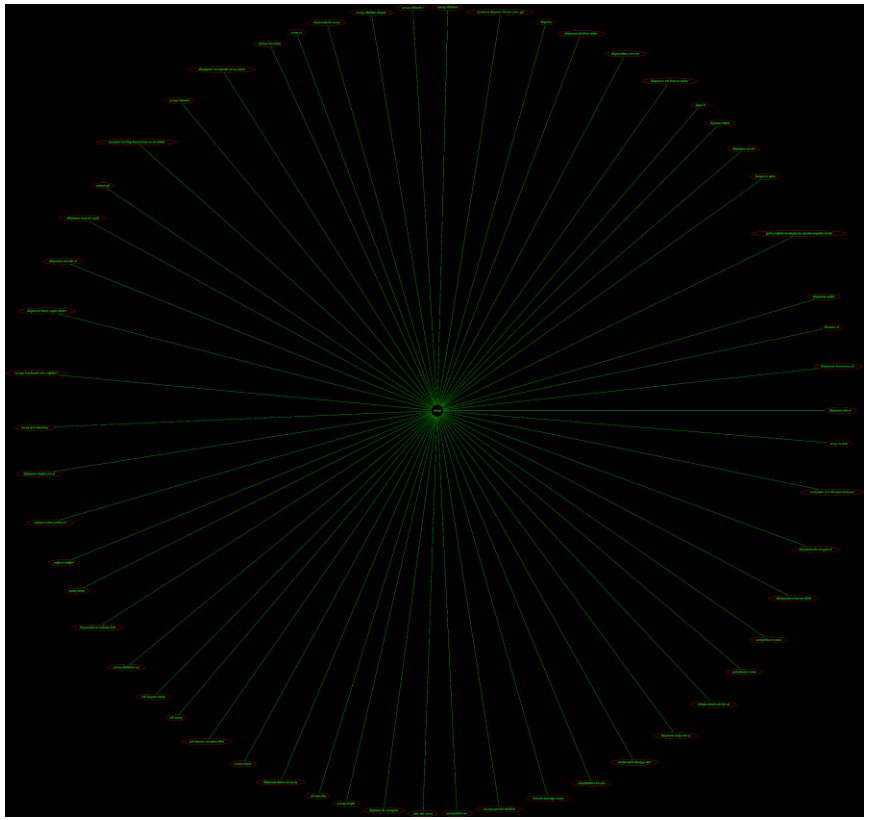

Number of N-Grams
44
33
14
3
2
2-GRAMS
3-GRAMS
4-GRAMS
5-GRAMS
6-GRAMS

OATH (TR. ANT)

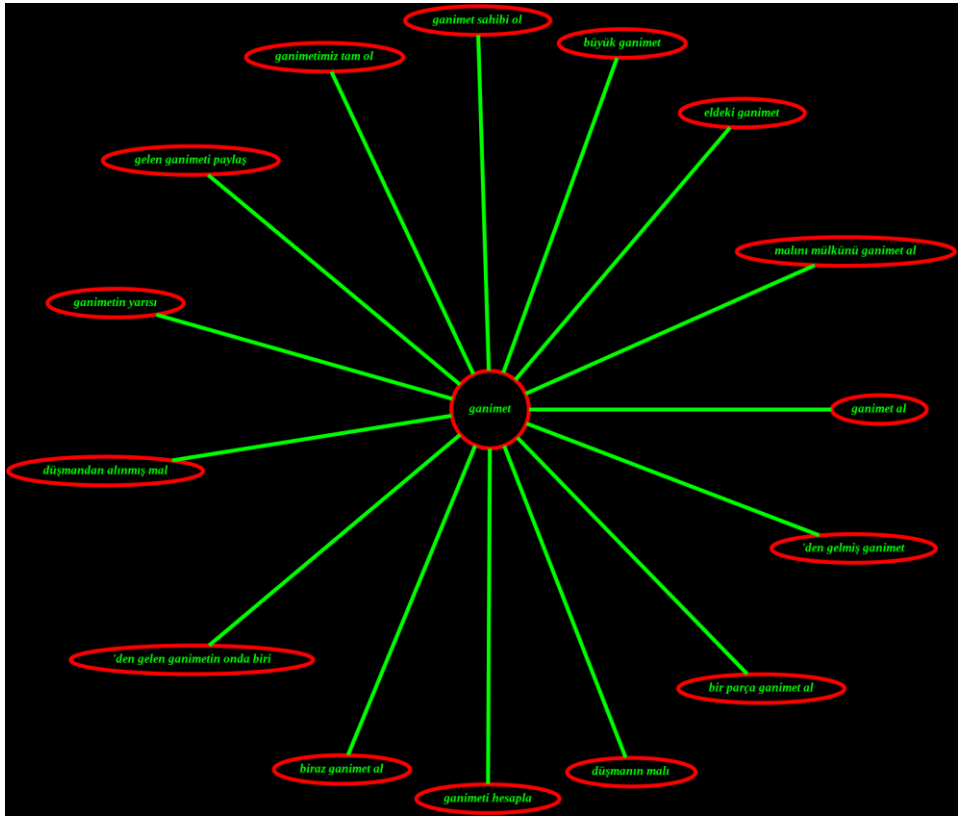


yemin ederek
alacak öcün
yüğülin yemini
kılıç yalayıp
içtiğim andım
yemin ederim
öcünü alacağım
kılıcın yüzünü yalatıp
dostluk andımız
yemin emirir
dostluk andı
yemin etmişiz
intikamını alacağım
öcünü alayım
intikamını aldıktan sonra
antlaşıp
ant içip
ant
öç alayım
benim intikamım
öç alıp
ta üzerine
öcünü alıp babanın
ant içer
söz alıp
eski öç
eski kin
intikam alıp
babanın intikamı
yemin eder
verdiği söz
intikamını aldın
intikamını alayım
intikam almacak düşmanın
baban Karbek'in öcünü
büyük yemin olsun
alamadan öcünü

WAR/FIGHT (TR. SAVAŞ)

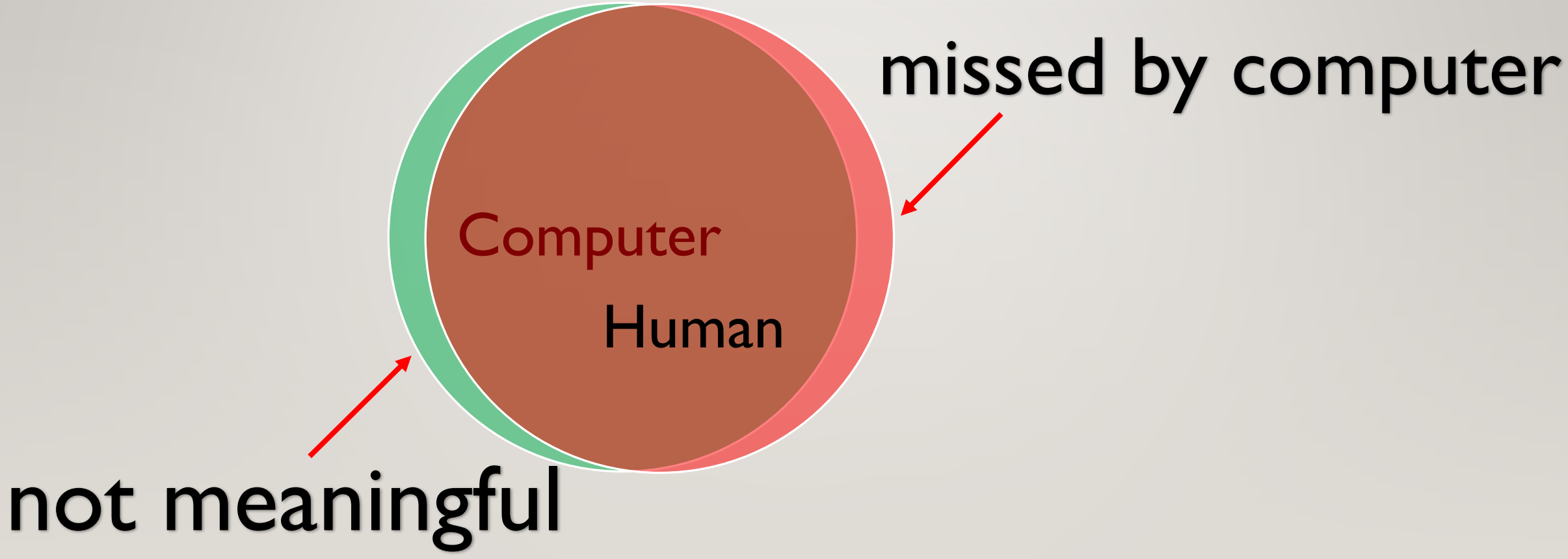


LOOT (TR. GANİMET)



SUCCESS RATE

- 96% of the computer detected collocations were meaningful.
- 95% of the manually detected collocations were detected among the meaningful collocations detected by the computer.


missed by computer
Computer
Human
not meaningful

BENEFITS OF THE STUDY

- Semi-automatic and automatic detection of collocations and their types.
- Better understanding of historical texts.
- Inter-period comparative studies.
- Automatic ontology creation.
- Preparation of collocation dictionaries for historical periods.

THANK YOU VERY MUCH

FOR LISTENING TO OUR PRESENTATION.

NOW PLEASE ASK ANY QUESTIONS YOU HAVE.