

An NLP method in the corpus analysis of Central Kurdish definiteness marker

Hiwa Asadpour

JSPS International Fellow, University of Tokyo, Japan
Associate Researcher, Goethe University of Frankfurt, Germany
Asadpour@lingua.uni-frankfurt.de

Arash Amani

Independent Researcher
amani.ara@gmail.com

SADILAR DH COLLOQUIUM

15 March 2023



REGEX
solution for
low-resource
languages

Definiteness



Sample
region and
language



TURKEY

ARMENIA

AZERBAIJAN

Caspian Sea

KURDISTAN

Mediterranean Sea

SYRIA

LEBANON

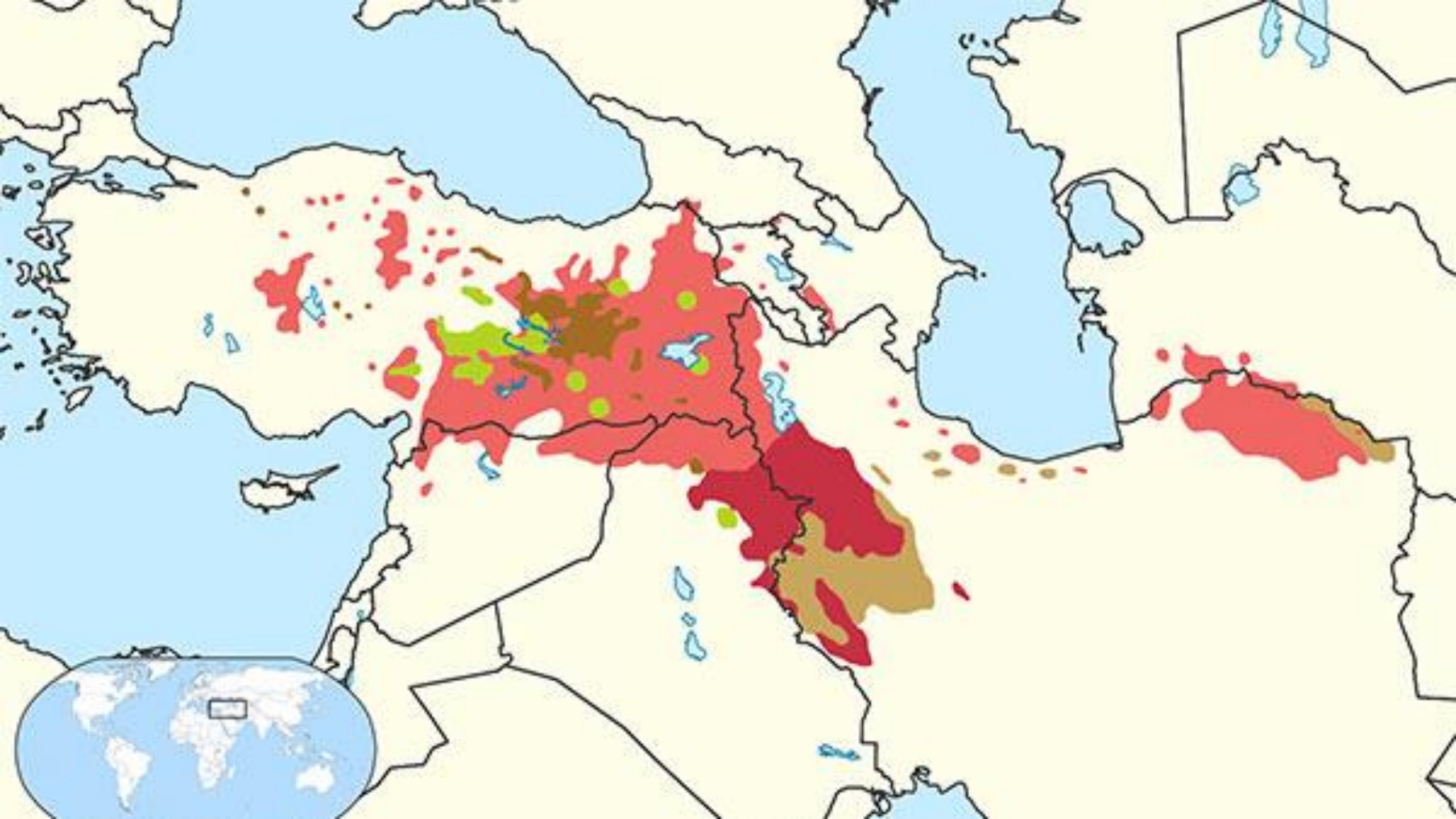
ISRAEL

IRAQ

IRAN

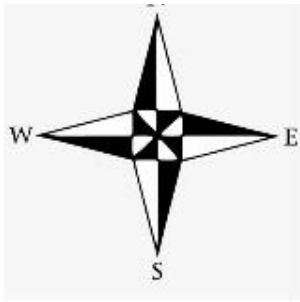
JORDAN

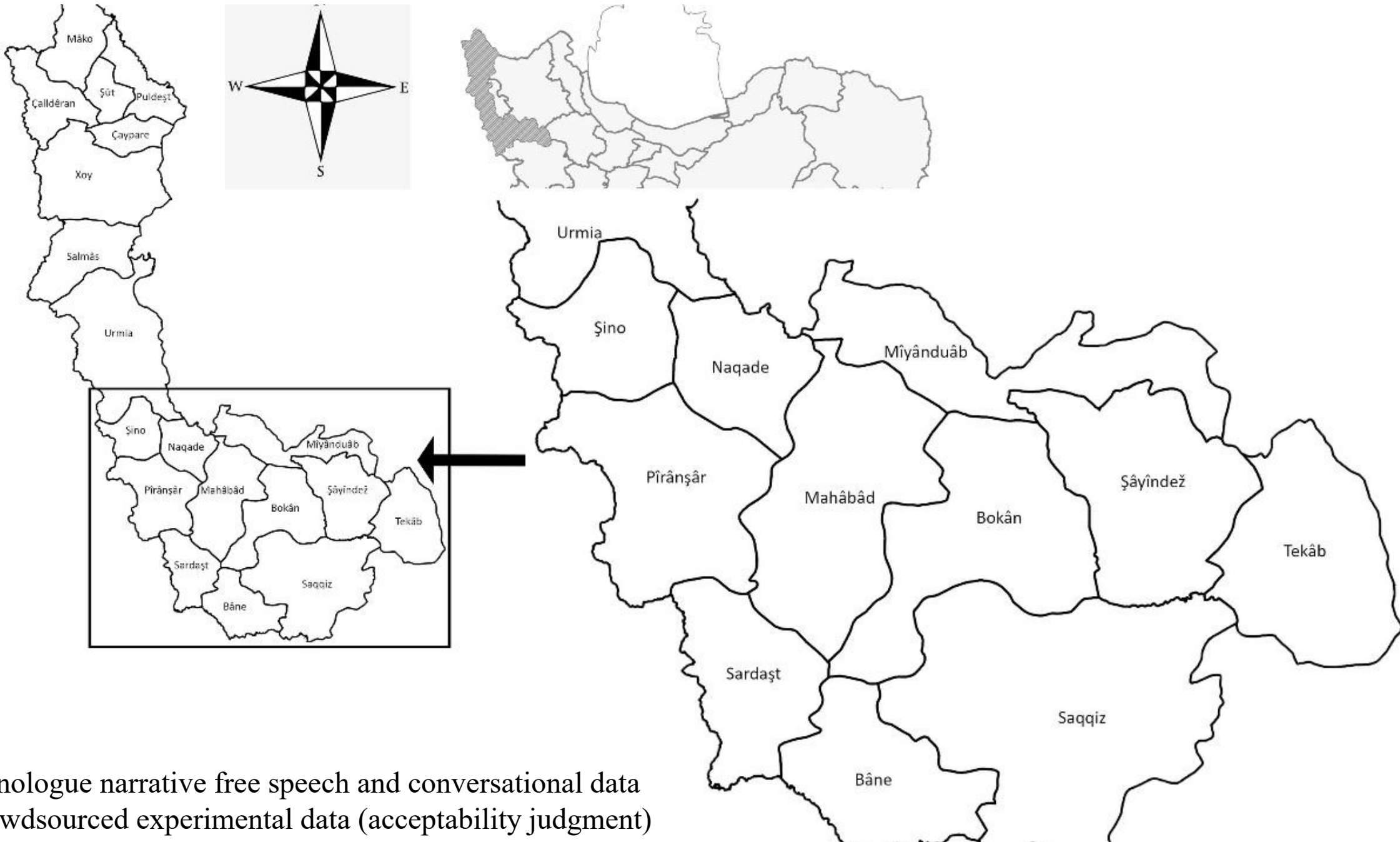
★
Baghdad





- Kurdish
 - Northeastern
 - Mukri
- Neo-Aramaic
 - Jewish
 - Christian
- Turkic Azeri
- Armenian





Data:

- a) Monologue narrative free speech and conversational data
- b) Crowdsourced experimental data (acceptability judgment)

Left-branching (OV)

Text processing of definite marking

English

Definite article: the

Indefinite article: a/an

Central Kurdish

Definite article: *-eke/-aka*

Indefinite article: *-ek/-ak/yek/yak*

Text processing of definite marking

English

Definite article: the

Indefinite article: a/an

Central Kurdish

Definite article: *-eke/-aka*

Indefinite article: *-ek/-ak/yek/yak*

<i>amən=iš</i>	<i>nānawā-yek=mān</i>	<i>habū</i>	<i>xom</i>	<i>da</i>	<i>bar</i>	<i>nānawā-y-aka-y</i>	<i>kərd</i>
1SG=ADD	bakery-INDF=PC.1PL.POSS	have.PST	self.1SG	at	on	bakery-GLID-DEF-OBL	do.PST

‘As for me, (we) had a bakery. (I) put myself under the bakery

[*lit.* I laid with the bakery].’

Text processing of definite marking

English

Definite article: the

Indefinite article: a/an

Central Kurdish

Definite article: *-eke/-aka*

Indefinite article: *-ek/-ak/yek/yak*

amənan=iš

1SG=ADD

nānawā-yek=mān

bakery-INDF=PC.1PL.POSS

habū

have.PST

xom

self.1SG

da

at

bar

on

nānawā-y-aka-y

bakery-GLID-DEF-OBL

kərd

do.PST

‘As for me, (we) had a bakery. (I) put myself under the bakery

[*lit.* I laid with the bakery].’

Text processing of definite marking

English

Definite article: the

Indefinite article: a/an

Central Kurdish

Definite article: *-eke/-aka*

Indefinite article: *-ek/-ak/yek/yak*

amən=iš

1SG=ADD

nānawā-yek=mān

bakery-INDF=PC.1PL.POSS

habū

have.PST

xom

self.1SG

da

at

bar

on

nānawā-y-aka-y

bakery-GLID-DEF-OBL

kərd

do.PST

‘As for me, (we) had a bakery. (I) put myself under the bakery

[*lit.* I laid with the bakery].’

ku bə-š-t=a

that SBJV.go.PRS-3SG=to

cə-y-ak-e

place-GLID-INDF-OBL.SG.F

‘[...] that (he) goes to somewhere.’

Text processing of definite marking

English

Definite article: the

Indefinite article: a/an

Central Kurdish

Definite article: *-eke/-aka*

Indefinite article: *-ek/-ak/yek/yak*

Central Kurdish

	sg.	pl.
Definite	<i>-e/-(e)k(e)</i>	<i>-ekan</i>
Indefinite	<i>-ēk/-ek</i>	<i>-ēkan/-anēk</i>
Bare	<i>-∅</i>	<i>-an</i>
“one”	<i>yek/yēk</i>	-

Mukri Kurdish

Subject	Object	Verb	Preposition	Target
<i>ew-ân</i> they-PL	<i>kitêb-êk</i> book-INDF	<i>=yân dâ</i> =PC.3PL give.PST	<i>be</i> to	<i>piyâw-ek-ân</i> man -DEF-PL

‘They gave a book to the men.’

Mukri Kurdish

Subject	Object	Verb	Preposition	Target
<i>ew-ân</i> they-PL	<i>kitêb-êk</i> book-INDF	= <i>yân</i> <i>dâ</i> =PC.3PL give.PST	<i>be</i> to	<i>piyâw-ek-ân</i> man -DEF-PL

‘They gave a book to the men.’

kut=î, *kuř-a* *wa* *peř=am* *kawət*
say.PST=PC.3SG.AGR boy-DEF to front=PC.1SG fall.PST

‘(He) said, the boy fell down in front of me.’

Marking NPs in Mukri:

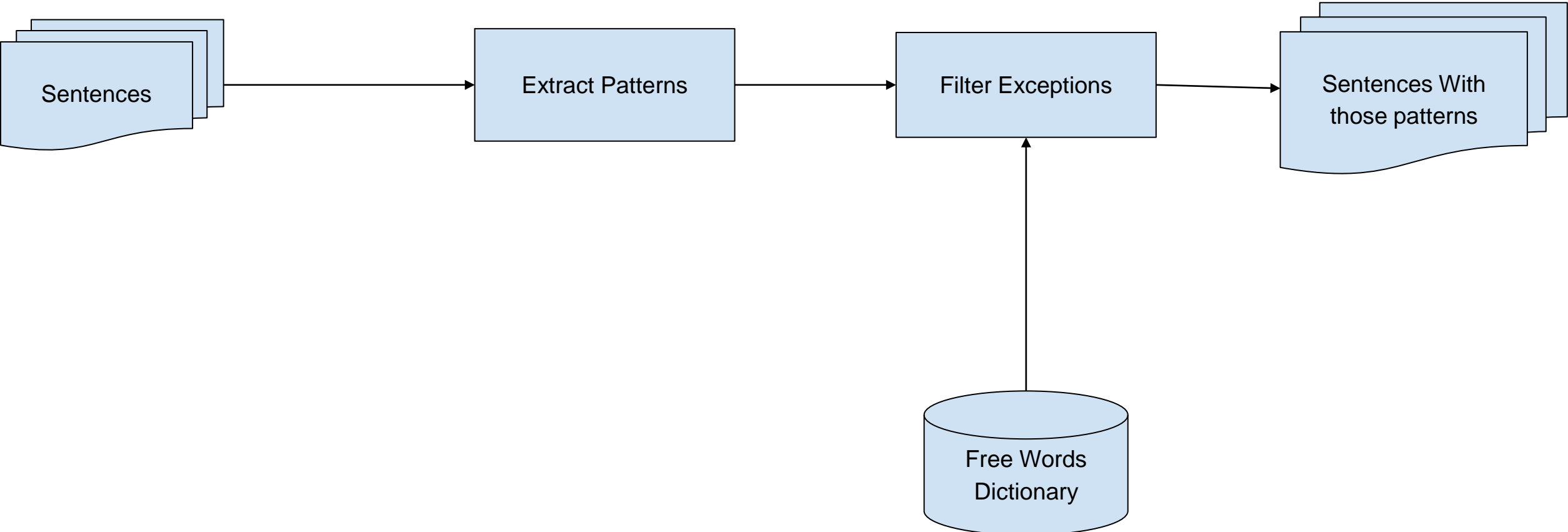
- 1- NP + demonstrative (exophoric, endophoric +/- distance)
e.g.: *ew piyaw=e* ‘that man’
- 2- NP + demonstrative + case
e.g.: *ew piyaw=e=y* ‘that man (OBL)’
- 3- NP + demonstrative + definite article
e.g.: *ew piyaw-ek(=)e* ‘that man’
- 4- NP + demonstrative + definite article + case
e.g.: *ew piyaw-ek(=)e=y* ‘that man (OBL)’
- 5- NP + demonstrative + indefinite article + case
e.g.: *ew piyaw-ēk=ī* ‘that man’ (specificity + unidentifiable)

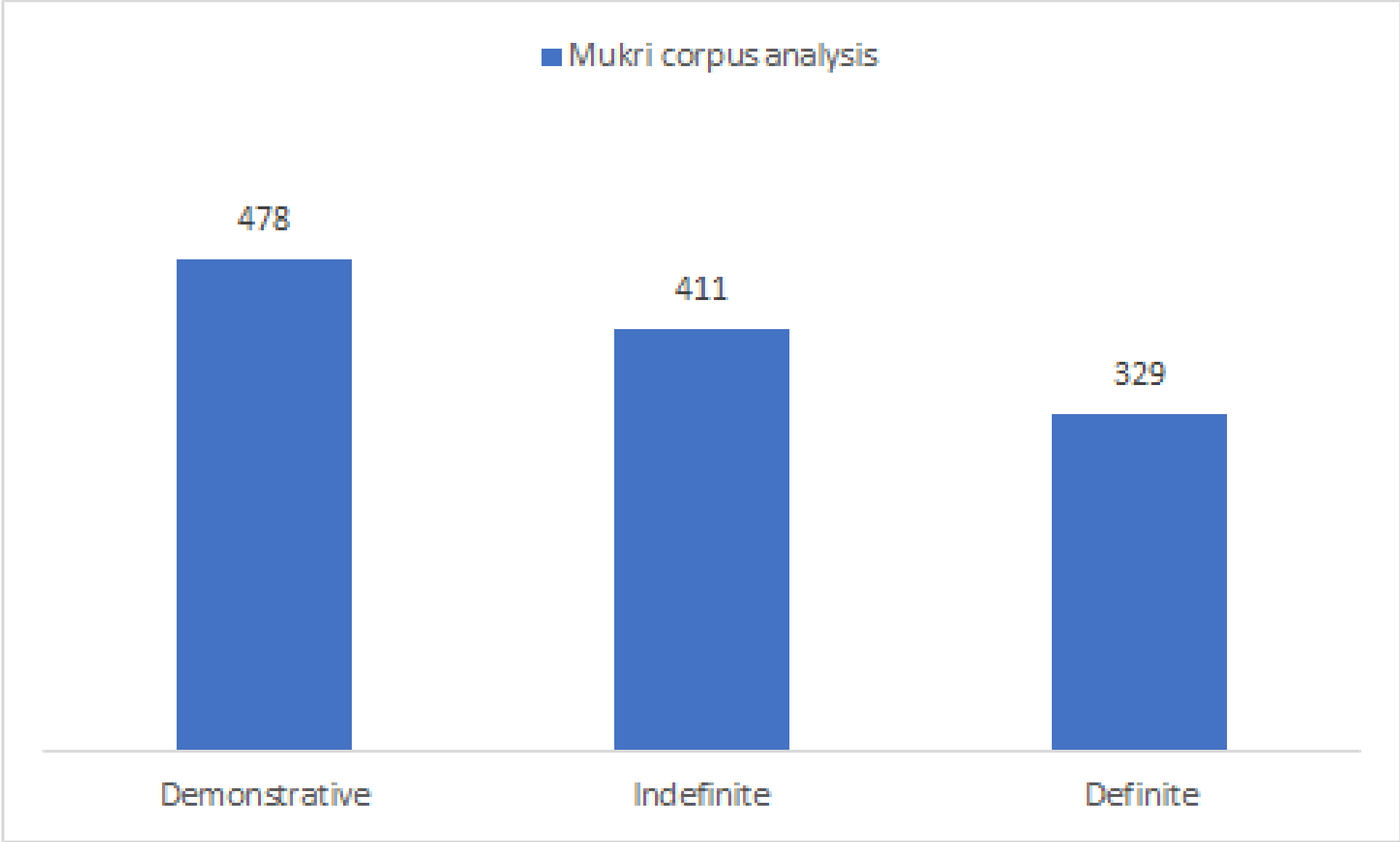
ČN.111	<i>ew sundūqe le hīč jēgāyēkī nefesī nebē</i>	(that) this coffer, should not have breath from anywhere (i.e. should allow air to enter).
ČN.112	<i>ew mindāley dew sundūqey denēyn, serī lē dādexeyn</i>	We shall put this child in that coffer, close the lid over him,
ČN.113	<i>dew čomey dāwēyn bā āw bībā</i>	(and) throw it into this river so that the water takes him away.
ČN.114	<i>deybā bo memleketēkī dīke</i>	It will take him to another country.”
ČN.115	<i>wilā čūn le kin nejāřēkī</i>	By God, they went to a carpenter.
ČN.116	<i>řerzī to bikem</i>	Let me tell you.
ČN.117	<i>sundūqekyān be durust kirdin dā eger le hīč kö āwē nedā</i>	They had a coffer built that would not take in water from anywhere.
ČN.118	<i>mindāleyān de sundūqē hāwīšt</i>	They threw the child into the coffer.
ČN.119	<i>be čomēyān dādā āw birdī</i>	They left it in the river, (and) the water took it.
ČN.120	<i>kutyān deřwā bo memleketekī dīke</i>	They said “It will go to another land.
ČN.121	<i>tāze, xudā yā dexinkē yā deřwā bo memleketekī dīke</i>	Now that, by God, either he drowns or he goes to another country.”
ČN.122	<i>wilā ew čome debū qoł qoł her jogeley be lāyēkī dečū</i>	By God, this river had many branches, one creek took one direction,

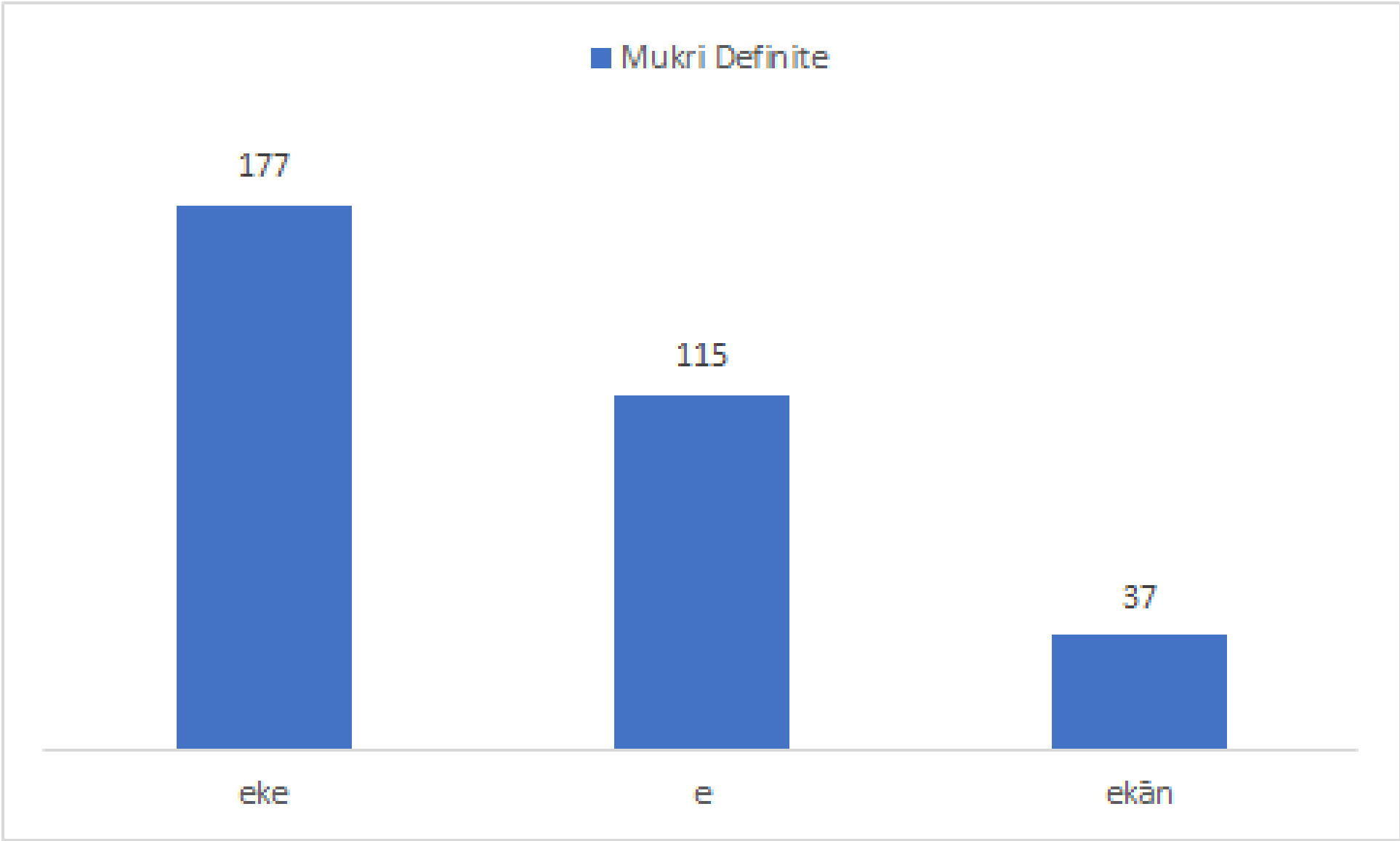
ČN.111	<i>ew sundūqe le hīč jēgāyēkī nefesī nebē</i>	(that) this coffer , should not have breath from anywhere (i.e. should allow air to enter).
ČN.112	<i>ew mindāley dew sundūqey denēyn, serī lē dādexeyn</i>	We shall put this child in that coffer , close the lid over him,
ČN.113	<i>dew čomey dāwēyn bā āw bībā</i>	(and) throw it into this river so that the water takes him away.
ČN.114	<i>deybā bo memleketēkī dīke</i>	It will take him to another country .”
ČN.115	<i>wilā čūn le kin nejāřēkī</i>	By God, they went to a carpenter .
ČN.116	<i>řerzī to bikem</i>	Let me tell you.
ČN.117	<i>sundūqekyān be durust kirdin dā eger le hīč kö āwē nedā</i>	They had a coffer built that would not take in water from anywhere.
ČN.118	<i>mindāleyān de sundūqē hāwīšt</i>	They threw the child into the coffer .
ČN.119	<i>be čomēyān dādā āw birdī</i>	They left it in the river , (and) the water took it.
ČN.120	<i>kutyān deřwā bo memleketēkī dīke</i>	They said “It will go to another land.
ČN.121	<i>tāze, xudā yā dexinkē yā deřwā bo memleketēkī dīke</i>	Now that, by God, either he drowns or he goes to another country .”
ČN.122	<i>wilā ew čome debū qoř qoř her jogelej be lāyēkī dečū</i>	By God, this river had many branches , one creek took one direction ,

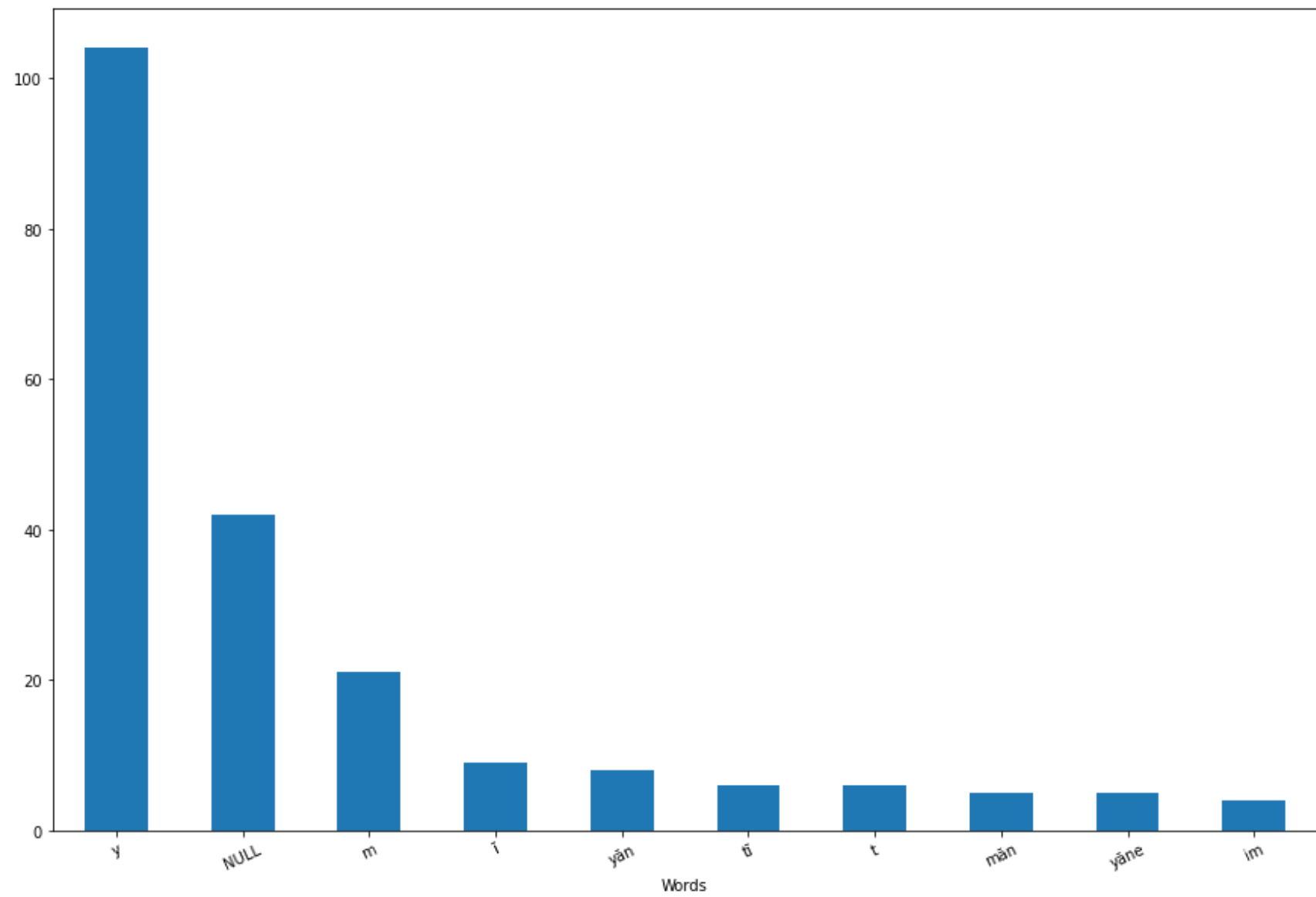
Corpus analysis using regex tool

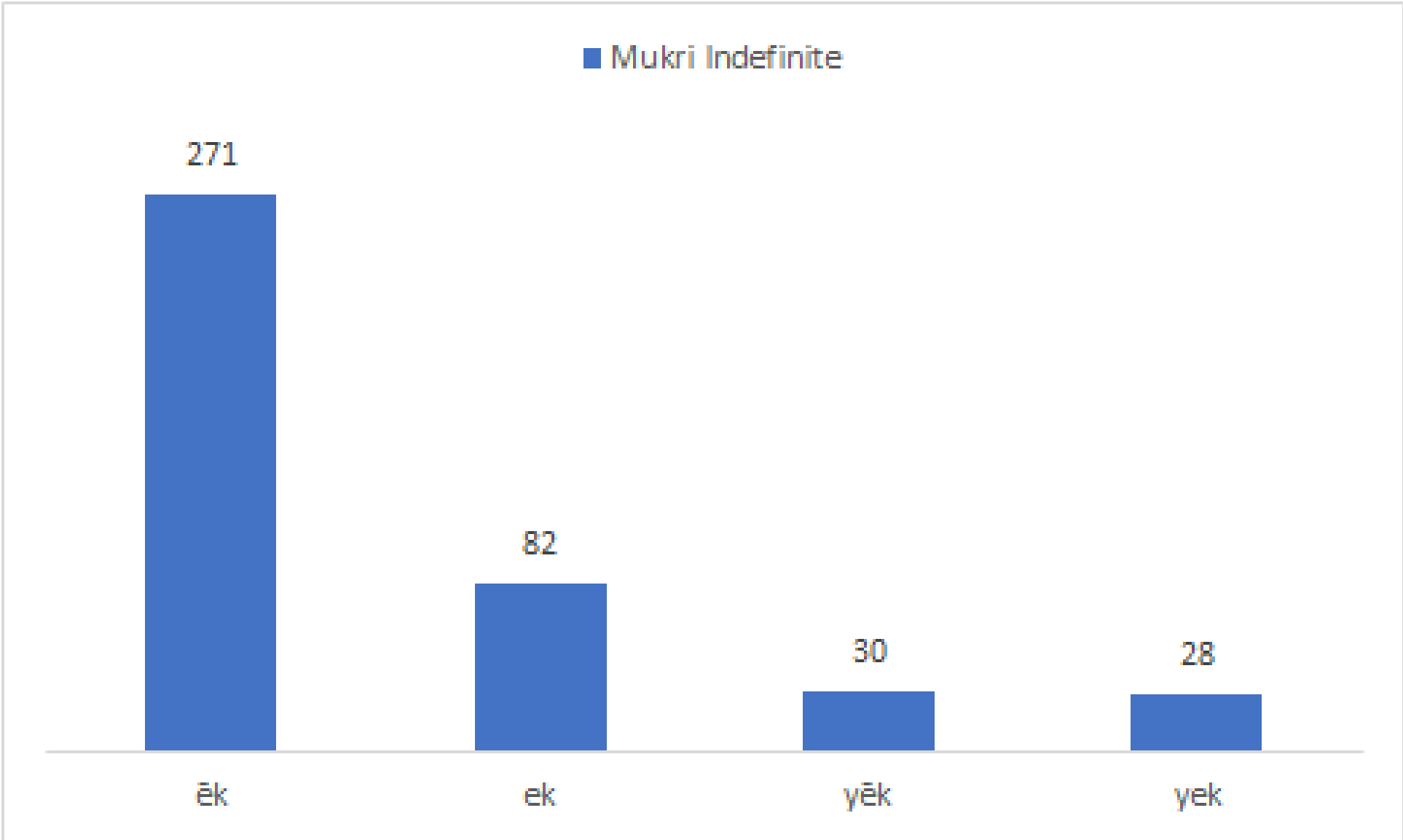
Extraction Process

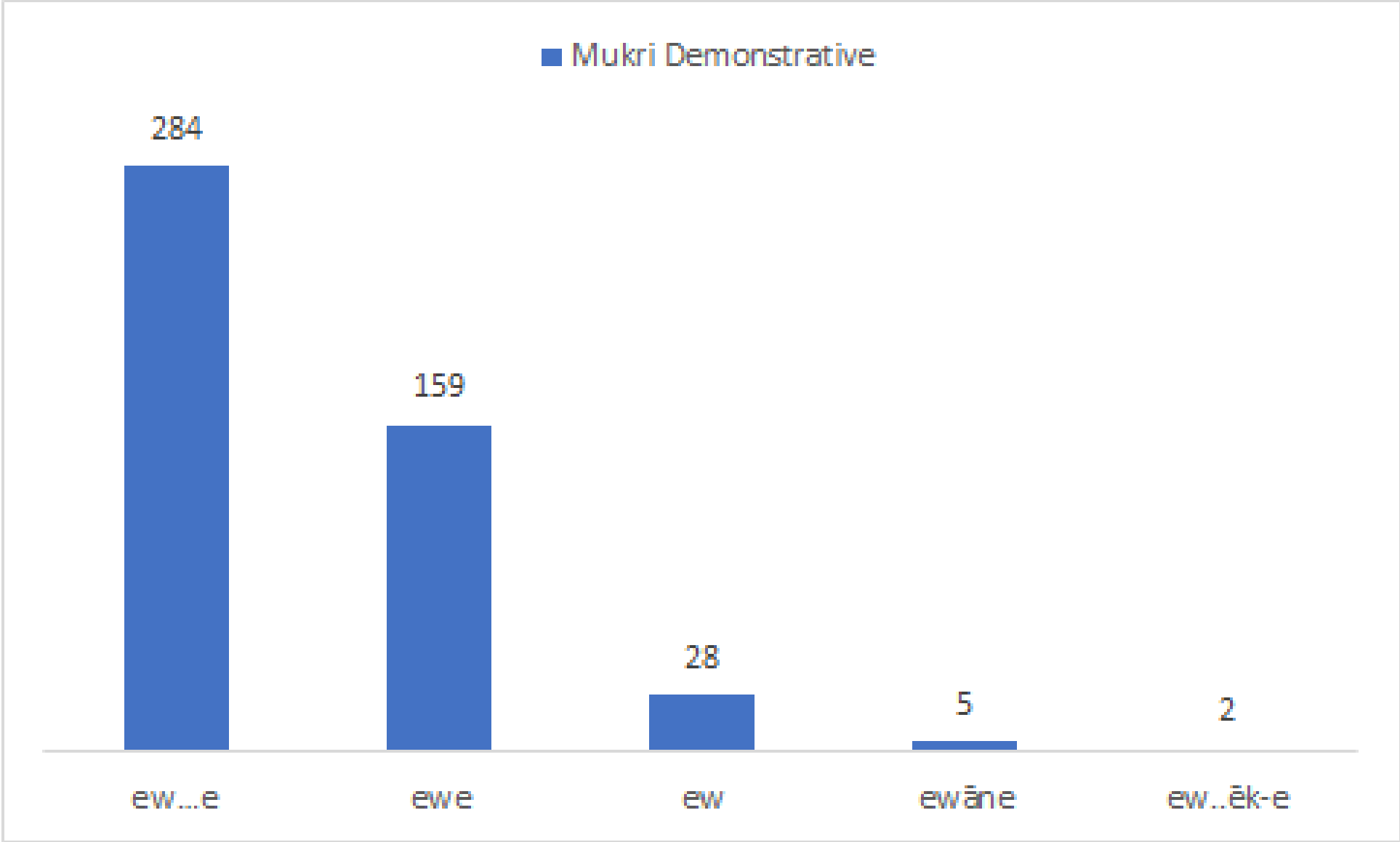






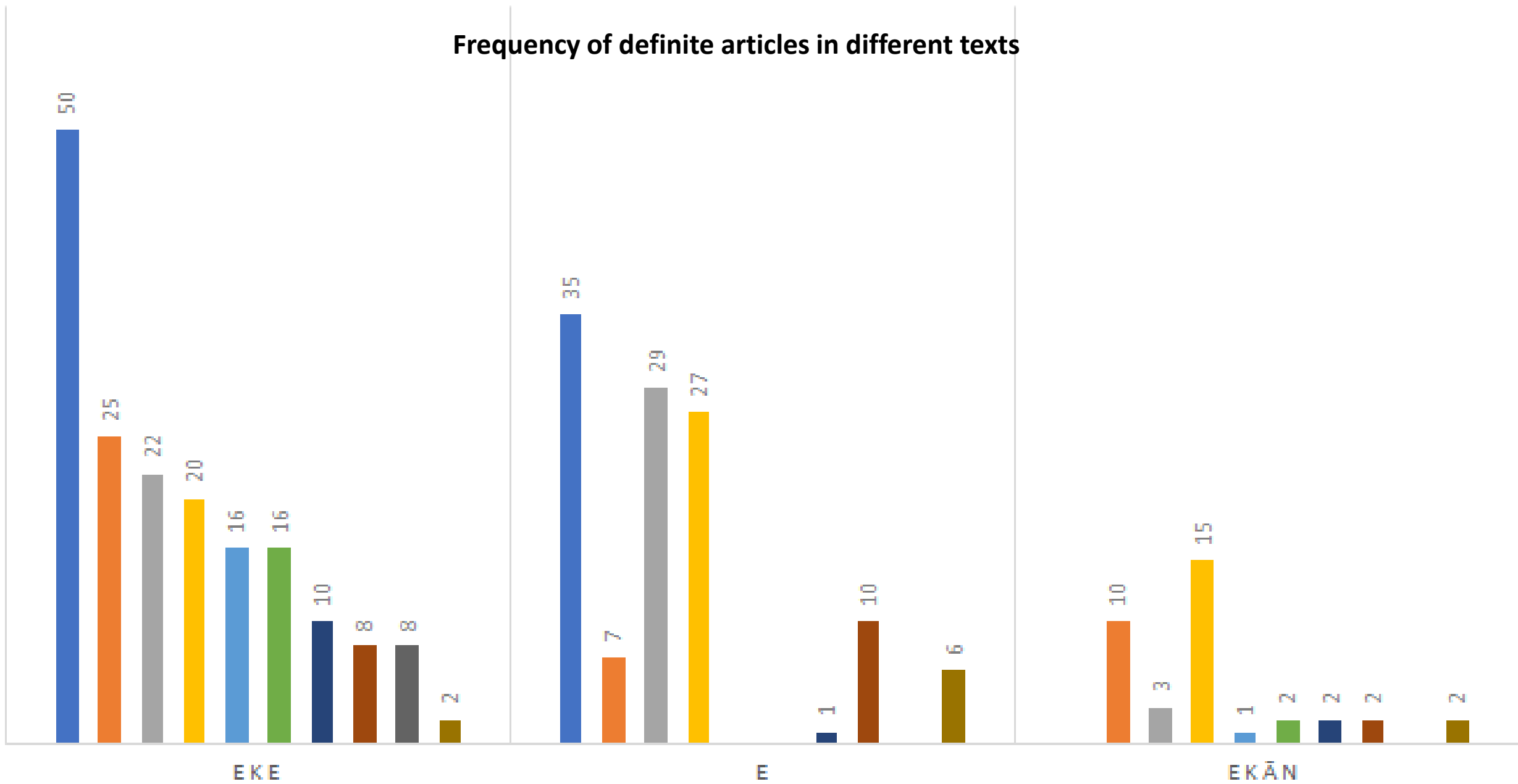






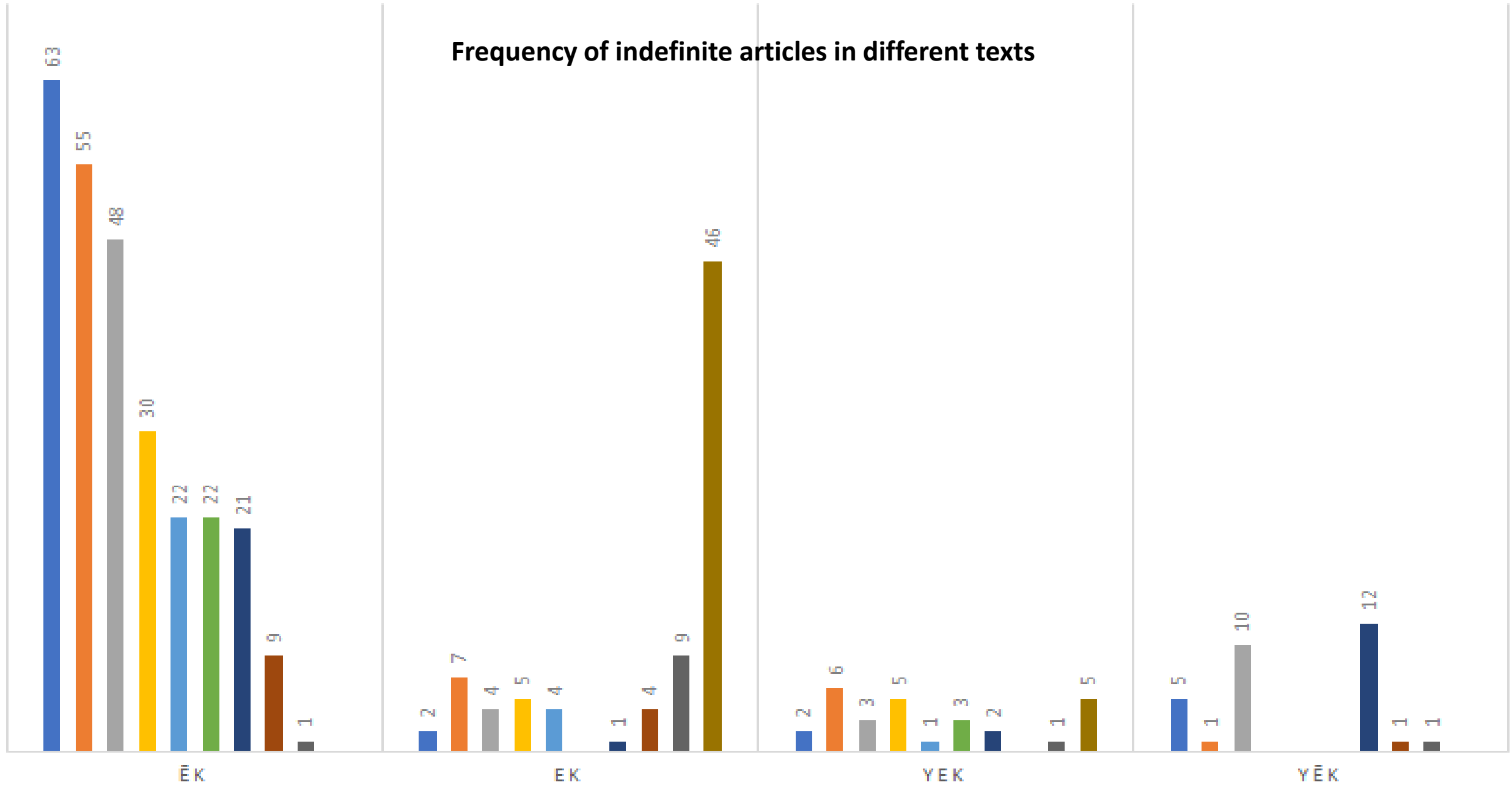
■ Text 8 ■ Text 5 ■ Text 3 ■ Text 10 ■ Text 4 ■ Text 7 ■ Text 6 ■ Text 9 ■ Text 1 ■ Text 2

Frequency of definite articles in different texts

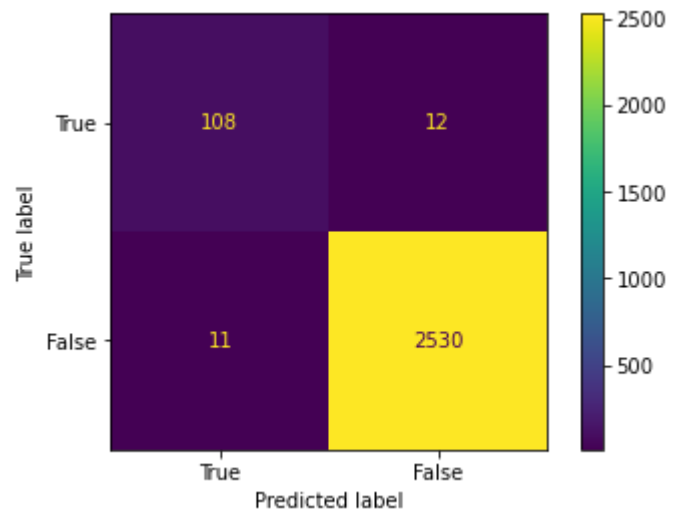


■ Text 3 ■ Text 8 ■ Text 5 ■ Text 10 ■ Text 6 ■ Text 4 ■ Text 1 ■ Text 9 ■ Text 7 ■ Text 2

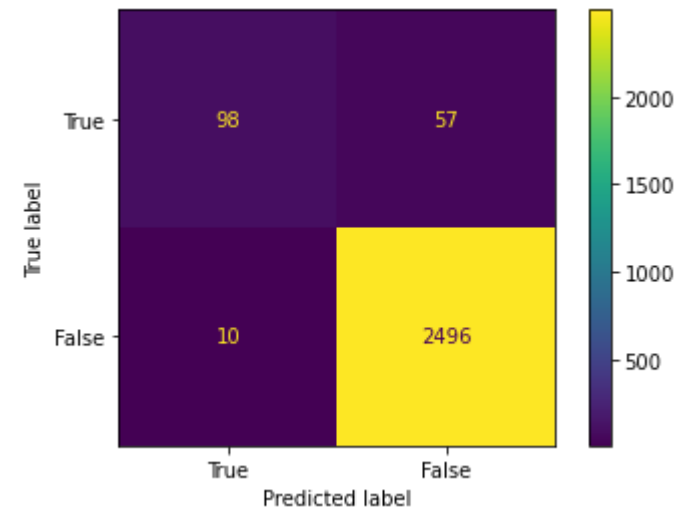
Frequency of indefinite articles in different texts



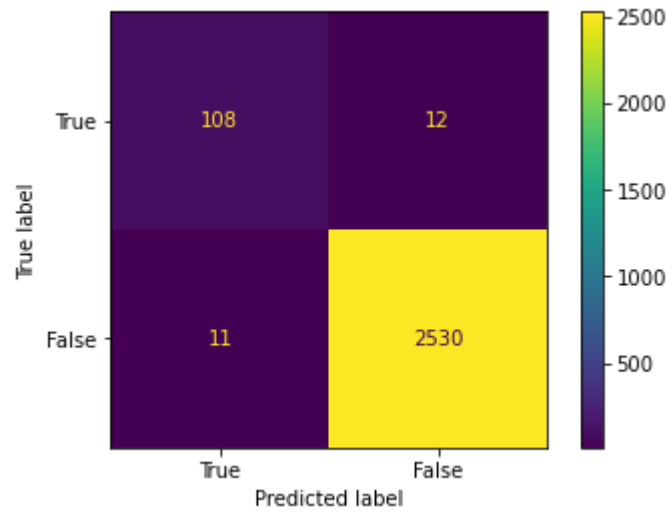
Indefinite



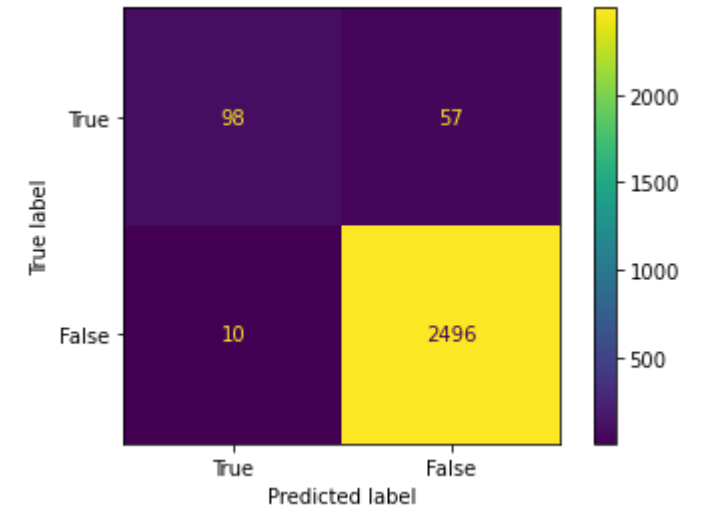
Definite



Indefinite



Definite



	Precision	Recall	F1
INDEFINITE	0.91	0.90	0.90
DEFINITE	0.91	0.63	0.75
Demonstration	0.94	0.97	0.96