

Esmeralda Mustafić

UMJETNA INTELIGENCIJA U LINGVISTIČKIM ISTRAŽIVANJIMA BOSANSKOG JEZIKA

Sažetak

Umjetna inteligencija je danas pronašla svoju primjenu u mnogim aspektima života – od upotrebe u poljoprivredi do stvaranja pametnih „samovozećih“ auta. Samim time, sasvim je prirodno da se upotreba umjetne inteligencije u nekom trenutku počela javljati i u lingvistici i lingvističkim istraživanjima. Upotreba umjetne inteligencije je odavno u upotrebi prilikom prevođenja tekstova sa jednog jezika na drugi, dakle u upotrebi stranih jezika. S druge strane, ovaj rad nastoji prikazati da li je i koliko je moguće umjetnom inteligencijom objasniti lingvistiku bosanskog jezika, dakle, lingvistiku maternjeg jezika. Lingvistička istraživanja pomoću umjetne inteligencije će se prikazati uz pomoć platforme *ChatGPT*. Budući da su ovakva istraživanja izuzetno obimna, rad će se fokusirati samo na sintaksu proste rečenice u bosanskom jeziku.

Ključne riječi: umjetna inteligencija. *ChatGPT*, sintagma, sintaksema

Uvod

Umjetna inteligencija se pojavila kao moćno sredstvo u lingvističkim istraživanjima. Uz pomoć umjetne inteligencije, istraživački su u mogućnosti da analiziraju i razumiju različite aspekte jezika uz veliku preciznost. Integracija umjetne inteligencije u lingvistička istraživanja nudi mnoštvo mogućnosti. Otvara se mogućnost da kompjuter razumije, interpretira i generira tekstove, prepoznaje određene obrasce kao i da prepozna kompleksne lingvističke uzroke i semantičke strukture (Groenwald, E. et al, 2024:1254). *ChatGPT* umjetnu inteligenciju definira kao:

A: Šta je umjetna inteligencija?

ChatGPT: Umjetna inteligencija (UI) je grana računarstva koja se bavi razvojem sistema i algoritama koji mogu obavljati zadatke koji obično zahtijevaju ljudsku inteligenciju. To uključuje učenje iz podataka, donošenje odluka, rješavanje problema, prepoznavanje govora i

slike, razumijevanje prirodnog jezika, pa čak i kreativne procese poput pisanja ili komponovanja muzike.

Dakle, u samim postavkama umjetne inteligencije je razumijevanje prirodnog jezika i prepoznavanje govora.

U lingvistici bosanskog jezika se malo pisalo o upotrebi umjetne inteligencije za istraživanja, dok su se, s druge strane, u lingvistici engleskog jezika napravila opsežna istraživanja o upotrebi umjetne inteligencije u lingvističkim istraživanjima. Jedan od razloga za ovakvu razliku je i u tome što se engleski jezik danas posmatra globalnim. Prema Davidu Crystalu (2003:3) globalni jezik nastaje kada on razvije posebnu ulogu koja je prepoznata u skoro svakoj državi. Budući da je engleski jezik odavno prerastao samo upotrebu u onim državama gdje je maternji, i postao je sredstvom komunikacije u skoro svim aspektima društvenog života, koristi se u poslovnoj komunikaciji, službeni je jezik mnogih evropskih organizacija, jezik je interneta itd. S obzirom na sve navedeno, nije neobično da su se lingvističkim istraživanjima u okviru umjetne inteligencije ponajviše bavile lingviste engleskog govornog područja. Samim time, broj jedinica i metapodataka koje posjeduje *ChatGPT* o engleskom jeziku je zasigurno toliki da se lingvistička istraživanja bazirana na njima mogu vjerodostojno provesti. S druge strane, bosanski jezik sa neuporedivo manjim brojem govornika, vjerovatno ne posjeduje dovoljan broj jedinica niti metapodataka kako bi se lingvistička istraživanja neometano vršila. Cilj ovog rada je prikazati da li *ChatGPT*, kao jedan od najpoznatijih alata umjetne inteligencije, može biti upotrijebljen za lingvistička istraživanja bosanskog jezika, nastoje se prikazati mogućnosti, ali i ograničenja upotrebe umjetne inteligencije.

Umjetna inteligencija u lingvističkim istraživanjima bosanskog jezika

Kako bi se dobila prava slika o mogućnostima i ograničenjima u radu će se prvo postaviti okvir pomoću kojeg će se dobiti standardni parametri o poznavanju bosanskog jezika od strane platforme *ChatGPT*. Samo istraživanje je fokusirano na postavljenje pitanja, te dobijanja odgovora. Pitanja su u početku vrlo jednostavna, te se postepeno usložnjavaju. Od općenitih pitanja o bosanskom jeziku doći će se do sintakse i rečenične analize gdje je potrebno razumijevanje i upotreba svih dotada naučenih jezičnih obrazaca. Time će se dobiti prava slika o mogućnostima ili ograničenjima upotrebe platforme *ChatGPT* za istraživanje bosanskog jezika.

A: Šta je bosanski jezik?

ChatGPT: Bosanski jezik je standardizirani južnoslovenski jezik kojim govore prvenstveno Bošnjaci, ali i drugi narodi u Bosni i Hercegovini, kao i dijelovima Srbije, Crne Gore, Hrvatske i dijaspore. To je jedan od tri službena jezika Bosne i Hercegovine, zajedno sa srpskim i hrvatskim jezikom. Bosanski jezik pripada štokavskoj skupini dijalekata, konkretno njezinoj istočnohercegovačko-krajiškoj varijanti, koja je također osnova za srpski i hrvatski jezik.

A: Koliko glasova postoji u bosanskom jeziku?

ChatGPT: U bosanskom jeziku postoji **30 glasova**, koji su predstavljeni sa **30 slova** u bosanskoj latinici i ćirilici.

A: Šta je fonem?

ChatGPT: Fonem je najmanja jedinica jezika koja ima razlikovnu funkciju. To znači da pomoću fonema razlikujemo značenja riječi.

Ovi glasovi su osnovni i svaki od njih ima jednoznačan grafemski oblik (jedno slovo za jedan glas).

A: Šta je morfem?

ChatGPT: Morfem je najmanja značenjska jedinica jezika. Za razliku od fonema, koji ima razlikovnu funkciju, morfem nosi značenje ili ima gramatičku funkciju.

Vrste morfema

Morfemi se dijele na:

1. Leksički morfemi – nose osnovno značenje riječi.
 - o Na primjer, u riječi "pisati", pis- je leksički morfem jer nosi značenje povezano s pisanjem.
2. Gramatički morfemi – dodaju gramatičko značenje riječi.
 - o Na primjer, u riječi "pisati", nastavak -ati je gramatički morfem koji označava infinitiv glagola.
 - o U riječi "pisaću", nastavak -ću označava prvo lice jednine budućeg vremena.

Primjeri morfema:

U riječi "kućica":

- o kuć- je leksički morfem (osnova riječi, znači "kuća"),
- o -ic- je gramatički morfem koji označava deminutiv (umanjenicu),
- o -a je gramatički morfem koji označava nominativ jednine.

Funkcija morfema:

Morfemi se koriste za oblikovanje riječi, dodavanje značenja ili određivanje gramatičkih kategorija poput vremena, broja, roda i padeža.

Zaključak: Morfem je najmanji dio riječi koji ima značenje ili gramatičku funkciju.

A: Šta je sintagma?

ChatGPT: Sintagma je skup od dvije ili više riječi koje su gramatički povezane i čine smislenu jedinstvo unutar rečenice. Za razliku od rečenice, sintagma ne mora imati potpuni smisao sama za sebe, ali doprinosi izgradnji rečenice.

A: Koji su rečenični članovi u bosanskom jeziku i kako ih dijelimo?

ChatGPT: Rečenični članovi u bosanskom jeziku su gramatičke jedinice koje grade rečenicu i imaju različite funkcije u njoj. Dije se na osnovne i dopunske (sporedne) rečenične članove. Osnovni rečenični članovi u bosanskom jeziku su subjekat i predikat. Oni čine osnovu rečenice i neophodni su za njenu potpunost. Uz njih postoje i izvedeni rečenični članovi kao što su objekat, atribut, apozicija i priloške odredbe.

Na osnovu navedenih primjera pitanja i odgovora može se zaključiti da na neke od osnovnih pojmova koji su vezani za lingvistiku bosanskog jezika *ChatGPT* odgovora korektno i bez grešaka.

A: Uradi analizu rečenice: Azra je pročitala knjigu.

ChatGPT: Konačna analiza rečenice: Azra – subjekat, je pročitala – predikat (glagolski), knjigu – objekat.

A: Uradi analizu rečenice: Azra, studentica, je pročitala novu knjigu.

ChatGPT: Konačna analiza rečenice: Azra – subjekat, studentica – apozicija (dopuna subjekta), je pročitala – predikat (glagolski), novu – atribut (opisuje imenicu "knjigu"), knjigu – objekat.

Na osnovu navedene analize ove dvije rečenice zaključuje se da *ChatGPT* može adekvatno analizirati jednostavne rečenice u kojima nema funkcionalnih preobrazbi, to jest, rečenice u kojima sintakseme odgovaraju rečeničnim funkcijama koje vrše.

A: Uradi analizu rečenice: Majka je neumorno izbacivala guste kolutove dima iz kuće.

ChatGPT:

1. Subjekt

- Majka – subjekat rečenice.
- Odgovara na pitanje: Ko je izbacivao? → Majka.
- Subjekt je imenica u nominativu i označava vršioca radnje.

2. Predikat

- je izbacivala – glagolski predikat.
- Sastoji se od pomoćnog glagola "je" i glagola "izbacivala" u glagolskom pridjevu radnom.
- Označava radnju koja se odvijala u prošlosti.

3. Priloška odredba za način

- neumorno – priloška odredba za način.
- Odgovara na pitanje: Kako je izbacivala? → Neumorno.
- Bliže određuje predikat "izbacivala".

4. Objekat

- guste kolutove dima – objekat rečenice.
- Odgovara na pitanje: Šta je izbacivala? → Guste kolutove dima.
- Objekat je imenica u akuzativu sa svojim atributima.
 - o guste – atribut koji opisuje imenicu kolutove (Kakve kolutove?).
 - o dima – atribut (u genitivu) koji opisuje imenicu kolutove (Čije kolutove? → dima).

5. Priloška odredba za mjesto

- iz kuće – priloška odredba za mjesto.
- Odgovara na pitanje: Odakle je izbacivala? → Iz kuće.
- Dopunjuje predikat i označava mjesto radnje.

Konačna analiza rečenice: Majka – subjekat, je izbacivala – predikat (glagolski), neumorno – priloška odredba za način, guste – atribut (opisuje "kolutove"), kolutove dima – objekat, iz kuće – priloška odredba za mjesto.

U početnom obrazloženju rečenice gdje se analizira svaki član pojedinačno riječ *dim* je tačno analizirana kao atribut u genitivu, iako se ne daju detalji o tome kakav je to atribut. Međutim, u konačnoj analizi rečenice, sintagma *kolutove dima* je iz nekog razloga objedinjena u objekat.

A: Zašto je *kolutove dima* objekat?

ChatGPT: Objekat je rečenicni član koji odgovara na pitanje Šta? ili Koga? (za direktni objekat) u odnosu na glagol u rečenici.

- U ovoj rečenici postavljamo pitanje:
 - Šta je izbacivala? → Kolutove dima.
- Dakle, "kolutove dima" je direktan objekat jer je to predmet koji trpi radnju izbacivanja.

Razrada sintagme:

- Kolutove je glavna riječ (imenica u akuzativu).
- Dima je atribut (imenica u genitivu) koji dodatno određuje riječ kolutove i objašnjava čije kolutove.
- Zajedno čine imenicu sa atributom koja funkcioniše kao objekat.

I u dodatnom obrazloženju je situacija slična prethodno navedenoj. Konstrukcija *kolutove dima* jest jedna sintagma koja je sastavljena od dviju sintaksema koje su povezane funkcionalno i gramatički, međutim, prema sintaksi bosanskog jezika svaka punoznačna sintaksema vrši jednu funkciju u datoj rečenici. Riječ *dima* je sintaksema u genitivu, i ona vrši funkciju nekongruentnog atributa dodatno objašnjavajući riječ *kolutove*, konkretno, radi se o posesivnom genitivu.

A: Šta je nekongruentni atribut?

ChatGPT: Nekongruentni atribut je vrsta atributa (dopune imenici) koji ne slaže svoj oblik sa imenicom koju određuje u rodu, broju i padežu. Za razliku od kongruentnog atributa koji se uvijek slaže sa imenicom (npr. *crvena haljina*), nekongruentni atribut je obično imenica u genitivu ili prijedloška konstrukcija.

Za primjer nekongruentnog atributa u daljem objašnjenju su navedeni primjeri: *haljina od svile, kuća na obali, čovjek sa sela* itd. To znači da je *ChatGPT* u mogućnosti da definira nekongruentni atribut kao takav, ali pri analizi rečenice se nailazi na određeni problem.

A: Uradi analizu rečenice: Ulazeći tiho u svoje domove ljudi su čuli zvuk zvona i poziv na molitvu.

ChatGPT:

1. Subjekat

- ljudi – subjekat rečenice.
- Odgovara na pitanje: Ko je čuo? → Ljudi.
- Subjekat je imenica u nominativu i označava vršioce radnje.

2. Predikat

- su čuli – glagolski predikat.
- Glavna radnja rečenice. Sastoji se od pomoćnog glagola "su" i glavnog glagola "čuli" u glagolskom pridjevu radnom.

3. Objekat

- zvuk zvona i poziv na molitvu – složeni objekat.
 - o Odgovara na pitanje: Šta su čuli? → Zvuk zvona i poziv na molitvu.
 - o zvuk zvona:
 - zvuk – glavna riječ (imenica u akuzativu).
 - zvona – atribut u genitivu (određuje imenicu "zvuk").
 - o poziv na molitvu:
 - poziv – glavna riječ (imenica u akuzativu).
 - na molitvu – atribut (prijedloška konstrukcija koja bliže određuje "poziv").

4. Priloška odredba za način

- tiho – priloška odredba za način.
- Odgovara na pitanje: Kako su ulazili? → Tiho.
- Određuje glagolski prilog "Ulazeći".

5. Priloška odredba za mjesto

- u svoje domove – priloška odredba za mjesto.
- Odgovara na pitanje: Gdje su ulazili? → U svoje domove.

6. Glagolski prilog sadašnji

- Ulazeći – glagolski prilog sadašnji.
- Izražava dodatnu radnju koja se događa istovremeno sa glavnom radnjom (čuli).
- Određuje subjekat "ljudi" (koji su ulazili).

7. Konačna analiza rečenice:

1. Subjekat: ljudi.
2. Predikat: su čuli.
3. Objekat: zvuk zvona i poziv na molitvu.
 - o zvuk – glavna riječ,

- o zvona – atribut,
 - o poziv – glavna riječ,
 - o na molitvu – atribut.
4. Priloška odredba za način: tiho.
 5. Priloška odredba za mjesto: u svoje domove.
 6. Glagolski prilog sadašnji: ulazeći (opisuje radnju istovremenu sa glavnom radnjom).

I u malo kompleksnijoj rečenici se dolazi do sličnih problema, u ovom slučaju *ChatGPT* nije prepoznao ni glagolski prilog sadašnji kao prilošku odredbu za vrijeme, iako je prepoznao da opisuje radnju koja je istovremena sa glavnom radnjom. Sintagme *zvuk zvona* i *poziv na molitvu* su ponovo generalizirane i uvrštene u funkciju objekta, iako su samo sintakseme *zvuk* i *poziv* objekti, dok su sintakseme *zvona* i *na molitvu* nekongruentni atributi, koje u detaljnijoj analizi i definira kao attribute ali se na kraju sve svode na značenje objekta. Sintagma *u svoje domove* je u potpunosti definirana kao priloška odredba mjesta, iako se sastoji od dvije sintakseme koje su obje punoznačne i mogu samostalno vršiti funkciju u rečenici. Sintaksema *svoje* u ovom slučaju ima funkciju kongruentnog atributa.

Rezultat ovakve analize dovodi do zaključka da platforma *ChatGPT* nije u mogućnosti detaljno analizirati rečenicu, to jest, da postoje određena ograničenja pri analizi rečenica. Uzrok tome se vjerovatno nalazi u prethodno navedenoj tezi da sama platforma ne posjeduje dovoljno jedinica i metapodataka kojima bi se poslužila pri analizi rečenice.

Pozitivno je što se platformi može ukazati na greške, te se te greške pri analizi iste rečenice automatski koriguju. Ali da bi se moglo ukazati na greške, potrebno je znanje o istom. Nakon ukazivanja grešaka, *ChatGPT* je ponudio sljedeću analizu:

ChatGPT:

1. Subjekat

- ljudi – subjekat rečenice.
- Odgovara na pitanje: Ko je čuo? → Ljudi.
- Subjekat je imenica u nominativu i označava vršioca radnje.

2. Predikat

- su čuli – glagolski predikat.
 - Sastoji se od pomoćnog glagola "su" i glagola "čuli" u glagolskom pridjevu radnom. •
- Označava radnju koju su izvršili ljudi.

3. Nekongruentni atributi

- zvona – nekongruentni atribut u genitivu.
 - o "Zvona" je imenica koja određuje imenicu "zvuk" i u ovom kontekstu je u genitivu (čije zvuke? → zvuk zvona).
- na molitvu – nekongruentni atribut (prijedloška konstrukcija).
 - o "Na molitvu" je prijedloška konstrukcija koja određuje imenicu "poziv" (poziv na šta? → poziv na molitvu).

4. Kongruentni atribut

- svoje – kongruentni atribut.
 - o "Svoje" je pridjev koji se slaže u rodu, broju i padežu sa imenicom "domove".
 - o "Domove" je u akuzativu množine, a pridjev "svoje" je takođe u akuzativu množine, što pokazuje kongruenciju.
 - o Odgovara na pitanje: Koji domove? → Svoje domove.

5. Priloška odredba za način

- tiho – priloška odredba za način.
- Odgovara na pitanje: Kako su ulazili? → Tiho.
- Bliže određuje glagol "ulazeći".

6. Priloška odredba za vrijeme

- ulazeći – priloška odredba za vrijeme.
- Označava radnju koja se dešavala istovremeno sa radnjom iz glavne rečenice ("su čuli").
- Odgovara na pitanje: Kada su čuli? → Ulazeći.
- Povezan je sa subjektom "ljudi" i opisuje radnju koja se događa tokom druge radnje (slušanje zvuka i poziva).

7. Priloška odredba za mjesto

- u svoje domove – priloška odredba za mjesto.
- Odgovara na pitanje: Gdje su ulazili? → U svoje domove.

8. Konačna ispravljena analiza rečenice:

- Subjekat: ljudi
- Predikat: su čuli

- Nekongruentni atributi:
 - o zvona – atribut u genitivu (zvuk zvona),
 - o na molitvu – atribut u prijedloškoj konstrukciji (poziv na molitvu).
- Kongruentni atribut:
 - o svoje – pridjev u akuzativu množine koji se slaže sa "domove".
- Priloška odredba za način: tiho
- Priloška odredba za vrijeme: ulazeći
- Priloška odredba za mjesto: u svoje domove

Zaključak

Ideja iza ovog rada je bila prikazivanje mogućnosti i ograničenja koje *ChatGPT* može imati prilikom upotrebe za istraživanje lingvistike bosanskog jezika. Samo istraživanje je provedeno komunikacijom autora i platforme. Postavljena su prvo jednostavnija pitanja kako bi se dobili parametri tačnosti. Zatim se rad fokusirao na sintaksu bosanskog jezika, tačnije na sintagmatsku analizu rečenice i određivanje rečeničnih članova.

Došlo se do određenih zaključaka. Platforma vrlo dobro funkcioniše kada su u pitanju osnovne definicije i pojmovi koje treba definirati. Međutim, na određene probleme se nailazi pri dubljoj analizi rečenice i rečeničnih članova gdje platforma jednostavno ne prepoznaje određene sintakse kao zasebne sintaksičke jedinice. To dovodi do zaključka da se *ChatGPT* može koristiti u određene svrhe pri izučavanju bosanskog jezika, ali je to dosta ograničeno, budući da je potrebno i određeno znanje kako bi se isključile greške koje se javljaju.

Literatura:

1. Crystal, D. (2003), *English as a global language*, Cambridge University Press, New York
2. Sibonghanoy Groenewald, E., et al. (2024), Artificial Intelligence in Linguistics Research: Applications in Language Acquisition and Analysis u *Naturalista Campano*, Vol. 28, Issue 1, ISSN:1827-7160, 1253-1262
3. Jahić, Dževad, Halilović, Senahid, Palić, Ismail (2000), *Gramatika bosanskoga jezika*, Zenica, Dom Štampe.
4. Palić, Ismail (2024), *Gramatika bosanskoga jezika (sintaksa)*, Institut za jezik, Sarajevo.

ARTIFICIAL INTELLIGENCE IN LINGUISTIC RESEARCH OF THE BOSNIAN LANGUAGE

Abstract

Artificial intelligence has found its application in many aspects of life today – from its use in agriculture to the creation of smart "self-driving" cars. Therefore, it is natural that the use of artificial intelligence eventually began to appear in linguistics and linguistic research. The use of artificial intelligence has long been applied in the translation of texts from one language to another, that is, in the use of foreign languages. On the other hand, this paper aims to explore whether and to what extent artificial intelligence can explain the linguistics of the Bosnian language, that is, the linguistics of the native language. Linguistic research using artificial intelligence will be presented with the help of the *ChatGPT* platform. Since such research is extremely extensive, this paper will focus only on the syntax of simple sentences in the Bosnian language.

Keywords: artificial intelligence, *ChatGPT*, phrase, syntagma