VÁRADI TAMÁS

ELTE Nyelvtudományi Kutatóközpont
varadi.tamas@nytud.hu
https://orcid.org/0000-0001-5765-3908

Váradi Tamás: Mit tudnak valójában a nagy nyelvmodellek?
Nyelvészeti kérdések az LLM-ek értelmezéséhez
Alkalmazott Nyelvtudomány, Különszám, 2026/1. szám, 144–162.
doi: http://dx.doi.org/10.18460/ANY.K.2026.1.008

# Mit tudnak valójában a nagy nyelvmodellek? Nyelvészeti kérdések az LLM-ek értelmezéséhez

What do Large Language Models actually know? Linguistic questions for the interpretation of LLMs

The release of ChatGPT in late 2022 brought large language models (LLMs) to public prominence with remarkable speed. Although this phenomenal overnight success was due in large part to the system's impressive linguistic capabilities, these were all too quickly taken for granted as attention shifted toward using LLMs mostly as seemingly omniscient information systems. This paper does not address the general intelligence of such systems or their practical applications; instead, it focuses on a narrower question: how should their performance be interpreted from a linguistic perspective?

The paper argues that LLMs do not operate on traditional linguistic units such as morphemes, words, or explicit grammatical categories. Their input consists of statistically derived tokens, and their core learning task is next-token prediction. Their internal operation relies on distributed vector representations and attention-based relational processing rather than explicitly encoded symbolic rules. From this starting point, the paper asks what kind of linguistic knowledge may legitimately be attributed to such systems.

It is argued that traditional linguistic categories can often be recovered from model representations, but this does not in itself prove that such categories are explicitly present in the system. Rather, distributional learning may create stable patterns that can be interpreted in linguistic terms. This claim is illustrated with examples from syntax, semantics, and discourse. LLMs can handle many phenomena traditionally described in structural terms, including agreement, long-distance dependencies, context-sensitive meaning, anaphora, and discourse coherence.

At the same time, the paper emphasizes the limits of a purely distributional explanation. Meaning, reference, grounding, compositionality, systematic generalization, and interpretability remain open problems. The final part addresses the methodological difficulty of interpreting LLMs linguistically: unlike human speakers, these systems do not provide anything comparable to speaker intuitions. For this reason, the paper suggests that current empirical methods should be complemented by elicitation-based approaches analogous to those used in field linguistics. The conclusion is that if LLMs continue to display forms of linguistic behavior that, in human speakers, would count as evidence of substantial grammatical knowledge, this may require a reconsideration of some basic assumptions of theoretical linguistics.

Keywords: large language models; linguistic competence; distributional learning; neural representations; language interpretation

## 1. Bevezetés

Az utóbbi években a nagy nyelvmodellek (large language models, LLM-ek) nemcsak a mesterséges intelligencia kutatásának, hanem a szélesebb nyilvánosság figyelmének középpontjába is kerültek. (A mesterséges intelligencia és a neurális hálózatok történetéhez lásd Héja, 2024, 2026.). Az OpenAI által fejlesztett ChatGPT 2022 novemberi megjelenése rövid idő alatt példátlan felhasználói növekedést hozott: néhány hónapon belül több tízmilliós, majd százmilliós nagyságrendű felhasználói bázis alakult ki, ami jól mutatja, hogy a rendszer által kínált lehetőségek milyen gyorsan és széles körben ragadták meg a felhasználók fantáziáját, és mekkora lelkesedést váltott ki ez a „beszélő", társalkodó rendszer.

A kezdeti benyomás az volt, hogy ezek a rendszerek elsősorban nyelvi feladatok — szövegalkotás, fordítás, összefoglalás — megoldásában nyújtanak új lehetőségeket. Rövid időn belül azonban világossá vált, hogy szerepük ennél jóval általánosabb. A nagy nyelvmodellek ma már nem csupán nyelvi alkalmazásokban jelennek meg, hanem olyan területeken is, amelyeknek közvetlenül semmi közük a nyelvhez — például orvosi képalkotásban vagy közlekedésirányításban –, mégis a természetes nyelv válik a rendszerek irányításának és használatának alapvető eszközévé.

Ebben az értelemben a természetes nyelv – különösen az angol – egyre inkább a különböző mesterséges intelligencia rendszerek közös „működési/működtési nyelvévé" válik: nemcsak a felhasználók kommunikálnak rajta keresztül a rendszerekkel, hanem a fejlesztők is ilyen módon irányítják és hangolják őket.

Az első eufória után az emberek viszonylag gyorsan napirendre tértek a ChatGPT és a gombamód szaporodó rivális modellek nyelvi teljesítménye felett. Ez részben abból fakadt, hogy ezeket a rendszereket nem csupán a fent említett nyelvi feladatokra kezdték használni, hanem egyfajta mindentudó társalgási asszisztenst láttak bennük, de olyan elvárásokat is megfogalmaztak velük szemben, mint a minden témára kiterjedő tárgyi tudás.

Ebben az összefüggésben váltak láthatóvá azok a problémák is, amelyeket ma „hallucinációként" emlegetünk, illetve a generált szövegekben megjelenő torzítások és előítéletek kérdése.

Ez a tanulmány kizárólag a nagy nyelvmodellek teljesítményének nyelvészeti értelmezésére vállalkozik. Nem a modellek általános mesterségesintelligencia-képességeivel és nem a gyakorlati alkalmazásokkal foglalkozik, a mögöttes technológia részleteit pedig csak annyiban érinti, amennyiben ez segíti a központi kérdés megértését: mit jelent az, hogy ezek a rendszerek „nyelvtudással" rendelkeznek?

A kiindulópont az a megfigyelés, hogy a nagy nyelvmodellek nyelvi teljesítménye sok tekintetben meggyőző. A kérdés azonban nem az, hogy mire

képesek nyelvileg, hanem az, hogy ez a teljesítmény hogyan értelmezhető nyelvészeti szempontból.

Mielőtt nekilátnánk az elemzésnek, fontos tisztázni, hogy nyelvészeti szempontból mi tekinthető releváns kérdésnek. A nagy nyelvmodellek teljesítményével kapcsolatban ugyanis számos olyan elvárás is megfogalmazódik, amelyek kívül esnek a nyelvtudás kérdésén.

Igaz ugyan, hogy a nyelvtudás és a világtudás szorosan összefonódik, azt azonban világosan kell látnunk, hogy a tárgyi tévedés, a félrevezetés vagy akár a hazugság nem nyelvtudás kérdése. A nyelvészet az ilyen jelenségekkel legfeljebb annyiban foglalkozik, hogy a beszédaktus-elmélet keretében meghatározza, milyen feltételek mellett tekinthető egy kijelentés például hazugságnak vagy ígéretnek.

Ebben az értelemben nyelvészeti szempontból félrevezető az a megfogalmazás is, hogy a mesterséges intelligencia „hazudik”. Hasonlóképpen, az olyan jelenségek, mint a durva vagy udvariatlan hangnem, nem a nyelvtudás szintjét minősítik.

Ezek a példák arra mutatnak rá, hogy különbséget kell tennünk a nyelvi viselkedés különböző aspektusai között. A továbbiakban a nyelvtudást a klasszikus értelemben vett nyelvi kompetenciával hozzuk összefüggésbe, és elválasztjuk azoktól a performatív vagy pragmatikai jelenségektől, amelyek – Chomsky terminológiájával élve – a nyelvi performancia körébe tartoznak.

## 2. A nagy nyelvmodellek működése mint nyelvészeti kiindulópont

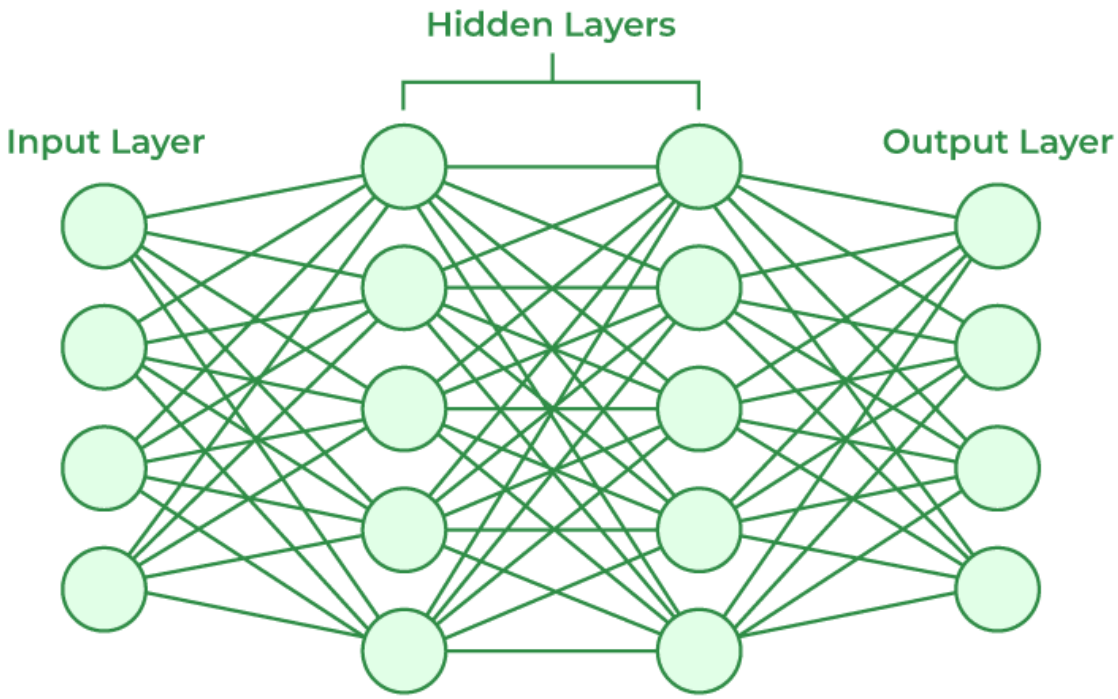
Ebben a fejezetben a nagy nyelvmodellek működésének néhány alapvető aspektusát tekintjük át. (A nagy nyelvmodellek működésének közérthető áttekintéséhez lásd Karpathy, 2025; magyar nyelven pedig Prószéky, 2024; Ligeti-Nagy, 2026; és Yang Zijian, 2026 munkáit.) A cél nem a technológiai részletek teljes körű ismertetése, hanem azoknak az elemeknek a kiemelése, amelyek nyelvészeti szempontból relevánsak, és amelyek nélkül a későbbi értelmezés nem lenne megalapozható.

Mielőtt rátérnénk a tanulási folyamat részleteire, érdemes nagyon vázlatosan megérteni, hogyan épül fel az a számítógépes rendszer, amellyel a nyelvmodellt előállítják.

A nagy nyelvmodellek neurális hálózatok: olyan számítási rendszerek, amelyek sok egymásra épülő rétegen keresztül alakítják át a bemenetet kimenetté.

A tanulás bemeneteként szolgáló nagyméretű szövegkorpusz tokenei a tanulás kezdetén nagy dimenziójú, valós számokat tartalmazó vektorokká alakulnak, és a továbbiakban a tanulás ezekkel a vektorokkal végzett átalakítások révén történik (1. ábra).

1. ábra. A neurális hálózat sematikus ábrája: a bemeneti szöveg reprezentációja több rejtett rétegen keresztül alakul át, amelyekben a modell fokozatosan megtanulja a nyelvi mintázatokat, végül pedig egy valószínűségi kimenetet ad a következő tokenre. (Forrás: W1)




Maga a hálózat nagyszámú mesterséges neuronból áll, amelyek rétegekbe szerveződve kapcsolódnak egymáshoz. A kapcsolatokhoz tartozó súlyok határozzák meg, hogy az egyes bemenetek milyen mértékben befolyásolják a további feldolgozást. A tanulás során ezek a – szintén vektorokból álló – súlyok módosulnak, és ennek eredményeként alakulnak át fokozatosan a tokenek vektorreprezentációi, és ezekben a reprezentációkban jelenik meg az, amit a modell „nyelvi tudásának” nevezhetünk.

## 2.1. A tanulási folyamat vázlatos áttekintése

A tanulás bemeneteként nagy méretű szövegkorpusz szolgál, amelyet a rendszer először kisebb egységekre, úgynevezett tokenekre bont – ezek a feldolgozás alapegységei, amelyek nem esnek egybe a hagyományos nyelvi kategóriákkal, és amelyekről részletesebben a későbbiekben lesz szó.

A tanulási feladat rendkívül egyszerűen megfogalmazható: a rendszernek minden egyes lépésben azt kell megtanulnia, hogy egy adott szövegkörnyezet után melyik token következik a legnagyobb valószínűséggel.

A tanulás során a modell minden lépésben becslést ad arra, hogy az adott kontextus után az egyes tokenek milyen valószínűséggel következhetnek. Ezt az

előrejelzést összeveti a tényleges folytatással, majd a különbség alapján módosítja belső paramétereit annak érdekében, hogy a következő lépésben pontosabb becslést adjon.

A folyamat iteratív módon ismétlődik, és addig tart, amíg a modell megfelelő pontossággal képes a következő tokenek előrejelzésére.

## 2.2. A modell előtanítása

A nyers korpuszon végzett tanítás eredményét előtanított vagy alapmodellnek nevezzük. A tanulási feladat természetéből adódóan az ilyen modell „tudása” arra irányul, hogy a megelőző kontextus alapján a lehető legpontosabban megbecsülje a következő token valószínűségét, vagyis egy úgynevezett promptként adott indító szöveget folytasson.

Ebben az értelemben, némileg provokatív módon, indokolt úgy fogalmazni, hogy az alapmodell egyfajta „internet-szöveg szimulátor” (Karpathy, 2025): olyan rendszer, amely elsajátítja a tanulási adatokban rejlő, látens disztribúciós mintázatokat, és ezek alapján állít elő szövegfolytatásokat.

Fontos tisztán látnunk, hogyan értendő ebben az összefüggésben a szimuláció fogalma. A modell tanítása során nem az a cél, hogy a rendszer a tanító szövegeket szó szerint felidézze. Valójában túltanulásnak hívják, és kifejezetten kerülendőnek tartják a fejlesztők, ha a rendszer csak a tanító korpusz szövegeit tanulja meg, mintegy változatlan formában megjegyezve azokat.

A tanulás célja ezzel szemben az, hogy a rendszer új szövegeket legyen képes előállítani, amelyek a tanító korpusz nyelvi sajátosságait tükrözik: hasonló mintázatokat követnek szintaktikai, szemantikai és stiláris szempontból. Ebben az értelemben beszélhetünk arról, hogy a modell „szimulálja” az interneten megjelenő szövegeket.

A szimuláció „szorossága” a modell használata során szabályozható. Az úgynevezett hőmérséklet paraméter beállításával befolyásolható, hogy a generált szöveg mennyire kövesse szorosan a tanult disztribúciós mintázatokat, illetve mennyire térjen el azoktól kreatívabb megoldások irányába.

## 2.3. Alapmodell és asszisztens modell

A betanított alapmodell már képes jól formált és koherens szövegek előállítására, de működése alapvetően szövegfolytatásra korlátozódik: egy adott promptból kiindulva generál további szöveget. Ez önmagában is jelentős előrelépést jelentett a korábbi nyelvfeldolgozó rendszerekhez képest, de használata korlátozott marad.

A modell nem arra van megtanítva, hogy a felhasználó utasításait hajtsa végre, például hogy egy szöveget összefoglaljon, vagy hogy konkrét kérdésekre válaszoljon, hanem minden esetben a bemenet legvalószínűbb folytatását állítja elő. Ennek következtében a válaszok gyakran nem illeszkednek a felhasználói

szándékhoz, vagy nem abban a formában jelennek meg, ahogyan azt egy interaktív rendszer esetében elvárnánk.

Ez a korlát vezetett ahhoz, hogy a modellek további tanítására volt szükség: olyan eljárások kidolgozására, amelyek révén a rendszer viselkedése jobban igazítható a felhasználói elvárásokhoz és a konkrét feladatokhoz.

## 2.4. Az alignment probléma

A fenti korlát általánosabban az úgynevezett alignment problémához vezet. Ennek lényege, hogy a modell tanulási célja – a következő token valószínűségének minél pontosabb becslése – nem esik egybe azzal, amit a felhasználó egy valódi asszisztenstől vár.

A felhasználó tipikusan nem pusztán szövegfolytatást vár, hanem értelmes, releváns és a kommunikációs helyzethez illeszkedő választ. Az alapmodell azonban nem rendelkezik ilyen értelemben vett célokkal: működését kizárólag a tanulás során elsajátított disztribúciós mintázatok határozzák meg.

Ez a különbség azt eredményezi, hogy a modell viselkedése gyakran eltér a felhasználói szándéktól. Előfordulhat, hogy a válasz formailag helyes, de nem releváns; vagy éppen meggyőző, de tartalmilag pontatlan.

Az alignment probléma tehát nem egyszerűen technikai kérdés, hanem annak a következménye, hogy a tanulási cél és a használati cél nem azonos. A további fejlesztések célja éppen ennek az eltérésnek a csökkentése.

Technikailag itt sem jelenik meg gyökeresen új mechanizmus: továbbra is a 2.1. részben leírt tanulási eljárásról van szó, csak más típusú adatokon — párbeszédeken, instrukciókon és az ezekre adott válaszokon. Míg az alapmodell az interneten megjelenő szövegek folytatásait tanulja meg szimulálni, addig az asszisztensmodell emberi szakértők által előállított válaszmintákon tanul, és ebben az értelemben a szakértői válaszadás nyelvi és pragmatikai mintázatait sajátítja el.

Fontos hangsúlyozni: a modell nem válik szakértővé abban az értelemben, hogy explicit tudással rendelkezne, hanem példák alapján sajátít el válaszmintázatokat. Ugyanakkor ezekből a párbeszédes adatokból olyan, hagyományosan pragmatikai jellegű tényezőket is megtanul kezelni, mint a regiszterek, stílusok vagy beszédaktusok.

## 2.5. Megerősítéses tanulás és preferenciák

A finomhangolás egy további lépése a megerősítéses tanulás emberi visszacsatolással (reinforcement learning from human feedback, RLHF), amely során a modell azt tanulja meg, hogy az emberek milyen válaszokat preferálnak. Ennek eredményeként a rendszer viselkedése nemcsak az asszisztensi mód

betanításához használt korpuszban megfigyelhető nyelvi mintázatokhoz, hanem az emberi értékelésekhez is igazodik.

Ez fontos következménnyel jár: a modell által generált válaszok nem pusztán valószínűségi alapon, hanem preferenciák szerint optimalizáltak. Jól definiált, zárt feladatokban az ilyen tanulási eljárások akár az emberi teljesítményt is meghaladhatják, mert a rendszer maga is ellenőrizni tudja, hogy helyes-e a megoldása – klasszikus példa erre a go. A nyitottabb nyelvi feladatok esetében azonban nem ez a helyzet: itt a reális cél nem valamiféle objektív optimum elérése, hanem az, hogy a modell válaszai minél jobban közelítsék az emberi preferenciákat.

## 2.6. Mi a modell és hogyan használjuk

A betanítás után előálló modell technikai értelemben egy rögzített állapot: egy nagyméretű paraméterhalmaz, amely vektorreprezentációk formájában kódolja a tanult mintázatokat. A nyelvi modell „tudása” ebben a paramétertérben testesül meg.

A nyelvi modell nem úgy működik, mint a hagyományos szoftverek, amelyek a programkódban előre meghatározott utasítások végrehajtásával oldanak meg konkrét feladatokat. A modell nem tartalmaz explicit módon megfogalmazott szabályokat vagy eljárásokat.

A használathoz a betanítás során meghatározott paramétereket egy neurális hálózatba kell betölteni, amely ezután minden egyes lépésben kiszámítja, hogy az adott bemenet alapján mely kimenetek a legvalószínűbbek. A működés ebben az értelemben hasonlít a tanulás során végzett számításokra, azzal a lényeges különbséggel, hogy ilyenkor a paraméterek már nem változnak.

Fontos hangsúlyozni, hogy a modell minden egyes használat során ehhez a rögzített állapothoz tér vissza. A rendszer nem halmoz fel tartós tudást a használat során: bár egy adott párbeszéden belül képes információkat „megjegyezni”, ezek a hatások nem épülnek be tartósan a modell paramétereibe.

# 3. Reprezentáció és nyelvi struktúra

A nagy nyelvmodellek megértésében kulcsszerepet játszik az a kérdés, hogy milyen egységekben és milyen formában reprezentálják a nyelvet. A technológiai megoldások ezen a ponton különösen élesen eltérnek a nyelvészet hagyományos fogalmaitól, és ez az eltérés alapvető következményekkel jár a későbbi értelmezésre nézve.

## 3.1. A tokenizáció mint alapvető reprezentációs döntés

A tokenizáció nem egyszerű előfeldolgozási lépés, hanem a modell szempontjából a nyelv alapegységeinek meghatározása. Mielőtt a tanulás

egyáltalán megkezdődne, eldől, hogy milyen egységek lesznek azok, amelyekkel a rendszer operálni fog. Ez a döntés nem pusztán technikai jellegű: nyelvészeti és nyelvfilozófiai szempontból is alapvető következményekkel jár.

A nyelvészet klasszikus leírása hierarchikus szerkezetet feltételez. A beszéd hangokból vagy betűkből épül fel, ezekből morfémák és szavak, majd nagyobb szerkezetek jönnek létre. A nyelv kettős artikulációja (Martinet, 1960) éppen abban áll, hogy egy véges számú, önmagában jelentéssel nem bíró elem kombinációjából potenciálisan végtelen számú jelentéssel bíró egység állítható elő.

A nagy nyelvmodellek esetében azonban a nyelv nem ilyen értelemben vett egységekre bomlik. A modell számára a nyelv egy egydimenziós szekvencia, amelynek elemei nem a nyelvészeti elemzés kategóriáiból származnak, hanem a feldolgozás hatékonyságának szempontjai szerint kerülnek kialakításra.

## 3.2. Byte-ok, szekvenciák és a tokenkészlet kialakulása

A nyers szöveg a modell számára kezdetben karaktersorozat, pontosabban byte-ok sorozata. Egy ilyen reprezentáció esetében a szekvencia minden egyes pontján 256 lehetséges karakter fordulhat elő, ami egy rendkívül hosszú és számítási szempontból nehezen kezelhető struktúrát eredményez.

A feldolgozás hatékonysága érdekében ezért a karaktereket statisztikai alapon összevonják. Az egyik elterjedt eljárás a byte pair encoding (BPE), amely iteratív módon egyesíti a leggyakrabban együtt előforduló elemeket (Sennrich et al., 2016). Először a leggyakoribb byte-párokból lesznek új egységek, majd ezek az egységek további összevonásokban vesznek részt: byte–byte, majd byte–token, végül token–token kombinációk hoznak létre új tokeneket.

Ez a folyamat addig folytatódik, amíg el nem érünk egy előre meghatározott szótárméretet, amely tipikusan több tízezer elemet tartalmaz. A folyamat során a szekvencia hossza csökken, miközben a szótár mérete nő: a modell egy kompromisszumot keres a feldolgozási hatékonyság és a reprezentáció részletessége között.

Ennek eredményeként a modell bemenete nem karakterek, és nem is szavak sorozata, hanem egy olyan tokenekből álló szekvencia, amely statisztikai alapon kialakított egységekből épül fel.

## 3.3. A tokenek nyelvészeti státusza

A létrejövő tokenek csak részben fedik le a nyelvészeti értelemben vett egységeket. Egyes tokenek valóban teljes szavakat reprezentálnak, mások szóelemeket, megint mások több szó részleteit vagy gyakori karakterkombinációkat. A tokenek tehát nem tekinthetők sem morfémáknak, sem szavaknak, sem más hagyományos nyelvi egységeknek.

Ez a megfigyelés önmagában még nem lenne különösebben problematikus. A kritikus pont az, hogy a modell kizárólag ezekkel az egységekkel dolgozik. A nyelv a rendszer számára nem más, mint tokenek sorozata, és minden további reprezentáció erre a szintre épül.

Ennek következtében a modell alapegységeinek jelentős része nem rendelkezik önálló, stabil jelentéssel. A tokenkészlet nagy hányada olyan egységekből áll, amelyek nem feleltethetők meg a nyelvészeti értelemben vett nyelvi jelnek, amelynek alakja és jelentése is van. Ha a modell mégis képes jelentéssel bíró egységeket vagy szerkezeteket kezelni, akkor ezek nem előfeltételek, hanem a tanulás során kialakuló emergens jelenségek.

Ez a pont különösen fontos a további érvelés szempontjából. A nyelvészeti elemzés hagyományosan olyan egységekből indul ki, mint a morfémák és a szavak, amelyek jelentéssel bírnak. A nagy nyelvmodellek ezzel szemben olyan reprezentációból indulnak ki, amelyben az alapegységek jelentős része nem ilyen természetű. Ez felveti a kérdést, hogy milyen értelemben beszélhetünk egyáltalán nyelvi reprezentációról ezekben a rendszerekben.

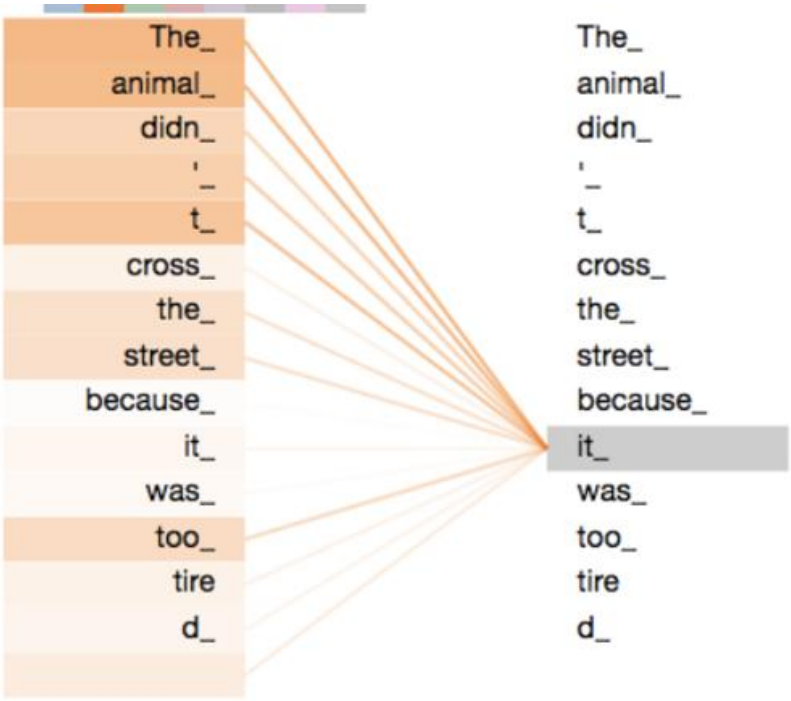
## 3.4. A figyelmi mechanizmus (attention) mint a relációk tanulásának helye

A fenti lépések után eljutunk ahhoz a ponthoz, ahol a modell működésének nyelvészeti szempontból talán legfontosabb eleme jelenik meg. Ha a tokenizáció meghatározza, hogy milyen egységekkel dolgozik a rendszer, akkor az attention mechanizmus határozza meg azt, hogy ezek között az egységek között milyen viszonyokat képes felismerni és kihasználni. Ebben az értelemben itt történik az, amit – bizonyos megszorításokkal – nyelvtanulásnak nevezhetünk.

A modell tanulási célja formálisan változatlanul egyszerű: a következő token előrejelzése. Ennek a feladatnak a sikeres megoldása azonban csak akkor lehetséges, ha a rendszer képes figyelembe venni, hogy az aktuális token milyen kapcsolatban áll a kontextus más elemeivel. Ezek a kapcsolatok nem korlátozódnak a közvetlen szomszédságra: gyakran távoli elemek között jönnek létre, és többféle reláció egyidejű kezelését igénylik.

Az attention mechanizmus éppen ezt teszi lehetővé. A modell minden feldolgozási lépésben súlyozza a kontextus különböző pontjait, és ezek alapján integrálja az információt. Így a feldolgozás nem lineáris értelemben történik, hanem relációs módon: a rendszer nem egyszerűen egymás után következő elemekkel dolgozik, hanem egy olyan hálózatot épít fel, amelyben a szekvencia különböző pontjai különböző mértékben hatnak egymásra (Vaswani et al., 2017).

2. ábra. Az attention mechanizmus működése egy példamondaton. Az „it” token reprezentációja a feldolgozás során a kontextus különböző elemeihez kapcsolódik. A súlyozott kapcsolatok mintázata azt mutatja, hogy a modell nem lineáris szekvenciaként, hanem relációs hálózatként dolgozza fel a szöveget. (Forrás: W2)




Ez a folyamat ráadásul dinamikus. A modell egymást követő rétegeiben a tokenekhez tartozó vektorreprezentációk folyamatosan módosulnak: új információk épülnek be, a korábbiak átértékelődnek, és a releváns kapcsolatok fokozatosan megerősödnek. Ennek eredményeként a reprezentáció egyre inkább a teljes kontextus által meghatározott állapotot tükrözi. A jelentés ebben az értelemben nem egyetlen lépésben jelenik meg, hanem a feldolgozás során, rétegről rétegre alakul ki.

Ezt a működést szemlélteti a 2. ábra, amely különböző attention-fejek mintázatait mutatja. Az ábrán jól látható, hogy egyes mechanizmusok inkább lokális kapcsolatokra érzékenyek, míg mások távolabbi elemek között hoznak létre összefüggéseket. A modell tehát nem egyetlen, jól elkülöníthető szerkezetet épít fel, hanem párhuzamosan többféle relációs mintázatot alakít ki, amelyek együtt teszik lehetővé a komplex nyelvi viselkedést.

A tanulás eredménye nem explicit szabályok formájában jelenik meg, hanem a tokenekhez tartozó vektorreprezentációkban. Ezek a reprezentációk nagy dimenziós, folytonos terekben helyezkednek el, és a különböző relációk hatására alakulnak. A modell tehát nem „tárolja” külön azt, hogy egy adott elem milyen

nyelvtani vagy szemantikai szerepet tölt be, hanem olyan reprezentációt hoz létre, amelyben ezek a különbségek implicit módon kódolódnak.

A fenti működés azt mutatja, hogy a nyelvi információ nem előre adott kategóriák formájában jelenik meg, hanem a feldolgozás során, relációk hálózataként alakul ki. A modell által létrehozott reprezentációk így nem közvetlenül feleltethetők meg a nyelvészet hagyományos egységeinek, hanem azok egy lehetséges, elosztott realizációjaként értelmezhetők. A következőkben azt vizsgáljuk meg, hogy mit jelent egyáltalán „nyelvészeti kategóriákról" beszélni ezeknek a modelleknek az esetében, és hogy ezek a kategóriák a rendszer belső szerkezetének vagy inkább az elemzés módjának tulajdoníthatók-e.

## 4. Nyelvi reprezentáció és értelmezés az LLM-ekben

Az előző fejezetben láttuk, hogy a nagy nyelvmodellek működésének kulcsa a nyelvi relációk tanulása: a tokenek reprezentációja a feldolgozás során fokozatosan alakul ki a kontextushoz való kapcsolódások révén. Az attention mechanizmus lehetővé teszi, hogy a rendszer a szekvencia különböző pontjai között dinamikusan építsen ki kapcsolatokat, és ezek alapján módosítsa a reprezentációkat.

Ez a kép azonban önmagában még nem válaszolja meg azt a kérdést, amely nyelvészeti szempontból a legfontosabb: milyen természetű az a tudás, amely ezekben a reprezentációkban megjelenik?

### 4.1. Egy konkrét eset: koreferencia-feloldás

A kérdés megközelítéséhez érdemes egy egyszerű példát megvizsgálni. Tekintsük a 2. ábrában szereplő két mondatot:

*The animal didn't cross the street because it was too tired.*
*The animal didn't cross the street because it was too wide.*

Az „it" névmás értelmezése nem triviális: a mondat több lehetséges referenst is tartalmaz, és a helyes értelmezés a kontextus finom különbségeitől függ, például a „tired" és a „wide" jelzők eltérő szelekciós korlátaiból adódóan.

A modell nem egyetlen lépésben „dönt" a koreferenciáról. A modell tanításának korai szakaszában az „it" reprezentációja még alig különbözik más elemekétől. A későbbi rétegekben azonban fokozatosan erősödnek azok a kapcsolatok, amelyek a releváns kontextuselemekhez kötik. A reprezentáció így egyre inkább összhangba kerül azzal az értelmezéssel, amely a teljes mondat alapján a legvalószínűbb.

Ez a folyamat jól mutatja, hogy a modell működése nem diszkrét döntések sorozata, hanem folyamatos reprezentációk dinamikus átalakulása. A „jelentés"

ebben az értelemben nem egy előre adott tulajdonság, hanem a kontextus által formált állapot. Ez a megfigyelés közvetlenül felveti a következő kérdést: vajon ez a viselkedés egy mögöttes, szimbolikus szerkezet következménye, vagy elegendő hozzá a disztribúciós információ ilyen típusú feldolgozása?

## 4.2. Mit jelent itt a „tudás”?

A fenti példa alapján könnyen megfogalmazható az a benyomás, hogy a modell „tudja”, hogy az ‘it’ token névmásként viselkedik. Ez a megfogalmazás azonban többértelmű, és könnyen félrevezető lehet.

Amikor ilyen állításokat teszünk, valójában két különböző dolgot állíthatunk:

a modell reprezentációi tartalmaznak olyan információt, amely alapján a helyes értelmezés előállítható;

a modell explicit módon reprezentálja azokat a kategóriákat és szabályokat, amelyekkel ezt az értelmezést a nyelvészet leírja.

A két állítás nem ekvivalens. Az empirikus eredmények az elsőt erősen alátámasztják, a másodikat viszont nem igazolják egyértelműen.

## 4.3. Mit mutatnak a vizsgálati módszerek?

A modellek belső reprezentációinak vizsgálatára az egyik legelterjedtebb módszer az úgynevezett szondázás (probing). Ennek lényege, hogy a modellt változatlanul hagyva egy egyszerű, külső modellt illesztünk a belső reprezentációkra, amely egy adott tulajdonságot – például szófaji kategóriát vagy szintaktikai szerepet – képes előre jelezni, és amelynek teljesítménye alapján következtethetünk arra, hogy az információ mennyire van jelen a reprezentációkban.

Ha ez az egyszerű modell, a szonda, jól teljesít, az arra utal, hogy az adott információ valamilyen formában jelen van a reprezentációban. Fontos azonban hangsúlyozni, hogy ez a módszer nem azt mutatja meg, hogy a modell ténylegesen használja-e ezt az információt, hanem azt, hogy az mennyire könnyen nyerhető ki (Belinkov et al., 2017; Hewitt & Manning, 2019).

Ez a különbség alapvető. A szondázás eredményei azt mutatják, hogy a modellek reprezentációi gazdagok és strukturáltak, de nem döntik el azt a kérdést, hogy ezek a struktúrák a rendszer belső működésének részei-e, vagy inkább az elemzés során konstruált leírások.

## 4.4. Disztribúciós tanulás és/vagy szimbolikus reprezentáció?

Ezzel eljutunk a jelen tanulmány központi elméleti kérdéséhez.

A klasszikus nyelvészeti megközelítés szerint a nyelv diszkrét szimbolikus kategóriák és szabályok rendszere. A disztribúciós megközelítés ezzel szemben azt hangsúlyozza, hogy a nyelvi jelenségek a használat mintázataiból vezethetők le (Harris, 1954; Firth, 1957).

A nagy nyelvmodellek viselkedése mindkét értelmezést látszólag alátámasztja. Egyfelől képesek olyan műveletekre, amelyek szimbolikus jellegűeknek tűnnek: például referenciák követésére vagy strukturális összefüggések kezelésére. Másfelől azonban működésük alapja egyértelműen disztribúciós: a tanulás során kizárólag a tokenek együttelőfordulási mintázatait használják.

A jelenlegi eredmények alapján úgy tűnik, hogy a modellek reprezentációiból viszonylag egyszerű leképezésekkel visszanyerhetők a hagyományos nyelvészeti kategóriák. Ez azonban nem feltétlenül jelenti azt, hogy ezek a kategóriák explicit módon jelen lennének a rendszerben. Inkább arról lehet szó, hogy a disztribúciós tanulás olyan reprezentációs teret hoz létre, amelyben stabil mintázatok jelennek meg, amelyek hagyományos nyelvészeti kategóriákként és szerkezetekként értelmezhetők (vö. Rigotti et al., 2013).

Ez a megfigyelés több fontos következménnyel jár. Egyrészt arra utal, hogy a nyelvészeti kategóriák nem feltétlenül a rendszer alapegységei, hanem olyan leíró konstrukciók, amelyek bizonyos szinten jól megragadják a viselkedést. Másrészt felveti annak a lehetőségét, hogy a szabályszerűnek tűnő nyelvi jelenségek részben disztribúciós alapokon is rekonstruálhatók. Ugyanakkor a kérdés nyitva marad: nem egyértelmű, hogy a modell működése kimeríti-e mindazt, amit a nyelvészet a nyelvről állít, vagy csak annak bizonyos aspektusait ragadja meg.

A kérdés így pontosabban az, hogy a nyelvészeti kategóriák a modell működésének belső elemei-e, vagy olyan mintázatok, amelyeket a disztribúciós reprezentációból utólag azonosítunk, azaz mi vetítjük rá őket a modell viselkedésére.

Ez a kérdés egyszerre empirikus és elméleti. Empirikus abban az értelemben, hogy a modellek viselkedése vizsgálható és mérhető. Elméleti pedig abban, hogy a kapott eredmények értelmezése nem független attól a fogalmi kerettől, amelyben a nyelvet leírjuk.

A fenti kérdés nem csupán módszertani természetű, hanem közvetlen következményekkel jár arra nézve is, hogyan értelmezzük a nyelvi struktúrát ezekben a rendszerekben. A következőkben ezt a kérdést a disztribúciós mintázatok és a nyelvi szerkezet viszonyának szempontjából vizsgáljuk meg.

## 5. Disztribúciós mintázatok és nyelvi struktúra

Az előző fejezetben megfogalmazott állítás – miszerint a nyelvi kompetencia jelentős része disztribúciós információból modellezhető – csak akkor tartható fenn, ha meg tudjuk mutatni, hogy ez az információ valóban képes olyan jelenségek kezelésére, amelyeket a nyelvészet hagyományosan strukturális vagy szabályalapú kategóriákban ír le. A kérdés tehát nem az, hogy a modellek *működnek-e*, hanem az, hogy mit magyaráz meg a működésükből fakadó reprezentáció.

## 5.1. Szintaktikai jelenségek

A nyelvészeti irodalomban a szintaxis gyakran a szabályalapú leírás egyik legfontosabb terepe. Egyeztetés, beágyazottság, hosszú távú függőségek – ezek mind olyan jelenségek, amelyek látszólag explicit szerkezeti reprezentációkat feltételeznek.

A nagy nyelvmodellek azonban számos esetben képesek ezeknek a jelenségeknek a kezelésére anélkül, hogy explicit módon reprezentálnák a szerkezetet. Például helyesen kezelik az alany–állítmány egyeztetést még olyan mondatokban is, ahol az egyeztetési viszonyt zavaró, közbeékelődő elemek nehezítik. Hasonlóképpen képesek kezelni a beágyazott szerkezeteket és a hosszú távú dependenciákat.

Ezek a képességek összhangban állnak a transzformer architektúra működésével, amely lehetővé teszi, hogy a modell a szekvencia különböző pontjai között fennálló relációkat figyelembe vegye. A self-attention mechanizmus révén a rendszer nem lokális szabályokat alkalmaz, hanem dinamikusan súlyozza a kontextus elemei közötti kapcsolatokat.

A kulcskérdés itt az, hogy ezek a relációk mennyiben tekinthetők a hagyományos értelemben vett szintaktikai struktúrák megfelelőinek. Bár a modell nem épít explicit fákat vagy kategóriákat, viselkedése sok esetben kompatibilis az ilyen struktúrák jelenlétével.

## 5.2. Szemantikai viszonyok

A disztribúciós megközelítés ereje talán a szemantikában mutatkozik meg a legvilágosabban. A word2vec típusú modellek már korán megmutatták, hogy a szavak jelentése részben megragadható a kontextusok hasonlósága révén: a hasonló jelentésű szavak a vektortérben egymáshoz közel helyezkednek el, és bizonyos relációk — például analógiák — vektorműveletekkel is leírhatók (Mikolov et al., 2013).

A nagy nyelvmodellek ezt az elképzelést kiterjesztik kontextusfüggő reprezentációkra. Egy szó jelentése nem egyetlen fix vektor, hanem a konkrét kontextus függvényében alakul. Ez lehetővé teszi például a poliszémia kezelését, valamint finomabb szemantikai különbségek megragadását.

Fontos azonban hangsúlyozni, hogy a szakirodalomban nincs egyértelmű bizonyíték arra, hogy ezek a reprezentációk explicit definíciókat vagy fogalmi struktúrákat tartalmaznának.

A jelentés ebben a megközelítésben nem önálló, előre adott tulajdonságként jelenik meg, hanem a tokenek közötti relációk mintázataiban ragadható meg. Ez a megközelítés illeszkedik Ludwig Wittgenstein késői filozófiájához és a disztribúciós szemantika klasszikus téziséhez (Harris, 1954; Firth, 1957).

## 5.3. Diskurzus és koherencia

A nyelvi kompetencia nem merül ki a mondatszintű jelenségekben. A szövegkoherencia, az anafora kezelése vagy a tematikus folytonosság fenntartása olyan képességek, amelyek hagyományosan magasabb szintű reprezentációkat feltételeznek.

A nagy nyelvmodellek e tekintetben meglepően jól teljesítenek. Képesek követni a referenciákat több mondaton keresztül, fenntartani egy diskurzus témáját, és koherens válaszokat adni komplex kérdésekre, ami arra utal, hogy a modell reprezentációi nem korlátozódnak lokális mintázatokra. Technikailag a modell a nyelvet tokenek egydimenziós szekvenciájaként kezeli, ahol szóközök természetesen vannak, de mondathatárt jelölő tokenek nincsenek. A szekvencia hosszát az aktuális bemeneti szekvencia mérete szabja meg, és amelyet a modell maximális kontextusablaka felülről korlátoz. A mai rendszerekben a kontextusablak általában több tízezer tokenre terjed ki. A diskurzus szintű koherencia technológiai feltétele tehát eleve adott: a modell számára a kontextus nem mondatok elkülönült sorozata, hanem egyetlen, egymással sűrű relációban álló tokenlánc.

## 5.4. Mit magyaráz a disztribúció?

A fenti példák alapján egy fontos következtetés adódik. A modellek nem explicit szabályok alkalmazásával kezelik a nyelvi jelenségeket, hanem a korábban látott szövegekből elsajátított valószínűségi mintázatok révén. Ezek a disztribúciós mintázatok olyan felszíni nyelvi jelenségekhez vezetnek, amelyeket a nyelvészeti leírás strukturális összefüggésekkel ír le.

A szintaktikai, szemantikai és diskurzusszintű példák egyaránt arra utalnak, hogy a disztribúciós tanulás nem pusztán lokális vagy felszíni szabályszerűségek rögzítésére képes. A modellek olyan relációs mintázatokat sajátítanak el, amelyek lehetővé teszik számukra, hogy többféle nyelvi szinten is koherens és sok esetben szabályok által vezéreltnek látszó nyelvi megnyilatkozásokat hozzanak létre.

A disztribúciós megközelítés erős empirikus teljesítménye nem jelenti azt, hogy önmagában teljes magyarázatot adna a nyelvi viselkedés minden aspektusára. A jelentés és referencialitás problémája, a világba ágyazottság hiánya, és a reprezentációk interpretálhatóságának nehézségei továbbra is nyitott kérdések maradnak. Az sem világos, hogy az LLM-ek a jelentést valóban kompozicionálisan kezelik-e, illetve hogy nyelvi tudásukat valóban szisztematikusan tudják-e kiterjeszteni új esetekre, vagy pedig ezekben a feladatokban is főként a tanult disztribúciós mintázatokra támaszkodnak. Ezek az ellenérvek nem feltétlenül azt mutatják, hogy a disztribúciós megközelítés téves, hanem inkább azt, hogy nem teljes: a nyelv bizonyos aspektusai további magyarázatot igényelnek. Ebben az értelemben a nagy nyelvmodellek nem végső

választ adnak, hanem empirikus tesztesetként segítenek feltárni, hogy a nyelvi viselkedés mely aspektusai magyarázhatók pusztán a használati mintázatok alapján, és melyek igényelnek további elméleti eszközöket.

## 6. Módszertani nehézségek az LLM-ek nyelvészeti értelmezésében

A nagy nyelvmodellek nyelvészeti értelmezésének egyik sajátos nehézsége ugyanakkor az, hogy esetükben nem áll rendelkezésünkre az a fajta hozzáférés, amely az emberi nyelv vizsgálatában – minden korlátja ellenére – mégis fontos szerepet játszik. Az emberi nyelv értelmezésében a beszélői intuíció, bármennyire fuzzy, fokozatos és részben idioszinkratikus is, mégiscsak valamilyen közvetlen támpontot nyújt ahhoz, hogy a megfigyelhető nyelvi viselkedés mögött milyen grammatikai tudást, kategóriákat vagy megszorításokat feltételezzünk. Ez a hozzáférés nem problémamentes, de létezik, és a nyelvészeti hagyomány jelentős része épít rá. (Chomsky, 1965)

Az LLM-ek esetében azonban semmi hasonló feltáró eszközünk nincs. Nem férünk hozzá a rendszer esetleges, az emberi beszélői intuícióhoz hasonlítható belső állapotaihoz, és az általa adott válaszokat sem kezelhetjük ugyanabban az értelemben introspektív ítéletekként, mint az emberi beszélők megnyilatkozásait.

Éppen ezért az értelmezés során könnyen belecsúszunk abba, hogy a rendszerre a saját értelmezési sémáinkat vetítjük rá. Azt keressük, hogy hol vannak benne a szófaji kategóriák, a szimbolikus jegyek, a szabályok vagy a nyelvészeti címkék; vagyis ugyanazokat az objektumokat próbáljuk „megtalálni” a rendszer belsejében, amelyeket a leírás során mi magunk használunk. Ezt a késztetést könnyű megérteni: a nyelvészeti és NLP-s beidegződések szinte óhatatlanul ebbe az irányba tolnak bennünket. Módszertanilag azonban ez értelmezési csapda. Nem arról van szó, hogy bárki tudatosan introspekcióval akarná vizsgálni az LLM-eket; a probléma inkább az, hogy az értelmezés során hallgatólagosan mégis ezt a mintát követjük, és antropomorfizáló módon olyan belső nyelvi tárgyakat keresünk, amelyek létezésére nincs közvetlen bizonyítékunk. Ez a csapda közvetlenül összefügg azzal a korábban tárgyalt jelenséggel is, amikor a modell viselkedéséből túl gyorsan következtetünk a nyelvészeti kategóriák tényleges belső jelenlétére.

Ebből következően célszerű lehet az LLM-ek nyelvészeti kutatásában a jelenleg alkalmazott empirikus eljárásokat – például a probingot és a nyelvészeti benchmarkok használatát (Belinkov et al., 2017; Hewitt & Manning, 2019) – olyan módszerekkel is kiegészíteni, amelyek közelebb állnak a terepmunkán alapuló nyelvészet gyakorlatához. A field linguistics klasszikus helyzete éppen az, hogy a kutató nem fér hozzá közvetlenül a vizsgált rendszer belső szerkezetéhez, hanem gondosan megtervezett elicitációs technikákkal, minimálpárokkal, kontextusmanipulációval, célzott kérdezéssel és kontrollált

viselkedési próbákkal próbálja feltárni, hogy milyen általánosítások, oppozíciók és megszorítások tulajdoníthatók neki. Az LLM-ek esetében is valami ilyesmire volna szükség: nem belső címkék és explicit szimbolikus objektumok keresésére, hanem olyan rafinált elicitációs eljárásokra, amelyekből a rendszer outputja alapján következtethetünk arra, milyen nyelvészeti tudást tulajdoníthatunk neki.

Ez a javaslat ugyanakkor nem jelentene ontológiai elköteleződést. Nem kellene azt állítanunk, hogy a nyelvészeti leírásban használt konstruktumok – kategóriák, szabályok, szimbolikus reprezentációk – ténylegesen léteznek az LLM-ekben. A jelenlegi ismereteink alapján inkább az óvatosság indokolt ezen a ponton. Ugyanakkor az elicitációs vizsgálatok alapján nagyon is megalapozott lehet olyan állításokat tenni, hogy a rendszer viselkedése egy adott jelenség tekintetében megfelel annak, amit emberi beszélő esetében a szóban forgó nyelvészeti tudás birtoklásának tekintenénk. Más szóval: nem azt állítanánk, hogy a modellben „van” egy adott szabály vagy kategória, hanem azt, hogy olyan outputot produkál, amelyet emberi nyelvhasználó esetében ennek bizonyítékaként értelmeznénk.

Ebben az értelemben az elicitációs megközelítés nem a nyelvészeti fogalmak elvetését jelentené, hanem éppen ellenkezőleg: továbbra is ezek a részben introspekcióra épülő, illetve azzal validált fogalmak, kategóriák és elméleti struktúrák adnák a vizsgálat leíró keretét. A különbség az volna, hogy ezeket nem a rendszer belső ontológiájának elemeiként kezelnénk, hanem elemzési eszközként, amelyek segítségével viselkedési alapon lehet megállapítani, milyen tudást tulajdoníthatunk neki. Ez a módszertani fegyelem talán segíthetne abban, hogy az LLM-eket ne emberi beszélőként „olvassuk”, és ne is klasszikus szimbolikus rendszerként próbáljuk megfejteni, hanem olyan nyelvi ágensekként vizsgáljuk, amelyek meggyőző teljesítményt nyújtanak, miközben belső működésük ontológiai státusza továbbra is nyitott kérdés marad.

Az LLM-ek értelmezésének egy további kérdése az, hogy tulajdonképpen mit modelleznek a nagy nyelvmodellek. Bár nyelvi viselkedésük sok esetben közel áll az emberi nyelvhasználathoz, nem magától értetődő, hogy ebből milyen típusú modellkövetkeztetések vonhatók le. Kérdés, hogy ezek a rendszerek kizárólag a nyelvi performancia modelljeiként értelmezhetők-e, vagy a klasszikus értelemben vett kompetencia, illetve a langue szintjén is relevánsak lehetnek. Hasonlóképpen nyitott kérdés az is, hogy a nyelvmodellek inkább a mentális reprezentációk, vagy az ezeket megvalósító neurális folyamatok modelljeiként foghatók-e fel. A jelenlegi empirikus evidenciák alapján mindenesetre megalapozottabbnak tűnik őket a nyelvi performancia modelljeiként értelmezni, míg a nyelvi kompetencia modelljeként való értelmezés egyelőre spekulatívabb.

## 7. Konklúzió

A nagy nyelvmodellek nyelvi teljesítménye ma már nem szorul különösebb bizonyításra: több nyelven – így magyarul is – képesek jól formált, koherens és pragmatikailag adekvát szövegek előállítására. Teljesítményük ráadásul messze nem korlátozódik a mondatok grammatikai helyességére: sok esetben stilárisan is adekvát szövegeket hoznak létre, érzékenyek a beszédhelyzetre, képesek igazodni a kommunikációs szerepekhez, sőt akár a megszólalók feltételezett karakteréhez, életkorához vagy társadalmi helyzetéhez is. Ezek a jelenségek részben már túl is mutatnak a nyelvészet klasszikus fókuszán, és éppen ezért még inkább figyelemre méltó, hogy egy disztribúciós tanulásra épülő rendszer ilyen gazdag nyelvi viselkedést képes produkálni.

Ha az LLM-ek nyelvi teljesítménye eléri vagy megközelíti azt a szintet, amit emberi beszélők esetében a mentális grammatika működésének tulajdonítanánk, akkor ennek már túl kell mutatnia maguknak a modelleknek a leírásán. Ebben az esetben nemcsak az válik kérdéssé, hogy mit tulajdoníthatunk az LLM-eknek, hanem az is, hogy nem szükséges-e felülvizsgálat alá venni az elméleti nyelvészet néhány alapvető kiindulópontját. Ide tartozhat a disztribúciós megközelítés érvényének újraértékelése, a nyelv tanulhatóságáról alkotott elképzeléseink pontosítása, valamint a nyelvészeti kategóriák és magyarázó konstrukciók státuszának újragondolása. Más szóval: ha a disztribúciós tanulásra épülő rendszerek olyan nyelvi megnyilatkozásokat mutatnak, amelyeket emberi beszélők esetében a mentális grammatika bizonyítékának tekintenénk, akkor ez szükségképpen felveti az elméleti nyelvészet néhány alapvetésének felülvizsgálatát.

Ebben az értelemben a nagy nyelvmodellek nemcsak új nyelvtechnológiai eszközök, hanem a nyelvészeti gondolkodás számára is új kihívást jelentenek.

Köszönetnyilvánítás

A tanulmány nyelvi és alaki megformálásában igénybe vettem a ChatGPT (5.4 thinking model) (W3) segítségét. A végleges szövegért, annak tartalmáért és az esetlegesen fennmaradó hibákért természetesen teljes mértékben én vállalok felelősséget.

# Irodalom


Belinkov, Y., Durrani, N., Dalvi, F., Sajjad, H. & Glass, J. (2017). What do neural machine translation models learn about morphology? In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (861–872). Vancouver, Canada: Association for Computational Linguistics. doi:10.18653/v1/P17-1080. Letöltés: https://aclanthology.org/P17-1080/

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.

*Firth, J. R. (1957). A synopsis of linguistic theory, 1930–1955. In Philological Society (szerk.), Studies in linguistic analysis* (1–32). Oxford: Blackwell.

Harris, Z. S. (1954). Distributional structure. *Word, 10*(2–3), 146–162. doi:10.1080/00437956.1954.11659520

Hewitt, J. & Manning, C. D. (2019). A structural probe for finding syntax in word representations. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long and Short Papers) (4129–4138). Minneapolis, Minnesota: Association for Computational Linguistics. doi:10.18653/v1/N19-1419. Letöltés: https://aclanthology.org/N19-1419/

Héja Enikő (2024). A ChatGPT története. *Magyar Tudomány, 185*(6), 815–828. doi: 10.1556/2065.185.2024.6.11 (Letöltés: https://www.youtube.com/watch?v=7xTG NNLPyMI)

Héja Enikő (2026). A mesterséges intelligencia története. *IPM Magazin, 2026*(1), 1–15.

Karpathy, A. (2025). *Deep dive into LLMs like ChatGPT*. (Letöltés: https://www.youtube.com/watch?v=7xTG NNLPyMI)

Ligeti-Nagy Noémi (2026). Lehull a lepel. *IPM Magazin, 2026*(1), 20–23.

Martinet, A. (1960). *Elements of general linguistics*. Chicago: University of Chicago Press.

Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013). *Efficient estimation of word representations in vector space*. Letöltés: https://arxiv.org/abs/1301.3781

Prószéky Gábor (2024). *Magyar nyelvtechnológiai eredmények a mesterséges intelligencia korában*. – Prószéky Gábor közgyűlési díszelőadása videón. Letöltés: https://mta.hu/kozgyules2024/magyar-nyelvtechnologiai-eredmenyek-a-mesterseges-intelligencia-koraban-proszeky-gabor-kozgyulesi-diszeloadasa-videon-113663

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K. & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature, 497*(7451), 585–590. doi:10.1038/nature12160. Letöltés: https://www.cns.nyu.edu/wanglab/publications/pdf/rigotti_2013.pdf

Sennrich, R., Haddow, B. & Birch, A. (2016). Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (1715–1725). Berlin, Germany: Association for Computational Linguistics. doi:10.18653/v1/P16-1162. Letöltés: https://aclanthology.org/P16-1162/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008. (Letöltés: https://arxiv.org/html/1706.03762v7)

Yang Zijian Győző (2026). Szövegből számok. *IPM Magazin, 2026*(1), 24–28.


# Források

W1 = https://tinyurl.com/3d2e5knd
W2 = https://jalammar.github.io/illustrated-transformer
W3 = ChatGPT Desktop verzió 2026, 5.4 thinking modell