

Laki László János

Mesterséges intelligencia a gépi fordításban

1. Bevezetés

A nyelvtechnológia egyik legfontosabb feladata a nyelvi diverzitás okozta akadályok áthidalása, vagyis a számítógépek alkalmassá tétele különböző nyelvek közti fordítások megvalósítására. Az elmúlt évtizedben az információtechnológia robbanásszerű fejlődése lehetővé tette a számítógépes nyelvészet számára, hogy megoldást nyújtson erre a problémára. A feladat mindig is kiemelt fontosságú volt az emberiség számára, ezért pénzt, erőt és energiát nem kímélve igyekeztek megoldást találni a természetes szövegek számítógéppel történő fordítására. Ennek ellenére az elmúlt több mint hatvan év fejlesztései során nem sikerült megvalósítani a tökéletes mesterséges fordítórendszert, viszont a mesterséges intelligencia egyre szélesebb körű alkalmazásával egyre inkább megjelennek az ember számára elfogadható minőségű fordítások. A számítógépes fordítórendszereket a fordítástámogatás területén sikerült több-kevesebb sikerrel alkalmazni.

A folyamatosan fennálló nehézségek ellenére a napjainkban megjelent legújabb számítógépes erőforrásoknak (videokártya/GPU), valamint a neurálhálózat-alapú gépi tanulásnak köszönhetően jelentős változás figyelhető meg a legtöbb tudományág területén. A mesterséges intelligencia viszonylag nagy pontossággal képes felismerni és megkülönböztetni a tárgyakat és a személyeket képekről, valamint a térfigyelő kamerák képeiről. Napjainkban győzte le a mesterséges intelligencia az emberiség legnehezebb játékának tartott GO-ban az aktuális

nagymestert,¹ valamint a sakkban is olyan stratégiákat talál ki, melyeket a sakknagymesterek próbálnak értelmezni és megtanulni tőle.²

Ez a jelenség és tendencia a gépi fordításban is megfigyelhető. A legújabb mélytanulós módszereknek köszönhetően a gépi fordítás eredményessége megduplázódott, és a legfontosabb nyelvpárok között (angol–spanyol, angol–német, angol–francia) sosem látott minőséget produkál. Ez még nem jelenti azt, hogy a humán fordítók elveszítik munkájukat, csupán a jobb minőségű előfordító rendszereknek köszönhetően sokkal eredményesebben képesek dolgozni.

Jelen írásomban bemutatom a gépi fordítás történetét, majd ismertetem napjaink piacvezető technológiájának számító neurális hálózat-alapú gépi fordítórendszer működését. Végül bemutatom ezen rendszerek néhány érdekes alkalmazási területét, úgymint a képhez történő szövegenerálás vagy a hang leiratozása.

2. A gépi fordítás története

A nyelvtechnológia egyik jelentős területe a soknyelvűség támogatása, amire a napjainkban igen hangsúlyos globalizációs törekvések miatt egyre növekvő igény van. Ebben nyújt támogatást a gépi fordítás, melynek módszerei nem csupán a nyelvek közötti transzformáció megvalósításáról szólnak, hanem tartalmazzák a szövegek elő- és utófeldolgozását, valamint a fordítások minőségének előzetes becslését, illetve azok kiértékelését is. Az elmúlt közel hatvan év során számos megközelítés született a természetes nyelvek közötti fordítás megoldására, amelyek közül a legfontosabb állomásokat mutatom be. A gépi fordítás fejlődését nagymértékben befolyásolja a számítógépek teljesítményének növekedése is, hiszen a személyi számítógépek fejlesztése során

¹ Silver–Schrittwieser és mtsai 2017.

² Silver–Hubert és mtsai 2017.

megfigyelhető technológiai (generációs) váltásokat rövid idő eltelével követték a gépi fordítás területén a jelentős módszertani változtatások.

2.1. A gépi fordítás korai története

A gépi fordítás tudománya egyidős az első számítógépek megjelenésével, és mind a mai napig a számítógépes nyelvészet egyik leginkább kutatott területe. Az egyik legelső fordítórendszernek Alan Turing és csapata által létrehozott elektromechanikus rendszer tekinthető, mely segítségével sikerült feltörni a II. világháború során a németek legfejlettebb titkosítási algoritmusát, az úgynevezett Enigma-kódot.

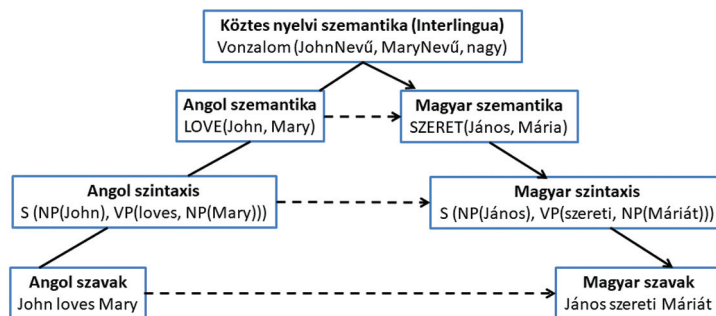
A világháború után a gépi fordítás történetének mozgóterülete a hidegháború volt. Az amerikai kormány jelentős mértékben támogatta a kutatókat, hogy megoldást találjanak az angol és az orosz nyelvek közötti automatikus fordításra. A támogatás eredményeképp 1954-ben a Georgetown Egyetem és az IBM közösen mutatta be az első gépi fordítórendszert. Az elért eredmények alapján a kutatók jelentős része úgy érezte, hogy a gépi fordítás pár éven belül megoldott probléma lesz. Azonban a kezdeti sikerek és a hatalmas anyagi támogatások ellenére sem történt meg a várva várt áttörés. Az amerikai kormány vizsgálóbizottságot hívott össze, melynek feladata a gépi fordítás hatékonyságának kiértékelése volt. Így 1966-ban elkészült az ALPACH-jelentés, melyben nagyon lesújtó képet festettek a számítógép általi fordítás jelenéről és jövőjéről. A jelentés hatására töredékére esett vissza a fordításra szánt állami támogatások összege. A megcsappant támogatások ellenére a fejlesztések nem álltak le, így sorra jöttek létre az első üzleti megoldások; először 1976-ban a Météo és a Systran, majd 1980-ban a Logo, a METAL, valamint a Trados rendszerei.

2.2. A szabályalapú gépi fordítórendszer

A 80-as, 90-es években a számítógépek teljesítménye nagymértékben javult a korábbi évekhez képest. Egyrészt alkalmasak lettek nagyobb méretű szótárak tárolására és használatára, valamint képessé váltak komolyabb mennyiségű szöveges tartalom feldolgozására, másrészt pedig a személyi számítógép megjelenésével a fejlesztőknek lehetőségük nyílt komolyabb programkódok megírására is. Ebben az időben jelentek meg a gépi fordítás támogatására alkalmazott első számítógépes nyelvfeldolgozó rendszerek, melyek integrációjával megalkották az első szabályalapú gépi fordítórendszereket is.

A szabályalapú gépi fordítórendszer (RBMT – Rule-Based Machine Translation) alapötlete, hogy a fordítandó szövegből kinyerhető legtöbb információt használja fel a fordítás során. A legegyszerűbb első implementációk az úgynevezett direkt fordítórendszerek voltak. A módszer lényege, hogy a fordítandó szöveget egy szótár alapján szóról szóra fordítja le, majd a megfelelő sorrendbe rendezi. Ennek előnye, hogy viszonylag könnyen megvalósítható. Hátránya azonban, hogy nem képes komplex nyelvtani szerkezetek kezelésére, illetve ebből eredendően csak gyenge fordításminőséget képes elérni. A későbbi, bonyolultabb rendszerek a fordítandó szövegből elemzés segítségével állítanak elő egy köztes reprezentációt, amit előre definiált átviteli szabályok segítségével alakítanak át egy absztrakt célnyelvi reprezentációra. Végül ebből a reprezentációból generálják a célnyelvi szóalakokat. Ezeket a rendszereket az elemzés és generálás mélysége, valamint az átvitel helye alapján osztályozhatjuk, melyet a Vauquois-háromszög szemléltet (1. ábra). Az ábrán látható, hogy minél mélyebb a nyelvi elemzés mértéke a fordítás során, annál közelebb áll egymáshoz a két nyelv reprezentációja, amelyek közt a transzformációt végre kell hajtani. Egy precízen megírt szabályokkal rendelkező szabályalapú gépi fordítórendszer nagy pontosságú fordítást képes előállítani, de az átviteli szabályok létrehozásához elengedhetetlen a jó minőségű szintaktikai és/vagy szemantikai elemző, ami csak nagyon kevés nyelv esetén áll ren-

delkezésére. Továbbá, mivel ezek a szabályok nyelvspecifikusak, minden nyelvpárra külön-külön kell definiálni ezeket, ami megnehezíti a rendszer kiterjesztését újabb nyelvekre.



1. ábra. Vauquois-háromszög³

2.3. Statisztikai gépi fordítás

Az 1990-es évek elejére a digitálisan elérhető párhuzamosan lefordított szövegek száma ugrásszerűen megnőtt ezzel újabb lehetőségeket nyitva a fordítástámogató rendszerek számára. A szabályalapú rendszerekkel párhuzamosan megjelentek az adatmotivált módszerek. Az első ilyen rendszernek a *példaalapú fordítórendszer* számít, melynek alapötlete, hogy az aktuális fordításhoz felhasználja a már korábban lefordított mondatokat. A rendszer egy előre létrehozott fordítómemóriából kiválasztja a fordítandó mondat részeinek eltárolt fordításait, melyek egyesítésével megkapjuk a kívánt fordítást.⁴ Annak ellenére, hogy a rendszer nem tartalmaz komplex, nyelvspecifikus modulokat, fordítási minősége nem sokkal marad el szabályalapú társaitól. Alapvető hiányossága azonban, hogy a fordítómemóriában tárolt

³ Indurkha–Damerau 2010; Koehn 2010.

⁴ Hutchins 2005.

szegmenspárok elemi egységei (például morfémák, szavak, kifejezések stb.) nincsenek összekötve. Ennek következtében hiába tudja a fordítórendszer, hogy a memóriában tárolt forrásnyelvi szegmens melyik részében különbözik a fordítandó szegmenstől, azt viszont nem tudja meghatározni, hogy a célnyelvi oldalon ez melyik szavakra van hatással.

A *statisztikai gépi fordítórendszer* (SMT – Statistical Machine Translation) a példaalapú fordítórendszer általánosított változatának tekinthető, mivel képes javaslatokat tenni a fordítómemóriában nem szereplő szegmensek fordítására is. A statisztikai gépi fordítás alapötlete, hogy a rendszer párhuzamos kétnyelvű tanítóanyag segítségével felügyelt módon tanulja meg a fordításhoz szükséges modelleket. A párhuzamos kétnyelvű korpusz egy olyan, mondatpárokból álló szöveges adathalmaz, melyben a forrásnyelvi mondatokhoz hozzá van rendelve azok célnyelvi fordítása. Az algoritmus könnyű és gyors implementálhatósága, valamint nyelvfüggetlen alkalmazhatósága nagymértékben hozzájárult ahhoz, hogy a módszer napjainkra a legtöbbet hivatkozott gépi fordító architektúra legyen.

Az SMT alapjait az IBM T. J. Watson Research Center munkatársai fektették le,⁵ akik a fordítás feladatát a beszédtechnológiában használatos Shannon-féle zajoscsatorna-modell⁶ segítségével közelítették meg. A későbbi kutatások eredményeként napjainkra a zajoscsatorna-modell kiegészített változatát alkalmazzák, az úgynevezett log-lineáris modellt.⁷ A következőkben bemutatom ezt a két modellt, valamint ezek kapcsolódását a gépi fordításhoz.

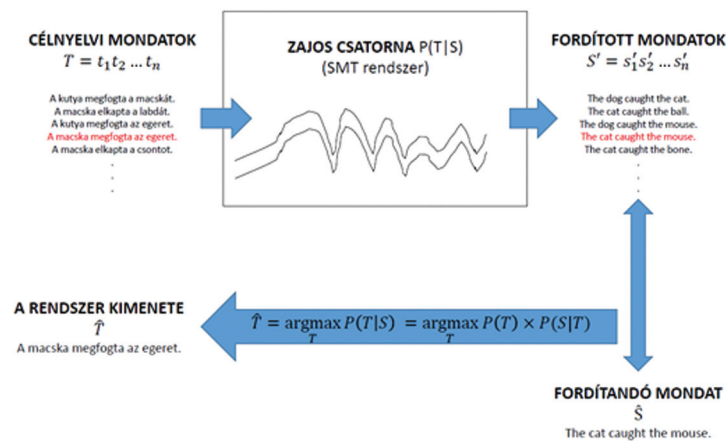
⁵ Brownsé mtsai 1993.

⁶ Shannon 1948a, 1948b.

⁷ Och–Ney 2003, 2004.

2.3.1. Zajoscsatorna-modell

A statisztikai gépi fordítás feladata megfogalmazható a Shannon-féle zajoscsatorna-modell⁸ segítségével, amit a 2. ábra mutat. Az elmélet alapja, hogy a fordítás során az egyetlen biztosan ismert információ a fordítandó szöveg. A fordítás úgy történik, mintha a célnyelvi szövegek halmazát egy zajos csatornán átengedve a csatorna kimenetén összehasonlítanánk a forrásnyelvi szöveggel. Az a célnyelvi mondat lesz a rendszer kimenete, amelyik a csatornán való áthaladás után a legjobban hasonlít a fordítandó mondatra.



2. ábra. Zajoscsatorna-modell

Formálisan az SMT-módszer a fordítás feladatát úgy tekinti, mint a fordítás pontosságát, valamint görbülékenységét reprezentáló modellek kombinációja által elérhető maximális valószínűségi értékhez tartozó szöveg meghatározása. A fordítás fel-

⁸ Shannon 1948a, 1948b.

adata tehát úgy fogalmazható meg, hogy keressük azt a célnyelvi mondatot (\hat{T}), amelyik a célnyelvi mondatok halmazából (T) a legvalószínűbb fordítása a forrásnyelvi mondatnak (S).

$$\hat{T} = \operatorname{argmax}_T P(T|S)$$

A $P(T|S)$ valószínűség azonban közvetlenül nem számolható, viszont önmagában modellezhető részekre bontható. A Bayes-tétel alapján az (1) egyenlet átalakítható a következőképpen:

$$\hat{T} = \operatorname{argmax}_T P(T|S) = \operatorname{argmax}_T \frac{P(S|T)P(T)}{P(S)} = \operatorname{argmax}_T P(S|T)P(T)$$

Mivel T függvényében $P(S)$ konstans, ezért elhagyható. Az így kapott egyenlet két komponensből áll:

- $P(T)$ a nyelvmodell, ami a gördülékenyséért felelős
- $P(T|S)$ a fordítási modell, ami a fordítás pontosságát biztosítja

A két modell kombinációjának maximális értékét a fordítórendszer dekódernek nevezett komponense határozza meg. A (2) egyenlet legfontosabb jellemzője, hogy a fordítás feladatát kiszámítható egységekre bontja. Ráadásul e komponensek becslése egy- és kétnyelvű korpuszok segítségével automatikusan történik.

A zajoscsatorna-modell használata számos megszorítást vezet be, amik korlátozzák a fordítórendszer minőségét. Ilyen megszorítás például, hogy ez a modell egy szónak vagy egy kifejezésnek a fordítását a környező szavaktól függetlennek tekinti, illetve hogy a nyelvmodell csak néhány megelőző szót vesz figyelembe a fordítás során. Sajnos ezek a korlátok jelentősen csökkentik a fordítórendszer minőségét. Emiatt szükségessé vált, hogy a fordításhoz a nyelvmodellen és a fordítási modellen kívül egyéb tudás is felhasználható legyen. Erre ad megoldást a log-lineáris modell, melyet az alábbiakban mutatok be.

2.3.2. *Log-lineáris modell*

A log-lineáris modell a gépi tanulás tudományágának egyik gyakran használt módszere. Lényege, hogy egy feladatot egymástól független jellemzők ($h_i(\chi)$) súlyozott szorzatával (λ_i) írja le. Formálisan a modell a következő:

$$P(x) = \exp \sum_{i=1}^n \lambda_i h_i(x)$$

ahol $h_i(\chi)$ az i . jellemző függvény, míg λ_i a hozzá tartozó súly.⁹ A log-lineáris modell bevezetésével a gépi fordítórendszer tanítása egy újabb lépéssel bővül. Első lépésben be kell tanítani a rendszer modelljeit külön-külön, majd egy úgynevezett paraméter optimalizációs (tuning) lépésben keressük azokat az λ_i értékeket, amivel a fordítórendszer a legjobb minőségű fordítást állítja elő.

Ez a modell a zajoscsatorna-modell kiegészítésének tekinthető, mivel a nyelvmodell és a fordítási modell mellett tetszőleges számú komponenssel bővíthető. Ezek a komponensek lehetővé teszik, hogy a fordítórendszer mélyebb nyelvi (főleg szintaktikai vagy grammatikai) ismeretekkel rendelkezzen, és ezeket felhasználva javítson a fordítások minőségén. A gyakorlatban elterjedt modellek közös jellemzője, hogy mindegyik egyszerű nyelvfüggetlen megszorításokkal igyekszik a fordítási állapotot csökkenteni, melynek köszönhetően sikerült javítani a fordítások minőségét az angol–spanyol, az angol–francia és az angol–német nyelvpárok között. A biztató eredmények ellenére az angol nyelvre kiválóan működő megszorítások a magyar nyelv esetén kártékonynak bizonyultak. A teljesség igénye nélkül nézzünk meg néhány kiegészítő modellt.

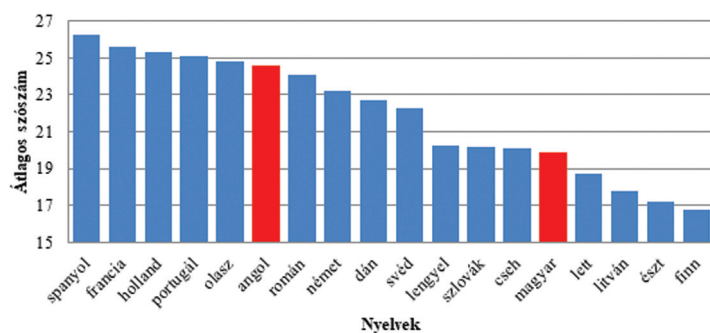
⁹ Och–Ney 2003, 2004.

2.3.2.1. SZÓRENDBELI KÜLÖNBSÉGET BÜNTETŐ MODELL

A szórendbeli különbséget büntető modell (distortion model) feladata a fordítás során a szavak helyes sorrendjének a meghatározása, valamint a keresési gráf méretének a csökkentése. Ez az átrendezési modell azzal a megközelítéssel él, hogy a célnyelvi mondat szavainak sorrendje hasonlít a forrásnyelvi mondat szavainak sorrendjéhez. Ez a feltevés igaz az egymáshoz grammatikailag közel álló nyelvekre, viszont problémát okoz például angol–német nyelvpárok közötti fordítás esetén, ahol a mellékmondat végén szereplő ige fordítását a modell olyan súlyos mértékben bünteti, hogy az SMT-rendszer egyszerűen nem fordítja le azt. Hasonló probléma jelentkezik az angol–magyar vagy a japán–angol fordítás esetében is.

2.3.2.2. A MONDATHOSSZ-HARMONIZÁCIÓS MODELL

A mondathossz-harmonizációs modell azzal a feltételezéssel él, hogy az eredeti mondat és a lefordított mondat szavainak száma hasonló. Ennek megfelelően a modell feladata, hogy kiszűrje a szószámában jelentősen eltérő fordítási javaslatokat. Könnyen belátható, hogy a modell által használt feltételezés túlságosan erős feltétel. A 3. ábra egy példát mutat be, ahol az Európarl



3. ábra. A mondatok átlagos szószáma

korpusz¹⁰ alapján kiszámoltam, hogy átlagosan hány szóból áll egy mondat a korpuszban szereplő nyelvekben. A diagram jól szemlélteti, hogy a különböző típusú nyelvek között jelentős eltérés mutatkozik a mondatok átlagos szószámában, emiatt a mondathossz-harmonizációs modell hangsúlyozott figyelembevétele ronthatja a fordítórendszer minőségét.

2.3.2.3. A LEXIKALIZÁLT ÁTRENDÉZÉSI MODELL

A lexikalizált átrendezési modell (phrase based lexicalized reordering model) lényege, hogy egy kifejezés fordítása után három úton folytatódhat a fordítás folyamata: 1. balról jobbra történő monoton fordítás; 2. a soron következő kifejezés átugrása; 3. a megelőző kifejezés fordítása.

Például az „*I can count the stars visible.*” mondat esetén a *the* szó fordítása után nem a soron következő *stars* szó fordítása következik, hanem az azt következő *visible* szóé, majd ezután visszalép a megelőző szóra. Így kapjuk meg a „*Meg tudom számolni a látható csillagokat.*” fordítást.

Doktori disszertációmban az ezen modellek okozta minőségromlást igyekeztem csökkenteni saját, magyar nyelvre specifikus nyelvfüggetlen modulok létrehozásával. Munkám során sikerült különböző technikákkal (szintaktikai elemzés és generálás, szórendi átrendezések) a magyar és az angol nyelvű mondatokat grammatikailag közelebb hozni egymáshoz, és ezáltal javulást értem el a gépi fordítások minőségében.

2.4. Neurális hálózat-alapú gépi fordítás

Az évtized elejére az SMT-alapú gépi fordítással a kutatóknak nem sikerült jelentősebb minőségjavulást elérni, így indultak el

¹⁰ Koehn 2005.

a kutatások újabb és újabb technikák irányába. Ezekben az években kezdte meg hódító útját a mesterséges neurális hálózatok használata a gépi tanulás területén. A módszer egyáltalán nem tekinthető újnak csupán a számítógépek teljesítményének korlátossága akadályozta ezen technológia használatát. A beszédtechnológiában elért komoly sikereken felbuzdulva a kutatók megvizsgálták a neurális hálózatok alkalmazhatóságát a gépi fordítás feladatának megoldására.

Philipp Koehn a neurális hálózat-alapú gépi fordítás (NMT) fejlődését bemutató könyvének¹¹ bevezetőjében rámutatott arra, hogy ez a technológia már a 80-as, 90-es években is rendelkezésünkre állt, és más tudományterületeken sikerrel alkalmazták. Az 1990-es években a kutatók megpróbálták kiaknázni a neurális hálózatokban rejlő perspektívákat a gépi fordítás területén is. Forcada és Neco¹², valamint Castaño és társai¹³ már 1997-ben létrehoztak olyan neurális hálózat-alapú rendszereket, melyek fordítási minősége megközelítette a létező domináns rendszerek minőségét. Ennek ellenére, mivel az akkoriban rendelkezésre álló erőforrások még nem voltak alkalmasak nagy adathalmazok kezelésére, az egyik ilyen modellt sem tudták megfelelő módon tanítani ahhoz, hogy számottevő eredményeket tudjanak elérni. Ebben az időben ugyanis a rendszerek számítási komplexitása messzemenően meghaladta a korszak számítási erőforrásait, így ezzel a módszerrel lényegében két évtizedig senki sem foglalkozott. Ez alatt a nyugalmi időszak alatt az olyan adatközpontú megoldások, mint például a frázisalapú statisztikai fordítórendszerek, nagy fejlődésen mentek keresztül, és hasznos segítséget nyújtottak számos fordítási probléma megoldásában, illetve növelték a szakfordítók teljesítményét is.

A hagyományos SMT-rendszerek neurális modellekkel történő kiegészítése a neurális módszerek iránti érdeklődés feltáma-

¹¹ Koehn 2017.

¹² Forcada–Neco 1997.

¹³ Castaño és társai 1997.

dását eredményezte. Az áttörést Schwenk¹⁴ hozta, amikor az ily módon felépített hibrid rendszerével jelentős fordításminőség-javulást ért el. Ezek a törekvések azonban a számítási komplexitás miatt csak lassan nyertek teret. Több kutatócsoport is dolgozott a neurálisháló-alapú rendszerek GPU-kon történő tanításának megvalósításán, hiszen a GPU-kban eddig még kiaknázatlan lehetőségek rejlettek.

Az ugrásszerű fejlesztéseknek köszönhetően egy-két év leforgása alatt a gépi fordítás területén is a neurálisháló-alapú kutatások lettek népszerűek a kutatók körében. Összehasonlításképp, az évente megrendezett Conference on Machine Translation (WMT) megmérettetésén 2015-ben még csak egy neurálisháló-alapú fordítórendszert mutattak be, amely versenyképes volt ugyan, de a hagyományos rendszerek még jobbak voltak. 2016-ban már minden nyelvpár-kategóriában a neurálisháló-alapú rendszerek fölényeskedtek, 2017-ben pedig már majdnem az összes benevezett fordítórendszer alapja a neurális hálózat volt. Ez a fejlődés még mindig tart, és az elkövetkezendő években több irányvonal fog megjelenni a rendszerek tökéletesítése céljából.

3. Mesterséges neurálishálózatok bemutatása és alkalmazása

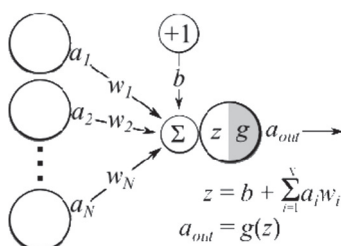
A továbbiakban a neurálishálózat-alapú gépi fordítás módszerével ismerkedhetünk meg.

3.1. Mesterséges neuronok működése

Ahogy azt a 2.3.2. fejezetben említettem, a statisztikai gépi fordítórendszerek a log-lineáris modell segítségével valósítják meg a

¹⁴ Schwenk 2007.

fordítás feladatát. A fordítandó mondatból kinyerhető, valamint a rendszer tanítása során felépített modellekből számolt jellemzők (angolul *feature*) súlyozott összege segítségével számoljuk ki a rendszer kimenetén megjelenő célnyelvi mondatot. A log-lineáris modellt legegyszerűbben egy hálózattal lehet szemléltetni, ahol a bemeneten a jellemzők, az éleken az azokhoz tartozó súlyok, míg a kimeneten ezek összege szerepel. Ezt a reprezentációt a 4. ábra mutatja. A gyakorlatban ez a neuron bemeneti és kimeneti vektorokból (a_i, a_{out}), a súlyok vektorából (w_i) (és kettő szorzatából tevődik össze).



4. ábra. Mesterséges neuron sematikus rajza¹⁵

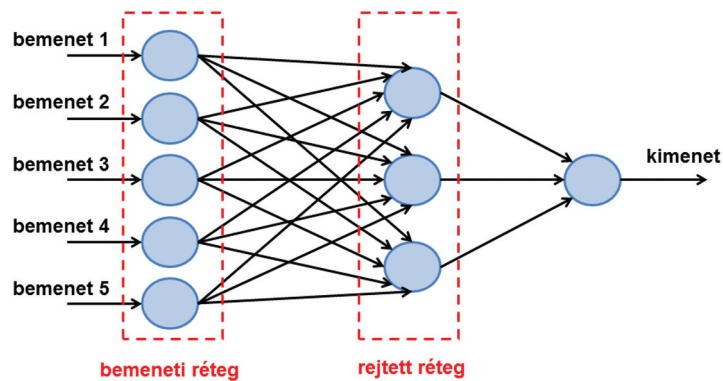
A neurális hálózatok a lineáris modelleket kétféle szemszögből változtatják meg. Az első, hogy a bemeneti értékekből direkt módon számolt kimenetek helyett többrétegű architektúrát alkalmaznak. Ezt a köztes réteget rejtett rétegnek nevezzük, hiszen csak a rendszer bemenetét és kimenetét látjuk, azonban a köztes belső kapcsolatokat nem. A neurális hálózatok másik erőssége a log-lineáris modellel szemben az úgynevezett nem linearitás bevezetése. Ez azt jelenti, hogy a log-lineáris modell egyenletén végrehajtunk egy nem lineáris transzformációt, aminek a következő hatásai vannak a rendszerre: 1. képes egy neuront teljesen kikapcsolni; 2. lehetséges egy neuront félig bekapcsolt állapotba hozni; valamint 3. lehetséges a neuront bekapcsolt állapotban

¹⁵ <https://theclevermachine.files.wordpress.com/2014/09/perceptron2.png>

tartani. Ezeknek a tulajdonságoknak köszönhetően a rejtett réteg képes megtalálni a fontos jellemzőket, és csökkenteni a nem releváns vagy ártó jellemzők hatását.

3.2. Előreccsatolt és rekurrens neurális hálózatok

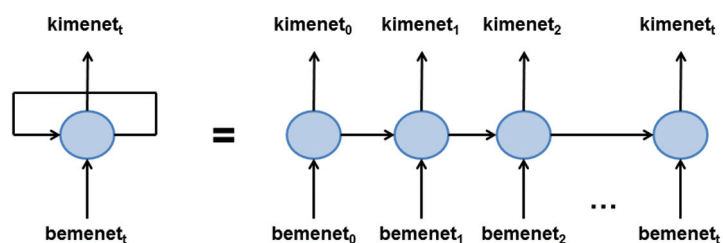
Az 5. ábra egy rejtett réteggel rendelkező neurális hálózatot ábrázol, aminek öt független bemeneti és egy darab kimeneti értéke van. Ezt az architektúrát előreccsatolt neurális hálózatnak nevezük. Az ilyen típusú hálózatok rendkívül jól teljesítenek osztályozási feladatok elvégzésére, mint például a képről tárgy-, arc- vagy személyfelismerés. Mivel az előreccsatolt neurális hálózatok paramétereit és jellemzőit (tehát a különböző rétegek neuronjainak a számát) a tanításuk elején rögzíteni kell, emiatt az ilyen típusú hálózatok nem alkalmasak folytonos hang, illetve szöveg feldolgozására, mivel ezeknél a feladatoknál nem ismerjük előre a bemeneti elemek számát.



5. ábra. Előreccsatolt neurális hálózat egy rejtett réteggel

Erre a feladatra jelent megoldást a rekurrens neurális hálózatok (RNN – Recurrent Neural Network) használata. A rekurrens

hálózat lényege, hogy a mesterséges neuron kimenetét visszacsatoljuk ugyanannak a neuronnak a bemenetére. A szövegfeldolgozás során a vizsgált szöveg szavai egymás után lesznek a neuron bemenetei. A kimenet visszacsatolásának köszönhetően minden egyes szó feldolgozására hatással vannak az őt megelőző szavak is. Az 6. ábra egy önmagába visszacsatolt neuront ábrázol. A rekurrens neurális hálózatokat szokás kifejtett módon is ábrázolni, ahol szemléletesebben látszik a tényleges működése.



6. ábra. Rekurrens neurális hálózat sematikus rajza

A neurálishálózat-alapú gépi fordítás során a fordítandó mondat feldolgozása során ilyen rekurrens neurális hálózatokat használunk. Mielőtt részletesebben leírom a napjaink legjobb minőségű NMT rendszerét az úgynevezett enkóder-dekóder alapú NMT-t, be kell mutatnom a szavak kódolására használatos szóbeágyazás modell nevű technikát is.

3.3. Szóbeágyazás-modell

A nyelvtechnológia egyik fontos célja, hogy a dokumentumokat, a szövegeket, illetve a szavakat olyan reprezentációban tudjuk eltárolni, amely a gép számára könnyen feldolgozható. Fontos, hogy ez a reprezentáció magában foglalja a lehető legtöbb információt, és ezeket a lehető legkisebb helyen tárolja. A kérdés az, hogy melyik reprezentáció képes a szavak felszíni alakja mellett



7. ábra. Két hasonló jelentésű mondat szavainak leképezése a szóbeágyazási modell segítségével

azok morfoszintaktikai, szintaktikai vagy akár szemantikai információit is megőrizni.¹⁶

Napjainkban a hagyományos disztribúciós szemantikai modellek helyett a számítógép számára sokkal kényelmesebb folytonos vektoros reprezentációk alkalmazása terjedt el.¹⁷ A módszer lényege, hogy a rendszerben a lexikai elemek egy valós sokdimenziós vektortér egyes pontjai, melyek úgy helyezkednek el az adott térben, hogy az egymáshoz szemantikailag hasonló szavak egymáshoz közel, míg a különböző szavak egymástól távol kerülnek egymástól. A reprezentáció további előnye, hogy értelmezhető két szóvektor különbsége, ami a szavak jelentésbeli hasonlóságát vizsgálja, valamint a szóvektorok összege is, ami a közös jelentésüket tárolja.

A szóbeágyazás-modell tanítása a statisztikai gépi fordító-rendszerekhez hasonlóan nagyméretű szöveges korpusz segítségével történik oly módon, hogy a tanítóanyag minden egyes szavát, valamint a vizsgált szó környezetében lévő szavakat egy rejtett rétegű előrecsatolt neurális hálózat bemenetére helyezzük. A tanítás végén a szavak vektoros reprezentációja a rejtett réteg vektorállapotaiból olvasható ki. A 7. ábrán két azonos jelentésű mondat, szavainak vektoros ábrázolása olvasható. Látható, hogy

¹⁶ Siklósi–Novák 2016a.

¹⁷ Mikolov–Chen és mtsai 2013; Mikolov–Yih és mtsai 2013.

a hasonló jelentésű valamint a szinonima szavak a vektortérben közel helyezkednek el (például *finom* és az *ízletes*), míg a különböző jelentésűek távolabb helyezkednek el (például *fővárosi* és a *fagylalt*).

Az ebben a fejezetben bemutatott példamondat szavait a Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar Nyelvtechnológiai Kutatócsoportja által létrehozott szóbeágyazási-modell¹⁸ segítségével helyeztem el a vektortérben. A rendszernek létezik egy lekérdező felülete, amely segítségével bármelyik tetszőleges szóhoz megtudhatjuk a hozzá legközelebbi szavak listáját. A 8. ábrán az előző ábrán (7. ábrán) szereplő példamondat szavai láthatóak, valamint a vektortérben a hozzájuk legközelebbi elemek szerepelnek. Például az első táblázatból kiderül, hogy a *fagylalt* szóhoz a *jégkrém*, a *fagyi*, valamint a *sütemény* szavak hasonlítanak legjobban, de a *szörp* és a *lekvár* szavak is viszonylag hasonló környezetben szerepelnek, mint a *fagylalt* szó.

4. Neurálishálózat-alapú gépi fordítás működése

Ahogy azt a 2.4. fejezetben megemlítettem, napjainkra a kutatók több különböző felépítésű NMT-implementációt hoztak létre, amelyek kisebb-nagyobb eltéréssel, de nagyjából hasonló minőségben végzik a fordítást a különböző nyelvpárok között. Ezen megvalósítások közül a leggyakrabban alkalmazott technika az úgynevezett enkóder-dekóder architektúra, ami napjaink legjobb minőségű gépi fordító megvalósításának számít. A módszer lényege, hogy a fordítás feladatát két elkülöníthető egységgel valósítja meg, ahol az első rész a fordítandó szöveget alakítja át egy szóbeágyazás-alapú reprezentációba, míg a második rész ebből a köztes állapotból generálja ki a célnyelvi szavak felszíni alakját. A módszert Kalchbrenner és Blunsom¹⁹ alkalmazták először az

¹⁸ Siklósi–Novák 2016b.

¹⁹ Kalchbrenner–Blunsom 2013.

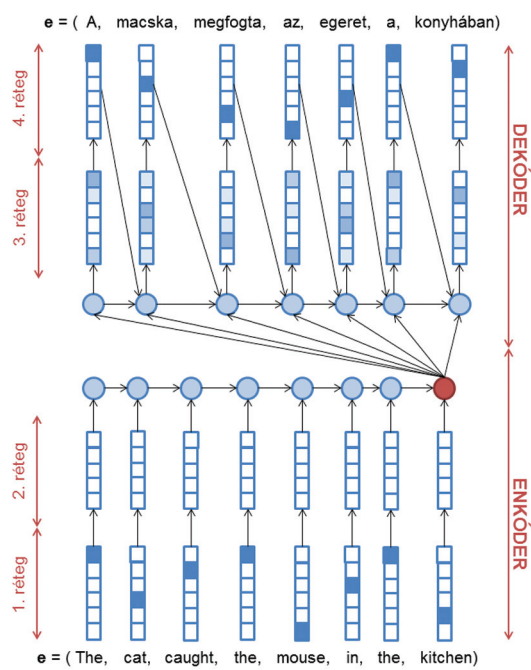
0	fogyólt	1	9133	0	fővárosi	1	129437	0	ízletes	1	18182
1	jéghélm	0.8835	4488	1	kerülőb	0.7771	77525	1	zamatos	0.8263	4732
2	fagyf	0.8783	17395	2	budapesti	0.7759	306934	2	lakatos	0.8262	2812
3	sülemény	0.8597	36852	3	zugói	0.7585	6819	3	fincsiklandó	0.8040	2584
4	desszert	0.8337	18511	4	pécsi	0.7557	128393	4	fincsiklandozó	0.7790	1710
5	csokoládé	0.8295	29077	5	miskolci	0.7473	47096	5	gustuzos	0.7487	3726
6	édesség	0.8200	38281	6	kaposvári	0.7456	36611	6	ízletes	0.7433	302
7	szendvics	0.8194	23288	7	városi	0.7431	214867	7	omlós	0.7410	3160
8	szörp	0.8169	5054	8	jászévhároi	0.7398	5117	8	évszgygyfejlesztő	0.7390	2151
9	likvár	0.8095	16148	9	szombathelyi	0.7331	44923	9	ehető	0.7362	10577
10	teaszemély	0.8055	619	10	székesfehérvári	0.7221	23085	10	lelős	0.7311	2105

8. ábra. A fagyfalt, fővárosi és ízletes szavakhoz legjobban hasonlító szavak listája

Oxfordi Egyetemen, majd idővel számos kutatócsoport²⁰ folytatta a rendszer tökéletesítését. A 9. ábra mutatja be az enkóder-dekóder architektúra sematikus rajzát.

4.1. Az enkóder működése és felépítése

Az enkóder feladata a forrásmondat elhelyezése a megfelelő vektoros reprezentációba, amit úgy valósít meg, hogy a mondat szavait egymás után megmutatja az enkóder belsejében található rekurrens neurális hálónak. A folyamatot a 9. ábra mutatja be,

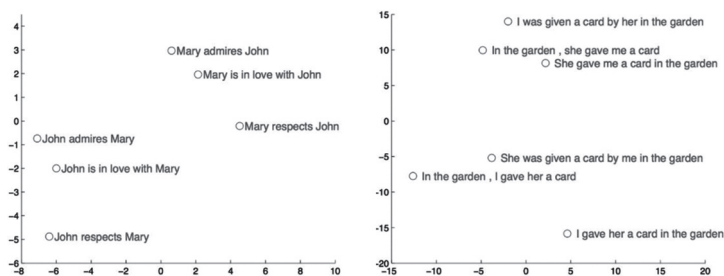


9. ábra. Enkóder-dekóder architektúrájú gépi fordítórendszer sematikus rajza

²⁰ Bahdanau és mtsai 2014; Cho és mtsai 2014.

ahol megfigyelhető a forrásoldali szavak vektoros ábrázolása. Az 1. rétegben a szavakat az úgynevezett „one-hot” reprezentációban tároljuk, ami azt jelenti, hogy a bemeneti szóképlet méretű vektorban egy érték kivételével mindegyik 0 értékű. Ennek a reprezentációnak az előnye, hogy könnyen implementálható, viszont a szavak felszíni alakján kívül semmilyen többletinformációt nem tartalmaznak. Emiatt van szükség a második rétegre, ami a 3.3. fejezetben bemutatott szóbeágyazási modellt valósítja meg.

Miután a mondat összes szava feldolgozásra került, az RNN-neuronban lényegében a forrásmondat vektoros reprezentációja szerepel (a 9. ábrán a piros karikával rajzolt *node* jelöli). Ezzel a módszerrel a szöveget leképeztük egy 300-500 dimenziós vektorba, amik a szóbeágyazási modellhez hasonló tulajdonságokkal rendelkeznek. Erre láthatunk példát a 9. ábrán, ahol a hasonló jelentésű angol és magyar mondatok egymáshoz közel helyezkednek el a térben, míg a különbözőek távolabbi csoportokat alkotnak. Ennek köszönhetően képes ez az NMT-architektúra fordítást végezni, mivel a folytonos vektorreprezentáció hatására a különböző nyelvű mondatok azonos vektortérbe kerülnek.



10. ábra. Példa az NMT-rendszer belsejében szereplő vektoros mondat ábrázolásra²¹

²¹ <https://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-gpus-part-2/>

4.2. A dekóder működése és felépítése

A fordítórendszer második fő komponense a dekóder, ami a forrásnyelvi mondat vektoros reprezentációja alapján egy RNN-hálózat segítségével állítja elő a célnyelvi mondatot. Az RNN-hálózat az előző szó felszíni alakja, a fordítandó mondat vektoros reprezentációja, valamint az eddigi fordítások önmagába visszacsatolt állapota segítségével rangsorolja a célnyelvi szótár szavait, hogy melyik szó következik legnagyobb valószínűséggel a fordítás során.

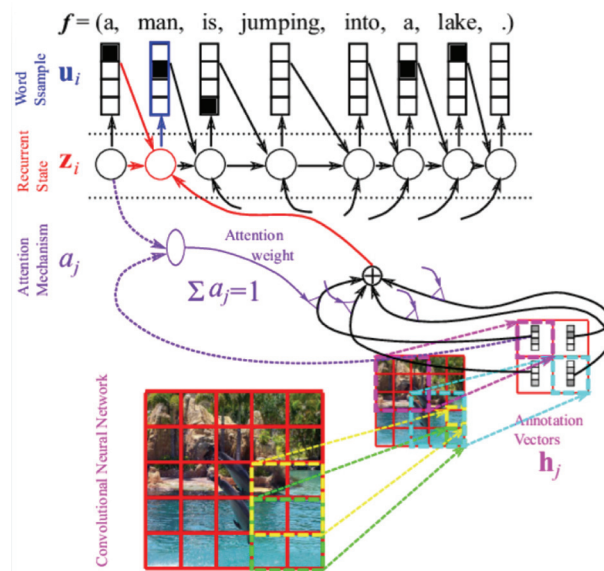
Ezen a ponton merül fel az a kérdés, hogy egy darab 300-500 dimenziós vektor valóban alkalmas arra, hogy tetszőleges hosszúságú mondatnak képes legyen eltárolni oly módon a jelentését, hogy abból egyértelműen előállítható legyen a célnyelvi mondat fordítása? A kérdéssel foglalkozó kutatások²² erre azt a választ hozták, hogy ebben a formájában ez az architektúra nem képes erre. A megoldást Bahdanau és társai²³ mutatták be 2014-ben. Munkájuk során az enkóder-dekóder architektúrát kiegészítették egy figyelmi modellnek nevezett neurális hálózattal, amely a dekódolást végző RNN-neuron bemenetére az aktuális szó fordításához szükséges forrásnyelvi szavakat is figyelembe veszik.

5. Érdekességek a technológia további alkalmazhatóságára

A neurálishálózat-alapú rendszerek óriási előnye, hogy nemcsak a nyelvek szintjére korlátozódnak. Ez azt jelenti, hogy bármilyen adattípust képesek kezelni bemenetként, amennyiben létezik annak megfelelő folytonos vektoriális reprezentációja. Ilyen lehet akár a hang, a videó, a kép stb.

²² Bahdanau és mtsai 2014; Cho és mtsai 2014; Sutskever és mtsai 2014.

²³ Bahdanau és mtsai 2014.

11. ábra. Példa a képleírás generátor működésére²⁴

Nemrégiben a Montreali Egyetemen és a Torontói Egyetemen végzett kutatások bebizonyították, hogy lehetséges olyan figyelemalapú enkóder-dekóder modell kialakítása, amely egy kép alapján szöveges formátumban leírást készít annak tartalmáról. A képet az enkóder konvolúciós neurális hálózattal történő helyettesítésével írja le, majd a figyelmi modell segítségével állítják elő a leírásnak számító mondatot. Ezt a folyamatot mutatja be a 10. ábrán. Hasonló munkákat²⁵ is bemutatott.

Tovább feszegették a határokat Li és munkatársai, amikor 2015-ben hasonló figyelemalapú rendszerekkel kísérleteztek videókra. A video leírásához a dekóder a videóból kinyert ideiglenes struktúrákat használta fel. A közelmúltban több kutatás

²⁴ <https://devblogs.nvidia.com/paralleforall/introduction-neural-machine-translation-gpus-part-3/>

²⁵ Donahue és mtsai 2014; Fang és mtsai 2014; Karpathy–Li 2014.

foglalkozott a videók enkóder-dekóder architektúra segítségével készített leírásával.²⁶ A 12. ábra egy példát mutat, ahol egy főzéssel kapcsolatos videofelvétel képkockái alapján készül el a következő leírás angolul: „Valaki olajban halat süt.” Ezt a négy képkockából kinyert információk alapján állította össze.



12. ábra. Példa filmfelirat generálására²⁷

Végül Chan és munkatársai²⁸ kutatásuk során létrehoztak egy hangból leíratot előállító rendszert. A beszédtechnológiából ismert eljárások segítségével állították elő a hangfelvétel vektoros reprezentációját, amiből karakteralapú enkóder-dekóder hálózat segítségével generálták ki a szöveges kimeneti formátumot.

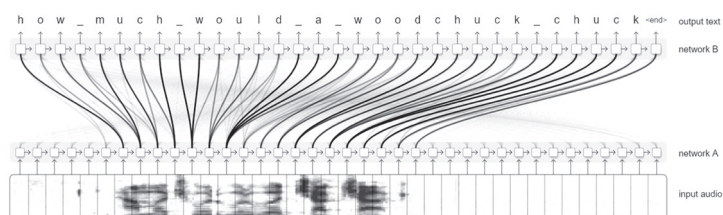
²⁶ Li és mtsai 2015; Venugopalan és mtsai 2015.

²⁷ <https://devblogs.nvidia.com/paralleforall/introduction-neural-machine-translation-gpus-part-3/>

²⁸ Chan és mtsai 2015.

6. Összefoglaló

A Magyar Tudomány Ünnepe keretében elmondott előadásomban bemutattam a mesterséges intelligencia alkalmazását a gépi fordítás területén. Napjainkban a mesterséges intelligencia térnyerésével jelentős változások figyelhetők meg a legtöbb tudományág területén, és ez nincs másként a nyelvtudományokkal sem. A bemutatott új technológiák és architektúrák alkalmazásával sikerült áttörést elérni az eddigi piacvezető kifejezésalapú gépi fordítási módszerekkel szemben és nagymértékű minőségjavulást elérni, még az olyan nehéz nyelvpárok esetén is, mint az angol és a magyar. Azt most nehéz megmondani, hogy mire lesznek képesek a számítógépek a jövőben, és hogy bármikor is sikerül-e létrehozni a tökéletes fordítórendszert, de a tudomány számára nincs lehetetlen.



13. ábra. Példa a hangból szöveg generálásra

Irodalom

- Bahdanau, Dzmitry – Cho, Kyunghyun – Bengio, Yoshua 2014. Neural Machine Translation by Jointly Learning to Align and Translate. *Computing Research Repository (CoRR)*, abs/1409.0473.
- Brown, Peter F. – Pietra, Vincent J. Della – Pietra, Stephen A. Della – Mercer, Robert L. 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation, *Computational Linguistics*, 19 (2): 263–311.

- Castaño, M. – Asunción, Casacuberta, Francisco – Vidal, Enrique 1997. *Machine Translation using Neural Networks and Finite-State Models* 160–167.
- Chan, William – Jaitly, Navdeep, Le – Quoc V. – Vinyals, Oriol 2016. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition, in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 4960–4964.
- Cho, KyungHyun – Merriënboer, Bart van – Bahdanau, Dzmitry – Bengio, Yoshua 2014. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches, *CoRR*, abs/1409.1259.
- Donahue, Jeff – Hendricks, Lisa – Anne, Guadarrama, Sergio – Rohrbach, Marcus – Venugopalan, Subhashini – Saenko, Kate – Darrell, Trevor 2014. Long-term Recurrent Convolutional Networks for Visual Recognition and Description, *CoRR*, abs/1411.4389.
- Fang, Jing – Nevin, Philip – Kairys, Visvaldas – Venclovas, Česlovas – Engen, John R. – Beuning, Penny J. 2014. Conformational Analysis of Processivity Clamps in Solution Demonstrates that Tertiary Structure Does Not Correlate with Protein Dynamics. *Structure* 22 (4): 572–581.
- Forcada, Mikel L. – Neco, Ramón P. 1997. Recursive hetero-associative memories for translation, in Mira, José, Moreno-Díaz, Roberto, and Cabestany, Joan Eds., *Biological and Artificial Computation: From Neuroscience to Technology: International Work-Conference on Artificial and Natural Neural Networks, IWANN'97 Lanzarote, Canary Islands, Spain, June 4–6, 1997 Proceedings*, Berlin, Heidelberg: Springer. 453–462.
- Hutchins, John 2005. Towards a Definition of Example-Based Machine Translation. In *Proceedings of Workshop on Example-Based Machine Translation, MT Summit X*, Phuket, Thailand. 63–70.
- Indurkha, Nitin – Damerau, Fred J. 2010. *Handbook of Natural Language Processing*, 2nd ed. Boca Raton, FL, Chapman & Hall/CRC.
- Kalchbrenner, Nal – Blunsom, Phil 2013. Recurrent Continuous Translation Models, in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle: Association for Computational Linguistics.
- Karpathy, Andrej – Fei-Fei, Li 2015. Deep Visual-Semantic Alignments for Generating Image Descriptions, in *The IEEE Confe-*

- rence on Computer Vision and Pattern Recognition (CVPR) Boston, USA.
- Koehn, Philipp 2005. Europarl: A Parallel Corpus for Statistical Machine Translation, in *Conference Proceedings: the tenth Machine Translation Summit*. Phuket, Thailand: AAMT. 79–86.
- Koehn, Philipp 2010. *Statistical Machine Translation*. New York: Cambridge University Press.
- Koehn, Philipp 2017. Neural Machine Translation, *CoRR*, abs/1709.07809.
- Li, Yao – Torabi, Atousa – Cho, Kyunghyun – Ballas, Nicolas – Pal, Christopher – Larochelle, Hugo – Courville, Aaron 2015. Describing videos by exploiting temporal structure. *Proceedings of the IEEE international conference on computer vision*. 4507–4515.
- Mikolov, Tomas – Chen, Kai – Corrado, Greg – Dean, Jeffrey 2013. Efficient Estimation of Word Representations in Vector Space, in *Proceedings of Workshop at ICLR Scottsdale, AZ*.
- Mikolov, Tomas – Yih, Wen-tau – Zweig, Geoffrey 2013. Linguistic Regularities in Continuous Space Word Representations. *HLT-NAACL*. 746–751.
- Och, Franz Josef – Ney, Hermann 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics* 29 (1): 19–51.
- Och, Franz Josef – Ney, Hermann 2004. The Alignment Template Approach to Statistical Machine Translation. *Computational Linguistics* 30 (4): 417–449.
- Schwenk, Holger (2007) Continuous Space Language Models. *Comput. Speech Lang* 21 (3): 492–518.
- Shannon, Claude Elwood 1948a. A mathematical theory of communication. *Bell System Technical Journal* 27 (July): 379–423.
- Shannon, Claude Elwood 1948b. A mathematical theory of communication. *Bell System Technical Journal* 27 (October): 623–656.
- Siklósi Borbála – Novák Attila 2016a. Beágyázási modellek alkalmazása lexikai kategorizációs feladatokra. In Tanács Attila – Varga Viktor – Vincze Veronika (eds.): *XII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2016)*. Szeged: Szegedi Tudományegyetem. 3–14.
- Siklósi Borbála – Novák Attila 2016. Közeli rokonunk, az autó. In Tanács Attila – Varga Viktor – Vincze Veronika (eds.): *XII. Magyar*

Számítógépes Nyelvészeti Konferencia (MSZNY 2016). Szeged: Szegedi Tudományegyetem. 27–36.

- Silver, David – Hubert, Thomas – Schrittwieser, J. – Antonoglou, I. – Lai, M. – Guez, A. – Lanctot, M. – Sifre, L. – Kumaran, D. – Graepel, T. – Lillicrap, T. – Simonyan, K. – Hassabis, D. 2017. Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm. *ArXiv e-prints*.
- Silver, David – Schrittwieser, Julian – Simonyan, Karen – Antonoglou, Ioannis – Huang, Aja – Guez, Arthur – Hubert, Thomas – Baker, Lucas – Lai, Matthew – Bolton, Adrian – Chen, Yutian – Lillicrap, Timothy – Hui, Fan – Sifre, Laurent – van den Driessche, George – Graepel, Thore – Hassabis, Demis 2017. Mastering the game of Go without human knowledge. *Nature* 550: 354–359.
- Sutskever, Ilya – Vinyals, Oriol – Le, Quoc V. 2014. Sequence to Sequence Learning with Neural Networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Montreal: MIT Press. 3104–3112.
- Venugopalan, Subhashini – Rohrbach, Marcus – Donahue, Jeff – Mooney, Raymond J. – Darrell, Trevor – Saenko, Kate 2015. Sequence to Sequence – Video to Text. *CoRR*, abs/1505.00487.