Nyelvi rekeszelés, avagy hogyan érzékeltessük írásban a hangsúlyozást avatárok számára

(The role of the structuring in the interpretation of languages/sentences or how it is possible to initialize verbal effects in written form)

Pitlik László (MY-X team)

Kivonat: Az idegen nyelvek megértése kapcsán sokszor felmerül a tanulóban, hogy azt sem érti, egyes szavak miért (és/vagy miért éppen ott / olyan sorrendben) szerepelnek egy-egy mondatban. A rekeszelés szó tehát arra utal, hogy az idegen nyelvek tanulása során az emberi agynak az anyanyelvétől eltérő szó-összetartozásokat meg kell tanulni egyre inkább és egyre komplexebb módon felismerni. A helyzetet csak bonyolítja az, ha a leírt mondat mást jelent eltérő hangsúlyozások esetén. Ebben a cikkben arra láthatunk példát, mit is kezdenek manapság (2021. tavaszán) az automatikus fordítási szolgáltatások azokkal a potenciális félreértésekkel, értelmezési alternatívákkal, melyek az emberi agy számára is gondot okoznak. Itt és most nem lehet arról beszélni, vajon melyik szoftver a jobb. A szoftverek eltérő viselkedése sokkal inkább arra hívhatja fel a gondosan fogalmazni akarók figyelmét – automatizálható gyanúgenerálás formájában, hogy adott mondat nem elég robosztus még. A mondatok robosztussága mindezek alapján automatikusan és valós időben értelmezhető akár döntés-támogató keretek között is hasonlóságelemzésre támaszkodva.

Kulcsszavak: Google-Translate, DeepL, gyanú-generálás

Abstract: The understanding of foreign languages may catalyse emotions in the learners that e.g. certain words are unnecessary in a sentence or even the ranking of the used words can quite not be interpreted. The delineation (as keyword) is responsible for that processes of the human brain – how we identify words with strong connections to each other in case of foreign languages. Let alone, it is also possible, that the meaning of a sentence will be an other one if the accentuation is changed. The paper presents examples where the automatic translations services (2021 - spring) should have problems with the interpretation alternatives as in case of the human brains. The question will not be whether a translation service seems to be better than an other service. The different translation results are rather a sign for a situation where the robustness of a basic sentence should be seen as suspected in an automated way. The robustness-index of the sentences can be derived based on similarity analysis even in an automated way and/or in real-time.

Keywords: Google-Translate, DeepL, suspicion generating

# Bevezetés

Az, hogy az emberi nyelvben a tagolás (rekeszelés, szórend és a hangsúlyok) hatással bírnak, triviális: vö. <https://hu.wikipedia.org/wiki/Reginam_occidere> - ahogy az a Bánk Bán kapcsán ismert példával is jól érzékeltethető:

**A királynét megölni nem kell félnetek jó lesz ha mindenki egyetért én nem ellenzem.**

A szöveg kétféle, egymással ellentétes értelme csupán a vesszők alkalmazásával áll elő:

A királynét megölni nem kell, félnetek jó lesz, ha mindenki egyetért, én nem, ellenzem.

(Reginam occidere bonum est timere, nolite. Et si omnes consenserint, ego non. Contradico.)

A királynét megölni nem kell félnetek, jó lesz, ha mindenki egyetért, én nem ellenzem.

(Reginam occidere bonum est timere nolite et si omnes consenserint ego non contradico.)

A mondatok jelentésének kockázataival foglalkozik a 20q-alapú online játékra támaszkodva az a gondolat (vö. <https://miau.my-x.hu/miau2009/index.php3?x=e0&string=20q>), miszerint minél több és minél gyakoribb az egyre magas absztrakciós szintű szavak aránya egy szövegben, annál kevésbé lehet értelmet tulajdonítani ezen szövegnek.

A másik mondat-értelmezési koncepció értelmében (vö. <https://miau.my-x.hu/mediawiki/index.php/QuILT-Robot-English>) egy akármilyen nyelvű alapmondatot, ha úm. körbe kergetünk automatikus fordításokon nyelvről nyelve támaszkodva az előző fordításra, akkor az a mondat a robosztus mondat, mely minél több nyelven keresztül végül quasi sértetlenül adja vissza az eredeti nyelvre fordítás után az ugyanezen nyelven megadott alapmondatot.

További háttér-dokumentumok egy nyelvi IQ-teszthez, ahol a mutató névmást értelmezési alternatívái adják az adott személy gondolkodásmódjának megítéléséhez az alapot pl. munkaerő-felvételt érintő elbeszélgetésekben:

* <https://miau.my-x.hu/mediawiki/index.php/QuILT-1> (egg + floor + crack)
* <https://miau.my-x.hu/miau/quilt/Definitions_of_knowledge.docx> (Ei Boden)

# A kiindulási helyzet

Vegyünk végre immár egy látszólag egyszerű német mondatot és ennek zárójelekkel vagy kötőjelekkel finomhangolt verzióit:

1. DE: Die Fenster waren wegen des Lärms draußen geschlossen.
2. DE: Die Fenster waren wegen des Lärms von draußen geschlossen.
3. DE: Die Fenster waren wegen des Lärms - draußen geschlossen.
4. DE: Die Fenster waren wegen des Lärms - von draußen geschlossen.
5. DE: Die Fenster waren - wegen des Lärms - draußen geschlossen.
6. DE: Die Fenster waren - wegen des Lärms - von draußen geschlossen.
7. DE: Die Fenster waren (wegen des Lärms) draußen geschlossen.
8. DE: Die Fenster waren (wegen des Lärms) von draußen geschlossen.
9. DE: Die Fenster waren (wegen des Lärms draußen) geschlossen.
10. DE: Die Fenster waren (wegen des Lärms von draußen) geschlossen.
11. DE: Die Fenster waren - wegen des Lärms draußen - geschlossen.
12. DE: Die Fenster waren - wegen des Lärms von draußen - geschlossen.
13. DE: Die Fenster waren wegen des Lärms draußen - geschlossen.
14. DE: Die Fenster waren wegen des Lärms von draußen - geschlossen.

(A vesszővel való manipulálás itt kizárásra kerül – de a robot-fordítók vélhetően a vesszője is reagálnak)

A kérdés csak az, vajon az automatikus fordítók (DEEPL és Google-Translate) miként reagálnak a kötőjelekre vagy a zárójelekre:

# Fordítások

A következőkben kerüljenek a fenti (DE:1-14) mondatok lefordításra a Google Translate (<https://translate.google.hu/>) és a DeepL (<https://www.deepl.com/translator>) online fordítási szolgáltatásain keresztül. A fordításhoz a fenti (DE:1-14) lista kerül bemásolásra mindkét fordító robot input-mezőjébe, mert arra is kíváncsiak vagyunk, hogy tételesen mit kezd a két szolgáltatás pl. az ország-jelölésekkel és/vagy a sorszámokkal (együtt):

## HU

### Google-Translate (DE🡪HU)

1. (DE)HU: Az ablakokat bezárták a kinti zaj miatt.
2. (DE)HU: Az ablakokat bezárták a kívülről érkező zaj miatt.
3. (DE)HU: Az ablakokat bezárták a zaj miatt - kint.
4. (DE)HU: Az ablakokat bezárták a zaj miatt - kívülről.
5. (DE)HU: A zaj miatt kint voltak bezárva az ablakok.
6. (DE)HU: Az ablakokat kívülről bezárták a zaj miatt.
7. (DE)HU: Kint bezárták az ablakokat (a zaj miatt).
8. (DE)HU: Kintről bezárták az ablakokat (a zaj miatt).
9. (DE)HU: Az ablakokat becsukták (a kinti zaj miatt).
10. (DE)HU: Az ablakok be voltak csukva (a kívülről érkező zaj miatt).
11. (DE)HU: Az ablakokat bezárták a kinti zaj miatt.
12. (DE)HU: Az ablakokat bezárták a kívülről érkező zaj miatt.
13. (DE)HU: Az ablakokat bezárták a kinti zaj miatt.
14. (DE)HU: Az ablakokat bezárták a kívülről érkező zaj miatt.

Gyors értelmezések:

* A „HU:” nem került lefordításra.
* A kezdeti kötőjelek száma (12) jelentősen csökkent (2).
* A páros kötőjelek száma (4) változott (0).
* A zárójelek száma (8) nem változott (vö. 8).
* A zárójeleken belüli tartalom 2- helyett 3-alakú.
* A kezdő szó változott.
* A mondatkezdet mindenkor nagybetű maradt.

### DeepL (DE🡪HU)

1. HU: Az ablakok zárva voltak a kinti zaj miatt.
2. az ablakok zárva voltak a kinti zaj miatt.
3. az ablakok zárva voltak a zaj miatt - kint.
4. az ablakok zárva voltak a kinti zaj miatt.
5. az ablakok - a zaj miatt - kint zárva voltak.
6. az ablakok - a zaj miatt - kívülről zárva voltak.
7. az ablakok (a zaj miatt) kint zárva voltak.
8. HU 8: Az ablakok (a zaj miatt) kívülről zárva voltak.
9. DE 9: Az ablakok zárva voltak (a kinti zaj miatt).
10. 10. az ablakok zárva voltak (a kinti zaj miatt).
11. 11. az ablakok zárva voltak (a kinti zaj miatt).
12. 12. Az ablakok zárva voltak a kinti zaj miatt.
13. 13. Az ablakok zárva voltak a kinti zaj miatt.
14. Az ablakok zárva voltak - a kinti zaj miatt.

Gyors értelmezések:

* A „DE:” részlegesen helyesen került lefordításra, ill. nem került lefordításra, valamint elhagyásra került.
* A kezdeti kötőjelek száma (12) csökkent (6).
* A páros kötőjelek száma (4) változott (2).
* A zárójelek száma (8) megváltozott, ráadásul nőtt (vö. 10).
* A zárójeleken belüli tartalom 2-alakú 2-alakú kiindulási helyzetben.
* A kezdő szó nem változott.
* A mondatkezdet nem mindig nagybetű.

## EN

### Google Translate (DE🡪EN)

1. (DE)EN: The windows were closed because of the noise outside.

2. (DE)EN: The windows were closed because of the noise from outside.

3. (DE)EN: The windows were closed because of the noise - outside.

4. (DE)EN: The windows were closed because of the noise - from outside.

5. (DE)EN: The windows were closed outside because of the noise.

6. (DE)EN: The windows were closed from outside because of the noise.

7. (DE)EN: The windows were closed (because of the noise) outside.

8. (DE)EN: The windows were closed (because of the noise) from outside.

9. (DE)EN: The windows were closed (because of the noise outside).

10. (DE)EN: The windows were closed (because of the noise from outside).

11. (DE)EN: The windows were closed because of the noise outside.

12. (DE)EN: The windows were closed because of the noise from outside.

13. (DE)EN: The windows were closed because of the noise outside.

14. (DE)EN: The windows were closed because of the noise from outside.

Gyors értelmezések:

* A „DE:” nem került lefordításra, de elhagyásra sem került.
* A kezdeti kötőjelek száma (12) jelentősen csökkent (2).
* A páros kötőjelek száma (4) változott (0).
* A zárójelek száma (8) nem változott (vö. 8).
* A zárójeleken belüli tartalom 2- helyett 3-alakú.
* A kezdő szó nem változott.
* A mondatkezdet mindig nagybetű.

### DeepL (DE🡪EN)

1. 1st EN: The windows were closed because of the noise outside.
2. EN: The windows were closed because of the noise from outside.
3. 3rd EN: The windows were closed because of the noise - outside.
4. 4th EN: The windows were closed because of the noise - from outside.
5. 5th EN: The windows were closed - because of the noise - outside.
6. the windows were closed - because of the noise - from outside.
7. 7th EN: The windows were closed (because of the noise) outside.
8. EN 8: The windows were closed (because of the noise) from outside.
9. 9th EN: The windows were closed (because of the noise outside).
10. EN 10: The windows were closed (because of the noise from outside).
11. 11th EN: The windows were closed - because of the noise outside.
12. the windows were closed because of the noise outside.
13. EN 13: The windows were closed - because of the noise outside.
14. the windows were - closed because of the noise from outside.

Gyors értelmezések:

* A „DE:” mindenkor kezelésre került, sőt lefordításra, valamint elhagyásra is került.
* A kezdeti kötőjelek száma (12) csökkent (9).
* A páros kötőjelek száma (4) változott (2).
* A zárójelek száma (8) nem változott (vö. 8).
* A zárójeleken belüli tartalom 2- helyett 3-alakú.
* A kezdő szó nem változott.
* A mondatkezdet nem mindig nagybetű.

# Alapvető konklúziók

* A kötőjelek részlegesen eltűnnek.
* A páros kötőjelek
  + teljesen eltűnhetnek,
  + egyszeres kötőjelre redukálódhatnak,
  + megmaradhatnak.
* A szórend változhat.
* A zárójelek száma változhat (nőhet – ami vélhetően egy fajta öröklődési hatás az egyben kezelt 14 példán belül).
* A „DE:” részlegesen (helyesen) kezelésre kerül.

Minél több változás érhető tetten minél több fordítási szolgáltatásban, ill. minél nagyobb az eltérés a fordítási szolgáltatások azonos inputra vonatkozó eredményei között, annál nagyobb az adott alapmondat kockázata. (Itt az esetleges körbe-fordítások önálló gyanújel-termelő erőtérként értelmezhetők.)

A fenti minél-annál értelmezési konstrukciók megalapozzák egy-egy mondat kockázatának mesterséges intelligencia-alapú értelmezhetőségét – automatizált módon (vö. <https://miau.my-x.hu/myx-free/> - COCO Y0).

Ha egy példamondat variánsai helyett egy bekezdést veszünk alapul, ahol minden mondat egy-egy objektum az OAM-ban, akkor bárminemű fordítási változást leíró mutatószám attribútumként lesz értelmezhető. Így a mondatok egymáshoz képest önálló kockázat-indexet tudnak kapni a lehet-e-minden-mondat-másként-egyformán-kockázatos anti-diszkriminatív elemzés keretében.

A mondatok kockázat-indexértékeinek egymásutánisága pedig egy fajta dinamikát képes érzékeltetni, ahol az egyre romló fogalmazási minőség pl. több műsorvezetővel/riporterrel/előadóval (vö. jam-session, kerekasztal) dolgozó produkciókban lehet akár real-time vészjelzés is arra vonatkozóan – kitől kell elvenni úm. a szót (hacsak nem az a cél, hogy pl. humoros műsorokban éppen a többértelműségre legyen kihegyezve a kommunikáció)…

# Több nyelvű nézetek

DE: Die Fenster waren wegen des Lärms draußen geschlossen.

*EN: The windows were closed because of the noise outside.*

*HU: Az ablakokat bezárták a kinti zaj miatt.*

*RU: Окна были закрыты из-за шума снаружи.*

*NL: De ramen waren gesloten vanwege het lawaai buiten.*

*SE: Fönstren stängdes på grund av bullret utanför.*

DE: Die Fenster waren wegen des Lärms - draußen geschlossen.

*EN: The windows were closed because of the noise - outside.*

*HU: Az ablakokat bezárták a zaj miatt - kint.*

*RU: Окна были закрыты из-за шума - снаружи.*

*NL: De ramen waren gesloten vanwege het lawaai - buiten.*

*SE: Fönstren stängdes på grund av bullret - utanför.*

(Forrás: Google-Translate)

# További konklúziók I.

Egyedül magyarul tesz a Google érdemi (írásban is félreérthetetlenül létező, strukturális) különbséget – noha a német inkább rokona a többi véletlenszerűen választott nyelvnek, mint a magyarnak…

A zaj lehet tehát **kinti** és az ablakok becsukása történhet éppenséggel valahol **kint**.

A „draußen” szócska helyzete ugyanis lehet egy fajta „jelző” a zajt illetően és lehet az igéhez kapcsolódó pontosítás („határozó”) is…

Hogy mit is akar az eredeti mondat mondani, azt a beszélt nyelvben a hangsúly, az írott nyelvben az hangsúlyt helyettesítő kötőjel képes lehet tehát kifejezni…

## Módosított példa

DE: Die Fenster waren wegen des Lärms von draußen geschlossen.

*EN: The windows were closed because of the noise outside.*

*EN: The windows were closed because of the noise from outside.*

*HU: Az ablakokat bezárták a kinti zaj miatt.*

*HU: Az ablakokat a kinti zaj miatt becsukták. (szó-sorrend)*

*RU: Окна были закрыты из-за шума снаружи.*

*RU: Окна были закрыты из-за шума снаружи. (azonos)*

*NL1: De ramen waren gesloten vanwege het lawaai buiten.*

*NL2: De ramen waren gesloten vanwege het lawaai van buitenaf.*

*SE1: Fönstren stängdes på grund av bullret utanför.*

*SE2: Fönstren stängdes på grund av bullret från utsidan.*

DE: Die Fenster waren wegen des Lärms – von draußen geschlossen.

*EN: The windows were closed because of the noise - outside.*

*EN: The windows were closed because of the noise - from outside.*

*HU: Az ablakokat bezárták a zaj miatt - kint.*

*HU: Az ablakokat bezárták a zaj miatt - kívülről.*

*RU: Окна были закрыты из-за шума - снаружи.*

*RU: Окна были закрыты из-за шума - снаружи. (azonos)*

*NL: De ramen waren gesloten vanwege het lawaai - buiten.*

*NL: De ramen waren gesloten vanwege het lawaai - van buitenaf.*

*SE: Fönstren stängdes på grund av bullret - utanför.*

*SE: Fönstren stängdes på grund av bullret - från utsidan.*

(Forrás: Google-Translate)

# További példák

Google Translate - <https://translate.google.hu/>

Die Fenster waren wegen des Lärms von draußen - geschlossen.

* The windows were closed because of the noise from outside.
* Az ablakokat a kinti zaj miatt becsukták.

Die Fenster waren wegen des Lärms draußen - geschlossen.

* The windows were closed because of the noise outside.
* Az ablakokat bezárták a kinti zaj miatt.

DEEPL - <https://www.deepl.com/translator>

Die Fenster waren wegen des Lärms draußen - geschlossen.

* The windows were because of the noise outside - closed.
* Az ablakok a kinti zaj miatt zárva voltak.

Die Fenster waren wegen des Lärms draußen geschlossen.

* The windows were closed because of the noise outside.
* Az ablakok zárva voltak a kinti zaj miatt.

A három (2+2 – 1 közös mondattal) alapmondat és a két fordító szolgáltatás eredőjeként:

* A magyar nyelvi megoldás 4 eltérő eredményre vezetett, de a kezdő szó (ablak) minden esetben azonos maradt, noha a korábbi vizsgálatokban ez sem volt triviálisan adott (vö. Google: DE🡪HU: 7 és 8)
* A közös alapmondat angol megfelelője sem vezetett azonos eredményre, hiszen változott a szórend a kötőjel ignorálása vs. megtartás okán.

# Jövőkép

A mondatok nyelvi robosztusságának letapogatása mellett kötőjelek és/vagy zárójelek segítségével, ezen jelek racionálisan használhatók a beszélő robotok hangsúlyozásának tudatos befolyásolására is (vö. <https://miau.my-x.hu/miau/quilt/2020/quilt2/launching2020III25/part0.html>).

# Referenciák

…szövegközben…